

LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL

Retos jurídicos, éticos y sociales
en México y el mundo



DERECHOS
HUMANOS



ÉTICA
ALGORÍTMICA



GOBERNANZA
DIGITAL



RESPONSABILIDAD
JURÍDICA



IMPACTO
SOCIAL

COORDINADORES

Oscar Lozano Carrillo
Alfredo Garibay Suárez
Francisco Javier Arias Vargas

**La Regulación
de la Inteligencia Artificial**

Primera edición, 2026

Coordinación general del trabajo:

Oscar Lozano Carrillo

Alfredo Garibay Suárez

Francisco Javier Arias Vargas

Integración editorial y digital primera fase

Grecia Josafat Jarillo Olague

Editado por:

Red Mexicana de Investigadores en

Estudios Organizacionales

(REMINEO), A.C.

América 126-A int. 202, Col. Parque San Andrés, Coyoacán n, Ciudad de México,

CP 04040

ISBN: 978-607-8049-40-0

El contenido, las opiniones y la originalidad de los capítulos publicados en este libro son responsabilidad exclusiva de sus autores y no reflejan necesariamente el punto de vista de los coordinadores, del editor o de los avales académicos.

Editado en México.

La Regulación de la Inteligencia Artificial

Coordinación general del trabajo:

Oscar Lozano Carrillo

Alfredo Garibay Suárez

Francisco Javier Arias Vargas

Red Mexicana de Investigadores en Estudios
Organizacionales A.C. (REMINEO)



CONTENIDO

<i>Prólogo</i>	10
<i>Capítulo I – La Regulación Global de la Inteligencia Artificial</i>	13
Autores: Oscar Lozano Carrillo, Alfredo Garibay Suarez, Francisco Javier Arias Vargas	
<i>Capítulo II – La Gobernanza Global de la Inteligencia Artificial desde el Sur Global: Una visión crítica y propositiva</i>	57
Autora: Cinthia Noemi Zacatenco Arellano	
<i>Capítulo III – Gobernanza Ética y Regulación de la Inteligencia Artificial en la Educación Superior: Propuesta Normativa para la Universidad del Caribe</i>	90
Autores: Miguel Ángel Olivares Urbina, Ricardo Sonda de la Rosa, Rosiluz Ceballos Povedano	
<i>Capítulo IV – Gobernanza de la Inteligencia Artificial y Trabajo: Desafíos regulatorios ante el desplazamiento laboral en el siglo XXI</i>	116
Autora: Didre Estefani Morales Rodríguez	
<i>Capítulo V – Regulación de la Inteligencia Artificial y Sostenibilidad competitiva de las MIPYMES digitales en Guayaquil</i>	165
Autores: María Paulina Brito Ochoa, Fabricio Alfonso Vasco Mora	
<i>Capítulo VI – Ética en el uso de la Inteligencia Artificial: Entre la innovación y la responsabilidad</i>	185
Autores: Sergio Cruz Hernández, Didre Estefani Morales Rodríguez	

Capítulo VII – <i>La Regulación de la Inteligencia Artificial en la Administración Pública: Retos Jurídicos, Criterios Jurisprudenciales, y Perspectivas de Gobernanza</i>	218
Autor: Miguel Francisco Barrera Aldama	
Capítulo VIII – <i>La implementación de la IA en prácticas de gobernanza para la democratización del conocimiento</i>	237
Autora: Erandi Nina Velarde Torres	
Capítulo IX – <i>Inteligencia Artificial, ética y accesibilidad: desafíos en contextos de discapacidad y brecha digital</i>	262
Autora: Nadia Estefania Dalcol	
Capítulo X – <i>Inteligencia artificial, trabajo e innovación en el contexto de la empresa familiar en México</i>	275
Autores: Juan Cayetano Niebla Zatarain, Xiomara Penelo Zaldivar Colado, José Guadalupe Robles Hernández, José Efren Leyva Duarte	
Capítulo XI – <i>Transparencia algorítmica: Por qué explicar la inteligencia artificial es más importante que regularla</i>	310
Autor: Rafael Angeles García	
Capítulo XII – <i>La Inteligencia Artificial y la Autoría Pedagógica: El Libro de texto humanizado en la educación media superior</i>	333
Autores: César Augusto Arellano Rodríguez, Ivanna Arellano Reyes	
Capítulo XIII – <i>Regulación jurídica de la Inteligencia Artificial frente a los Derechos Humanos</i>	361
Autores: Jesús Manuel López Marroquín, Gisela Cota Yucupicio, Edna Karina Miranda Mondaca	

Capítulo XIV – <i>Regulación de la Inteligencia Artificial a Nivel Global: Enfoques y Modelos contemporáneos</i>	412

Autora: Marlene Atenea Grimaldo Ramírez

Capítulo XV – <i>Ética, investigación e Inteligencia Artificial Generativa en la Educación Superior latinoamericana: fundamentos filosóficos para una orientación de política universitaria</i>	438

Autores: Jorge Alberto Rosas Castro, Gabriela Camacho Deonicio

Capítulo XVI – <i>La ilusión de la neutralidad: inteligencia artificial, poder y sesgo en la configuración de la realidad digital</i>	482

Autor: Víctor Daniel Santos Hortelano

Capítulo XVII – <i>Gobernanza de la IA con Propósito: Un Marco para el Impacto Social Universitario</i>	515

Autora: Janette Brito Laredo

Capítulo XVIII – <i>Regulación activa del individuo frente a la inteligencia artificial: el sujeto como primer agente regulador en la era de la IA generativa</i>	544

Autor: Fernando Javier Vega Riveros

Capítulo XIX – <i>Elementos para la construcción de una gobernanza digital situada: un análisis de la regulación de la IA en entornos legislativos</i>	580

Autor: Luis Angel García Covarrubias

Capítulo XX – <i>La Inteligencia Artificial Aplicable En Los Procesos Jurisdiccionales Mexicanos</i>	601

Autor: Jesús Alberto Pérez Rodríguez

Capítulo XXI – *Ética, sesgos algorítmicos y responsabilidad social* 630
Autor: Cristian Arturo Plaza Cuadras

Capítulo XXII – *La Inteligencia Artificial y la sustitución de procesos humanos en el derecho* 658
Autora: Grecia Josafat Jarillo Olague

Capítulo XXIII – *Inteligencia Artificial responsable:
Un análisis sistemático* 684
Autores: Edgar Alfonso Sansores Guerrero, Juana Edith Navarrete Marneou

Capítulo XXIV – *Ética, Sesgos Algorítmicos y Derechos Humanos: Hacia una Gobernanza Global de la Inteligencia Artificial* 706
Autoras: Bianca Melissa Soto Rivera, Sandra Paola Morales Flores

Capítulo XXV – *Análisis de la norma ISO/IEC 42001 en contextos organizacionales emergentes: una lectura desde la racionalización y el isomorfismo institucional* 732
Autora: Virginia López Nevárez

Capítulo XXVI – *NotebookLM a la Luz de los Principios de Chapultepec* 764
Autora: Elizabeth Ramírez Navarro

Capítulo XXVII – *Gobernanza pedagógica de la inteligencia artificial en educación superior: modelo metodológico, análisis comparado y aplicación en el programa PIAESS* 786
Autora: Leslie Bridshaw Araya

Capítulo XXVIII – *Aportes para la regulación de una IA sostenible* 811

Autores: Karen Liliana Chip Domínguez, Jorge Iván González Gómez

Capítulo XXIX – *Gobernanza algorítmica y trabajo digno: hacia un marco regulatorio de la inteligencia artificial en las relaciones laborales contemporáneas* 846

Autora: Aylin Torres Ramírez

Capítulo XXX – *Centro de Inteligencia de Mercados y Analítica del Consumidor (CIMEC-IUE): concepto, funciones estratégicas y líneas de acción en el marco de la regulación de la inteligencia artificial* 879

Autores: Francisco Javier Arias Vargas, Oscar Lozano Carrillo, Rafael Alejandro Betancourt Durango, Valery Gabriela Galvis Monterroza, Valentina Chavarria Calle, Laura Echeverry Sierra

Capítulo XXXI – *Entre algoritmos y emociones: La regulación de la inteligencia artificial en el deporte y sus límites humanos* 916

Autora: Marlene Atenea Grimaldo Ramírez

Capítulo XXXII – *Políticas culturales en la era de la inteligencia artificial: regulación, diversidad y trabajo creativo* 941

Autores: Guadalupe Zaragoza Rodríguez, José Antonio Martínez Barajas

Capítulo XXXIII – *Dignidad, base de toda regulación de inteligencia artificial: hacia un constitucionalismo tecnológico de la vida compartida* 962

Autores: Filberto Eduardo R. Manrique Molina, Mauricio Iván Vargas Mendoza

Capítulo XXXIV – *Regulación de la inteligencia artificial en empresas: Gobernanza, marcos internacionales y vacíos normativos en México* 990

Autora: Wendy Sebastiana Hernández del Puerto

Capítulo XXXV – *Regulación de la IA en México Práctica docente y de Investigación Universitaria* 1014

Autores: Arcadio Zebadúa Sánchez, Zoily Mery Cruz Sánchez, Karla Beatriz García Arteaga, Consuelo Guadalupe Morales Flores, Cruz Susana Estrada Castellanos.

Capítulo XXXVI – *Inteligencia Artificial General: Entre La Expansión De Capacidades Y Los Límites De Su Regulación* 1040

Autores: Oscar Lozano Carrillo, Aylin Torres Ramírez, Cinthia Noemi

Zacatenco Arellano

Prólogo

En el umbral de una nueva era tecnológica, la inteligencia artificial se ha posicionado como una de las fuerzas más transformadoras de nuestro tiempo. Sin embargo, su irrupción no solo plantea desafíos técnicos o económicos, sino interrogantes profundamente humanos, sociales y civilizatorios. El libro *La Regulación de la Inteligencia Artificial* emerge precisamente en ese punto de tensión, donde la innovación se encuentra con la responsabilidad, y donde el futuro aún está en disputa.

Tal como se plantea desde sus primeras páginas, la inteligencia artificial no puede entenderse únicamente como una herramienta neutral, sino como un fenómeno que reconfigura la manera en que producimos conocimiento, tomamos decisiones y organizamos la vida social, en este sentido, regular la IA no es simplemente establecer límites técnicos o jurídicos, sino intervenir en la forma misma en que se construye la realidad contemporánea. Una de las principales virtudes de esta obra es reconocer que los riesgos asociados a la inteligencia artificial operan en múltiples niveles, por un lado, se encuentran aquellos riesgos visibles e inmediatos —como la vulneración de la privacidad, la desinformación o el impacto en el empleo— que demandan respuestas urgentes desde los marcos normativos existentes , pero, por otro lado, el libro nos invita a mirar más profundamente, hacia una dimensión menos evidente pero más determinante: aquella en la que la IA transforma estructuras de poder, redefine la autonomía humana y reconfigura las bases mismas de la sociedad.

Este desplazamiento analítico es fundamental, dado que por mucho tiempo, la regulación tecnológica ha operado en la superficie de los fenómenos, atendiendo sus manifestaciones más visibles, no obstante, como se argumenta en esta obra, limitarse a este nivel implica dejar intactas las dinámicas estructurales que sostienen dichos problemas, en términos conceptuales, podríamos decir que la regulación ha estado mirando la “punta del iceberg”, ignorando los procesos más profundos que configuran el sentido, la cultura y el poder.

Desde esta perspectiva, la regulación de la inteligencia artificial debe entenderse como un proceso multidimensional, capaz de articular lo técnico con lo social, lo jurídico con lo ético, y lo inmediato con lo estructural, no se trata únicamente de controlar la tecnología, sino de orientar su desarrollo en función de valores que prioricen la vida, la dignidad y la equidad. El libro ofrece, además, una contribución especialmente relevante al situar este debate en el contexto latinoamericano y del Sur Global, escenario dominado por grandes corporaciones tecnológicas y marcos regulatorios desarrollados en el norte global, esta obra plantea la necesidad de construir una gobernanza situada, capaz de responder a las realidades, desigualdades y aspiraciones de nuestras sociedades y esta mirada no solo enriquece el debate, sino que lo vuelve políticamente significativo.

Asimismo, el texto pone sobre la mesa una tensión central de nuestro tiempo: la relación entre autonomía humana y dependencia tecnológica; A medida que los sistemas de inteligencia artificial adquieren mayor capacidad de decisión, se vuelve imprescindible cuestionar no solo qué pueden hacer las máquinas, sino qué estamos dejando de hacer los humanos, la erosión de capacidades cognitivas, la automatización de la desigualdad y la concentración del poder tecnológico no son efectos colaterales, sino procesos que deben ser comprendidos y atendidos desde una lógica de intervención consciente.

En este sentido, este libro no es únicamente un compendio de análisis jurídicos o técnicos y es, ante todo, una invitación a pensar la inteligencia artificial como un fenómeno social total, que atraviesa todas las dimensiones de la vida humana su lectura exige, por tanto, una postura crítica, reflexiva y comprometida.

Finalmente, la obra nos recuerda una idea fundamental: la inteligencia artificial solo tiene sentido si está al servicio de la vida y esta afirmación, lejos de ser retórica, constituye el núcleo ético y político de toda propuesta regulatoria en un mundo donde la tecnología avanza a una velocidad sin precedentes, la verdadera pregunta no es qué podemos hacer con la IA, sino qué tipo de humanidad queremos construir con ella.

Oscar Lozano Carrillo

CAPÍTULO I

LA REGULACIÓN GLOBAL DE LA INTELIGENCIA ARTIFICIAL

Global Regulation of Artificial Intelligence

Oscar Lozano Carrillo¹

exato@azc.uam.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0001-6166-3033>

Alfredo Garibay Suárez²

garibay@azc.uam.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0002-2159-2864>

Francisco Javier Arias Vargas³

fjarias@correo.iue.edu.co

Institución Universitaria de Envigado

<https://orcid.org/0000-0002-4483-1741>

Resumen

El capítulo analiza la inteligencia artificial (IA) como una tecnología transformadora que genera riesgos inmediatos y estructurales para la sociedad. En primer lugar, estudia riesgos visibles como la vulneración de la propiedad intelectual, el uso indebido de datos personales, la desinformación mediante deepfakes, el impacto en el empleo y la falta de transparencia algorítmica. Estos problemas afectan derechos fundamentales, la privacidad, la estabilidad democrática y las dinámicas

¹ Profesor-investigador del Departamento de Administración de la UAM Azcapotzalco

² Profesor-investigador del Departamento de Administración de la UAM Azcapotzalco

³ Profesor-investigador de la Institución Universitaria de Envigado

económicas, evidenciando las limitaciones de los marcos jurídicos tradicionales frente al avance tecnológico.

Posteriormente, el texto profundiza en riesgos estructurales menos visibles, pero de mayor alcance, como la pérdida de control humano sobre sistemas autónomos, la concentración de poder tecnológico, la automatización de desigualdades y la erosión de capacidades humanas como el pensamiento crítico y la autonomía. Se sostiene que estos riesgos se encuentran interrelacionados y pueden transformar profundamente la organización social, política y cultural.

Asimismo, el capítulo examina los principales modelos regulatorios internacionales, regionales, nacionales e institucionales, destacando instrumentos como el AI Act de la Unión Europea, los marcos éticos de la UNESCO y diversas iniciativas en México. Se concluye que la regulación actual sigue siendo predominantemente reactiva y centrada en riesgos inmediatos, dejando insuficientemente atendidos los riesgos estructurales. Por ello, se propone una regulación integral, multinivel y orientada a preservar la autonomía humana, la equidad, la democracia y el interés colectivo frente al creciente poder tecnológico.

Palabras clave: Inteligencia artificial; regulación; riesgos estructurales; privacidad; desinformación; deepfakes; transparencia algorítmica; autonomía humana; desigualdad; poder tecnológico.

Abstract

The chapter discusses artificial intelligence (AI) as a transformative technology that creates immediate and structural risks for society. First, it studies visible risks such as intellectual property infringement, misuse of personal data, disinformation through deepfakes, impact on employment, and lack of algorithmic transparency. These problems affect fundamental rights, privacy, democratic stability, and economic dynamics, evidencing the limitations of traditional legal frameworks in the face of technological advances.

The text then delves into less visible, but more far-reaching, structural risks, such as the loss of human control over autonomous systems, the concentration of

technological power, the automation of inequalities, and the erosion of human capacities such as critical thinking and autonomy. It is argued that these risks are interrelated and can profoundly transform social, political, and cultural organization. The chapter also examines the main international, regional, national and institutional regulatory models, highlighting instruments such as the AI Act of the European Union, UNESCO's ethical frameworks and various initiatives in Mexico. It is concluded that the current regulation continues to be predominantly reactive and focused on immediate risks, leaving structural risks insufficiently addressed. For this reason, a comprehensive, multilevel regulation aimed at preserving human autonomy, equity, democracy and collective interest in the face of growing technological power is proposed.

Keywords: Artificial intelligence; regulation; structural risks; privacy; misinformation; deepfakes; algorithmic transparency; human autonomy; inequality; technological power.

Introducción

La inteligencia artificial (IA) se ha consolidado como una de las tecnologías más influyentes del siglo XXI, con un impacto transversal en prácticamente todos los ámbitos de la vida social, económica y política. Su capacidad para procesar grandes volúmenes de datos, automatizar decisiones y generar contenido ha abierto oportunidades sin precedentes para la innovación, pero también ha generado un conjunto complejo de riesgos que desafían los marcos regulatorios tradicionales, en este contexto, la regulación de la IA se vuelve no solo pertinente, sino urgente.

En el debate contemporáneo, los riesgos asociados a la IA suelen abordarse desde una perspectiva inmediata y operativa, siendo de los más recurrentes se encuentran la vulneración de la propiedad intelectual, el uso indebido de datos personales, la proliferación de desinformación mediante deepfakes, el impacto en

el empleo y la falta de transparencia en los sistemas algorítmicos, estos problemas han captado la atención de legisladores, empresas y organismos internacionales, en tanto afectan directamente derechos individuales y sectores económicos clave. Por ejemplo, la tensión entre el entrenamiento de modelos y los derechos de autor refleja un conflicto estructural entre innovación tecnológica y marcos legales existentes (Rodríguez, 2020), de igual manera, la opacidad de los sistemas de IA plantea desafíos significativos para la rendición de cuentas, ya que dificulta comprender cómo y por qué se toman ciertas decisiones automatizadas (Copeland, 1993).

Asimismo, la preocupación por la privacidad y el uso de datos personales ha adquirido una centralidad particular en el diseño de políticas públicas, especialmente en regiones como la Unión Europea, donde se ha promovido un enfoque regulatorio basado en la protección de derechos fundamentales (Gamero Casado, 2024). En paralelo, el avance de tecnologías capaces de generar contenido sintético altamente realista ha intensificado los riesgos de manipulación informativa, afectando la confianza pública y la estabilidad democrática, estos riesgos, si bien críticos, tienden a ser abordados desde una lógica reactiva, centrada en mitigar efectos visibles y de corto plazo.

Sin embargo, limitar la regulación de la IA a estos problemas más evidentes resulta insuficiente. Existe una segunda dimensión de riesgos, menos visibles, pero significativamente más profunda, que afecta las estructuras mismas de la sociedad y la condición humana. Entre estos destacan la posible pérdida de control humano sobre sistemas autónomos, la concentración extrema de poder en actores tecnológicos, la automatización de desigualdades sociales y la erosión de capacidades cognitivas humanas. Estos riesgos no solo implican fallas técnicas o legales, sino transformaciones estructurales que pueden redefinir la relación entre tecnología y humanidad.

Diversos autores han advertido sobre este horizonte. Lassalle (2024) señala que el desarrollo de una “civilización artificial” podría derivar en escenarios donde la autonomía humana se vea subordinada a lógicas algorítmicas, especialmente si estas son controladas por corporaciones o regímenes con escasa rendición de cuentas. Por su parte, Sadin (2022) advierte que la expansión de sistemas digitales puede consolidar formas de poder que erosionan el espacio común y fortalecen dinámicas individualistas y autorreferenciales. En este sentido, la IA no solo actúa como herramienta, sino como un agente transformador de las estructuras sociales y culturales.

Además, la idea de que los sistemas de IA pueden evolucionar sin una comprensión plena de sus procesos internos refuerza la necesidad de cautela. Como sugiere Larson (2021), existe una tendencia a sobreestimar las capacidades de la IA y a subestimar sus limitaciones, lo que puede conducir a decisiones regulatorias inadecuadas o tardías. Este desfase entre desarrollo tecnológico y comprensión humana amplifica los riesgos de delegar funciones críticas a sistemas cuya lógica no es completamente transparente ni controlable.

En consecuencia, una regulación efectiva de la inteligencia artificial debe reconocer esta doble dimensión del riesgo. Por un lado, es necesario atender los problemas inmediatos que ya afectan derechos y mercados. Pero, por otro, resulta indispensable anticipar y contener aquellos riesgos estructurales que pueden comprometer la autonomía, la equidad y la sostenibilidad de las sociedades humanas. Ignorar esta segunda dimensión implicaría regular la superficie del problema, dejando intactas sus raíces más profundas.

I. Riesgos más comunes en el uso de la IA

La inteligencia artificial (IA) se ha consolidado como una de las tecnologías más disruptivas del siglo XXI, no solo por su capacidad de automatizar procesos, sino

por su potencial para transformar la manera en que las sociedades producen conocimiento, toman decisiones y estructuran sus dinámicas económicas y políticas. A diferencia de otras innovaciones tecnológicas, la IA no se limita a mejorar herramientas existentes, sino que interviene directamente en funciones tradicionalmente asociadas a la cognición humana, como el razonamiento, la creatividad y la interpretación del entorno.

Esta característica ha generado un cambio cualitativo en la forma en que se perciben los riesgos tecnológicos. Mientras que en el pasado los riesgos estaban asociados principalmente a fallas mecánicas o errores humanos, en el caso de la IA los riesgos se relacionan con sistemas capaces de operar con cierto grado de autonomía, aprender de datos históricos y generar resultados que pueden afectar directamente a individuos y comunidades. En consecuencia, el debate contemporáneo ha identificado una serie de riesgos inmediatos que requieren atención urgente.

Entre los riesgos más comúnmente mencionados se encuentran la vulneración de la propiedad intelectual, el uso indebido de datos personales, la proliferación de desinformación, el impacto en el empleo y la falta de transparencia en los sistemas algorítmicos. Estos riesgos destacan no solo por su frecuencia en el debate público, sino también por su impacto directo en derechos fundamentales y estructuras económicas.

El análisis de estos riesgos permite comprender las tensiones entre innovación tecnológica y protección social, así como los límites de los marcos normativos tradicionales. Como señala Larson (2021), existe una tendencia a sobreestimar las capacidades de la IA y a subestimar sus implicaciones, lo que dificulta una evaluación crítica de sus efectos reales. En este contexto, resulta fundamental examinar de manera rigurosa los riesgos más visibles, no como fenómenos aislados, sino como parte de un sistema tecnológico en expansión.

1. Riesgo de vulneración de la propiedad intelectual

La relación entre inteligencia artificial y propiedad intelectual constituye uno de los campos más conflictivos en el desarrollo tecnológico actual. Los sistemas de IA, especialmente los modelos generativos, requieren grandes volúmenes de datos para su entrenamiento. Estos datos suelen incluir obras protegidas por derechos de autor, lo que plantea interrogantes fundamentales sobre la legalidad y legitimidad de su uso.

Rodríguez (2020) señala que la IA opera mediante la reutilización masiva de información, lo que desafía las nociones tradicionales de autoría. A diferencia de los procesos creativos humanos, donde existe una intención consciente, los sistemas de IA generan contenido a partir de patrones estadísticos aprendidos. Esto dificulta establecer una relación directa entre el resultado y una fuente específica. Uno de los principales problemas radica en el uso de contenido sin consentimiento. Muchos modelos han sido entrenados utilizando bases de datos que incluyen textos, imágenes y obras artísticas sin autorización de sus creadores. Este hecho ha generado conflictos legales en múltiples jurisdicciones, donde artistas y autores han denunciado el uso indebido de su trabajo.

Además, surge la cuestión de la autoría del contenido generado. Cuando un sistema produce una imagen o un texto, ¿quién es el autor? Esta pregunta no tiene una respuesta clara en los marcos jurídicos actuales. Algunos argumentan que la autoría corresponde al usuario que introduce las instrucciones, mientras que otros consideran que el desarrollador del sistema debe ser responsable.

Otro aspecto crítico es la imitación de estilos. Los modelos generativos pueden reproducir el estilo de artistas específicos con gran precisión, lo que ha sido interpretado como una forma de apropiación. Este fenómeno ha generado preocupación en las industrias creativas, que ven amenazada su sostenibilidad

económica. Desde una perspectiva más amplia, este riesgo refleja una tensión estructural entre innovación y derechos. La IA necesita datos para funcionar, pero el uso de estos datos puede vulnerar derechos existentes. Como advierte Copeland (1993), el desarrollo tecnológico debe ser evaluado no solo por su eficiencia, sino también por sus implicaciones éticas.

2. Riesgo en la privacidad y protección de datos

La privacidad se ha convertido en uno de los ejes centrales en el debate sobre la inteligencia artificial. La dependencia de los sistemas de IA en grandes volúmenes de datos implica que gran parte de la información utilizada es de carácter personal, lo que genera riesgos significativos para los individuos. Uno de los principales problemas es la falta de consentimiento informado. En muchos casos, los usuarios no son conscientes de que sus datos están siendo recopilados y utilizados para entrenar modelos de IA. Esto plantea cuestiones éticas sobre la transparencia y la legitimidad de estos procesos, Gamero (2024) por ejemplo, destaca que la protección de datos se ha convertido en un pilar fundamental en algunos enfoques regulatorios, especialmente en Europa. Sin embargo, incluso en estos contextos, la implementación de estos principios enfrenta dificultades debido a la complejidad técnica de los sistemas.

Otro riesgo importante es la desidentificación de datos, porque, aunque los datos sean anonimizados, existen técnicas que permiten reconstruir la identidad de los individuos a partir de patrones, lo que aumenta el riesgo de filtraciones y uso indebido de información, asimismo, la integración de la IA en sistemas de vigilancia ha generado preocupaciones sobre el control social. Tecnologías como el reconocimiento facial permiten monitorear a la población de manera continua, lo que plantea desafíos en términos de derechos civiles. Sadin (2022) advierte que estas tecnologías pueden consolidar formas de poder basadas en el control de la información, lo que transforma la relación entre individuos y sistemas tecnológicos, los riesgos asociados a la privacidad no solo afectan a los individuos, sino que

también tienen implicaciones más amplias para la sociedad, ya que pueden alterar la forma en que se ejerce la libertad y la autonomía.

3. Riesgo de desinformación y manipulación mediante inteligencia artificial

3.1 Transformación del ecosistema informativo

Uno de los riesgos más visibles y preocupantes en el uso contemporáneo de la inteligencia artificial es su capacidad para transformar radicalmente la producción, distribución y consumo de información. A diferencia de etapas anteriores de la comunicación digital, donde la desinformación dependía en gran medida de la intervención humana, la IA permite automatizar la generación de contenido a gran escala, con niveles de calidad y verosimilitud sin precedentes. Los modelos generativos actuales son capaces de producir textos, imágenes, audios y videos que imitan con gran precisión la realidad. Esto ha dado lugar a un fenómeno particularmente problemático: los *deepfakes*. Estas representaciones sintéticas pueden reproducir la voz, el rostro o el comportamiento de una persona de manera convincente, dificultando la distinción entre lo real y lo artificial.

Este cambio implica una alteración profunda en el ecosistema informativo. La información deja de ser un reflejo más o menos fiable de la realidad para convertirse en un producto potencialmente fabricado, manipulable y replicable a gran escala. En este sentido, la IA no solo amplifica la desinformación existente, sino que introduce una nueva dimensión en la construcción de la realidad social, como señala Sadin (2022), la digitalización contemporánea no se limita a transmitir información, sino que redefine las condiciones mismas de producción de sentido. En este contexto, la inteligencia artificial se convierte en un actor central en la mediación entre los hechos y su interpretación.

3.2 Deepfakes y erosión de la confianza

El problema de los *deepfakes* no radica únicamente en la existencia de contenido falso, sino en su capacidad para erosionar la confianza en la información en general. Cuando cualquier imagen, audio o video puede ser manipulado con facilidad, la credibilidad de las evidencias visuales y sonoras se debilita. Este fenómeno genera lo que algunos autores han denominado “crisis de la evidencia”. Tradicionalmente, los registros audiovisuales han sido considerados pruebas fiables en contextos judiciales, periodísticos y sociales. Sin embargo, la proliferación de contenido sintético cuestiona esta premisa, creando un entorno donde la autenticidad se vuelve incierta. Además, los *deepfakes* pueden ser utilizados de manera estratégica para desacreditar a individuos o instituciones. Por ejemplo, la difusión de un video falso puede dañar la reputación de una persona, influir en procesos electorales o generar conflictos sociales. Incluso cuando el contenido es posteriormente desmentido, el daño ya puede haber sido causado, Larson (2021) también advierte que la sobre confianza en las capacidades tecnológicas puede llevar a subestimar sus implicaciones sociales. En el caso de los *deepfakes*, esto se traduce en una falta de preparación para enfrentar sus efectos en la percepción pública.

3.3 Automatización de la desinformación

Otro aspecto crítico es la capacidad de la IA para automatizar la producción de desinformación. A diferencia de campañas tradicionales, que requerían recursos humanos significativos, los sistemas actuales permiten generar grandes volúmenes de contenido en poco tiempo y con bajo costo, esto facilita la creación de redes de desinformación que pueden operar de manera coordinada y persistente. Estas redes pueden utilizarse para influir en la opinión pública, polarizar a la sociedad o generar confusión en torno a temas específicos.

La automatización también permite personalizar la desinformación. Mediante el análisis de datos, los sistemas pueden adaptar los mensajes a las características y preferencias de los individuos, aumentando su efectividad. Este fenómeno,

conocido como microsegmentación, amplifica el impacto de la desinformación, desde una perspectiva estructural, este riesgo no solo afecta a la calidad de la información, sino también a la cohesión social. La proliferación de narrativas contradictorias puede fragmentar el espacio público, dificultando la construcción de consensos.

3.3 Implicaciones políticas y sociales

La desinformación impulsada por IA tiene implicaciones particularmente graves en el ámbito político. La capacidad de manipular la percepción de los votantes puede afectar la integridad de los procesos democráticos. Esto incluye desde la difusión de noticias falsas hasta la creación de contenido que simula declaraciones de figuras públicas, pues la desinformación puede utilizarse para desestabilizar sociedades, generar desconfianza en las instituciones y exacerbar conflictos existentes.

En este sentido, la IA no solo representa un riesgo tecnológico, sino también geopolítico, Sadin (2022) sostiene que la expansión de tecnologías digitales está vinculada a nuevas formas de poder que operan a través de la información. La desinformación, en este contexto, se convierte en una herramienta de influencia que puede ser utilizada por actores estatales y no estatales.

3.4 Límites del control y complejidad del fenómeno

Uno de los principales desafíos asociados a este riesgo es la dificultad para controlarlo. La naturaleza descentralizada de internet, combinada con la velocidad de propagación de la información, hace que la detección y eliminación de contenido falso sea compleja y existen tensiones entre la regulación de la desinformación y la protección de la libertad de expresión. Cualquier intento de controlar el contenido puede ser percibido como una forma de censura, lo que limita las opciones de intervención.

Desde un punto de vista técnico, aunque se han desarrollado herramientas para detectar *deepfakes*, estas tecnologías suelen ir por detrás de los avances en generación de contenido. Esto crea una dinámica en la que la capacidad de producir desinformación supera la capacidad de detectarla.

4. Riesgo de impacto en el empleo y transformación del trabajo

4.1 Automatización y cambio en la naturaleza del trabajo

El impacto de la inteligencia artificial en el empleo constituye uno de los temas más discutidos en el debate contemporáneo. La capacidad de los sistemas de IA para automatizar tareas ha generado preocupaciones sobre la posible sustitución de trabajadores en diversos sectores, a diferencia de revoluciones tecnológicas anteriores, donde la automatización se concentraba en tareas manuales, la IA tiene la capacidad de reemplazar funciones cognitivas, esto incluye actividades como el análisis de datos, la redacción de textos y la toma de decisiones básicas. Lassalle (2024) señala que esta transformación no solo afecta al empleo, sino también a la estructura misma de la economía. La automatización de procesos cognitivos redefine el valor del trabajo humano, desplazando algunas habilidades y valorizando otras.

4.2 Desplazamiento laboral y vulnerabilidad

Uno de los principales riesgos es el desplazamiento laboral. A medida que las empresas adoptan tecnologías de IA, ciertos puestos de trabajo pueden volverse obsoletos. Esto afecta especialmente a trabajos repetitivos o basados en reglas, sin embargo, el impacto no es uniforme. Mientras algunos sectores experimentan pérdidas de empleo, otros pueden beneficiarse de nuevas oportunidades. Esta desigualdad en el impacto genera tensiones sociales y económicas.

Los trabajadores con menor acceso a educación tecnológica son particularmente vulnerables, y la falta de habilidades digitales puede dificultar su adaptación a un mercado laboral en transformación, aumentando el riesgo de exclusión.

4.3 Transformación de habilidades y redefinición del valor humano

Más allá del desplazamiento, la IA está transformando las habilidades necesarias en el mercado laboral. Las competencias técnicas, como la programación y el análisis de datos, adquieren mayor relevancia, mientras que otras habilidades pueden perder valor, al mismo tiempo, se revalorizan habilidades humanas como la creatividad, la empatía y el pensamiento crítico, sin embargo, estas habilidades no siempre son fácilmente transferibles o reconocidas en términos económicos. Larson (2021) advierte que existe una tendencia a sobreestimar la capacidad de la IA para replicar la inteligencia humana, lo que puede llevar a subestimar el valor de las habilidades humanas, percepción que puede influir en la manera en que se diseñan las políticas laborales.

4.4 Desigualdad y concentración de beneficios

Otro aspecto crítico es la concentración de los beneficios de la IA. Las empresas que desarrollan y controlan estas tecnologías pueden obtener ventajas significativas, mientras que los trabajadores enfrentan incertidumbre, este fenómeno puede ampliar las desigualdades económicas, tanto dentro de los países como a nivel global. Las economías con mayor capacidad tecnológica tienen más posibilidades de beneficiarse, mientras que otras pueden quedar rezagadas, también Lassalle (2024) señala que esta dinámica puede conducir a una reconfiguración del poder económico, donde el control de la tecnología se convierte en un factor determinante.

4.5 Impacto en la identidad y sentido del trabajo

El trabajo no solo tiene una dimensión económica, sino también social y psicológica. La automatización puede afectar la identidad de las personas, especialmente en contextos donde el trabajo es una fuente central de reconocimiento y sentido, la pérdida de empleo o la transformación de roles puede generar incertidumbre, estrés y desafección. Esto plantea desafíos no solo económicos, sino también sociales, Sadin (2022) advierte que la tecnología puede contribuir a una forma de

individualismo que debilita los vínculos sociales. En el contexto laboral, esto se traduce en una mayor precariedad y fragmentación.

4.6 Complejidad del fenómeno y dificultad de anticipación

Uno de los desafíos principales en relación con este riesgo es su carácter dinámico. La evolución de la IA hace difícil prever con precisión cuáles serán los efectos a largo plazo en el empleo, además, el impacto depende de múltiples factores, incluyendo políticas públicas, estructuras económicas y decisiones empresariales. Esto hace que el fenómeno sea complejo y difícil de abordar de manera uniforme.

5. Riesgo de falta de transparencia y explicabilidad en los sistemas de inteligencia artificial

5.1 La opacidad como característica estructural

Uno de los riesgos más relevantes y persistentes en el desarrollo de la inteligencia artificial es la falta de transparencia en el funcionamiento de los sistemas. Muchos modelos actuales, especialmente aquellos basados en aprendizaje profundo (*deep learning*), operan como lo que comúnmente se denomina “cajas negras”. Esto significa que, aunque es posible observar los datos de entrada y los resultados de salida, los procesos internos que conducen a dichas decisiones no son fácilmente comprensibles ni siquiera para sus propios desarrolladores.

Esta opacidad no es un accidente, sino una consecuencia directa de la complejidad de los modelos, ya que los sistemas de aprendizaje automático funcionan mediante redes neuronales con múltiples capas y millones (o incluso miles de millones) de parámetros, estas estructuras generan relaciones internas altamente complejas que no pueden ser interpretadas de manera intuitiva, Copeland (1993) por ejemplo, ya advertía que uno de los principales desafíos de la inteligencia artificial es precisamente la comprensión de sus procesos internos. A medida que los sistemas se vuelven más sofisticados, esta dificultad se incrementa, generando una brecha entre el funcionamiento técnico y la comprensión humana.

5.2 Problemas de confianza y legitimidad

La falta de explicabilidad tiene implicaciones directas en la confianza que las personas depositan en los sistemas de IA. Cuando un algoritmo toma una decisión que afecta a un individuo —por ejemplo, la aprobación de un crédito, la selección en un proceso de contratación o la evaluación en un sistema educativo—, es fundamental que exista una justificación comprensible, no obstante, en muchos casos, las decisiones algorítmicas no pueden ser explicadas de manera clara. Esto genera incertidumbre y desconfianza, especialmente cuando los resultados son negativos o inesperados.

La confianza en los sistemas tecnológicos no depende únicamente de su precisión, sino también de su transparencia. Un sistema altamente preciso pero incomprensible puede ser percibido como arbitrario o injusto, en este sentido, la opacidad no solo es un problema técnico, sino también social. Larson (2021) señala que existe una tendencia a atribuir a los sistemas de IA una racionalidad que en realidad no poseen, esta percepción puede llevar a aceptar decisiones sin cuestionarlas, lo que incrementa el riesgo de dependencia acrítica.

5.3 Dificultades en la asignación de responsabilidad

Uno de los efectos más problemáticos de la falta de transparencia es la dificultad para asignar responsabilidad. En los sistemas tradicionales, cuando ocurre un error, es posible identificar al responsable: una persona, una institución o un proceso específico. Sin embargo, en el caso de la IA, la cadena de responsabilidad se vuelve difusa, cuando un sistema algorítmico comete un error, ¿quién es responsable? ¿El desarrollador que diseñó el modelo? ¿La empresa que lo implementó? ¿El usuario que lo utilizó? ¿O el propio sistema?, esta ambigüedad genera lo que algunos autores denominan “vacío de responsabilidad”. La incapacidad de identificar claramente a un responsable dificulta la reparación de daños y debilita los mecanismos de rendición de cuentas. Sadin (2022) advierte que esta situación puede favorecer la desresponsabilización institucional, donde las decisiones se

atribuyen a sistemas tecnológicos en lugar de a actores humanos, esto no solo reduce la accountability, sino que también puede erosionar la confianza en las instituciones.

5.4 Sesgos ocultos y decisiones injustas

La opacidad de los sistemas de IA también dificulta la identificación de sesgos. Los modelos aprenden a partir de datos históricos, los cuales pueden contener prejuicios y desigualdades existentes en la sociedad. Si estos sesgos no son detectados, pueden ser reproducidos e incluso amplificados por los sistemas, por ejemplo, un sistema de selección de personal entrenado con datos históricos puede favorecer a ciertos grupos sobre otros, replicando patrones discriminatorios, si el funcionamiento del modelo no es transparente, estos sesgos pueden pasar desapercibidos.

Este problema es particularmente grave porque los sistemas de IA tienden a ser percibidos como objetivos y neutrales, así lo señala Lassalle (2024), la tecnología nunca es completamente neutral, ya que refleja las condiciones sociales en las que se desarrolla, la falta de explicabilidad impide cuestionar estas decisiones, lo que puede llevar a la normalización de prácticas injustas, en este sentido, la opacidad no solo oculta errores, sino que también puede legitimar desigualdades.

5.5 Complejidad técnica y límites de la explicabilidad

Aunque la explicabilidad es un objetivo deseable, su implementación enfrenta limitaciones técnicas. En algunos casos, simplificar el funcionamiento de un modelo para hacerlo comprensible puede implicar una pérdida de precisión. Esto genera una tensión entre rendimiento y transparencia, además, no todas las explicaciones son igualmente útiles. Una explicación técnica detallada puede ser incomprensible para la mayoría de los usuarios, mientras que una explicación simplificada puede no reflejar adecuadamente el funcionamiento del sistema, Copeland (1993) sugiere que la comprensión de los sistemas de IA requiere un equilibrio entre formalización

y accesibilidad, este equilibrio es difícil de alcanzar, especialmente en contextos donde los usuarios no tienen formación técnica.

5.6 Dependencia tecnológica y pérdida de autonomía

La falta de transparencia también contribuye a la dependencia tecnológica. Cuando las personas no comprenden cómo funcionan los sistemas, tienden a confiar en ellos de manera automática, esto puede reducir la capacidad crítica y la autonomía en la toma de decisiones, en contextos donde la IA se utiliza de manera intensiva, como en la medicina o las finanzas, esta dependencia puede tener consecuencias significativas y la delegación de decisiones en sistemas opacos puede limitar la capacidad de intervención humana.

Sadin (2022) advierte que esta tendencia puede conducir a una forma de subordinación tecnológica, donde las decisiones humanas son reemplazadas por procesos algorítmicos sin suficiente supervisión.

II. Riesgos menos visibles, pero más peligrosos en el uso de la IA.

Si bien gran parte del debate contemporáneo sobre la inteligencia artificial se ha centrado en riesgos inmediatos —como la privacidad, la propiedad intelectual o la desinformación— existe una segunda capa de análisis que resulta significativamente más crítica. Esta dimensión no se enfoca en los efectos visibles y de corto plazo, sino en las transformaciones estructurales que la IA puede generar en la organización de la sociedad, la distribución del poder y la propia condición humana, estos riesgos, a diferencia de los anteriores, no siempre se manifiestan de forma evidente o inmediata. Sin embargo, su potencial de impacto es mucho mayor, ya que afectan los fundamentos sobre los cuales se construyen las instituciones, la autonomía individual y la vida colectiva, en este sentido, no se trata simplemente de problemas técnicos o regulatorios, sino de desafíos de carácter civilizatorio.

Como señala Lassalle (2024), el desarrollo de la inteligencia artificial puede conducir hacia una “civilización artificial” en la que los sistemas algorítmicos no solo

asisten al ser humano, sino que comienzan a sustituirlo en funciones clave. Este proceso implica una reconfiguración profunda de la relación entre tecnología y humanidad, donde la capacidad de decisión puede desplazarse progresivamente hacia sistemas automatizado, por su parte Sadin (2022) advierte que la expansión de las tecnologías digitales está vinculada a nuevas formas de poder que operan a través de la automatización, la predicción y el control de la conducta, con ello la inteligencia artificial no solo es una herramienta, sino un dispositivo que reestructura las dinámicas sociales y políticas y en esta perspectiva existen múltiples y graves riesgos a saber.

1. Riesgo de pérdida de control humano sobre sistemas autónomos

1.1 Naturaleza del problema

Uno de los riesgos más debatidos en el ámbito de la inteligencia artificial avanzada es la posibilidad de que los sistemas adquieran un grado de autonomía que supere la capacidad humana de supervisión y control, riesgo que no implica necesariamente que las máquinas desarrollen conciencia o intencionalidad, sino que su complejidad operativa y su capacidad de aprendizaje generen comportamientos difíciles de predecir o corregir.

Los sistemas de IA actuales, especialmente aquellos basados en aprendizaje automático, funcionan mediante la identificación de patrones en grandes volúmenes de datos, a partir de estos patrones, generan decisiones o recomendaciones que pueden influir en procesos críticos y a medida que estos sistemas se vuelven más complejos, la capacidad de comprender su funcionamiento disminuye.

Larson (2021) advierte que existe una brecha significativa entre el rendimiento de los sistemas de IA y la comprensión humana de sus procesos internos. Esta brecha

se amplía a medida que los modelos incorporan más variables y relaciones, lo que dificulta anticipar sus comportamientos en contextos no previstos.

1.2 Automatización de decisiones críticas

El riesgo de pérdida de control se intensifica cuando la inteligencia artificial se integra en sistemas que operan en entornos críticos, lo que incluye sectores como la salud, las finanzas, el transporte y la seguridad. En estos contextos, las decisiones tomadas por sistemas de IA pueden tener consecuencias directas sobre la vida de las personas, por ejemplo, un error en un sistema médico automatizado puede afectar un diagnóstico, mientras que una falla en un sistema financiero puede desencadenar crisis económicas.

La automatización de decisiones críticas implica una transferencia de responsabilidad desde los humanos hacia los sistemas. Sin embargo, esta transferencia no siempre está acompañada de mecanismos adecuados de supervisión, lo que incrementa el riesgo de errores sistémicos, Copeland (1993) señala que la delegación de funciones cognitivas a las máquinas debe ser evaluada con cautela, especialmente cuando estas funciones tienen implicaciones éticas o sociales. La automatización no elimina la responsabilidad, sino que la redistribuye de manera compleja.

1.2 Escalamiento de errores y efectos sistémicos

Uno de los aspectos más preocupantes de la pérdida de control es la capacidad de los sistemas de IA para escalar errores de manera rápida y amplia, a diferencia de los errores humanos, que suelen ser limitados en alcance, los errores algorítmicos pueden propagarse a gran escala en poco tiempo y esto es especialmente relevante en sistemas interconectados, donde múltiples algoritmos interactúan entre sí, en estos entornos, un error en un componente puede generar efectos en cadena, afectando a todo el sistema, Lassalle (2024) advierte que la creciente dependencia de sistemas automatizados puede generar vulnerabilidades estructurales, donde fallas tecnológicas tienen consecuencias desproporcionadas.

Este fenómeno se agrava cuando los sistemas operan con poca supervisión humana.

1.4 Dificultad de intervención humana

Otro aspecto crítico es la dificultad de intervenir en sistemas altamente automatizados, a medida que los sistemas operan de manera más autónoma, la capacidad de los humanos para detener o modificar su funcionamiento puede verse limitada, lo cual puede deberse a múltiples factores:

- Complejidad técnica
- Velocidad de operación
- Falta de comprensión del sistema

En situaciones donde se requiere una respuesta rápida, la intervención humana puede ser insuficiente. Esto plantea preguntas sobre la conveniencia de delegar ciertas funciones a sistemas automatizados, Sadin (2022) sugiere que la confianza excesiva en la tecnología puede llevar a una forma de dependencia que reduce la capacidad de acción humana. En este sentido, la pérdida de control no es solo un problema técnico, sino también cultural.

1.5 Autonomía sin comprensión

Es importante destacar que la autonomía de los sistemas de IA no implica comprensión. Los modelos actuales no “entienden” el mundo en el sentido humano, sino que operan mediante correlaciones estadísticas. Sin embargo, esta limitación no impide que sus decisiones tengan efectos reales, Larson (2021) critica la idea de que la inteligencia artificial esté cerca de replicar la inteligencia humana. Según el autor, esta percepción puede generar una falsa sensación de seguridad, donde se asume que los sistemas son más capaces de lo que realmente son y esta combinación de autonomía operativa y falta de comprensión constituye uno de los

riesgos más importantes, ya que puede llevar a confiar en sistemas que no están diseñados para manejar situaciones complejas o ambiguas.

2. Riesgo de concentración extrema de poder tecnológico

2.1 Centralización del conocimiento y la capacidad tecnológica

La inteligencia artificial requiere recursos significativos para su desarrollo, incluyendo datos, infraestructura computacional y talento especializado. Estos recursos no están distribuidos de manera equitativa, lo que ha llevado a una concentración de capacidades en un número reducido de actores. Las grandes empresas tecnológicas y algunos estados tienen acceso privilegiado a estos recursos, lo que les permite desarrollar sistemas de IA más avanzados, dicha concentración genera una asimetría de poder que puede tener implicaciones profundas para la sociedad. Lassalle (2024) señala que la IA se ha convertido en un elemento central en la geopolítica contemporánea, donde el control de la tecnología se traduce en poder económico y político.

2.2 Monopolios tecnológicos y dependencia estructural

La concentración de poder en el ámbito de la IA puede conducir a la formación de monopolios tecnológicos, estas entidades no solo controlan el desarrollo de la tecnología, sino también su acceso y uso, lo que genera una dependencia estructural, donde individuos, empresas y gobiernos dependen de sistemas desarrollados por un número reducido de actores, dependencia que puede limitar la autonomía y la capacidad de decisión. Sadin (2022) advierte que este fenómeno puede conducir a una forma de “totalitarismo suave”, donde el control no se ejerce de manera explícita, sino a través de la infraestructura tecnológica.

2.3 Control de datos y comportamiento

El poder en la era de la IA no se basa únicamente en la tecnología, sino también en el control de los datos, los datos permiten entrenar modelos, pero también comprender y predecir el comportamiento humano. Las plataformas digitales recopilan grandes cantidades de información sobre los usuarios, lo que les permite

influir en sus decisiones. Este control del comportamiento plantea cuestiones éticas y políticas.

2.4 Implicaciones democráticas

La concentración de poder tecnológico tiene implicaciones directas para la democracia. Cuando el control de la información y la tecnología está en manos de pocos actores, se reduce la capacidad de los ciudadanos para influir en las decisiones, además, la dependencia de sistemas tecnológicos puede limitar la transparencia y la rendición de cuentas.

3. Riesgo de automatización de la desigualdad estructural

3.1 Reproducción de patrones históricos

Uno de los riesgos más profundos de la inteligencia artificial es su capacidad para reproducir y amplificar desigualdades existentes. A diferencia de los sistemas humanos, que pueden cuestionar o corregir sesgos, los sistemas de IA aprenden a partir de datos históricos, los cuales reflejan estructuras sociales previas, incluyendo discriminaciones y asimetrías, esto implica que la IA no solo replica la realidad, sino que puede consolidarla. Si un modelo es entrenado con datos que contienen sesgos —por ejemplo, en procesos de contratación, acceso a crédito o decisiones judiciales—, es probable que reproduzca estos patrones en sus resultados. Larson (2021) advierte que los sistemas de IA no poseen capacidad de inferencia crítica en el sentido humano, lo que limita su capacidad para cuestionar los datos que procesan. En consecuencia, los modelos pueden perpetuar inequidades sin reconocerlas como tales.

3.2 Naturalización de la desigualdad

Un aspecto particularmente problemático es la tendencia a percibir las decisiones algorítmicas como objetivas. A diferencia de las decisiones humanas, que pueden ser cuestionadas por su subjetividad, los resultados de la IA suelen ser interpretados como neutrales o basados en evidencia, esta percepción contribuye a la naturalización de la desigualdad, cuando un sistema automatizado toma una

decisión discriminatoria, esta puede ser aceptada como legítima, precisamente porque proviene de un proceso tecnológico, la tecnología puede adquirir un carácter normativo, donde sus resultados son percibidos como estándares de verdad. En este contexto, la desigualdad deja de ser visible como problema y pasa a ser considerada una consecuencia “natural” de los datos (Lasalle, 2024).

3.3 Escalabilidad de la discriminación

Uno de los factores que agrava este riesgo es la capacidad de los sistemas de IA para operar a gran escala. Mientras que las decisiones humanas están limitadas en su alcance, los sistemas automatizados pueden aplicar los mismos criterios a miles o millones de casos simultáneamente, esto significa que un sesgo, incluso si es pequeño, puede tener un impacto significativo cuando se aplica de manera masiva, por ejemplo, un sistema de selección de personal que favorezca ligeramente a ciertos perfiles puede generar una exclusión sistemática de otros grupos, la automatización no elimina las desigualdades, sino que puede intensificarlas al hacerlas más eficientes y menos visibles (Sadín, 2022).

3.4 Desigualdad algorítmica y acceso a oportunidades

La inteligencia artificial también influye en la distribución de oportunidades. Los sistemas de recomendación, por ejemplo, determinan qué contenidos se muestran a los usuarios, qué productos se ofrecen o incluso qué oportunidades laborales se presentan, esto puede generar lo que se denomina “burbujas algorítmicas”, donde las personas son expuestas a información y oportunidades limitadas por sus perfiles, la IA no solo refleja la desigualdad, sino que también contribuye a su reproducción, además, el acceso a tecnologías avanzadas no está distribuido de manera equitativa, las empresas y países con mayores recursos tienen más capacidad para desarrollar y utilizar IA, lo que amplía las brechas existentes.

3.5 Impacto en grupos vulnerables

Los efectos de la automatización de la desigualdad no son homogéneos. Los grupos históricamente marginados son los más afectados por estos sistemas, ya que los datos que los representan suelen ser incompletos o sesgados, esto puede generar situaciones donde las decisiones algorítmicas afectan de manera desproporcionada a ciertos grupos, limitando su acceso a servicios, oportunidades o derechos, la tecnología debe ser evaluada no solo por su eficiencia, sino también por su impacto social, en el caso de la IA, este impacto incluye la forma en que se distribuyen los beneficios y los riesgos.

3.6 Dificultad de corrección

Otro desafío importante es la dificultad de corregir estos sesgos. Una vez que un sistema ha sido entrenado y desplegado, modificar su comportamiento puede requerir cambios significativos en los datos o en el modelo, además, identificar los sesgos no siempre es sencillo, especialmente en sistemas complejos. Esto implica que la desigualdad puede persistir incluso cuando existe la intención de corregirla.

4. Riesgo de erosión de capacidades humanas

4.1 Dependencia cognitiva

Uno de los riesgos menos visibles, pero más profundos de la inteligencia artificial es su impacto en las capacidades humanas. A medida que los sistemas automatizan funciones cognitivas, existe el riesgo de que las personas reduzcan su participación en estos procesos, esto puede generar una dependencia cognitiva, donde los individuos delegan tareas como la memoria, el análisis o la toma de decisiones en sistemas tecnológicos, aunque esta delegación puede ser eficiente, también puede debilitar las habilidades humanas y la inteligencia humana no puede reducirse a procesos computacionales, y que la dependencia excesiva de la IA puede limitar el desarrollo de capacidades críticas (Larssons, 2024).

4.2 Reducción del pensamiento crítico

La automatización de la información también afecta el pensamiento crítico. Cuando los sistemas generan respuestas de manera inmediata y aparentemente coherente, los usuarios pueden aceptar estos resultados sin cuestionarlos, esto reduce la necesidad de analizar, contrastar o interpretar la información, lo que puede afectar la capacidad de juicio, en contextos educativos, este fenómeno puede tener consecuencias significativas en la formación de habilidades, la tecnología puede fomentar una relación pasiva con el conocimiento, donde los individuos consumen información en lugar de producirla o cuestionarla (Sadín, 2022).

4.3 Externalización de la memoria y el conocimiento

La inteligencia artificial también contribuye a la externalización de la memoria. A medida que los sistemas almacenan y procesan información, los individuos dependen menos de su propia memoria, este fenómeno no es nuevo, pero la IA lo amplifica, la disponibilidad constante de información reduce la necesidad de retener conocimientos, lo que puede afectar la capacidad de aprendizaje, Lassalle (2024) sugiere que esta transformación puede alterar la relación entre conocimiento y experiencia, donde la acumulación de información pierde relevancia frente a su acceso inmediato.

4.4 Transformación de la creatividad

La IA también está modificando la creatividad humana. Los sistemas generativos pueden producir textos, imágenes y música, lo que plantea preguntas sobre el papel del ser humano en la creación, si bien la IA puede ser una herramienta para potenciar la creatividad, también existe el riesgo de que sustituya procesos creativos, reduciendo la participación humana, además, la disponibilidad de contenido generado puede influir en los estándares estéticos y culturales, homogenizando la producción artística.

4.5 Impacto en la autonomía

La erosión de capacidades humanas tiene implicaciones directas en la autonomía. Cuando las decisiones son delegadas a sistemas automatizados, la capacidad de los individuos para actuar de manera independiente se reduce, esto no implica una pérdida total de autonomía, pero sí una transformación en su ejercicio. La dependencia de sistemas tecnológicos puede limitar las opciones disponibles y condicionar el comportamiento, esta dinámica puede conducir a una forma de subordinación, donde las decisiones humanas están mediadas por sistemas que no son completamente comprendidos, Sadin (2022).

4.6 Implicaciones culturales y sociales

El impacto de la IA en las capacidades humanas no se limita al individuo, sino que también afecta a la cultura, la forma en que las sociedades producen conocimiento, arte y significado puede cambiar significativamente, lo que genera preguntas sobre la identidad humana en un contexto donde las máquinas participan activamente en procesos tradicionalmente humanos.

5. Interrelación de los riesgos estructurales

5.1 De riesgos aislados a un sistema de riesgos

Los riesgos analizados —pérdida de control humano, concentración de poder, automatización de la desigualdad y erosión de capacidades humanas— no deben entenderse como fenómenos independientes, por el contrario, constituyen un sistema interconectado donde cada uno refuerza y amplifica a los demás, la pérdida de control humano, por ejemplo, se ve intensificada en contextos donde el poder tecnológico está altamente concentrado, si los sistemas más avanzados están en manos de pocos actores, la capacidad de supervisión y corrección se reduce aún más, de manera similar, la automatización de la desigualdad se ve facilitada por la opacidad de los sistemas y la falta de capacidades críticas en los usuarios. Lassalle (2024) sugiere que la inteligencia artificial debe analizarse como parte de un proceso más amplio de reorganización social, donde las tecnologías no solo

introducen cambios puntuales, sino que reconfiguran las estructuras de poder, en este sentido, los riesgos no son acumulativos, sino sistémicos.

5.2 Retroalimentación entre tecnología y sociedad

Otro elemento clave es la relación de retroalimentación entre la tecnología y la sociedad, los sistemas de IA no operan en un vacío, sino que interactúan con contextos sociales específicos, los datos que utilizan reflejan realidades sociales, y sus resultados, a su vez, influyen en estas realidades, este ciclo de retroalimentación puede amplificar los riesgos, por ejemplo, un sistema que reproduce desigualdades puede generar datos que refuercen estas mismas desigualdades, creando un bucle difícil de romper. Sadin (2022) advierte que las tecnologías digitales tienden a intensificar dinámicas existentes, en lugar de transformarlas de manera neutral. Esto implica que los riesgos estructurales no son simplemente efectos colaterales, sino consecuencias inherentes al modo en que la tecnología se integra en la sociedad.

5.3 Dependencia estructural y pérdida de capacidad colectiva

La interrelación de los riesgos también se manifiesta en la creciente dependencia de la sociedad hacia los sistemas tecnológicos, esta dependencia no es solo individual, sino colectiva, las instituciones, las economías y los sistemas políticos dependen cada vez más de la inteligencia artificial para su funcionamiento y la dependencia puede limitar la capacidad de acción colectiva, cuando las decisiones están mediadas por sistemas automatizados, la posibilidad de intervención humana se reduce, lo que puede afectar la gobernanza democrática, la confianza en la tecnología debe ir acompañada de una comprensión crítica de sus limitaciones. Sin embargo, en muchos casos, esta comprensión es insuficiente, lo que aumenta la vulnerabilidad de los sistemas sociales (Larsson, 2021).

6. Diferencias entre riesgos comunes y riesgos estructurales

6.1 Visibilidad vs profundidad

Una de las principales diferencias entre los riesgos comunes y los estructurales radica en su visibilidad. Los riesgos comunes —como la privacidad o la desinformación— son fácilmente identificables y generan respuestas inmediatas, en cambio, los riesgos estructurales operan de manera más silenciosa y progresiva, lo que no significa que los riesgos comunes sean menos importantes, sino que su naturaleza permite una intervención más directa, por el contrario, los riesgos estructurales requieren un análisis más profundo y una comprensión de largo plazo.

6.2 Temporalidad del impacto

Los riesgos comunes suelen tener efectos a corto o mediano plazo, mientras que los riesgos estructurales se desarrollan a lo largo del tiempo. La pérdida de capacidades humanas, por ejemplo, no ocurre de manera inmediata, sino como resultado de procesos acumulativos. Lassalle (2024) advierte que estos procesos pueden pasar desapercibidos hasta que sus efectos se vuelven irreversibles. Esto plantea un desafío importante para la anticipación y la planificación.

6.3 Nivel de intervención

Los riesgos comunes pueden ser abordados mediante regulaciones específicas, mientras que los riesgos estructurales requieren cambios más profundos en la forma en que se concibe y utiliza la tecnología, lo que no implica que la regulación de la inteligencia artificial no pueda limitarse a soluciones técnicas, sino que debe incluir una reflexión sobre los valores y objetivos que guían su desarrollo.

7. Implicaciones para la humanidad

7.1 Reconfiguración de la autonomía

Uno de los efectos más significativos de los riesgos estructurales es la transformación de la autonomía humana, la delegación de decisiones en sistemas automatizados modifica la forma en que los individuos ejercen su libertad, esto no

implica necesariamente una pérdida total de autonomía, pero sí una reconfiguración de su ejercicio, las decisiones humanas pasan a estar mediadas por sistemas que no siempre son comprendidos ni controlados. Sadin (2022) advierte que esta mediación puede conducir a una forma de dependencia donde la capacidad de acción se ve condicionada por la lógica algorítmica.

7.2 Transformación de la democracia

La concentración de poder tecnológico y el control de la información tienen implicaciones directas para la democracia, cuando los sistemas que gestionan la información y las decisiones están en manos de pocos actores, la participación ciudadana puede verse limitada, además, la opacidad de los sistemas dificulta la rendición de cuentas, lo que debilita los mecanismos democráticos.

7.3 Cambio en la condición humana

Más allá de los aspectos políticos y económicos, los riesgos estructurales plantean preguntas sobre la condición humana, la inteligencia artificial no solo modifica lo que hacemos, sino también cómo pensamos, cómo aprendemos y cómo nos relacionamos, Lassalle (2024) sugiere que estamos ante una transición hacia una nueva etapa en la que la relación entre humanos y tecnología se redefine de manera profunda. El análisis de los riesgos estructurales de la inteligencia artificial permite comprender que los desafíos asociados a esta tecnología van más allá de los problemas visibles que dominan el debate contemporáneo. Si bien los riesgos comunes —como la privacidad o la desinformación— requieren atención inmediata, los riesgos estructurales representan una amenaza más profunda y duradera.

Estos riesgos no se manifiestan necesariamente como crisis inmediatas, sino como transformaciones progresivas que alteran las bases de la vida social. La pérdida de control humano, la concentración de poder, la automatización de la desigualdad y la erosión de capacidades humanas configuran un escenario en el que la inteligencia artificial puede redefinir la relación entre tecnología y humanidad.

III. La regulación de la inteligencia artificial

La regulación de la inteligencia artificial (IA) se ha convertido en uno de los desafíos centrales del siglo XXI, no solo por la velocidad de su desarrollo, sino por la naturaleza dual de los riesgos que genera, a diferencia de otras tecnologías, la IA produce efectos simultáneos en dos niveles: por un lado, riesgos inmediatos y tangibles que afectan derechos individuales y mercados; por otro, riesgos estructurales que pueden transformar la organización de la sociedad y la condición humana. Esta doble dimensión obliga a repensar los marcos regulatorios tradicionales. Como señala el enfoque europeo, el objetivo no es únicamente garantizar el funcionamiento del mercado, sino asegurar una IA “centrada en el ser humano” y compatible con los derechos fundamentales, sin embargo, este enfoque, aunque avanzado, enfrenta tensiones importantes cuando se confronta con los riesgos más profundos.

Desde una perspectiva crítica, puede afirmarse que gran parte de las iniciativas regulatorias actuales tienden a concentrarse en los riesgos visibles —privacidad, propiedad intelectual, transparencia— mientras que los riesgos estructurales reciben menor atención, esto genera una brecha entre lo que se regula y lo que realmente está en juego, la gobernanza de la IA requiere una aproximación ética integral que no se limite a mitigar daños inmediatos, sino que anticipe impactos sistémicos Floridi y otros. (2018), en la misma línea, Bostrom (2014) sostiene que el desarrollo de sistemas avanzados plantea riesgos que no pueden ser abordados únicamente mediante marcos regulatorios convencionales.

En este contexto, la regulación de la IA debe entenderse como un proceso multidimensional que articule distintos niveles de intervención. No se trata solo de controlar la tecnología, sino de orientar su desarrollo en función de valores sociales.

1. Primera dimensión: regulación de riesgos inmediatos

3.1 Enfoque predominante en la regulación actual

Las regulaciones existentes tienden a concentrarse en los riesgos inmediatos, es decir, aquellos que afectan directamente a los individuos y pueden ser identificados con relativa claridad. Este enfoque responde a la necesidad de generar respuestas rápidas frente a problemas concretos, como la protección de datos o la desinformación.

La “Ley de Inteligencia Artificial de la Unión Europea (2024)” representa el ejemplo más avanzado de este enfoque. Su modelo basado en niveles de riesgo establece obligaciones diferenciadas según el impacto potencial de los sistemas, este enfoque permite priorizar la regulación de aplicaciones consideradas de “alto riesgo”, como aquellas utilizadas en salud, educación o justicia, sin embargo, este modelo también refleja una lógica reactiva, donde la regulación se construye en función de aplicaciones específicas, en lugar de abordar la tecnología en su conjunto.

3.2 Protección de datos y derechos fundamentales

Uno de los pilares de la regulación contemporánea es la protección de datos personales, organismos como la UNESCO han enfatizado la necesidad de garantizar la privacidad como condición para el desarrollo ético de la IA .

Este enfoque se basa en principios como:

- consentimiento informado
- transparencia
- minimización de datos

No obstante, diversos autores han señalado las limitaciones de este enfoque (Zuboff, 2019), la economía digital se dice, ha generado un modelo de “capitalismo de vigilancia” donde la recolección de datos es estructural, lo que dificulta su regulación efectiva, en este sentido, la protección de datos, aunque necesaria,

resulta insuficiente para abordar las dinámicas más profundas del poder tecnológico.

3.3 Propiedad intelectual y economía creativa.

Otro eje central de regulación es la propiedad intelectual. La proliferación de modelos generativos ha generado conflictos en torno al uso de datos para entrenamiento y la autoría de contenido, Lessig (2006) ya advertía que el entorno digital desafía las estructuras tradicionales de propiedad, lo que se intensifica con la IA y actualmente, las discusiones giran en torno a:

- derechos de autor sobre datos de entrenamiento
- propiedad del contenido generado
- compensación a creadores

Sin embargo, estas discusiones suelen limitarse a aspectos jurídicos, sin abordar el impacto estructural en la producción cultural.

3.4 Transparencia y rendición de cuentas

La transparencia es otro elemento clave en las regulaciones actuales. El objetivo es garantizar que los sistemas de IA sean comprensibles y auditables. El modelo europeo incluye requisitos de explicabilidad para sistemas de alto riesgo, mientras que iniciativas como “Los Principios de Chapultepec” destacan la importancia de la rendición de cuentas en el uso de la IA, no obstante, como señala Pasquale (2015), la opacidad de los sistemas algorítmicos no es solo técnica, sino también política y la falta de transparencia puede ser utilizada como mecanismo de poder.

3.5 Limitaciones del enfoque actual

A pesar de sus avances, la regulación centrada en riesgos inmediatos presenta limitaciones importantes:

- i. Es reactiva y no anticipa riesgos futuros
- ii. Se enfoca en aplicaciones, no en estructuras
- iii. Subestima las dinámicas de poder

Brynjolfsson y McAfee (2014) señalan que las transformaciones tecnológicas suelen superar la capacidad de adaptación institucional, lo que genera desajustes entre regulación y realidad.

3. Segunda dimensión: regulación de riesgos estructurales

3.1 Un vacío en la gobernanza actual

Los riesgos estructurales —como la pérdida de control humano o la concentración de poder— reciben menos atención en los marcos regulatorios actuales. Esto se debe, en parte, a su complejidad y a la dificultad de traducirlos en normas concretas. Sin embargo, su impacto potencial es mucho mayor. Como advierte Bostrom (2014), los riesgos asociados a sistemas avanzados pueden ser de carácter existencial, lo que requiere una aproximación preventiva.

3.2 Gobernanza del poder tecnológico

Uno de los principales desafíos es la concentración de poder en manos de grandes corporaciones tecnológicas. Este fenómeno no está adecuadamente abordado por las regulaciones actuales, Srnicek (2017) describe cómo las plataformas digitales han consolidado modelos económicos basados en el control de datos y la infraestructura, lo que se amplifica con la IA.

En este contexto, la regulación debería incluir:

- mecanismos antimonopolio
- control sobre infraestructuras críticas
- soberanía tecnológica

Sin embargo, estos elementos suelen quedar fuera de las regulaciones específicas de IA.

3.3 Autonomía humana y dependencia tecnológica

Otro aspecto crítico es la erosión de la autonomía humana. Las regulaciones actuales reconocen la importancia de la supervisión humana, pero no abordan

plenamente el problema de la dependencia tecnológica, Haidt (2024) muestra cómo las tecnologías digitales pueden afectar las capacidades cognitivas y sociales, lo que tiene implicaciones profundas para la autonomía individual, esto plantea la necesidad de una regulación que no solo proteja derechos, sino que también preserve capacidades humanas.

3.4 Desigualdad y justicia social

La automatización de la desigualdad es otro riesgo estructural que requiere atención. Aunque algunos marcos éticos lo reconocen, su integración en la regulación es limitada. El decálogo de la UAM, por ejemplo, enfatiza la importancia de la equidad y la inclusión, lo que representa un avance en la integración de estos principios, sin embargo, su implementación práctica sigue siendo un desafío.

3.5 Hacia una regulación integral

La coexistencia de ambas dimensiones de riesgo sugiere la necesidad de un enfoque integral. Como señala Floridi y otros. (2018), la ética de la IA debe articular principios técnicos, sociales y políticos, esto implica:

- regular riesgos inmediatos
- anticipar riesgos estructurales
- integrar múltiples niveles de gobernanza

4. Análisis crítico de marcos regulatorios existentes

a. El modelo europeo: liderazgo normativo con límites estructurales

El Reglamento de Inteligencia Artificial de la Unión Europea (2024) constituye el esfuerzo más avanzado para regular la IA a nivel global. Su enfoque basado en riesgos establece categorías (riesgo inaceptable, alto, limitado y mínimo) que permiten diferenciar niveles de intervención. Este modelo representa un avance significativo en términos de seguridad jurídica, protección de derechos fundamentales y armonización normativa, sin embargo, este liderazgo normativo también presenta limitaciones importantes, en primer lugar, el modelo europeo se

centra predominantemente en los riesgos inmediatos, especialmente aquellos relacionados con el uso de la IA en contextos específicos (por ejemplo, sistemas biométricos o aplicaciones en servicios públicos), esto implica que la regulación tiende a abordar casos concretos, pero no necesariamente las dinámicas estructurales del poder tecnológico.

En segundo lugar, existe una tensión entre regulación e innovación, este enfoque europeo puede generar desventajas competitivas frente a regiones con marcos más flexibles. Esta tensión refleja un dilema central: regular demasiado puede frenar el desarrollo tecnológico, pero regular poco puede amplificar los riesgos (Casado, 2024). Desde una perspectiva crítica, puede afirmarse que el modelo europeo representa una forma de “regulación técnica avanzada”, pero aún insuficiente para abordar los riesgos civilizatorios de la IA.

b. UNESCO: ética global sin mecanismos coercitivos

La UNESCO ha desarrollado uno de los marcos más completos en términos éticos, destacando principios como la equidad, la inclusión, la transparencia y la protección de derechos humanos, su enfoque es particularmente relevante porque introduce una dimensión global y humanista en la gobernanza de la IA, sin embargo, su principal limitación radica en su carácter no vinculante. A diferencia del modelo europeo, las recomendaciones de la UNESCO carecen de mecanismos coercitivos, lo que dificulta su implementación efectiva. Floridi y otros. (2018) señalan que los marcos éticos son fundamentales, pero deben ser acompañados por estructuras institucionales que garanticen su cumplimiento y que, en ausencia de estos mecanismos, existe el riesgo de que la ética se convierta en un elemento declarativo más que operativo.

c. Principios de Chapultepec: ética contextual latinoamericana

Los Principios de Chapultepec representan un esfuerzo relevante para contextualizar la regulación de la IA en América Latina, incorporando dimensiones

como la justicia social, la cultura de paz y la inclusión. Este enfoque es particularmente valioso porque reconoce que la tecnología no es neutral y que su impacto depende del contexto social, no obstante al igual que el marco de la UNESCO, estos principios enfrentan desafíos en su implementación, desde una perspectiva crítica, puede afirmarse que estos marcos aportan una visión ética robusta, pero carecen de mecanismos institucionales que permitan traducir estos principios en políticas concretas.

d. Decálogo UAM: institucionalización ética en contextos educativos

El decálogo de la UAM representa un caso interesante de institucionalización de principios éticos en un contexto específico, su énfasis en la dignidad, la equidad y la responsabilidad refleja una comprensión integral de los riesgos de la IA. Este enfoque tiene la ventaja de ser aplicable en contextos concretos, como la educación y la investigación, pero su alcance es limitado, ya que no se trata de un marco normativo general, sino solo un código de referencia moral.

6. Tensiones fundamentales en la regulación de la IA

a. Innovación vs protección

Una de las tensiones más evidentes es la que existe entre innovación tecnológica y protección de derechos. Mientras que la regulación busca mitigar riesgos, las empresas tecnológicas tienden a priorizar el desarrollo rápido de productos, Brynjolfsson y McAfee (2014) argumentan que la innovación tecnológica puede generar beneficios significativos, pero también disrupciones que requieren intervención institucional y en este sentido, la regulación no debe ser vista como un obstáculo, sino como un mecanismo de equilibrio.

b. Globalización vs soberanía

La inteligencia artificial es una tecnología global, pero su regulación es fragmentada lo que necesariamente genera tensiones entre la necesidad de estándares internacionales y la soberanía de los Estados, las plataformas digitales operan a

escala global, lo que dificulta su regulación a nivel nacional por lo que se requieren mecanismos de gobernanza internacional (Srnicek, 2017).

c. Ética vs economía

Otra tensión importante es la que existe entre los principios éticos y los intereses económicos. Las empresas tecnológicas tienen incentivos para maximizar beneficios, lo que puede entrar en conflicto con la protección de derechos.

Zuboff (2019) describe cómo el modelo económico basado en datos ha generado nuevas formas de explotación, lo que plantea desafíos para la regulación.

d. Regulación reactiva vs anticipatoria

La mayoría de las regulaciones actuales son reactivas, es decir, responden a problemas existentes. Sin embargo, los riesgos estructurales requieren un enfoque anticipatorio. Bostrom (2014) advierte que algunos riesgos asociados a la IA pueden ser irreversibles, lo que hace necesario actuar antes de que se materialicen.

6. Hacia un enfoque integral de regulación

6.1 Integración de dimensiones de riesgo

La regulación de la IA debe integrar las dos dimensiones de riesgo:

- riesgos inmediatos (operativos)
- riesgos estructurales (civilizatorios)

Esto implica superar la fragmentación actual y adoptar una visión más amplia.

6.2 Multinivel de gobernanza

Un enfoque integral requiere múltiples niveles de gobernanza:

- internacional (UNESCO, acuerdos globales)
- regional (UE)
- nacional
- institucional (UAM)

Cada nivel cumple una función específica, pero deben estar articulados.

6.3 Regulación del poder tecnológico

Uno de los aspectos más críticos es la regulación del poder. Esto incluye:

- control de monopolios
- acceso a infraestructura
- distribución de beneficios

Pasquale (2015) señala que la transparencia no es suficiente si no se abordan las estructuras de poder.

6.4 Protección de la autonomía humana

La regulación también debe enfocarse en preservar la autonomía humana. Esto implica limitar la dependencia tecnológica y promover el desarrollo de capacidades. Haidt (2024) muestra cómo las tecnologías digitales pueden afectar el desarrollo humano, lo que refuerza la necesidad de intervención.

7. Principales regulaciones, normativas y códigos actuales.

Instrumento / Iniciativa	Tipo	Alcance	Enfoque principal	Nivel de obligatoriedad	Estado	Observaciones
Recomendación sobre la Ética de la IA (UNESCO 2021)	Marco ético	Internacional	Derechos humanos, equidad, inclusión	No vinculante	Vigente	Base ética global adoptada ampliamente
Principios de la OCDE sobre IA (2019)	Lineamientos	Internacional	IA confiable, transparencia, crecimiento inclusivo	No vinculante	Vigente	Influye en políticas públicas
AI4People Framework (2018)	Marco ético	Internacional	Justicia, explicabilidad, bienestar social	No vinculante	Vigente	Referente académico
Declaración de Bletchley (2023)	Acuerdo político	Internacional	Riesgos avanzados, seguridad IA	No vinculante	Vigente	Enfoque en IA avanzada
AI Act (2024)	Regulación formal	Regional	Clasificación por riesgos, derechos fundamentales	Vinculante	Aprobada	Modelo más avanzado
GDPR (2018)	Regulación de datos	Regional	Privacidad, protección de datos	Vinculante	Vigente	Base para IA basada en datos
AI Bill of Rights (2022)	Marco político	Nacional	Derechos civiles, protección del usuario	No vinculante	Vigente	Enfoque en derechos digitales

Instrumento / Iniciativa	Tipo	Alcance	Enfoque principal	Nivel de obligatoriedad	Estado	Observaciones
Executive Order on AI (2023)	Orden ejecutiva	Nacional	Seguridad, innovación, control de riesgos	Parcialmente vinculante	Vigente	Marco estratégico
Regulación IA generativa China (2023)	Regulación estatal	Nacional	Control de contenido, seguridad	Vinculante	Vigente	Fuerte control estatal
Estrategia Nacional de IA (México)	Política pública	Nacional	Innovación, adopción tecnológica	No vinculante	Vigente	Primer marco nacional
Declaración de ética IA México	Marco ético	Nacional	Derechos humanos, transparencia	No vinculante	Vigente	Inspirado en UNESCO
Ley Federal de Protección de Datos (LFPDPPP)	Ley	Nacional	Privacidad, datos personales	Vinculante	Vigente	Principal regulación indirecta
Lineamientos INAI	Regulación	Nacional	Protección de datos	Vinculante	Vigente	Aplica a IA con datos
Iniciativas Ley IA Senado	Iniciativa legislativa	Nacional	Regulación sector público, ética IA	No vinculante	En discusión	Fragmentadas
Iniciativas deepfakes Senado	Iniciativa legislativa	Nacional	Desinformación, identidad digital	No vinculante	En discusión	Enfoque mediático
Iniciativas IA Cámara de Diputados	Iniciativa legislativa	Nacional	Datos, automatización	No vinculante	En discusión	Sin marco integral
Reformas digitales Cámara de Diputados	Iniciativa legislativa	Nacional	Servicios públicos automatizados	No vinculante	En discusión	Enfoque sectorial
Agenda Digital CDMX	Política pública	Local	Gobierno digital, uso de IA	No vinculante	Vigente	Uso sin regulación específica
Iniciativas IA CDMX	Iniciativa local	Local	Automatización servicios públicos	No vinculante	En desarrollo	Falta marco legal
Lineamientos datos abiertos CDMX	Regulación	Local	Transparencia de datos	Vinculante	Vigente	Impacto indirecto
Decálogo IA UAM	Código ético	Institucional	Dignidad, equidad, transparencia	No vinculante	Vigente	Referente académico fuerte
Lineamientos IA universidades	Marco ético	Institucional	Uso responsable IA	No vinculante	Vigente	Adopción creciente

Fuente: elaboración propia con base en los documentos.

Conclusiones

El análisis desarrollado a lo largo del texto permite afirmar que la inteligencia artificial no es únicamente una tecnología disruptiva, sino un fenómeno estructural que está reconfigurando de manera profunda las bases de la vida social, económica y política. En este sentido, su regulación no puede ser entendida como un ejercicio técnico limitado a corregir fallas puntuales, sino como un desafío de carácter integral que exige repensar los marcos normativos, éticos y políticos desde una perspectiva amplia y anticipatoria. Uno de los principales hallazgos es la existencia de una doble dimensión de riesgos claramente diferenciada pero interrelacionada. Por un lado, los riesgos inmediatos —como la vulneración de la propiedad intelectual, la afectación a la privacidad, la desinformación, el impacto en el empleo y la opacidad algorítmica— constituyen problemas visibles que ya están generando efectos concretos en los derechos individuales y en los sistemas económicos. Estos riesgos han motivado la mayor parte de las iniciativas regulatorias actuales, las cuales, aunque relevantes, tienden a adoptar un enfoque reactivo centrado en mitigar consecuencias específicas.

Sin embargo, el análisis también evidencia la existencia de una segunda dimensión de riesgos, menos visible pero significativamente más profunda. La pérdida de control humano sobre sistemas autónomos, la concentración de poder tecnológico, la automatización de la desigualdad y la erosión de capacidades humanas configuran un conjunto de amenazas de carácter estructural que trascienden el ámbito técnico. Estos riesgos no solo afectan el funcionamiento de sistemas particulares, sino que inciden directamente en la autonomía individual, la equidad social y la sostenibilidad de los sistemas democráticos.

Un elemento clave es que estos riesgos no operan de manera aislada, sino como un sistema interconectado que se retroalimenta. La concentración de poder, por ejemplo, amplifica la pérdida de control, mientras que la opacidad de los sistemas facilita la reproducción de desigualdades, esta interdependencia convierte a la inteligencia artificial en un fenómeno complejo cuya gobernanza requiere enfoques

sistémicos, más allá de soluciones fragmentadas. En este contexto, el análisis de los marcos regulatorios actuales revela una brecha significativa entre la magnitud de los riesgos y la capacidad de respuesta institucional. Si bien existen avances importantes —como el modelo europeo basado en niveles de riesgo o los marcos éticos promovidos por organismos internacionales—, estos enfoques presentan limitaciones al concentrarse principalmente en los aspectos más visibles del problema. La falta de mecanismos vinculantes en muchos marcos éticos, así como la escasa atención a las dinámicas de poder y a los efectos estructurales, evidencia la necesidad de evolucionar hacia modelos regulatorios más integrales.

Asimismo, se identifican tensiones fundamentales que atraviesan la regulación de la IA, como la relación entre innovación y protección, entre globalización y soberanía, y entre ética y economía, estas tensiones reflejan la complejidad del fenómeno y la dificultad de diseñar marcos normativos que sean al mismo tiempo efectivos, flexibles y legítimos. En consecuencia, una regulación verdaderamente eficaz de la inteligencia artificial debe articular ambas dimensiones de riesgo. No basta con atender los problemas inmediatos; es indispensable anticipar y gestionar los efectos estructurales que pueden comprometer el futuro de la sociedad. Esto implica adoptar un enfoque multinivel, integrar principios éticos con mecanismos jurídicos vinculantes y, sobre todo, colocar la protección de la autonomía humana y el interés colectivo en el centro del desarrollo tecnológico.

En última instancia, la regulación de la inteligencia artificial no es solo una cuestión normativa, sino una decisión sobre el tipo de sociedad que se desea construir en un contexto de creciente mediación tecnológica.

Bibliografía

- Bostrom, Nick. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Brynjolfsson, Erik & McAfee, Andrew. (2014). *The second machine age*. W. W. Norton & Company.
- Copeland, B. Jack. (1993). *Inteligencia artificial: Una introducción filosófica*. Alianza Editorial.
- Floridi, Luciano; Cows, Josh; Beltrametti, Monica; Chatila, Raja; Chazerand, Patrice; Dignum, Virginia; Luetge, Christoph; Madelin, Robert; Pagallo, Ugo; Rossi, Francesca & Schafer, Burkhard. (2018). *AI4People—An ethical framework for a good AI society*.
- Gamero Casado, Eduardo. (2024). *El enfoque europeo de la inteligencia artificial*.
- Haidt, Jonathan. (2024). *La generación ansiosa*. Penguin Random House.
- Larson, Erik J. (2021). *El mito de la inteligencia artificial: Por qué las máquinas no pueden pensar como nosotros*. Shackleton Books.
- Lassalle, José María. (2024). *Civilización artificial*. Arpa Editores.
- Lessig, Lawrence. (2006). *Code: Version 2.0*. Basic Books.
- Pasquale, Frank. (2015). *The black box society*. Harvard University Press.
- Rodríguez, Pablo. (2020). *Inteligencia artificial*.
- Sadin, Éric. (2022). *La era del individuo tirano*. Caja Negra.
- Srnicek, Nick. (2017). *Platform capitalism*. Polity Press.
- Zuboff, Shoshana. (2019). *The age of surveillance capitalism*. PublicAffairs

Referencias y fuentes:

- Brynjolfsson, Erik, & McAfee, Andrew. (2014). *La segunda era de las máquinas: Trabajo, progreso y prosperidad en una época de tecnologías brillantes*. W. W. Norton & Company.
<https://wwnorton.com/books/the-second-machine-age/>
- Bostrom, Nick. (2014). *Superinteligencia: Caminos, peligros, estrategias*. Oxford University Press.
<https://global.oup.com/academic/product/superintelligence-9780198739838>

Consejo de la Unión Europea & Parlamento Europeo. (2024). *Reglamento de Inteligencia Artificial (AI Act)*.
<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52021PC0206>

Unión Europea. (2016). *Reglamento General de Protección de Datos (GDPR)*.
<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32016R0679>

Executive Office of the President of the United States. (2023). *Orden Ejecutiva sobre el desarrollo y uso seguro, protegido y confiable de la inteligencia artificial*.
<https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>

Floridi, Luciano; Cows, Josh; Beltrametti, Monica; Chatila, Raja; Chazerand, Patrice; Dignum, Virginia; Luetge, Christoph; Madelin, Robert; Pagallo, Ugo; Rossi, Francesca & Schafer, Burkhard. (2018). AI4People—Un marco ético para una buena sociedad de IA. *Minds and Machines*, 28(4), 689–707.
<https://link.springer.com/article/10.1007/s11023-018-9482-5>

Gobierno de México. (s.f.). *Estrategia Nacional de Inteligencia Artificial*.
<https://www.gob.mx>

Haidt, Jonathan. (2024). *La generación ansiosa: Por qué las redes sociales están causando una epidemia de enfermedades mentales*. Penguin Random House.
<https://www.penguinrandomhouse.com/books/705011/the-anxious-generation-by-jonathan-haidt/>

Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI). (2010). *Ley Federal de Protección de Datos Personales en Posesión de los Particulares*.
<https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>

Lessig, Lawrence. (2006). *El código 2.0*. Basic Books.
<https://codev2.cc/>

Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2019). *Principios de la OCDE sobre inteligencia artificial*.
<https://oecd.ai/es/ai-principles>

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO). (2021). *Recomendación sobre la ética de la inteligencia artificial*.
https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Pasquale, Frank. (2015). *La sociedad de la caja negra: Los algoritmos que controlan el dinero y la información*. Harvard University Press.
<https://www.hup.harvard.edu/books/9780674368279>

República Popular China. (2023). *Medidas provisionales para la gestión de servicios de inteligencia artificial generativa*.
https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

Srnicek, Nick. (2017). *Capitalismo de plataformas*. Polity Press.
<https://politybooks.com/bookdetail/?isbn=9781509504872>

The White House Office of Science and Technology Policy. (2022). *Declaración de derechos de la inteligencia artificial (AI Bill of Rights)*.
<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

Universidad Autónoma Metropolitana (UAM). (2024). *Decálogo de ética para el uso de la inteligencia artificial*.
<https://www.uam.mx>

Zuboff, Shoshana. (2019). *La era del capitalismo de la vigilancia*. PublicAffairs.
<https://www.publicaffairsbooks.com/titles/shoshana-zuboff/the-age-of-surveillance-capitalism/9781610395694/>

CAPÍTULO II

LA GOBERNANZA GLOBAL DE LA INTELIGENCIA ARTIFICIAL DESDE EL SUR GLOBAL: UNA VISIÓN CRÍTICA Y PROPOSITIVA

Global Governance of Artificial Intelligence from the Global South: A
Critical and Propositive View

Cinthia Noemi Zacatenco Arellano⁴

noemizacatencoa@gmail.com

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0009-0000-7515-8310>

Resumen

La gobernanza global de la inteligencia artificial (IA) se configura en un contexto marcado por profundas asimetrías tecnológicas, económicas y políticas. Este capítulo analiza críticamente cómo la regulación de la IA está siendo liderada por un número reducido de actores —principalmente Estados Unidos, la Unión Europea y China—, lo que genera un modelo de gobernanza global desigual que limita la participación de los países en desarrollo. A partir de un enfoque teórico y analítico, se examinan las dinámicas de poder que estructuran la gobernanza de la IA, incluyendo la concentración de datos, la dependencia tecnológica y la exclusión del Sur Global en la definición de estándares internacionales. Asimismo, se identifican los principales retos que enfrentan estos países, tales como la falta de capacidades tecnológicas, debilidad institucional y vulnerabilidad ante riesgos éticos y sociales. No obstante, el capítulo también destaca oportunidades emergentes para avanzar hacia una gobernanza más inclusiva, a través del fortalecimiento de capacidades locales, la cooperación internacional, la soberanía tecnológica y el desarrollo de

⁴ Licenciada en administración por la Universidad Autónoma Metropolitana Unidad Azcapotzalco.

modelos alternativos de innovación. Finalmente, se proponen líneas estratégicas orientadas a construir una regulación equitativa de la IA, basada en principios de justicia tecnológica, inclusión y bienestar social. Se concluye que la inteligencia artificial no es una tecnología neutral, sino un campo de disputa política, cuyo impacto dependerá de la capacidad de los actores globales para construir un modelo de gobernanza más democrático y contextualizado.

Palabras clave: Gobernanza global de la inteligencia artificial, Desigualdad tecnológica, Soberanía digital, Sur Global, Regulación de la IA, CUMBRE

Abstract

The global governance of artificial intelligence (AI) is configured in a context marked by profound technological, economic and political asymmetries. This chapter critically analyzes how AI regulation is being led by a small number of actors—mainly the United States, the European Union, and China—generating an unequal global governance model that limits the participation of developing countries. From a theoretical and analytical approach, the power dynamics that structure AI governance are examined, including data concentration, technological dependence, and the exclusion of the Global South in the definition of international standards. It also identifies the main challenges faced by these countries, such as lack of technological capabilities, institutional weakness, and vulnerability to ethical and social risks. However, the chapter also highlights emerging opportunities to move towards more inclusive governance, through the strengthening of local capacities, international cooperation, technological sovereignty and the development of alternative models of innovation. Finally, strategic lines are proposed aimed at building an equitable regulation of AI, based on principles of technological justice, inclusion and social welfare. It is concluded that artificial intelligence is not a neutral technology, but a field of political dispute, the impact of which will depend on the ability of global actors to build a more democratic and contextualized governance model.

Keywords: Global governance of artificial intelligence, Technological inequality, Digital sovereignty, Global South, AI regulation, SUMMIT

1. Introducción

La inteligencia artificial (IA) se ha consolidado como una tecnología de propósito general con capacidad de transformar múltiples sectores, desde la economía hasta la gobernanza pública. Su impacto no solo es técnico, sino estructural, al reconfigurar relaciones de poder, modelos productivos y dinámicas sociales a escala global (Brynjolfsson & McAfee, 2017; Crawford, 2021). En este contexto, la necesidad de establecer marcos regulatorios adecuados ha adquirido una relevancia central en la agenda internacional.

Sin embargo, la regulación de la IA no se está desarrollando en un vacío neutral, sino en un escenario profundamente marcado por desigualdades geopolíticas, económicas y tecnológicas. Las principales iniciativas normativas provienen de un número reducido de actores con alta capacidad tecnológica, particularmente Estados Unidos, la Unión Europea y China. Estos actores no solo lideran el desarrollo de la IA, sino que también influyen de manera decisiva en la definición de estándares, principios éticos y marcos regulatorios (Jobin y otros., 2019).

Este fenómeno plantea una preocupación fundamental: la emergencia de una gobernanza global asimétrica, en la que las reglas del juego son definidas por quienes concentran los recursos, los datos y las capacidades tecnológicas. Como resultado, los países en desarrollo corren el riesgo de quedar relegados a un papel periférico, tanto en la producción tecnológica como en la formulación de políticas (UNDP, 2021).

Además, la IA no es una tecnología neutral. Diversos estudios han demostrado que puede reproducir y amplificar desigualdades existentes, especialmente cuando se entrena con datos sesgados o se implementa en contextos de vulnerabilidad institucional (Buolamwini & Gebru, 2018). En ausencia de marcos regulatorios

inclusivos y contextualizados, estos riesgos pueden intensificarse en países con menor capacidad de supervisión y regulación.

Otro elemento clave es el creciente poder de las grandes corporaciones tecnológicas, que concentran infraestructura, talento y datos a escala global. Este fenómeno ha sido conceptualizado como una forma de “capitalismo de vigilancia” o “colonialismo de datos”, en la que la información generada por usuarios de todo el mundo es apropiada y monetizada por un número limitado de empresas (Zuboff, 2019; Couldry & Mejias, 2019). Esta concentración de poder no solo tiene implicaciones económicas, sino también regulatorias, ya que estas empresas participan activamente en la definición de estándares y políticas.

En este escenario, la discusión sobre la gobernanza de la IA no puede limitarse a aspectos técnicos o éticos abstractos. Es necesario abordarla como un problema político global, en el que están en juego cuestiones de soberanía, justicia y desarrollo. En particular, resulta fundamental analizar cómo los países del Sur Global pueden participar de manera más activa en la configuración de este nuevo orden tecnológico.

A partir de lo anterior, el presente capítulo se propone analizar críticamente la gobernanza global de la inteligencia artificial desde una perspectiva del Sur Global. Se examinan las dinámicas de poder que configuran la regulación internacional, las asimetrías tecnológicas existentes y los principales retos que enfrentan los países en desarrollo. Asimismo, se identifican oportunidades y se plantean propuestas orientadas a construir una gobernanza más inclusiva, equitativa y orientada al bienestar social.

De esta manera, el capítulo busca contribuir a un debate emergente en el campo de la regulación de la IA: la necesidad de transitar desde un enfoque dominado por potencias tecnológicas hacia una gobernanza global más plural, democrática y contextualizada, capaz de responder a las realidades y necesidades del conjunto de la humanidad.

2. Marco conceptual: gobernanza global de la inteligencia artificial

La gobernanza global de la inteligencia artificial (IA) constituye un campo emergente que articula elementos de regulación tecnológica, ética digital y relaciones internacionales. En términos generales, se refiere al conjunto de normas, instituciones, principios y prácticas que buscan orientar el desarrollo, implementación y uso de la IA a escala transnacional (Floridi y otros., 2018).

A diferencia de otros ámbitos de gobernanza global —como el comercio o el cambio climático—, la IA presenta características particulares que complejizan su regulación. Entre ellas destacan su rápida evolución tecnológica, su carácter transversal (aplicable a múltiples sectores) y su dependencia de recursos altamente concentrados, como grandes volúmenes de datos y capacidad computacional (Kerry, 2020). Estas características dificultan la construcción de marcos regulatorios homogéneos y efectivos.

2.1 Gobernanza global: enfoque teórico

Desde una perspectiva teórica, la gobernanza global se entiende como un sistema de coordinación sin una autoridad central única, en el que participan múltiples actores estatales y no estatales (Rosenau, 1995). En el caso de la IA, este modelo se refleja en la interacción entre:

- Estados nacionales, que desarrollan marcos regulatorios y políticas públicas.
- Corporaciones tecnológicas, que diseñan, implementan y controlan gran parte de la infraestructura de IA.
- Organismos internacionales, como la UNESCO, la OCDE o la ONU, que promueven principios y recomendaciones.
- Sociedad civil y academia, que participan en debates éticos y críticos.

Este entramado configura una gobernanza policéntrica, en la que el poder está distribuido, pero no necesariamente de forma equitativa.

2.2 Fragmentación normativa y pluralidad de enfoques

Uno de los rasgos distintivos de la gobernanza de la IA es su fragmentación. Actualmente, no existe un tratado internacional vinculante que regule la IA de manera integral. En su lugar, se observa una proliferación de principios éticos, guías voluntarias y regulaciones regionales (Jobin y otros., 2019).

Por ejemplo, el análisis de Jobin y otros. (2019) identifica más de 80 marcos éticos de IA a nivel global, con principios recurrentes como transparencia, justicia, no maleficencia y responsabilidad. No obstante, estos principios suelen carecer de mecanismos de implementación efectivos, lo que limita su impacto real.

Esta fragmentación genera un entorno regulatorio heterogéneo, donde coexisten múltiples enfoques:

- Enfoque basado en derechos (Unión Europea): prioriza la protección de derechos fundamentales.
- Enfoque orientado al mercado (Estados Unidos): enfatiza la innovación y la autorregulación.
- Enfoque estatal centralizado (China): combina desarrollo tecnológico con control político.

Estas diferencias reflejan no solo visiones regulatorias distintas, sino también modelos de gobernanza y valores sociopolíticos divergentes (Roberts y otros., 2021).

2.3 El papel del poder en la gobernanza de la IA

Un elemento central en el análisis de la gobernanza de la IA es la distribución del poder. La literatura reciente subraya que la IA no solo es un campo tecnológico, sino también un espacio de disputa geopolítica y económica (Crawford, 2021).

El poder en este ámbito se manifiesta en tres dimensiones principales:

1. **Control de datos:** Los datos son el insumo fundamental para el desarrollo de sistemas de IA. Su concentración en manos de grandes corporaciones genera asimetrías significativas (Zuboff, 2019).
2. **Capacidad computacional:** El acceso a infraestructura avanzada (como centros de datos y GPUs) está altamente concentrado en países desarrollados.
3. **Capacidad normativa:** Los países con mayor poder económico y tecnológico tienen mayor influencia en la definición de estándares globales.

Estas dinámicas han dado lugar a lo que algunos autores denominan una “geopolítica de la IA”, en la que los Estados compiten por liderazgo tecnológico y normativo (Kerry, 2020).

2.4 Gobernanza, ética y regulación

La gobernanza de la IA también está estrechamente vinculada con la ética tecnológica. Iniciativas como la Recomendación de la UNESCO (2021) buscan establecer principios universales para el desarrollo de la IA, incluyendo respeto a los derechos humanos, inclusión y sostenibilidad.

No obstante, diversos autores han señalado una brecha entre los principios éticos y su implementación práctica, fenómeno conocido como *ethics washing* (Floridi, 2019). Es decir, la adopción de discursos éticos sin cambios estructurales en las prácticas tecnológicas.

Por ello, la gobernanza efectiva de la IA requiere no solo principios, sino también mecanismos de cumplimiento, supervisión y rendición de cuentas.

2.5 Hacia una gobernanza inclusiva

Finalmente, un aspecto clave del debate contemporáneo es la necesidad de avanzar hacia una gobernanza más inclusiva. Actualmente, los países del Sur Global tienen una participación limitada en la definición de estándares y marcos regulatorios, lo que genera un déficit de representatividad (UNESCO, 2021).

Una gobernanza inclusiva implica:

- Incorporar diversidad geográfica y cultural en la toma de decisiones.
- Reconocer las desigualdades estructurales en capacidades tecnológicas.
- Promover modelos regulatorios adaptados a contextos locales.

En este sentido, la gobernanza de la IA no debe entenderse únicamente como un proceso técnico, sino como un proyecto político orientado a garantizar justicia, equidad y sostenibilidad en la era digital.

3. Panorama internacional de la regulación de la inteligencia artificial

En la última década, la regulación de la inteligencia artificial (IA) ha pasado de ser un tema incipiente para ocupar un lugar central en la agenda política global. Este proceso ha estado marcado por la emergencia de distintos modelos regulatorios impulsados por las principales potencias tecnológicas, lo que ha dado lugar a un entorno normativo fragmentado y competitivo.

Lejos de configurarse como un sistema coherente de gobernanza global, la regulación de la IA refleja una pluralidad de enfoques que responden a intereses económicos, valores políticos y estrategias geopolíticas diferenciadas (Kerry, 2020). En este apartado se analizan tres modelos predominantes: el europeo, el estadounidense y el chino, así como el papel de los organismos internacionales.

3.1 El modelo europeo: regulación basada en riesgos y derechos fundamentales

La Unión Europea (UE) se ha posicionado como líder en la construcción de un marco regulatorio integral para la IA, destacando por su enfoque centrado en la protección de derechos fundamentales. Su principal iniciativa es el Reglamento de Inteligencia Artificial (AI Act), propuesto en 2021, que establece un sistema de clasificación de riesgos para los sistemas de IA (European Commission, 2021).

Este modelo distingue entre:

- Riesgo inaceptable (prohibido): como sistemas de vigilancia masiva o manipulación cognitiva.

- Alto riesgo: sistemas utilizados en ámbitos críticos (salud, justicia, empleo).
- Riesgo limitado y mínimo: con menores exigencias regulatorias.

El enfoque europeo busca equilibrar la innovación tecnológica con la protección de derechos humanos, en línea con su tradición normativa. Además, introduce obligaciones claras en materia de transparencia, trazabilidad y rendición de cuentas.

No obstante, este modelo también ha sido objeto de críticas. Algunos autores señalan que una regulación estricta podría limitar la competitividad tecnológica de la UE frente a Estados Unidos y China (Veale & Borgesius, 2021). Aun así, su influencia global es significativa, al establecer estándares que pueden ser adoptados por otros países, fenómeno conocido como el “efecto Bruselas”.

3.2 El modelo estadounidense: innovación y autorregulación

En contraste con la UE, Estados Unidos ha adoptado un enfoque más flexible y orientado al mercado. La regulación de la IA se ha caracterizado por la promoción de la innovación, con una menor intervención estatal directa.

Documentos como el *Blueprint for an AI Bill of Rights* (White House, 2022) establecen principios orientadores —como seguridad, privacidad y equidad—, pero no constituyen un marco regulatorio vinculante. Asimismo, el Instituto Nacional de Estándares y Tecnología (NIST) ha desarrollado marcos de gestión de riesgos para la IA, de carácter voluntario.

Este enfoque responde a la importancia estratégica del sector tecnológico en la economía estadounidense, donde empresas como Google, Microsoft y Amazon lideran el desarrollo de IA a nivel global.

Sin embargo, la falta de regulación estricta ha generado preocupaciones en torno a la protección de derechos y la concentración de poder corporativo. Diversos estudios advierten que la autorregulación puede ser insuficiente para abordar los riesgos sociales y éticos de la IA (Crawford, 2021).

3.3 El modelo chino: control estatal y desarrollo estratégico

China representa un tercer modelo, caracterizado por una fuerte intervención estatal y una integración estrecha entre desarrollo tecnológico y control político. El gobierno chino ha implementado regulaciones específicas sobre algoritmos, recomendación de contenidos y uso de datos, con el objetivo de mantener la estabilidad social y el control informativo (Roberts y otros., 2021).

A diferencia de Estados Unidos, donde predomina el sector privado, en China el Estado desempeña un papel central en la orientación del desarrollo tecnológico. La estrategia nacional de IA busca posicionar al país como líder global, combinando inversión masiva, planificación estatal y regulación.

Este modelo plantea importantes debates en torno a derechos humanos, privacidad y vigilancia. No obstante, también ha demostrado capacidad para implementar regulaciones de manera rápida y efectiva.



Ilustración 1 elaboración propia con apoyo de inteligencia artificial (DALL·E, OpenAI).

3.4 El papel de los organismos internacionales

Ante la fragmentación de enfoques nacionales, los organismos internacionales han intentado promover principios comunes para la gobernanza de la IA. Entre las iniciativas más relevantes se encuentran:

- La Recomendación sobre la Ética de la IA de la UNESCO (2021), que establece principios como equidad, inclusión y sostenibilidad.
- Los Principios de IA de la OCDE (OECD, 2019), centrados en innovación responsable y crecimiento inclusivo.
- Iniciativas del G20 y Naciones Unidas para fomentar la cooperación internacional.

Si bien estas iniciativas contribuyen a generar consenso normativo, su carácter no vinculante limita su efectividad. En la práctica, la gobernanza global de la IA sigue dependiendo en gran medida de decisiones nacionales y regionales.

3.5 Tensiones geopolíticas y competencia tecnológica

La regulación de la IA no puede entenderse sin considerar el contexto de competencia geopolítica. Estados Unidos y China, en particular, mantienen una rivalidad estratégica por el liderazgo en IA, que se manifiesta en inversiones, políticas industriales y regulación tecnológica.

Esta competencia ha sido descrita como una “carrera por la IA”, en la que los países buscan no solo ventajas económicas, sino también influencia política global (Kerry, 2020). En este contexto, la regulación se convierte en una herramienta estratégica, más que en un instrumento neutral.

Por su parte, la Unión Europea intenta posicionarse como un actor normativo global, promoviendo estándares basados en derechos. Sin embargo, su capacidad de competir tecnológicamente sigue siendo limitada.

3.6 Implicaciones para el Sur Global

La coexistencia de estos modelos genera un entorno complejo para los países en desarrollo. En muchos casos, estos países se ven obligados a adoptar marcos regulatorios externos, sin una participación activa en su diseño.

Esto plantea riesgos de dependencia normativa y tecnológica, así como la posible imposición de estándares que no responden a sus realidades locales (UNDP, 2021).

Además, la falta de coordinación global dificulta la construcción de estrategias propias.

En este sentido, el panorama internacional de la regulación de la IA no solo refleja diversidad de enfoques, sino también asimetrías de poder que condicionan las posibilidades de desarrollo tecnológico y autonomía regulatoria de los países del Sur Global.

En el contexto más reciente de la gobernanza global de la inteligencia artificial, la Cumbre de Impacto de la IA 2026, celebrada en Nueva Delhi, India, del 16 al 20 de febrero, representa un hito relevante en la evolución de los espacios multilaterales dedicados a esta tecnología. A diferencia de encuentros anteriores, predominantemente impulsados por potencias tecnológicas y centrados en la gestión de riesgos asociados a sistemas avanzados, esta cumbre introdujo un enfoque más amplio que articula la inteligencia artificial con el desarrollo sostenible, la inclusión digital y la equidad global. En particular, el evento evidenció un cambio discursivo hacia la necesidad de considerar la IA no solo como un desafío técnico o de seguridad, sino como una herramienta estratégica para abordar problemáticas estructurales en países en desarrollo, tales como la pobreza, la desigualdad y la brecha digital (Government of India, 2026).

Asimismo, la cumbre reflejó el papel emergente de India como actor clave en la reconfiguración de la gobernanza tecnológica global. A través de su liderazgo, se promovió una agenda que busca equilibrar la innovación con la justicia social, enfatizando la importancia de democratizar el acceso a la tecnología, fortalecer capacidades locales y garantizar que los beneficios de la IA se distribuyan de manera más equitativa (Government of India, 2026). Este posicionamiento resulta particularmente significativo en un contexto donde los marcos regulatorios han sido históricamente definidos por actores del Norte Global, lo que ha limitado la incorporación de perspectivas y necesidades propias del Sur Global.

No obstante, la relevancia de esta cumbre también radica en las tensiones que pone en evidencia. Por un lado, se observa una creciente apertura hacia modelos

de gobernanza más inclusivos y multipolares; por otro, persisten dinámicas de poder que continúan favoreciendo a países con mayor capacidad tecnológica y económica. En este sentido, la Cumbre de Impacto de la IA 2026 ilustra tanto los avances como los desafíos en la construcción de un sistema de gobernanza global más equilibrado, en el que coexisten visiones orientadas a la seguridad, la innovación y el desarrollo (Government of India, 2026). En consecuencia, este tipo de iniciativas refuerza la necesidad de avanzar hacia mecanismos de cooperación internacional que no solo mitiguen riesgos, sino que también promuevan una distribución más justa de los beneficios de la inteligencia artificial a nivel global.

4. Asimetrías tecnológicas y desigualdad global en la era de la inteligencia artificial

El desarrollo y la implementación de la inteligencia artificial (IA) no se producen en un contexto neutral, sino dentro de un sistema global profundamente desigual. Las capacidades tecnológicas, el acceso a datos y la infraestructura digital están altamente concentrados en un reducido número de países y corporaciones, lo que genera asimetrías estructurales que condicionan la participación de los países en desarrollo en la economía digital (Crawford, 2021).

Estas desigualdades no solo afectan la capacidad de producir tecnología, sino también la posibilidad de influir en su regulación, lo que configura un escenario de dependencia tecnológica y subordinación normativa.

4.1 Concentración del poder tecnológico

Uno de los rasgos más característicos de la economía digital contemporánea es la concentración del poder en grandes corporaciones tecnológicas. Empresas como Google, Amazon, Microsoft y Meta controlan una proporción significativa de los datos, la infraestructura en la nube y los recursos computacionales necesarios para el desarrollo de sistemas de IA.

Este fenómeno ha sido analizado como una nueva forma de acumulación basada en datos, en la que la información generada por los usuarios se convierte en un

recurso estratégico central (Zuboff, 2019). En este modelo, los datos no solo son un insumo técnico, sino también una fuente de poder económico y político.

Asimismo, la concentración de infraestructura —particularmente centros de datos y capacidad de cómputo de alto rendimiento— refuerza estas asimetrías, limitando la entrada de nuevos actores y dificultando el desarrollo autónomo de IA en países con menores recursos (Srnicek, 2017).

4.2 Colonialismo de datos y dependencia digital

Diversos autores han conceptualizado estas dinámicas como una forma de colonialismo de datos, en la que las relaciones de extracción y apropiación características del colonialismo histórico se reproducen en el ámbito digital (Couldry & Mejias, 2019).

En este esquema, los países del Sur Global aportan grandes volúmenes de datos —a través del uso de plataformas digitales—, pero no participan de manera equitativa en los beneficios económicos derivados de su explotación. Esta situación genera una relación de dependencia estructural, en la que los países en desarrollo se convierten en proveedores de datos y consumidores de tecnología.

Además, esta dependencia no es únicamente económica, sino también tecnológica y política. La adopción de plataformas y sistemas de IA desarrollados en el extranjero implica la incorporación de lógicas, valores y sesgos que pueden no ser compatibles con los contextos locales.



Ilustración 2 elaboración propia con apoyo de inteligencia artificial (DALL·E, OpenAI).

4.3 Brecha digital y desigualdad de capacidades

La brecha digital sigue siendo un factor determinante en la configuración de las desigualdades globales en la era de la IA. Esta brecha no se limita al acceso a internet, sino que incluye dimensiones más complejas como:

- Acceso a infraestructura tecnológica avanzada
- Disponibilidad de talento especializado
- Capacidades de investigación y desarrollo
- Acceso a financiamiento

Según el Programa de las Naciones Unidas para el Desarrollo (UNDP, 2021), los países en desarrollo enfrentan importantes limitaciones en estas áreas, lo que reduce su capacidad para aprovechar las oportunidades de la IA.

Esta situación se traduce en una división internacional del trabajo digital, en la que los países desarrollados concentran actividades de alto valor agregado (diseño, desarrollo, innovación), mientras que los países en desarrollo participan en tareas de menor valor, como etiquetado de datos o consumo de tecnologías (Graham y otros., 2019).

4.4 Sesgos algorítmicos y reproducción de desigualdades

Otro aspecto crítico es la forma en que los sistemas de IA pueden reproducir y amplificar desigualdades existentes. Los algoritmos se entrenan con datos históricos que reflejan estructuras sociales desiguales, lo que puede generar resultados discriminatorios.

El estudio de Buolamwini y Gebru (2018) evidenció cómo los sistemas de reconocimiento facial presentan mayores tasas de error en mujeres y personas racializadas, lo que demuestra que la IA no es neutral, sino que incorpora sesgos estructurales.

En contextos del Sur Global, donde existen mayores desigualdades sociales y menor regulación, estos riesgos pueden intensificarse. La implementación de sistemas de IA sin mecanismos adecuados de supervisión puede contribuir a la exclusión y discriminación de grupos vulnerables.

4.5 Dependencia normativa y exclusión en la toma de decisiones

Las asimetrías tecnológicas también se traducen en desigualdades en el ámbito normativo. Los países con mayor capacidad tecnológica tienen una influencia desproporcionada en la definición de estándares globales, mientras que los países en desarrollo tienen una participación limitada (UNESCO, 2021).

Esto genera lo que puede denominarse una dependencia normativa, en la que los marcos regulatorios adoptados por países del Sur Global suelen basarse en modelos externos, sin una adaptación adecuada a sus contextos específicos.

La falta de representación en espacios de gobernanza global implica que las preocupaciones, necesidades y realidades de estos países no siempre son consideradas en la formulación de políticas internacionales.

4.6 Implicaciones para el desarrollo

Las asimetrías descritas tienen implicaciones profundas para el desarrollo económico y social. En lugar de reducir desigualdades, la IA corre el riesgo de amplificarlas, consolidando una estructura global en la que el conocimiento, la innovación y el valor económico se concentran en un número reducido de actores.

No obstante, estas dinámicas no son inevitables. La forma en que se regulen y gobiernen las tecnologías de IA será determinante para definir si estas contribuyen a un desarrollo más equitativo o a la profundización de las desigualdades existentes.

En este sentido, el análisis de las asimetrías tecnológicas no solo permite comprender los desafíos actuales, sino también identificar la necesidad urgente de construir alternativas que promuevan una mayor justicia tecnológica a nivel global.

5. Retos para los países en desarrollo en la gobernanza de la inteligencia artificial

Los países en desarrollo enfrentan un conjunto de desafíos estructurales en relación con la inteligencia artificial (IA), que limitan su capacidad para participar activamente tanto en su desarrollo como en su regulación. Estos retos no son únicamente tecnológicos, sino también institucionales, económicos y políticos, y se encuentran profundamente vinculados a las asimetrías globales previamente descritas.

En este contexto, la gobernanza de la IA se convierte en un campo donde se reproducen desigualdades históricas, pero también donde se abren posibilidades para redefinir trayectorias de desarrollo.

5.1 Limitadas capacidades tecnológicas y científicas

Uno de los principales obstáculos para los países en desarrollo es la falta de capacidades en investigación y desarrollo (I+D) en inteligencia artificial. La inversión en ciencia y tecnología en estos países es significativamente menor en comparación con las economías avanzadas, lo que se traduce en una baja producción de innovación tecnológica (OECD, 2023).

Además, existe una escasez de talento especializado en áreas clave como ciencia de datos, aprendizaje automático y ética de la IA. Esta situación genera una dependencia de conocimiento externo y limita la capacidad de desarrollar soluciones adaptadas a contextos locales.

5.2 Dependencia de infraestructura y plataformas extranjeras

La mayoría de los países en desarrollo dependen de infraestructuras digitales controladas por grandes corporaciones tecnológicas, particularmente en servicios de computación en la nube, almacenamiento de datos y plataformas de IA.

Esta dependencia implica varios riesgos:

- Pérdida de soberanía tecnológica

- Vulnerabilidad ante cambios en políticas corporativas
- Limitaciones para desarrollar ecosistemas digitales propios

Como señalan Couldry y Mejias (2019), esta situación forma parte de una lógica más amplia de dependencia digital, en la que los países periféricos quedan subordinados a centros de poder tecnológico.

5.3 Débil desarrollo de marcos regulatorios

Muchos países en desarrollo carecen de marcos regulatorios específicos para la IA o presentan legislaciones fragmentadas y desactualizadas. En algunos casos, las regulaciones existentes en materia de protección de datos, privacidad o derechos digitales son insuficientes para abordar los desafíos que plantea la IA.

Esta debilidad institucional dificulta:

- La supervisión del uso de sistemas de IA
- La prevención de riesgos éticos y sociales
- La protección de derechos fundamentales

Asimismo, la falta de capacidades técnicas dentro de las instituciones públicas limita la implementación efectiva de políticas regulatorias (UNESCO, 2021).



Ilustración 3 elaboración propia con apoyo de inteligencia artificial (DALL·E, OpenAI).

5.4 Exclusión en la gobernanza global

Los países del Sur Global tienen una participación limitada en los espacios donde se definen estándares y normas internacionales sobre IA. Aunque organismos multilaterales promueven la inclusión, en la práctica las decisiones más influyentes siguen estando dominadas por países con mayor poder económico y tecnológico (UNDP, 2021).

Esta exclusión tiene consecuencias importantes:

- Falta de representación de intereses locales
- Imposición de estándares externos
- Reducción de la autonomía regulatoria

En muchos casos, los países en desarrollo adoptan marcos normativos diseñados en otros contextos, lo que puede generar desajustes con sus realidades sociales y económicas.

5.5 Insuficiente inversión y financiamiento

El desarrollo de la IA requiere inversiones significativas en infraestructura, talento y ecosistemas de innovación. Sin embargo, los países en desarrollo enfrentan restricciones presupuestarias que limitan su capacidad para competir en este ámbito.

Según la OECD (2023), la inversión global en IA está altamente concentrada en un pequeño grupo de países, lo que refuerza las brechas existentes. Esta desigualdad en el acceso a financiamiento reduce las posibilidades de desarrollar industrias tecnológicas locales.

5.6 Riesgos sociales y éticos en contextos vulnerables

La implementación de sistemas de IA en contextos caracterizados por desigualdad, informalidad y debilidad institucional puede generar riesgos específicos. Entre ellos destacan:

- Reproducción de sesgos y discriminación
- Vigilancia excesiva sin controles adecuados
- Exclusión de poblaciones vulnerables

Como han demostrado Buolamwini y Gebru (2018), los sistemas de IA pueden amplificar desigualdades preexistentes. En países con menor capacidad regulatoria, estos efectos pueden ser más pronunciados.

5.7 Fuga de talento y dependencia del conocimiento

Otro desafío relevante es la migración de talento altamente calificado hacia países desarrollados, fenómeno conocido como *brain drain*. Esto limita la consolidación de comunidades científicas locales y debilita los ecosistemas de innovación.

Al mismo tiempo, el acceso al conocimiento especializado en IA está mediado por barreras económicas, lingüísticas y tecnológicas, lo que refuerza la dependencia de centros de investigación ubicados en el Norte Global.

5.8 El caso de América Latina (enfoque emergente)

En América Latina, estos retos se manifiestan de manera particular. Aunque algunos países han desarrollado estrategias nacionales de IA —como México, Brasil o Chile—, la región en su conjunto enfrenta limitaciones estructurales en términos de inversión, coordinación regional y capacidades institucionales (CEPAL, 2022).

Además, la fragmentación de políticas públicas y la falta de integración regional dificultan la construcción de una agenda común en materia de gobernanza de la IA.

No obstante, también existen iniciativas prometedoras, como el desarrollo de marcos éticos regionales y el impulso de la IA para el desarrollo social, lo que sugiere la posibilidad de construir modelos alternativos basados en las necesidades locales.

5.9 Síntesis crítica

En conjunto, estos retos evidencian que la participación de los países en desarrollo en la gobernanza de la IA está condicionada por factores estructurales que trascienden el ámbito tecnológico. La falta de capacidades, la dependencia externa y la exclusión en la toma de decisiones configuran un escenario de vulnerabilidad.

Sin embargo, reconocer estos desafíos es el primer paso para diseñar estrategias que permitan transformar estas limitaciones en oportunidades. En este sentido, el siguiente apartado abordará precisamente las posibilidades de avanzar hacia una gobernanza más inclusiva y equitativa.

6. Oportunidades para una gobernanza inclusiva de la inteligencia artificial

A pesar de las profundas asimetrías que caracterizan el desarrollo y la regulación de la inteligencia artificial (IA), los países en desarrollo no se encuentran en una posición exclusivamente pasiva. Por el contrario, el contexto actual también abre una serie de oportunidades para repensar la gobernanza de la IA desde enfoques más inclusivos, contextualizados y orientados al bienestar social.

Estas oportunidades emergen tanto de la flexibilidad del campo regulatorio —aún en construcción— como de la posibilidad de articular estrategias alternativas que desafíen las lógicas dominantes del desarrollo tecnológico.



Ilustración 4 elaboración propia con apoyo de inteligencia artificial (DALL·E, OpenAI).

6.1 Cooperación internacional y multilateralismo

Uno de los principales espacios de oportunidad reside en el fortalecimiento de la cooperación internacional. Organismos multilaterales han comenzado a promover marcos éticos y de gobernanza que buscan integrar a países en desarrollo en la discusión global sobre IA.

Iniciativas como la Recomendación sobre la Ética de la IA de la UNESCO (2021) o los principios de la OCDE (2019) representan esfuerzos por establecer bases comunes para el desarrollo responsable de la tecnología. Aunque no son vinculantes, estos marcos ofrecen plataformas de diálogo y cooperación que pueden ser aprovechadas por el Sur Global.

Asimismo, la cooperación Sur-Sur se presenta como una estrategia clave para compartir conocimientos, recursos y experiencias entre países con contextos similares.

6.2 Desarrollo de capacidades locales

El fortalecimiento de capacidades tecnológicas y científicas constituye una condición fundamental para reducir la dependencia estructural. Esto implica invertir en:

- Educación en ciencia, tecnología e inteligencia artificial
- Formación de talento especializado
- Infraestructura digital y centros de investigación
- Ecosistemas de innovación

El desarrollo de capacidades locales no solo permite participar en la economía digital, sino también adaptar las tecnologías a las necesidades específicas de cada contexto.

Según el UNDP (2021), la construcción de capacidades es un elemento central para garantizar que la IA contribuya al desarrollo humano y no a la profundización de desigualdades.

6.3 Regulación contextualizada y soberanía tecnológica

A diferencia de los países desarrollados, los países del Sur Global tienen la oportunidad de diseñar marcos regulatorios desde etapas relativamente tempranas del desarrollo de la IA. Esto permite construir regulaciones más adaptadas a sus realidades sociales, económicas y culturales.

La noción de soberanía tecnológica cobra especial relevancia en este contexto. Implica la capacidad de los países para tomar decisiones autónomas sobre el desarrollo, uso y regulación de tecnologías digitales.

Esto no necesariamente significa aislarse del sistema global, sino participar en él desde una posición más estratégica, promoviendo:

- Estándares propios
- Uso de tecnologías abiertas
- Protección de datos locales
- Infraestructura digital soberana

6.4 Inteligencia artificial para el desarrollo social

Una de las oportunidades más relevantes es el uso de la IA como herramienta para el desarrollo social. En contextos de escasez de recursos, la IA puede contribuir a mejorar la eficiencia y el impacto de políticas públicas en áreas clave como:

- Salud (diagnóstico y prevención)
- Educación (personalización del aprendizaje)
- Agricultura (optimización de recursos)
- Gestión ambiental

Diversos estudios destacan el potencial de la IA para abordar problemas estructurales en países en desarrollo, siempre que se implemente de manera ética y contextualizada (UNDP, 2021).

Este enfoque implica desplazar la lógica de la IA como motor exclusivamente económico, hacia una visión orientada al bienestar colectivo.

6.5 Economía social y modelos alternativos de innovación

Más allá de los modelos dominantes basados en la maximización de beneficios, la IA también puede integrarse en esquemas de economía social y solidaria. Esto abre la posibilidad de desarrollar tecnologías orientadas a:

- Bienes comunes digitales
- Cooperativas tecnológicas
- Plataformas de propiedad colectiva
- Innovación abierta

Este enfoque permite democratizar el acceso y control de la tecnología, reduciendo la concentración de poder en grandes corporaciones (Srnicek, 2017).

Además, promueve una visión de la innovación centrada en el impacto social, en lugar de la rentabilidad exclusiva.

6.6 Alianzas regionales y articulación del Sur Global

La construcción de alianzas regionales representa otra oportunidad estratégica para los países en desarrollo. En el caso de América Latina, la cooperación regional puede fortalecer la capacidad de negociación frente a actores globales y facilitar el desarrollo de agendas comunes en materia de IA.

Organismos como la CEPAL han destacado la importancia de articular políticas regionales que permitan:

- Compartir infraestructura y conocimiento

- Desarrollar estándares comunes
- Fortalecer la soberanía digital
- Impulsar mercados regionales de tecnología

Estas alianzas pueden contribuir a reducir la fragmentación y aumentar la influencia del Sur Global en la gobernanza internacional.

6.7 Participación de la sociedad civil y enfoque democrático

Finalmente, una gobernanza inclusiva de la IA requiere la participación activa de múltiples actores, más allá de gobiernos y empresas. La sociedad civil, la academia y las comunidades locales desempeñan un papel fundamental en la definición de principios éticos y en la vigilancia del uso de la tecnología.

La inclusión de estos actores permite:

- Incorporar diversidad de perspectivas
- Identificar riesgos sociales y culturales
- Fortalecer la legitimidad de las políticas públicas

En este sentido, la gobernanza de la IA debe entenderse como un proceso democrático, en el que se negocian valores, intereses y visiones de futuro.

6.8 Síntesis propositiva

Las oportunidades identificadas muestran que, a pesar de las limitaciones estructurales, los países en desarrollo pueden desempeñar un papel activo en la configuración de la gobernanza global de la IA.

Aprovechar estas oportunidades requiere una combinación de estrategias:

- Fortalecimiento de capacidades internas
- Cooperación internacional y regional
- Desarrollo de marcos regulatorios propios

- Promoción de modelos alternativos de innovación

En este sentido, la gobernanza de la IA no debe ser concebida únicamente como un proceso de adaptación a normas externas, sino como una oportunidad para construir un modelo tecnológico más justo, inclusivo y orientado al bien común.



Ilustración 5 elaboración propia con apoyo de inteligencia artificial (DALL-E, OpenAI).

7. Propuestas para una regulación equitativa de la inteligencia artificial

El análisis de la gobernanza global de la inteligencia artificial (IA) pone de manifiesto la necesidad urgente de construir modelos regulatorios más inclusivos, equitativos y adaptados a la diversidad de contextos globales. Las asimetrías tecnológicas, la concentración del poder en actores dominantes y la limitada participación de los países en desarrollo en los procesos de toma de decisiones exigen la formulación de propuestas que permitan reconfigurar el actual orden tecnológico. En este sentido, avanzar hacia una gobernanza equitativa de la IA implica no solo corregir desequilibrios existentes, sino también sentar las bases para un modelo de desarrollo tecnológico orientado al bienestar colectivo.

Una de las principales líneas de acción consiste en la construcción de una gobernanza global verdaderamente inclusiva. Esto requiere garantizar la participación efectiva de los países del Sur Global en los espacios internacionales donde se definen normas, estándares y principios de la IA. No se trata únicamente

de ampliar la representación formal, sino de asegurar una capacidad real de incidencia en la toma de decisiones. En este sentido, resulta fundamental reformar los mecanismos de gobernanza en organismos multilaterales, promover la inclusión de actores diversos en foros técnicos y fortalecer plataformas de diálogo que integren múltiples perspectivas (UNESCO, 2021).

Paralelamente, el desarrollo de capacidades tecnológicas soberanas constituye una condición indispensable para reducir la dependencia estructural. La inversión en ciencia, tecnología e innovación debe ser una prioridad estratégica para los países en desarrollo, acompañada de políticas que fomenten la formación de talento especializado y la consolidación de ecosistemas de innovación. Asimismo, la creación de infraestructuras tecnológicas propias, como centros de datos y plataformas digitales, resulta clave para fortalecer la autonomía en el desarrollo y uso de la IA. Estas capacidades no solo permiten participar en la economía digital, sino también ejercer mayor control sobre las decisiones tecnológicas (UNDP, 2021).

Otro aspecto fundamental es el diseño de marcos regulatorios contextualizados. En lugar de adoptar de manera acrítica modelos desarrollados en otros contextos, los países en desarrollo deben construir regulaciones que respondan a sus realidades sociales, económicas y culturales. Esto implica adaptar principios globales a necesidades locales, priorizar problemáticas específicas y considerar las particularidades institucionales de cada país. Una regulación contextualizada permite abordar de manera más efectiva los riesgos asociados a la IA, al tiempo que maximiza sus beneficios potenciales (Jobin y otros., 2019).

En este mismo sentido, la promoción de la soberanía de datos se presenta como un eje estratégico central. Dado que los datos constituyen el principal insumo de la inteligencia artificial, su control y gestión adquieren una relevancia crítica. Los países en desarrollo deben fortalecer sus marcos legales de protección de datos, regular los flujos transfronterizos de información y fomentar infraestructuras de almacenamiento local. Asimismo, es necesario avanzar hacia modelos de

gobernanza de datos que prioricen el interés público y el bien común, evitando su concentración en manos de actores privados globales (Couldry & Mejias, 2019).

Además, resulta imprescindible impulsar modelos alternativos de innovación que contrarresten la concentración del poder tecnológico. El desarrollo de tecnologías abiertas, el fomento del software libre, la creación de cooperativas tecnológicas y la promoción de plataformas digitales de propiedad colectiva constituyen estrategias que pueden democratizar el acceso a la IA. Estos modelos permiten generar beneficios más equitativos y orientados al impacto social, desafiando las lógicas predominantes basadas en la maximización de beneficios económicos (Srnicek, 2017).

Por otra parte, el fortalecimiento de alianzas regionales emerge como una estrategia clave para aumentar la capacidad de acción de los países en desarrollo. En el caso de América Latina, la articulación de agendas comunes en materia de IA puede facilitar el intercambio de recursos, el desarrollo de estándares compartidos y la implementación de proyectos colaborativos. Estas alianzas no solo contribuyen a reducir la fragmentación, sino que también fortalecen la posición del Sur Global en el escenario internacional (CEPAL, 2022).

Asimismo, la integración de principios éticos en las políticas públicas es un elemento esencial para una gobernanza responsable de la IA. La ética no debe limitarse a declaraciones abstractas, sino traducirse en mecanismos concretos de implementación, como evaluaciones de impacto, sistemas de rendición de cuentas y garantías de transparencia. Incorporar estos principios en la práctica permite orientar el desarrollo tecnológico hacia la protección de derechos fundamentales y la promoción de la justicia social (Floridi y otros., 2018).

Finalmente, la participación de la sociedad civil, la academia y las comunidades locales resulta fundamental para democratizar la gobernanza de la IA. La inclusión de estos actores permite incorporar una mayor diversidad de perspectivas, identificar riesgos sociales y fortalecer la legitimidad de las decisiones públicas. En

este sentido, la gobernanza de la IA debe entenderse como un proceso participativo, en el que se negocian valores, intereses y visiones de futuro.

8. Conclusiones

La gobernanza global de la inteligencia artificial se encuentra en un momento decisivo, caracterizado por la rápida expansión tecnológica y la ausencia de un marco regulatorio verdaderamente inclusivo. A lo largo de este capítulo se ha demostrado que la configuración actual de dicha gobernanza está profundamente marcada por asimetrías estructurales, donde un número reducido de actores —principalmente potencias tecnológicas y grandes corporaciones— concentra la capacidad de desarrollar, regular y definir el rumbo de la IA a nivel global.

Este escenario no solo refleja desigualdades preexistentes, sino que también tiene el potencial de intensificarlas. La concentración de datos, infraestructura y capacidades tecnológicas, junto con la limitada participación del Sur Global en los espacios de toma de decisiones, configura un sistema en el que los países en desarrollo enfrentan condiciones de dependencia tecnológica y subordinación normativa. En este sentido, la inteligencia artificial no puede entenderse como una tecnología neutral, sino como un campo atravesado por relaciones de poder que influyen en sus impactos sociales, económicos y políticos (Crawford, 2021).

No obstante, el análisis también ha evidenciado que estas dinámicas no son inmutables. La naturaleza aún en construcción de la gobernanza de la IA abre un espacio de oportunidad para repensar sus fundamentos y avanzar hacia modelos más equitativos. Los países en desarrollo, lejos de ser actores pasivos, pueden desempeñar un papel activo mediante el fortalecimiento de capacidades locales, la construcción de marcos regulatorios contextualizados y la articulación de alianzas regionales y multilaterales.

En este sentido, una de las principales contribuciones de este capítulo radica en proponer un enfoque de gobernanza que trascienda la lógica dominante centrada en la competencia tecnológica, para situar en el centro principios de justicia, inclusión y bienestar social. Esto implica reconocer que la regulación de la IA no es

únicamente un desafío técnico, sino un proceso político que requiere la participación de múltiples actores y la incorporación de diversas perspectivas.

Asimismo, se ha planteado la necesidad de avanzar hacia una soberanía tecnológica que permita a los países en desarrollo ejercer mayor control sobre sus recursos digitales, especialmente los datos, y reducir su dependencia de plataformas extranjeras. Este enfoque no busca aislar a los países del sistema global, sino fortalecer su capacidad de negociación y participación en condiciones más equitativas.

De igual forma, la integración de modelos alternativos de innovación, basados en principios de economía social y tecnologías abiertas, representa una vía prometedora para democratizar el acceso y uso de la inteligencia artificial. Estos modelos permiten reorientar el desarrollo tecnológico hacia el bien común, desafiando las lógicas predominantes de concentración y maximización de beneficios.

En última instancia, el desafío central no radica únicamente en regular la inteligencia artificial, sino en definir qué tipo de sociedad se desea construir a partir de su implementación. La gobernanza de la IA debe ser concebida como un proyecto colectivo, en el que se negocian valores, derechos y visiones de futuro. En este proceso, garantizar la inclusión del Sur Global no es solo una cuestión de equidad, sino una condición necesaria para construir un sistema tecnológico verdaderamente sostenible y legítimo.

Así, la inteligencia artificial representa tanto un riesgo como una oportunidad. Su impacto dependerá, en gran medida, de las decisiones que se tomen en el presente. Apostar por una gobernanza global más inclusiva, democrática y contextualizada no solo permitirá mitigar las desigualdades existentes, sino también abrir caminos hacia un desarrollo tecnológico más justo, en el que la innovación esté al servicio de la humanidad en su conjunto.

REFERENCIAS

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15.

CEPAL. (2022). *Datos, algoritmos y políticas públicas: La inteligencia artificial en América Latina y el Caribe*. Comisión Económica para América Latina y el Caribe.

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*.

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1(6), 261–262. <https://doi.org/10.1038/s42256-019-0055-y>

Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>

Graham, M., Hjorth, I., & Lehdonvirta, V. (2019). Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer*, 25(2), 135–162. <https://doi.org/10.1177/1024258919836216>

Government of India. (2026). AI Impact Summit 2026: New Delhi, 16–20 February. Government of India.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>

Kerry, C. F. (2020). *The AI governance challenge: Aligning principles to practice*. Brookings Institution.

OECD. (2019). *OECD principles on artificial intelligence*. <https://oecd.ai/en/ai-principles>

OECD. (2023). *OECD AI policy observatory*. <https://oecd.ai>

Roberts, M. E., Cows, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & Society*, 36(1), 59–77. <https://doi.org/10.1007/s00146-020-00992-2>

Rosenau, J. N. (1995). Governance in the twenty-first century. *Global Governance*, 1(1), 13–43.

Srnicek, N. (2017). *Platform capitalism*. Polity Press.

UNDP. (2021). *Human development report 2021/2022: Uncertain times, unsettled lives*. United Nations Development Programme.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization.

Veale, M., & Borgesius, F. Z. (2021). Demystifying the draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97–112.

White House Office of Science and Technology Policy. (2022). *Blueprint for an AI Bill of Rights*.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

CAPÍTULO III

GOBERNANZA ÉTICA Y REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA EDUCACIÓN SUPERIOR: PROPUESTA NORMATIVA PARA LA UNIVERSIDAD DEL CARIBE

Ethical Governance and Regulation of Artificial Intelligence in Higher
Education: A Regulatory Proposal for the University of the Caribbean

Miguel Ángel Olivares Urbina

maolivares@ucaribe.edu.mx

Universidad del Caribe

<https://orcid.org/0000-0001-6300-1666>

Ricardo Sonda de la Rosa

rsonda@ucaribe.edu.mx

Universidad del Caribe

<https://orcid.org/0000-0002-6788-4758>

Rosiluz Ceballos Povedano

rceballos@ucaribe.edu.mx

Universidad del Caribe

<https://orcid.org/0000-0002-7389-9892>

Resumen

Esta nueva propuesta de resumen abandona la enumeración convencional para centrarse en la tensión dialéctica entre la innovación desmedida y la necesidad de una soberanía institucional ética, integrando la voz crítica del Dr. Olivares Urbina: Este análisis no se limita a proponer un ajuste burocrático, sino que denuncia la asincronía normativa que atraviesa la Universidad del Caribe, cuyo reglamento de 2005 ha quedado reducido a una reliquia frente a la irrupción de la IA como una "infraestructura silenciosa" que redefine la vida académica. Dr. Olivares Urbina sostiene que es imperativo romper con el mito de la neutralidad tecnológica; aceptar la IA sin regulación es aceptar una "colonización algorítmica" que impone modelos

extranjeros ajenos a la realidad cultural de América Latina. Bajo esta mirada crítica, regular se define como un "acto de cuidado" institucional. La propuesta ataca frontalmente la opacidad de la "caja negra" algorítmica mediante un esquema de gobernanza donde la responsabilidad humana (human-in-the-loop) es innegociable: ninguna decisión automatizada puede carecer de un rostro humano que responda por sus efectos. En el terreno pedagógico, la postura es defensiva frente a la automatización del pensamiento; se acepta a la IA como un "asistente", pero se combate como sustituto del juicio crítico, tipificando las "alucinaciones" y el fraude automatizado como transgresiones graves a la esencia intelectual de la educación.

Sustentado en el rigor del método comparativo de Bereday y el modelo TPACK, el estudio concluye que la verdadera vanguardia de Unicaribe no reside en la adopción de la herramienta, sino en su capacidad de construir una gobernanza participativa. Solo mediante una alfabetización digital crítica será posible transitar de la incertidumbre hacia una tecnología con "rostro humano", donde la innovación fortalezca, y no desplace, el desarrollo intelectual de la comunidad universitaria.

Palabras clave: Inteligencia artificial; gobernanza universitaria; soberanía tecnológica; ética digital; alfabetización digital crítica; transparencia algorítmica; human-in-the-loop; colonización algorítmica; pensamiento crítico; regulación institucional.

Abstract

This new summary proposal abandons the conventional enumeration to focus on the dialectical tension between excessive innovation and the need for ethical institutional sovereignty, integrating the critical voice of Dr. Olivares Urbina: This analysis is not limited to proposing a bureaucratic adjustment, but denounces the normative asynchrony that the University of the Caribbean is going through. whose 2005 regulations have been reduced to a relic in the face of the irruption of AI as a "silent infrastructure" that redefines academic life. Dr. Olivares Urbina argues that it is imperative to break with the myth of technological neutrality; to accept AI without

regulation is to accept an "algorithmic colonization" that imposes foreign models alien to the cultural reality of Latin America. Under this critical view, regulating is defined as an institutional "act of care". The proposal attacks head-on the opacity of the algorithmic "black box" through a governance scheme where human responsibility (human-in-the-loop) is non-negotiable: no automated decision can lack a human face that responds for its effects. In the pedagogical field, the posture is defensive against the automation of thought; AI is accepted as an "assistant", but it is fought as a substitute for critical judgment, typifying "hallucinations" and automated fraud as serious transgressions to the intellectual essence of education.

Based on the rigor of Bereday's comparative method and the TPACK model, the study concludes that Unicaribe's true vanguard does not lie in the adoption of the tool, but in its ability to build participatory governance. Only through critical digital literacy will it be possible to move from uncertainty to a technology with a "human face", where innovation strengthens, and does not displace, the intellectual development of the university community.

Keywords: Artificial intelligence; university governance; technological sovereignty; digital ethics; critical digital literacy; algorithmic transparency; human-in-the-loop; algorithmic colonization; critical thinking; institutional regulation.

Introducción

Más que una simple innovación, la Inteligencia Artificial (IA) se ha infiltrado en la educación superior como una infraestructura silenciosa que hoy reconfigura, sin pedir permiso, los procesos de enseñanza e investigación. Esta transición, acelerada por la explosión de la IA generativa a finales de 2022, no solo representa un cambio técnico, sino que ha desnudado una asincronía normativa alarmante: para la Universidad del Caribe, habitar un marco regulatorio que data de 2005 no es solo un rezago administrativo, es un desafío estructural que nos deja vulnerables ante una tecnología que evoluciona a una velocidad que nuestras leyes actuales no alcanzan a comprender.

Debemos desmontar la peligrosa falacia de la neutralidad tecnológica. Como bien sostiene la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (2026), "los avances tecnológicos nunca son socialmente neutros". Detrás de cada algoritmo palpitan decisiones humanas, intereses económicos y estructuras de poder que condicionan sus resultados; por ello, en este capítulo argumentamos que reglar es una forma de cuidar. Sin criterios claros, la comunidad universitaria queda expuesta a la opacidad de los sistemas, la reproducción de sesgos y la erosión de la privacidad. La regulación, por tanto, no se propone aquí como un trámite burocrático, sino como un acto de protección institucional frente a la incertidumbre.

Este análisis rompe con visiones idealizadas para confrontar la naturaleza real de la IA: un sistema de procesamiento probabilístico o, como se ha definido con rigor en la literatura, un "loro estocástico" que repite patrones sin comprensión alguna. Bajo esta premisa, el capítulo transita de la gobernanza internacional antropocéntrica hacia la cruda realidad del contexto latinoamericano, donde la dependencia tecnológica amenaza con una "colonización algorítmica". La pregunta que atraviesa este texto no es solo cómo innovar, sino cómo construir marcos propios que respondan a nuestra realidad social y cultural sin ceder nuestra autonomía a modelos desarrollados en otros contextos.

Finalmente, el capítulo aterriza en una propuesta normativa para la Universidad del Caribe que prioriza la responsabilidad humana (human-in-the-loop) sobre la autonomía de la máquina. No buscamos frenar el avance, sino orientarlo para que la tecnología sea un medio estratégico y no un sustituto del pensamiento crítico. La meta es clara: blindar una cultura académica donde la integridad, la transparencia y el sentido humano prevalezcan sobre la automatización desmedida.

Justificación: Hacia una soberanía intelectual en la era del algoritmo

La irrupción de la Inteligencia Artificial (IA) en la educación superior ha dejado de ser una promesa de futuro para convertirse en una infraestructura operativa que redefine, en tiempo real, la médula de la investigación y la producción de conocimiento. No estamos ante una simple herramienta digital; enfrentamos una

transformación estructural que, tras la explosión de la IA generativa a finales de 2022, ha dejado al descubierto la fragilidad de marcos normativos diseñados para una realidad analógica. En la Universidad del Caribe, operar bajo reglamentos de 2005 frente a este escenario no es solo un anacronismo, es un riesgo institucional.

La falacia de la neutralidad tecnológica

Es imperativo romper con la idea de que la tecnología es un lienzo en blanco. Todo algoritmo arrastra consigo decisiones humanas, intereses de mercado y estructuras de poder que condicionan sus resultados. Como bien advierte la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (2026), “los avances tecnológicos nunca son socialmente neutros” (p. 1). Bajo esta premisa, nuestra propuesta de regulación no nace de un afán burocrático, sino de la urgencia de blindar la dignidad humana frente a la mediación algorítmica. La ausencia de reglas claras no solo genera parálisis; expone a nuestra comunidad a riesgos de opacidad y discriminación sistémica.

La literatura académica es contundente: “la regulación de la IA no busca frenar la innovación, sino garantizar que ésta ocurra con sentido humano y equidad” (Cáceres Troche y otros., 2025, p. 41). En este sentido, la gobernanza que proponemos busca equilibrar el vértigo del avance técnico con los principios de justicia social que definen nuestra identidad universitaria.

Transparencia frente a la "caja negra"

Uno de los mayores desafíos éticos radica en la opacidad de los sistemas. Sin mecanismos de supervisión, la IA tiende a amplificar prejuicios históricos derivados de sus datos de entrenamiento. Por ello, debemos exigir que “los sistemas de IA protejan los derechos humanos y eviten decisiones automatizadas que... reproduzcan prejuicios sociales existentes” (Principios de Chapultepec, 2026, p. 2). Esta transparencia no es opcional; es la base de la confianza pública.

Para contrarrestar la dilución de responsabilidades donde parece que nadie es culpable cuando un algoritmo falla, defendemos el principio de responsabilidad explícita: “toda decisión apoyada por inteligencia artificial debe tener responsables humanos” (Secretaría de Ciencia, Humanidades, Tecnología e Innovación, 2026, p. 3). En el ámbito académico, donde una decisión impacta directamente en la trayectoria de un estudiante, el factor humano debe ser siempre el juez último.

Resiliencia normativa y soberanía cultural

Reconocemos que regular un campo tan volátil es un ejercicio complejo. Pérez-Ugena (2024) acierta al señalar que “los avances tecnológicos no son claramente previsibles” (p. 11), lo que nos obliga a diseñar marcos flexibles, capaces de evolucionar sin asfixiar la innovación. Además, para una universidad mexicana, esto es un acto de resistencia ante la “colonización algorítmica”; necesitamos normas propias que no dependan de modelos externos ajenos a nuestra realidad cultural.

En última instancia, el verdadero riesgo en nuestras aulas no es el plagio, sino la posible erosión del pensamiento crítico en favor de la respuesta inmediata y automatizada. La UNESCO (2024) subraya la necesidad de “etiquetar los documentos, informes, imágenes y vídeos generados por IAGen de forma apropiada y legal” (p. 22). Al final del día, la IA debe ser una aliada estratégica para potenciar “experiencias de aprendizaje inclusivas, críticas y significativas” (Pástor y otros., 2025, p. 118). Como bien concluye Breceda Pérez (2025), la transparencia debe ser el eje rector para evitar la opacidad y consolidar una cultura académica verdaderamente ética frente a la IA.

Marco teórico

La incorporación de la Inteligencia Artificial (IA) en los entornos universitarios no puede comprenderse como un fenómeno aislado o meramente instrumental; se trata de una transformación estructural que incide directamente en la producción,

validación y circulación del conocimiento. En este sentido, el despliegue acelerado de la IA generativa hacia finales de 2022 marcó un punto de inflexión que evidenció la distancia entre el avance tecnológico y la capacidad de respuesta institucional. Tal como señala la UNESCO (2024), “las herramientas de IA generativa de acceso público (IAGen) están surgiendo rápidamente, y el lanzamiento de versiones iterativas supera la velocidad de adaptación de los marcos regulatorios nacionales”, lo que ha dado lugar a escenarios caracterizados por incertidumbre normativa y vacíos en la regulación.

Esta brecha ha derivado en la configuración de zonas grises donde se ven comprometidos aspectos fundamentales como la privacidad de los datos, la autoría intelectual y la integridad académica. Desde una perspectiva teórica, resulta necesario desmitificar la naturaleza de estas tecnologías, entendiendo que no operan bajo esquemas de conciencia o razonamiento humano, sino mediante procesos probabilísticos de lenguaje. En este sentido, la literatura especializada ha recurrido a la noción de “loros estocásticos” para describir sistemas que “solo repiten patrones lingüísticos hallados en sus datos de entrenamiento... sin comprender su significado”, lo que obliga a replantear los criterios bajo los cuales se valida el conocimiento producido con su asistencia.

Bajo este contexto, la delimitación conceptual de la IA se vuelve un punto de partida indispensable para cualquier intento de regulación. Más allá de definiciones generales, es pertinente adoptar referentes técnicos que permitan acotar su alcance funcional. Pérez-Ugena (2024), retomando el Proyecto de Reglamento de la Unión Europea, plantea que “los sistemas de IA vendrían definidos por el software que se desarrolla, empleando ciertas técnicas o estrategias, y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones”. Esta definición permite comprender a la IA no solo como herramienta, sino como sistema de intervención en procesos de toma de decisiones.

A partir de esta base conceptual, la discusión se desplaza hacia los marcos de gobernanza internacional, donde destacan los esfuerzos por articular el desarrollo tecnológico con principios éticos y de derechos humanos. La Recomendación sobre la Ética de la IA de la UNESCO (2021) constituye un referente clave al proponer un enfoque antropocéntrico, en el que la tecnología debe estar subordinada al bienestar humano. En esta misma línea, el enfoque basado en riesgos promovido por la Unión Europea introduce criterios operativos relevantes para el ámbito universitario, particularmente al establecer restricciones como la prohibición de “sistemas de reconocimiento de emociones en el lugar de trabajo y en instituciones educativas”, con el propósito de salvaguardar la intimidad y la autonomía de las personas frente a prácticas de vigilancia algorítmica.

De manera complementaria, los Principios de Chapultepec (2026) aportan una base ética que resulta pertinente para la construcción de marcos institucionales. En ellos se enfatiza la necesidad de evitar que los sistemas automatizados “reproduzcan prejuicios sociales existentes”, así como de garantizar que “toda decisión apoyada por IA debe tener responsables humanos”, incorporando además el principio de explicabilidad como condición para legitimar su uso. Estos elementos configuran un marco de referencia que trasciende lo técnico y se inserta en la discusión sobre justicia social y responsabilidad institucional.

Sin embargo, la adopción de estos principios enfrenta desafíos particulares en el contexto latinoamericano. La región se caracteriza por una marcada heterogeneidad en el desarrollo de políticas públicas en materia de IA, lo que genera dinámicas de fragmentación normativa. Mientras algunos países han avanzado en la formulación de estrategias nacionales, otros presentan rezagos significativos. Incluso en los casos más avanzados, se reconoce que “la existencia de una política nacional no garantiza per se la adopción equitativa; se requiere una ‘traducción’ de estos principios a la realidad operativa de cada institución”. Esta condición evidencia que la regulación efectiva no depende únicamente de

lineamientos generales, sino de su capacidad de adaptación a contextos específicos.

En el caso mexicano, la ausencia de un marco normativo consolidado ha propiciado escenarios donde las instituciones operan bajo esquemas de regulación implícita, frecuentemente influenciados por plataformas tecnológicas de origen extranjero. Esta situación plantea el riesgo de adoptar modelos que no necesariamente responden a las particularidades culturales, sociales y educativas del país, configurando lo que algunos autores han denominado procesos de colonización algorítmica.

En este escenario, la universidad se convierte en un espacio estratégico para la construcción de respuestas institucionales. La incorporación de la IA en la docencia y la investigación no solo plantea retos operativos, sino también éticos y pedagógicos. Uno de los debates más relevantes se centra en la integridad académica y en la necesidad de redefinir los criterios de autoría. Desde esta perspectiva, se vuelve fundamental asumir que la IA debe operar como herramienta de apoyo y no como sustituto del proceso intelectual. En este sentido, se ha señalado que “la transparencia es esencial para garantizar que el uso de la inteligencia artificial sea claramente identificado y comprendido, evitando la opacidad” (UACJ, 2025), lo que implica prácticas como la declaración explícita de su uso y la trazabilidad de los procesos de generación de contenido.

De manera paralela, el rol del docente se encuentra en un proceso de reconfiguración. Más que transmisor de contenidos, se perfila como diseñador de experiencias de aprendizaje mediadas por tecnología. Bajo esta lógica, Pástor y otros. (2025) sostienen que “la IA generativa debe concebirse no como un fin en sí misma, sino como un medio estratégico para potenciar experiencias de aprendizaje inclusivas, críticas y significativas”, lo que plantea la necesidad de desarrollar competencias pedagógicas acordes con este nuevo entorno.

En el ámbito de la investigación, los desafíos se orientan hacia la preservación del rigor científico. Problemas como la generación de referencias inexistentes o la manipulación de datos obligan a reforzar los criterios de validación y ética académica, manteniendo al investigador humano como responsable último de la veracidad del conocimiento producido.

Finalmente, al situar este análisis en el contexto de la Universidad del Caribe, se hace evidente la necesidad de actualizar sus marcos normativos. El Reglamento Escolar vigente, cuya estructura responde a lógicas previas a la digitalización intensiva, no contempla aspectos relacionados con el uso de tecnologías emergentes ni con la protección de derechos digitales. Esta condición limita la capacidad institucional para responder de manera adecuada a los retos actuales.

En este sentido, la construcción de un marco regulatorio interno debe entenderse como un proceso orientado no a restringir, sino a encauzar el uso de la tecnología. Como señala Cáceres Troche (2025), “la regulación de la IA no busca frenar la innovación, sino garantizar que ésta ocurra con sentido humano y equidad”. Para la Universidad del Caribe, esto implica avanzar hacia esquemas de gobernanza participativa que integren a la comunidad académica en la definición de criterios éticos, así como en el diseño de mecanismos que aseguren la protección de la privacidad, la transparencia en el uso de herramientas digitales y el respeto a la dignidad de los estudiantes.

Metodología

La estrategia metodológica adoptada en esta investigación responde a la necesidad de comprender un fenómeno complejo desde múltiples ángulos, evitando reducciones simplistas. En este sentido, el estudio se sustenta principalmente en una revisión documental con enfoque cualitativo, articulada a partir de un análisis comparativo que permitió contrastar distintas perspectivas teóricas y normativas. Para ello, se realizó una búsqueda sistemática en bases de datos de alto impacto

como Scopus, Web of Science y SciELO, acotando el periodo de análisis entre 2019 y 2026. La solidez de este enfoque se fortaleció mediante la triangulación de fuentes, al poner en diálogo la doctrina jurídica con informes técnicos de organismos internacionales como la UNESCO y la OCDE, lo que facilitó la identificación de categorías emergentes, entre ellas los vacíos normativos y la justicia algorítmica.

De manera complementaria, la investigación incorpora un enfoque mixto con una lógica secuencial que permite aproximarse a la dimensión humana del fenómeno. No se trata únicamente de revisar documentos, sino de entender cómo estos marcos se interpretan y se viven en la práctica. En este punto, se retoma el método comparativo de Bereday, estructurado en las fases de descripción, interpretación, yuxtaposición y comparación, como una vía para analizar la percepción docente frente al uso de la inteligencia artificial. A ello se suma la implementación de diseños secuenciales explicativos, en los que primero se recuperaron datos cualitativos mediante entrevistas y aproximaciones desde la teoría fundamentada, para posteriormente someterlos a procesos de validación a través de análisis estadísticos apoyados en herramientas como SPSS y JASP.

En paralelo, el trabajo incorpora una perspectiva técnico-jurídica que permite dar sustento normativo al análisis. Para ello, se recurrió a una metodología jurídica especializada basada en la revisión de fuentes doctrinales y jurisprudenciales, con especial atención al desarrollo del Reglamento de Inteligencia Artificial de la Unión Europea como referente internacional. Este análisis no se mantuvo aislado, sino que se articuló con la revisión de modelos instruccionales como TPACK y ADDIE, con el propósito de valorar la viabilidad pedagógica de la IA en contextos educativos, particularmente en entornos híbridos.

Finalmente, la estrategia metodológica integra una evaluación de impacto y riesgos que coloca en el centro la dimensión ética del uso de la tecnología. Más allá de la

eficiencia técnica, el análisis se orienta a examinar en qué medida estas herramientas respetan valores fundamentales y cómo pueden reproducir sesgos si no son gestionadas adecuadamente. Bajo esta lógica, la investigación no solo busca generar conocimiento, sino también ofrecer una lectura crítica que dialogue tanto con los marcos normativos como con la realidad social de las instituciones educativas.

Análisis

Para avanzar hacia un análisis puntual, conviene partir de una idea central: la regulación de la Inteligencia Artificial (IA) en el ámbito académico no puede entenderse como un ejercicio aislado ni meramente normativo. Se trata, más bien, de un proceso que se construye en la intersección de marcos de gobernanza, desarrollos tecnológicos y prácticas institucionales concretas. En este cruce, distintos organismos internacionales y aportaciones académicas han ido delineando un conjunto de principios que buscan contener los riesgos de la automatización sin frenar su potencial transformador. Bajo esta lógica, el análisis que se presenta se organiza en torno a diversas dimensiones críticas que permiten ordenar estas posturas, desde la ética hasta la seguridad técnica.

Uno de los ejes más consistentes en la literatura es el enfoque centrado en el ser humano, que coloca a la dignidad individual como criterio rector del desarrollo tecnológico. En esta línea, la UNESCO (2024) advierte la necesidad de “garantizar una visión de estas nuevas tecnologías centrada en el ser humano” (p. 2), subrayando que la innovación no puede desvincularse de sus implicaciones sociales. Esta postura encuentra eco en la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (2026), que a través de los Principios de Chapultepec sostiene que “la inteligencia artificial solo es valiosa si genera bienestar para todas las personas” (p. 1). En conjunto, ambas perspectivas refuerzan la idea de que cualquier intento de regulación debe partir de un compromiso explícito con el bienestar colectivo.

En paralelo, la transparencia y la explicabilidad se posicionan como condiciones necesarias para contrarrestar la opacidad inherente a los sistemas algorítmicos. No se trata únicamente de hacer visible el uso de la IA, sino de garantizar que sus procesos y resultados puedan ser comprendidos y evaluados. Al respecto, Breceda Pérez (2025) señala que “la transparencia es esencial para garantizar que el uso de la inteligencia artificial sea claramente identificado y comprendido, evitando la opacidad” (p. 85), lo que introduce una dimensión de control que resulta indispensable en entornos académicos.

Este principio se vincula directamente con la integridad académica, en la medida en que el uso de herramientas tecnológicas no diluye la responsabilidad sobre los resultados generados. Por el contrario, la refuerza. En este sentido, Cáceres Troche y otros. (2025a) enfatizan que “toda decisión apoyada por inteligencia artificial debe tener responsables humanos” (p. 3), dejando claro que la intervención de la IA no sustituye el juicio crítico ni la rendición de cuentas.

Ahora bien, estos principios adquieren matices particulares cuando se observan desde el contexto latinoamericano. La región enfrenta limitaciones estructurales que se expresan tanto en la persistencia de vacíos normativos como en los riesgos asociados a la justicia algorítmica. La dependencia de desarrollos tecnológicos externos ha abierto la discusión sobre posibles procesos de “colonización algorítmica”, en los que se adoptan modelos que no necesariamente dialogan con las realidades locales. Frente a este escenario, las instituciones educativas se ven obligadas a asumir un papel más activo.

En esta dirección, la formación de capacidades emerge como un elemento estratégico. Más allá de la regulación formal, la alfabetización digital se posiciona como una herramienta clave para enfrentar los desafíos del entorno tecnológico. Como lo plantean Cárdenas Gándara y otros. (2025), “la capacitación es el primer

paso para incluir herramientas tecnológicas en el campo educativo” (p. 14), lo que sugiere que la apropiación crítica de la tecnología constituye, en sí misma, una forma de regulación desde la práctica.

Como resultado del análisis antes descrito se establecen las dimensiones que toda normatividad debe considerar para su diseño, observar Tabla 1.

Tabla.- 1. Dimensiones Críticas para la Gobernanza de la IA en la Educación Superior.

Ejes de Regulación	Fundamentación Ética y Operativa	Referentes de la Discusión Académica
I. Dignidad Humana y Ética Antropocéntrica	La IA se concibe como una extensión de las capacidades humanas, nunca como un sustituto de la esencia docente o la dignidad personal.	UNESCO (2024) y Sec. Ciencia (2026) priorizan la autonomía y el bienestar social sobre la eficiencia técnica. Pérez-Ugena (2024) advierte contra la manipulación de la conducta.
II. Transparencia Operativa	Es imperativo abrir la "caja negra" algorítmica para asegurar que cada resultado sea auditable y explicable.	Breceda Pérez (2025) propone la declaración obligatoria de <i>prompts</i> . Santos González (2017) enfatiza el derecho al acceso de datos de entrada.
III. Privacidad y Soberanía de Datos	Blindaje de la información sensible frente a la vigilancia algorítmica o el <i>proctoring</i> invasivo.	Cáceres Troche (2025b) y Torres Gastelú (2025) analizan el temor estudiantil a

Ejes de Regulación	Fundamentación Ética y Operativa	Referentes de la Discusión Académica
		filtraciones y la falta de regulación nacional.
IV. Integridad y Autoría Intelectual	El pensamiento crítico es el límite: se debe garantizar que la autoría siga siendo un proceso estrictamente humano.	Lagunes Domínguez (2025) sugiere topes de uso (50%) en investigación. Otero-Escobar (2025) alerta sobre el aprendizaje mecánico sin reflexión.
V. Alfabetización Crítica	La formación no es solo técnica, sino una competencia ética necesaria para el éxito de cualquier implementación.	Según Arias Méndez (2025), el 69% de las instituciones ven la falta de capacitación como su mayor reto. Cárdenas Gándara (2025) la sitúa como paso previo a la adopción.
VI. Justicia Social y Descolonización	Evitar que los sistemas repliquen sesgos del Norte Global o profundicen brechas de género e idioma.	Cáceres Troche (2025b) introduce el concepto de "Colonización Algorítmica". Sec. Ciencia (2026) apuesta por la inclusión de lenguas indígenas.
VII. Respuesta ante Vacíos Legales	Estrategias para cerrar el desfase entre la velocidad de la innovación y la parálisis normativa institucional.	Judikis Preller (2025) insta a generar lineamientos internos ante la omisión de las leyes nacionales. Arias Méndez

Ejes de Regulación	Fundamentación Ética y Operativa	Referentes de la Discusión Académica
		reporta que el 78% identifica la falta de leyes como desafío principal.
VIII. Seguridad y Resiliencia Técnica	Protección integral contra el secuestro de identidad, hackeo y ataques que superan las defensas tradicionales.	Torres Gastelú (2025) y la UNESCO (2024) destacan la urgencia de medidas técnicas mínimas ante la velocidad de los ciberataques actuales.

Elaboración Propia fuentes Varias.

Reflexiones

Lineamientos básicos para la integridad académica y el uso de la inteligencia artificial en la Universidad del Caribe

En el contexto actual, la incorporación de la Inteligencia Artificial en la educación superior exige algo más que adopción tecnológica: requiere criterios claros que orienten su uso dentro de la vida académica. En la Universidad del Caribe, este proceso se asume desde una lógica formativa y ética, en la que la tecnología se entiende como un medio para fortalecer el aprendizaje, la investigación y la gestión institucional, sin desplazar el papel central de la persona. A partir de ello, se establecen los siguientes lineamientos básicos:

En primer término, se reconoce que el desarrollo y uso de la inteligencia artificial debe mantener como eje central a la persona. Esto implica que toda herramienta tecnológica implementada en la institución debe evaluarse en función de su contribución al bienestar de la comunidad universitaria. Bajo este principio, no se permitirá el uso de sistemas que limiten la autonomía del estudiante o que

interfieran de manera indebida en sus procesos de aprendizaje, privilegiando siempre el fortalecimiento de sus capacidades.

De manera complementaria, se establece que toda aplicación de inteligencia artificial debe estar sujeta a responsabilidad humana. Ningún proceso automatizado sustituye el juicio ético ni la toma de decisiones informada. Por ello, en ámbitos como la evaluación académica, la admisión o la gestión administrativa, deberá existir siempre una instancia de supervisión que garantice la revisión de resultados y la posibilidad de corrección. La intervención humana no es opcional, sino una condición necesaria para la legitimidad de cualquier decisión apoyada en tecnología.

Asimismo, la transparencia se asume como un principio transversal. La comunidad universitaria tiene derecho a conocer cuándo y cómo se utilizan herramientas de inteligencia artificial en los distintos procesos académicos. En este sentido, se promueve la declaración explícita de su uso, así como la documentación de los procedimientos involucrados incluyendo, cuando corresponda, los prompts o instrucciones utilizadas con el propósito de asegurar la trazabilidad y la claridad en la generación de contenidos.

En relación con la integridad académica, se establece que la inteligencia artificial debe entenderse como un apoyo y no como un sustituto del trabajo intelectual. El valor formativo de la educación universitaria radica en el desarrollo del pensamiento crítico, por lo que prácticas como la presentación de contenidos generados por IA como producción propia, la fabricación de información o el uso indebido de estas herramientas se consideran contrarias a los principios institucionales. La autoría y la reflexión siguen siendo responsabilidad del estudiante y del docente.

Por otra parte, se reconoce la importancia de la protección de datos y la seguridad digital. La información generada en los procesos académicos debe ser resguardada

bajo criterios de confidencialidad, evitando prácticas invasivas o mecanismos de vigilancia que vulneren la privacidad de los integrantes de la comunidad. Cualquier uso de datos deberá sustentarse en el consentimiento informado y en políticas claras de protección de la información.

En este mismo sentido, se incorpora el principio de equidad en el uso de la inteligencia artificial. La Universidad del Caribe promoverá la revisión crítica de las herramientas tecnológicas con el fin de identificar y mitigar posibles sesgos que puedan derivar en prácticas discriminatorias. Asimismo, se buscará privilegiar soluciones que respondan al contexto cultural y social propio, evitando la adopción acrítica de modelos externos que no consideren la diversidad local.

Por lo que se plantea la necesidad de una gobernanza participativa en materia de inteligencia artificial. La construcción de lineamientos y criterios no debe ser un proceso unilateral, sino el resultado del diálogo entre estudiantes, docentes y autoridades. Para ello, se impulsarán espacios de reflexión, comités especializados y estrategias de alfabetización digital que permitan a la comunidad no solo utilizar estas herramientas, sino comprender sus implicaciones éticas y sus límites.

Bajo estos lineamientos, la Universidad del Caribe busca integrar la inteligencia artificial de manera responsable, asegurando que su implementación contribuya al fortalecimiento académico sin comprometer los valores fundamentales que sustentan la formación universitaria.

Conclusión

El sentido de una nueva norma: hacia una tecnología con rostro humano

Más que un conjunto de disposiciones se requiere de una convicción más profunda: la Inteligencia Artificial no representa únicamente una innovación tecnológica, sino una transformación en la manera en que aprendemos, pensamos y nos

relacionamos dentro de la universidad. Su incorporación no puede reducirse a un asunto operativo; implica, necesariamente, una toma de postura ética frente a una pregunta de fondo: cómo avanzar en la innovación sin perder de vista aquello que nos define como comunidad académica.

Desde esta perspectiva, la norma no se construye como un mecanismo de control, sino como una guía que orienta el uso de la tecnología con sentido. Se reconoce que la tecnología no es neutral. Los algoritmos reflejan decisiones, intereses y contextos específicos, por lo que asumirlos como herramientas objetivas resulta limitado. El reglamento debe hacer visible su funcionamiento, evitando que procesos relevantes queden ocultos en lógicas opacas que dificulten su comprensión y evaluación.

Al mismo tiempo, responde a una necesidad institucional evidente: cerrar la brecha entre la velocidad del cambio tecnológico y la capacidad normativa de la universidad. Operar con marcos diseñados para otro contexto genera incertidumbre y deja espacios sin regulación clara. La actualización normativa busca precisamente reducir esas zonas grises, particularmente en aspectos sensibles como la privacidad, la autoría intelectual y la integridad académica, ofreciendo mayor claridad a la comunidad.

Un elemento central de esta propuesta es la recuperación de la responsabilidad humana. La tecnología puede facilitar procesos, pero no sustituye el criterio ni la ética de quienes la utilizan. Por ello, se establece que toda decisión relevante deberá contar con supervisión humana, asegurando que los procesos automatizados no se conviertan en instancias definitivas sin posibilidad de revisión. La universidad, en este sentido, mantiene el principio de que las decisiones académicas tienen un rostro y una responsabilidad clara.

De igual forma, una nueva normativa deberá incorporar una reflexión sobre el contexto en el que se inserta la institución. En entornos como el mexicano y latinoamericano, la adopción de tecnologías suele estar mediada por desarrollos externos que no siempre dialogan con las realidades locales. Frente a ello, la Universidad del Caribe asume la necesidad de utilizar la IA desde una perspectiva situada, que reconozca su diversidad cultural y social, evitando reproducir esquemas que profundicen desigualdades o dependencias tecnológicas.

La propuesta debe partir de una idea sencilla pero fundamental: la inteligencia artificial debe fortalecer, no sustituir, el desarrollo intelectual. La universidad sigue siendo un espacio donde el pensamiento crítico, la creatividad y la ética constituyen el núcleo del aprendizaje. En este marco, la IA se entiende como un apoyo que puede potenciar estos procesos, pero nunca reemplazarlos.

La normatividad no busca limitar el uso de la tecnología, sino darle dirección. Representa un paso necesario para transitar de un escenario de incertidumbre a uno de mayor claridad institucional, donde la innovación avance de la mano de principios éticos, responsabilidad compartida y compromiso con la formación integral de la comunidad universitaria.

Propuesta de normativa

Para el presente capítulo se consideró altamente pertinente generar una propuesta académica para actualizar el marco normativo de la Universidad del Caribe (rezagado desde 2005) para enfrentar el desfase entre la innovación tecnológica y la capacidad de respuesta legal, asegurando una gobernanza con sentido humano y equidad.

Propuesta de Normativa para el Uso Ético y Responsable de la Inteligencia Artificial
Universidad del Caribe

Título Primero: Disposiciones Generales, Objeto y Definiciones

Artículo 1. Objeto. El presente reglamento tiene como propósito establecer los criterios que orientan el uso de la Inteligencia Artificial (IA) dentro de la Universidad del Caribe, particularmente en las funciones de docencia, investigación y gestión administrativa. Más que limitar su adopción, busca encauzarla bajo principios éticos que fortalezcan la vida académica.

Artículo 2. Definición. Para efectos de este reglamento, se entenderá por sistema de IA aquel software capaz de generar resultados como contenidos, predicciones, recomendaciones o decisiones a partir de objetivos previamente definidos por personas. Esta definición permite ubicar a la IA como una herramienta de apoyo, no como un agente autónomo.

Artículo 3. Naturaleza de la Tecnología. La Universidad reconoce que toda tecnología tiene implicaciones sociales. En este sentido, el uso de la IA deberá estar siempre alineado con el respeto a la dignidad humana, la equidad y el bienestar de la comunidad universitaria. Su implementación no es neutra, por lo que requiere criterios claros para su adopción.

Título Segundo: De los Principios Éticos Orientadores

Artículo 4. Principios Rectores. En concordancia con los Principios de Chapultepec (2026), el uso de la IA en la Universidad del Caribe se regirá por los siguientes principios:

Salvaguarda de derechos. Los sistemas de IA deberán proteger los derechos humanos y evitar reproducir sesgos o prácticas discriminatorias.

Responsabilidad humana. Ninguna decisión relevante podrá quedar exclusivamente en manos de un sistema automatizado. Siempre deberá existir una persona responsable de su supervisión y de sus efectos.

Transparencia y explicabilidad. La comunidad universitaria tiene derecho a saber cuándo interactúa con sistemas de IA y a contar con elementos suficientes para comprender sus resultados.

Justicia algorítmica. Se evitará el uso de herramientas que profundicen desigualdades o generen exclusión por razones de género, origen, creencias o condición social.

Título Tercero: De la Integridad Académica y la Docencia

Artículo 5. Función pedagógica. La IA generativa se reconoce como una herramienta de apoyo al aprendizaje. Su uso deberá orientarse a fortalecer el pensamiento crítico y la comprensión, sin sustituir el proceso formativo del estudiante.

Artículo 6. Declaración de uso. El estudiantado deberá declarar de manera explícita el uso de herramientas de IA en sus trabajos académicos. Cuando sea necesario, se incluirá un anexo que permita conocer el proceso seguido, incluyendo las instrucciones utilizadas, con el fin de garantizar transparencia.

Artículo 7. Responsabilidad en la investigación. En los procesos de investigación, la responsabilidad sobre la veracidad de la información recae en el investigador. Se prohíbe el uso de IA para generar referencias inexistentes, alterar datos o simular resultados.

Título Cuarto: De los Derechos Digitales y la Privacidad

Artículo 8. Protección de datos. La Universidad resguardará la información personal y académica bajo criterios de seguridad y confidencialidad, procurando su uso responsable en todo momento.

Artículo 9. Prohibiciones específicas. No se permitirá el uso de sistemas de IA que impliquen vigilancia invasiva o reconocimiento automatizado de emociones, cuando estos vulneren la privacidad o la autonomía de las personas.

Título Quinto: De las Faltas y Sanciones

Artículo 10. Régimen sancionatorio. El incumplimiento de este reglamento se atenderá conforme a lo establecido en la normativa institucional vigente, particularmente el Reglamento Escolar (2005) o el que lo sustituya.

Artículo 11. Infracciones a la integridad académica. Además de las faltas ya contempladas, se considerarán infracciones graves:

Presentar como propio contenido generado íntegramente por IA sin declararlo.

Utilizar IA para cometer fraude en evaluaciones.

Generar o utilizar información falsa, incluyendo citas o datos inexistentes.

Título Sexto: Gobernanza y Capacitación

Artículo 12. Diálogo institucional. La Universidad promoverá espacios de participación donde la comunidad académica pueda opinar y contribuir en la definición de criterios sobre el uso de tecnologías emergentes.

Artículo 13. Alfabetización digital. Se impulsarán programas permanentes de formación que permitan a estudiantes y docentes utilizar la IA de manera crítica, ética y profesional.

Esta propuesta no es limitada a la universidad general su construcción analítica y con perspectiva académica permite que pueda ser utilizada en todas las instituciones de educación superior de mexicanas.

Bibliografía Justificación

Arias Méndez, J. A., Borja Torresano, S. C., Strasser López, D. F., & Loor Rodríguez, C. D. (2025). Normativas y políticas educativas sobre inteligencia artificial aplicada en entornos de aprendizaje en educación superior. *Polo del Conocimiento*, 10(6), 1428-1440.

Breceda Pérez, J. A. (2025). Hacia la regulación ética del uso de inteligencia artificial en la investigación científica, lineamientos desde la Universidad Autónoma de Ciudad Juárez. *Revista de Bioética y Derecho*, (64), 81-97.

Cáceres Troche, D., Lagunes Domínguez, A., & Lagunes Domínguez, P. (2025). Análisis de marcos regulatorios de IA en educación superior: casos México y Paraguay en contexto latinoamericano. En C. A. Torres Gastelú y otros. (Eds.), *Uso, regulación y retos de la inteligencia artificial en la educación superior*. CIATA.

Cáceres Troche, D., Lagunes Domínguez, P., & Torres Gastelú, C. A. (2025). Privacidad y vigilancia algorítmica en universidades latinoamericanas. En C. A. Torres Gastelú y otros. (Eds.), *Uso, regulación y retos de la inteligencia artificial en la educación superior*. CIATA.

Cárdenas Gándara, T. J., Cornejo Álvarez, J. F., & Frausto Loera, M. Á. (2025). Inteligencia artificial generativa en el contexto universitario: un estudio de seguimiento sobre la percepción docente. *Transdigital*, 6(12), e484.

Judikis Preller, J. C., & Ruiz Fernández, N. (2025). Política, regulación y desafíos de la inteligencia artificial en la educación superior chilena. En C. A. Torres Gastelú y otros. (Eds.), *Uso, regulación y retos de la inteligencia artificial en la educación superior*. CIATA.

Lagunes Domínguez, A., Torres Gastelú, C. A., & Lagunes Domínguez, P. (2025). Rol de la Inteligencia Artificial en la investigación académica. En C. A.

Torres Gastelú y otros. (Eds.), *Uso, regulación y retos de la inteligencia artificial en la educación superior*. CIATA.

Otero-Escobar, A. D., & Suárez-Jasso, E. (2025). Responsabilidad ética del uso de ChatGPT en estudiantes universitarios. *Transdigital*, 6(11), e406.

Paredes Labra, J. (2025). Confianza del usuario en sistemas de IA. Revisión de literatura. En C. A. Torres Gastelú y otros. (Eds.), *Uso, regulación y retos de la inteligencia artificial en la educación superior*. CIATA.

Pástor, D., Bastidas Guacho, G. K., & Duque Vaca, M. A. (2025). Diseño instruccional potenciado por Inteligencia Artificial Generativa: Herramientas y estrategias para la enseñanza en línea. En C. A. Torres Gastelú y otros. (Eds.), *Uso, regulación y retos de la inteligencia artificial en la educación superior*. CIATA.

Pérez-Ugena, M. (2024). La inteligencia artificial: definición, regulación y riesgos para los derechos fundamentales. *Estudios de Deusto*, 72(1), 307-337.

Santos González, M. J. (2017). Regulación legal de la robótica y la inteligencia artificial: retos de futuro. *Revista Jurídica de la Universidad de León*, (4), 25-50.

Secretaría de Ciencia, Humanidades, Tecnología e Innovación. (2026). *Principios de Chapultepec: Declaración de ética y buenas prácticas para el uso y desarrollo de la inteligencia artificial*. Agencia de Transformación Digital y Telecomunicaciones.

Torres Gastelú, C. A., Lagunes Domínguez, A., & Diez Comi, M. (2025). Confianza y temores de los estudiantes hacia la Inteligencia Artificial. En C. A. Torres Gastelú y otros. (Eds.), *Uso, regulación y retos de la inteligencia artificial en la educación superior*. CIATA.

UNESCO. (2024). Guía para el uso de IA generativa en educación e investigación. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

CAPÍTULO IV

GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL Y TRABAJO: DESAFÍOS REGULATORIOS ANTE EL DESPLAZAMIENTO LABORAL EN EL SIGLO XXI

Artificial Intelligence Governance and Labor: Regulatory Challenges in the
Face of Job Displacement in the 21st Century

Didre Estefani Morales Rodriguez

estefi.mor28@gmail.com

Universidad Autónoma Metropolitana, Unidad Iztapalapa

<https://orcid.org/0009-0007-6958-9772>

Resumen

El presente capítulo analiza la relación entre inteligencia artificial y trabajo desde una perspectiva histórica, empírica y normativa, con el objetivo de responder a una pregunta central: *¿es posible regular la inteligencia artificial para limitar el desplazamiento de la mano de obra humana?* A partir de una revisión conceptual, se establece que la inteligencia artificial, particularmente en su vertiente generativa, representa un cambio cualitativo en los procesos de automatización al incidir no solo en tareas físicas, sino también en funciones cognitivas y creativas. Posteriormente, el análisis empírico muestra que, si bien la IA no está provocando una eliminación masiva del empleo, sí está generando una reconfiguración profunda del mercado laboral, caracterizada por la transformación de tareas, la polarización de habilidades y la intensificación de desigualdades.

En el plano regulatorio, el capítulo identifica límites estructurales derivados de la velocidad del cambio tecnológico, la globalización de la economía digital y la complejidad de los sistemas algorítmicos. A través de un análisis comparado, se examinan los enfoques de regulación en la Unión Europea, Estados Unidos, China y América Latina, evidenciando la ausencia de un modelo único y la influencia de

contextos políticos y económicos en la gobernanza de la inteligencia artificial. En el caso de México, se identifican avances incipientes y vacíos regulatorios significativos, particularmente en relación con el ámbito laboral.

Finalmente, el capítulo propone un marco regulatorio basado en principios de justicia social, transparencia y adaptabilidad, acompañado de instrumentos como evaluaciones de impacto laboral, fortalecimiento de derechos digitales y políticas de reconversión profesional. Se concluye que la inteligencia artificial no es un fenómeno inevitable e incontrolable, sino un proceso susceptible de ser orientado mediante decisiones políticas y mecanismos de gobernanza que permitan equilibrar innovación tecnológica y protección del trabajo humano.

Palabras clave: Inteligencia artificial; automatización; mercado laboral; desplazamiento laboral; regulación tecnológica; justicia social; derechos digitales; gobernanza algorítmica; reconversión profesional; trabajo humano.

Abstract

This chapter analyzes the relationship between artificial intelligence and work from a historical, empirical and normative perspective, with the aim of answering a central question: *is it possible to regulate artificial intelligence to limit the movement of human labor?* Based on a conceptual review, it is established that artificial intelligence, particularly in its generative aspect, represents a qualitative change in automation processes by affecting not only physical tasks, but also cognitive and creative functions. Subsequently, the empirical analysis shows that, although AI is not causing a massive elimination of employment, it is generating a profound reconfiguration of the labor market, characterized by the transformation of tasks, the polarization of skills and the intensification of inequalities.

On the regulatory level, the chapter identifies structural limits derived from the speed of technological change, the globalization of the digital economy and the complexity of algorithmic systems. Through a comparative analysis, regulatory approaches in

the European Union, the United States, China and Latin America are examined, evidencing the absence of a single model and the influence of political and economic contexts on the governance of artificial intelligence. In the case of Mexico, incipient advances and significant regulatory gaps are identified, particularly in relation to the labor field.

Finally, the chapter proposes a regulatory framework based on principles of social justice, transparency and adaptability, accompanied by instruments such as labor impact assessments, strengthening of digital rights and professional retraining policies. It is concluded that artificial intelligence is not an inevitable and uncontrollable phenomenon, but a process that can be guided by political decisions and governance mechanisms that allow balancing technological innovation and protection of human work.

Keywords: Artificial intelligence; automation; labor market; labor displacement; technological regulation; social justice; digital rights; algorithmic governance; professional retraining; human work.

Introducción

La irrupción de la inteligencia artificial en la vida económica y social contemporánea ha reconfigurado de manera acelerada las formas en que se produce, se organiza y se valora el trabajo. Aunque el desarrollo de esta tecnología tiene antecedentes que se remontan a varias décadas, su expansión reciente —particularmente a partir del auge de la inteligencia artificial generativa— ha intensificado el debate en torno a sus implicaciones laborales, generando tanto expectativas de productividad como preocupaciones sobre el desplazamiento de la mano de obra humana.

En este contexto, el trabajo, entendido tradicionalmente como el eje articulador de la vida económica y social, se encuentra en un proceso de transformación profunda. La capacidad de los sistemas de inteligencia artificial para ejecutar tareas cognitivas, generar contenido y participar en procesos de toma de decisiones

introduce una ruptura con las formas históricas de automatización, las cuales se habían concentrado principalmente en actividades físicas o rutinarias. Como resultado, sectores que anteriormente se consideraban protegidos frente a la sustitución tecnológica —como los servicios profesionales, la educación o la producción creativa— comienzan a experimentar cambios significativos en su estructura laboral.

Sin embargo, el impacto de la inteligencia artificial en el empleo no puede reducirse a una narrativa de sustitución total. La evidencia empírica muestra que, más que eliminar trabajos de manera masiva, la IA está reconfigurando el contenido del trabajo, transformando tareas, modificando competencias y generando nuevas formas de organización laboral. Este proceso, no obstante, no es homogéneo ni neutral, ya que tiende a beneficiar a ciertos sectores y perfiles laborales, mientras que incrementa la vulnerabilidad de otros, particularmente aquellos vinculados a tareas rutinarias o con menor acceso a formación tecnológica.

Frente a este escenario, surge una interrogante fundamental: ¿quién regula la inteligencia artificial en el ámbito laboral y hasta qué punto es posible establecer límites a su uso para evitar el desplazamiento de la mano de obra humana? Esta pregunta adquiere especial relevancia en un contexto en el que la innovación tecnológica avanza a un ritmo superior al de los marcos regulatorios, generando vacíos normativos y desafíos para las instituciones encargadas de garantizar la protección del trabajo.

El problema se complejiza aún más si se considera la naturaleza global de la economía digital. A diferencia de procesos tecnológicos anteriores, la inteligencia artificial se desarrolla en entornos transnacionales, lo que limita la capacidad de los Estados para regular de manera autónoma sus efectos. Asimismo, la concentración del desarrollo tecnológico en grandes corporaciones introduce nuevas dinámicas de poder que desafían los mecanismos tradicionales de gobernanza.

En este marco, el presente capítulo tiene como objetivo analizar el impacto de la inteligencia artificial en el trabajo y explorar las posibilidades y límites de su regulación, con especial énfasis en el contexto mexicano. Para ello, se adopta un enfoque multidimensional que integra tres niveles de análisis. En primer lugar, se desarrolla un marco conceptual que permite comprender la relación entre inteligencia artificial, trabajo y regulación, incorporando el papel central de la IA generativa como catalizador del cambio actual. En segundo lugar, se examina la evolución histórica de la relación entre tecnología y empleo, así como la evidencia empírica reciente sobre el impacto de la IA en el mercado laboral. Finalmente, se analiza el panorama internacional de la regulación, se estudia el caso de México y se propone un conjunto de lineamientos orientados a construir un marco de gobernanza de la inteligencia artificial centrado en el trabajo humano.

La relevancia de este análisis radica en la necesidad de superar enfoques deterministas que conciben la tecnología como una fuerza autónoma e inevitable. Por el contrario, se parte de la premisa de que la inteligencia artificial es un fenómeno socialmente construido, cuyo desarrollo y aplicación pueden ser orientados a través de decisiones políticas, institucionales y organizacionales. En este sentido, la regulación no debe entenderse como un obstáculo para la innovación, sino como una herramienta para garantizar que los beneficios del progreso tecnológico se distribuyan de manera equitativa y que sus costos no recaigan de forma desproporcionada en los sectores más vulnerables.

En consecuencia, el capítulo propone una reflexión crítica sobre el futuro del trabajo en la era de la inteligencia artificial, destacando la importancia de construir marcos regulatorios que no solo respondan a los cambios tecnológicos, sino que anticipen sus efectos y promuevan una integración más justa, inclusiva y sostenible de la IA en las dinámicas laborales contemporáneas.

2. Marco conceptual: inteligencia artificial, trabajo y regulación

2.1 ¿Qué entendemos por inteligencia artificial en el trabajo?

La comprensión contemporánea de la inteligencia artificial en el ámbito laboral exige trascender definiciones estrictamente técnicas para situarla como un fenómeno sociotécnico que reconfigura las dinámicas productivas, organizacionales y cognitivas del trabajo. En términos generales, la inteligencia artificial puede entenderse como un conjunto de sistemas computacionales capaces de ejecutar tareas que tradicionalmente requerían capacidades humanas, tales como el aprendizaje, el razonamiento, la toma de decisiones y la generación de contenido. Sin embargo, su relevancia en el entorno laboral no radica únicamente en su capacidad técnica, sino en su potencial para alterar las formas en que el trabajo se organiza, se distribuye y se valora.

Desde esta perspectiva, la IA no debe concebirse como una herramienta aislada, sino como un actor que interviene activamente en los procesos organizacionales. En este sentido, diversos estudios han señalado que la inteligencia artificial no solo optimiza tareas, sino que también reconfigura las relaciones de poder, los flujos de información y los mecanismos de control dentro de las organizaciones (Sansores Guerrero y otros., 2025, p. 14). Esta dimensión resulta fundamental para comprender que el impacto de la IA en el trabajo no se limita a la eficiencia productiva, sino que implica una transformación estructural de las dinámicas laborales.

En este marco, es pertinente distinguir entre diferentes tipos de inteligencia artificial. Por un lado, la denominada inteligencia artificial débil o estrecha (narrow AI) se refiere a sistemas diseñados para ejecutar tareas específicas dentro de un dominio limitado, como el análisis de datos, la clasificación de información o la automatización de procesos administrativos. Por otro lado, la inteligencia artificial general (AGI), aún no desarrollada plenamente, implicaría la capacidad de realizar cualquier tarea cognitiva humana con un nivel de autonomía y adaptabilidad

comparable al de las personas. No obstante, resulta crucial subrayar que los impactos actuales en el mercado laboral se derivan casi exclusivamente de la IA débil, lo cual pone en evidencia que no es necesario alcanzar niveles de inteligencia general para generar disrupciones significativas en el empleo.

Aunado a esta distinción, la literatura contemporánea ha enfatizado la diferencia entre automatización y augmentación como dos formas fundamentales de interacción entre la inteligencia artificial y el trabajo humano. Mientras que la automatización implica la sustitución directa de tareas previamente realizadas por personas, la augmentación refiere a la capacidad de la IA para complementar y potenciar las habilidades humanas. Esta distinción resulta clave, ya que el impacto de la IA en el empleo no es homogéneo, sino que depende de la forma en que se integre en los procesos productivos. En este sentido, Acemoglu y Restrepo (2020) sostienen que la automatización puede generar desplazamiento laboral, pero también abre espacios para la creación de nuevas tareas, configurando así un proceso dinámico de transformación del trabajo (p. 2190).

Sin embargo, en el contexto actual, la discusión sobre inteligencia artificial en el trabajo no puede comprenderse plenamente sin incorporar el análisis de la inteligencia artificial generativa, la cual representa un punto de inflexión en la evolución tecnológica reciente. A diferencia de las formas tradicionales de automatización, la IA generativa se caracteriza por su capacidad para producir contenido original, incluyendo texto, imágenes, código y otros formatos complejos, a partir de modelos entrenados con grandes volúmenes de datos.

La emergencia de la IA generativa, particularmente a partir de 2023, ha marcado un cambio cualitativo en la relación entre tecnología y trabajo. Como se señala en desarrollos recientes, esta tecnología no solo automatiza tareas rutinarias, sino que incide directamente en actividades cognitivas y creativas que históricamente habían sido consideradas exclusivas del trabajo humano (Lozano Carrillo, 2024, p. 12).

Este fenómeno introduce una ruptura con las lógicas previas de automatización, ya que amplía significativamente el espectro de ocupaciones potencialmente afectadas.

Además, la IA generativa ha contribuido a una democratización sin precedentes en el acceso a herramientas avanzadas, lo que ha acelerado su adopción en múltiples sectores laborales. El crecimiento exponencial de plataformas como ChatGPT, que alcanzaron un millón de usuarios en apenas cinco días, ilustra la velocidad con la que estas tecnologías se integran en la vida cotidiana y en los entornos de trabajo (Lozano Carrillo, 2024, p. 13). Este ritmo de adopción no solo intensifica su impacto económico, sino que también dificulta la capacidad de los marcos regulatorios para responder de manera oportuna.

Desde una perspectiva crítica, la IA generativa introduce nuevas tensiones en el ámbito laboral. Por un lado, incrementa la productividad individual al permitir que un solo trabajador realice tareas que anteriormente requerían equipos completos. Por otro lado, genera procesos de sustitución parcial que no siempre se traducen en despidos directos, sino en una reducción progresiva de la demanda laboral, lo que puede derivar en formas más sutiles de precarización. De acuerdo con el informe de Goldman Sachs (2023), se estima que hasta 300 millones de empleos a nivel global podrían verse afectados por la automatización impulsada por IA generativa, particularmente en sectores intensivos en conocimiento.

Asimismo, estudios recientes de la Organización Internacional del Trabajo (OIT, 2023) advierten que el impacto de la IA generativa no será uniforme, sino que afectará de manera diferenciada según el nivel educativo, el tipo de ocupación y el contexto económico, siendo los trabajos administrativos y de oficina los más expuestos a procesos de automatización parcial. Este enfoque coincide con la idea de que la IA no elimina el trabajo en su totalidad, sino que lo transforma, redistribuyendo tareas y redefiniendo competencias.

En este sentido, la inteligencia artificial generativa no solo redefine qué tareas pueden ser automatizadas, sino que también cuestiona los fundamentos mismos del valor del trabajo en las economías contemporáneas. Al desplazar funciones cognitivas hacia sistemas algorítmicos, se genera una tensión entre la eficiencia tecnológica y la dignidad del trabajo humano, lo que plantea interrogantes profundas sobre los límites éticos y regulatorios de su implementación.

En suma, la inteligencia artificial en el ámbito laboral debe entenderse como un fenómeno complejo y multidimensional que combina elementos tecnológicos, organizacionales y sociales. La incorporación de la IA generativa en este análisis permite reconocer que nos encontramos ante una transformación estructural del trabajo, caracterizada por la aceleración del cambio, la ampliación del espectro de automatización y la emergencia de nuevos desafíos regulatorios. En consecuencia, cualquier aproximación al estudio de la IA y el trabajo debe considerar no solo sus capacidades técnicas, sino también sus implicaciones en términos de poder, desigualdad y gobernanza.

2.2 Trabajo humano en transformación

El análisis del impacto de la inteligencia artificial en el ámbito laboral requiere, necesariamente, una revisión crítica del concepto mismo de trabajo, el cual ha experimentado transformaciones profundas a lo largo de la historia, pero que en la actualidad se encuentra en un punto de inflexión particularmente significativo. En este contexto, el trabajo deja de ser concebido exclusivamente como una actividad productiva vinculada a la generación de bienes y servicios, para entenderse como un fenómeno social complejo, atravesado por dimensiones simbólicas, cognitivas y tecnológicas.

Desde una perspectiva histórica, el trabajo ha estado íntimamente ligado al desarrollo tecnológico; sin embargo, la actual revolución impulsada por la inteligencia artificial introduce un elemento distintivo: la capacidad de las máquinas para intervenir no solo en procesos físicos, sino también en funciones cognitivas y

decisionales. Esta transición implica que el trabajo humano ya no se define únicamente por la ejecución de tareas, sino por su capacidad de adaptación, interpretación y generación de valor en entornos altamente automatizados.

En este sentido, diversos enfoques contemporáneos han señalado que nos encontramos ante una transición desde una economía basada en el conocimiento hacia una economía algorítmica, en la cual los sistemas inteligentes no solo procesan información, sino que participan activamente en la producción de conocimiento. Como se advierte en estudios recientes sobre organizaciones, la inteligencia artificial redefine la relación entre conocimiento y trabajo, desplazando parte de las funciones cognitivas hacia sistemas automatizados que operan con grandes volúmenes de datos (Sansores Guerrero y otros., 2025, p. 11). Esta transformación implica que el conocimiento deja de ser un recurso exclusivamente humano para convertirse en un activo compartido entre personas y máquinas.

A partir de este cambio, el trabajo humano se reconfigura en al menos tres dimensiones fundamentales. En primer lugar, se observa una creciente fragmentación de las tareas, donde las actividades laborales son descompuestas en unidades específicas susceptibles de ser automatizadas o asistidas por sistemas de inteligencia artificial. Este fenómeno, ampliamente documentado en la literatura sobre automatización, genera una redistribución del valor del trabajo, en la medida en que las tareas más rutinarias tienden a ser absorbidas por la tecnología, mientras que las actividades que requieren juicio crítico, creatividad o interacción social adquieren mayor relevancia.

En segundo lugar, se evidencia una transformación en las competencias laborales, caracterizada por una creciente demanda de habilidades híbridas que combinan capacidades técnicas, cognitivas y socioemocionales. En este contexto, el trabajador contemporáneo no solo debe dominar herramientas digitales, sino también desarrollar competencias como el pensamiento crítico, la adaptabilidad y la capacidad de aprendizaje continuo. Este proceso se encuentra estrechamente vinculado con lo que algunos autores denominan la “obsolescencia acelerada del

conocimiento”, en la cual las habilidades pierden vigencia en periodos cada vez más cortos debido al ritmo vertiginoso del cambio tecnológico (Lozano Carrillo, 2024, p. 14).

En tercer lugar, el trabajo se ve atravesado por una reconfiguración de las relaciones organizacionales, en la medida en que la inteligencia artificial introduce nuevas formas de control, supervisión y toma de decisiones. En este sentido, la incorporación de sistemas algorítmicos en la gestión laboral no solo optimiza procesos, sino que también redefine las dinámicas de poder dentro de las organizaciones. Tal como se ha señalado desde la perspectiva de la intervención organizacional, el uso de tecnologías inteligentes implica una reconfiguración de los significados, las relaciones y las estructuras que configuran la vida organizacional (Lozano Carrillo y otros., 2024, p. 2). De esta forma, el trabajo deja de ser un espacio exclusivamente humano para convertirse en un entorno híbrido en el que interactúan actores humanos y no humanos.

Ahora bien, uno de los elementos más relevantes en esta transformación es la emergencia de lo que podría denominarse un modelo de trabajo mediado por inteligencia artificial, en el cual las actividades laborales se realizan en interacción constante con sistemas inteligentes. Este modelo no solo redefine la forma en que se ejecutan las tareas, sino también la manera en que se construye la identidad profesional. En este sentido, el trabajador ya no se define únicamente por su conocimiento o experiencia, sino por su capacidad para interactuar eficazmente con tecnologías avanzadas.

Este fenómeno adquiere particular relevancia en el contexto de la inteligencia artificial generativa, la cual ha ampliado significativamente el alcance de la automatización hacia actividades que anteriormente eran consideradas intrínsecamente humanas. En efecto, la capacidad de estos sistemas para generar textos, diseños, código o estrategias implica que el trabajo creativo y cognitivo se encuentra cada vez más mediado por herramientas tecnológicas. Como se ha documentado en experiencias recientes en organizaciones y microempresas, la IA

puede funcionar como una “prótesis cognitiva” que amplifica las capacidades humanas, permitiendo a los trabajadores realizar tareas complejas con mayor eficiencia (Lozano Carrillo y otros., 2025, p. 12).

No obstante, esta ampliación de capacidades también introduce tensiones significativas. Por un lado, la interacción con sistemas inteligentes puede potenciar la productividad y abrir nuevas oportunidades laborales; por otro, puede generar procesos de dependencia tecnológica que reducen la autonomía del trabajador. Asimismo, la creciente mediación algorítmica en el trabajo plantea interrogantes sobre la pérdida de habilidades, la deshumanización de ciertos procesos laborales y la redefinición del valor del esfuerzo humano.

Desde una perspectiva global, diversos estudios empíricos han confirmado que la transformación del trabajo impulsada por la inteligencia artificial no se traduce únicamente en la eliminación de empleos, sino en una reconfiguración estructural del mercado laboral. El informe del World Economic Forum (2023) estima que, si bien la automatización podría desplazar alrededor de 83 millones de empleos para 2027, también podría generar aproximadamente 69 millones de nuevos puestos, lo que evidencia un proceso de transición más que de sustitución total. Sin embargo, esta transición no es neutral, ya que tiende a beneficiar a trabajadores con mayores niveles de educación y acceso a tecnología, profundizando así las desigualdades existentes.

En el caso de América Latina y particularmente de México, esta transformación se ve condicionada por factores estructurales como la informalidad laboral, la desigualdad educativa y la brecha digital. En este contexto, la incorporación de la inteligencia artificial en el trabajo no solo implica desafíos tecnológicos, sino también sociales y políticos, en la medida en que puede amplificar las desigualdades si no se acompaña de políticas públicas adecuadas.

En consecuencia, el trabajo humano en la era de la inteligencia artificial no desaparece, pero sí se transforma de manera profunda, desplazándose desde la ejecución de tareas hacia la gestión de procesos, la interpretación de información

y la interacción con sistemas inteligentes. Esta transición plantea la necesidad de repensar no solo las competencias laborales, sino también los marcos regulatorios, educativos y organizacionales que sustentan el trabajo en las sociedades contemporáneas.

En suma, la transformación del trabajo impulsada por la inteligencia artificial debe entenderse como un proceso multidimensional que involucra cambios en las tareas, las competencias, las relaciones organizacionales y las estructuras económicas. Lejos de representar una simple sustitución tecnológica, este fenómeno implica una reconfiguración profunda del significado mismo del trabajo, lo que obliga a reconsiderar su papel en la construcción de la identidad, la cohesión social y el desarrollo económico.

2.3 Regulación tecnológica: actores, enfoques y límites en la era de la inteligencia artificial

La creciente incorporación de la inteligencia artificial en el ámbito laboral no solo plantea desafíos técnicos y organizacionales, sino que introduce una problemática central de carácter normativo: la capacidad —y los límites— de las sociedades contemporáneas para regular tecnologías que evolucionan a un ritmo significativamente más rápido que los marcos jurídicos tradicionales. En este sentido, la regulación de la inteligencia artificial debe entenderse como un campo en construcción, caracterizado por tensiones entre innovación, control, derechos laborales y dinámicas económicas globales.

En términos generales, la regulación tecnológica puede definirse como el conjunto de normas, principios y mecanismos institucionales orientados a orientar, limitar o encauzar el desarrollo y uso de tecnologías en función de objetivos sociales, económicos y éticos. Sin embargo, en el caso específico de la inteligencia artificial, esta regulación enfrenta una dificultad estructural: su objeto de regulación no es estático, sino dinámico, adaptable y, en muchos casos, opaco. A diferencia de otras tecnologías industriales, la IA no solo ejecuta instrucciones predefinidas, sino que

aprende, evoluciona y toma decisiones en contextos cambiantes, lo que complica su supervisión y control.

Desde una perspectiva analítica, es posible distinguir al menos tres enfoques predominantes en la regulación de la inteligencia artificial: el enfoque preventivo, el enfoque reactivo y el enfoque basado en principios éticos. El enfoque preventivo busca anticipar riesgos y establecer límites antes de que la tecnología genere impactos negativos significativos; en contraste, el enfoque reactivo responde a daños ya ocurridos, ajustando la normativa a partir de la experiencia empírica; finalmente, el enfoque ético se centra en la formulación de principios orientadores —como la transparencia, la equidad o la responsabilidad— que guían el desarrollo y uso de la IA más allá de la regulación estrictamente jurídica.

No obstante, estos enfoques no operan de manera aislada, sino que coexisten y, en ocasiones, entran en tensión. Mientras que una regulación excesivamente preventiva puede inhibir la innovación, una regulación meramente reactiva puede resultar insuficiente frente a los efectos disruptivos de la inteligencia artificial. En este contexto, diversos estudios han señalado la necesidad de adoptar modelos híbridos que combinen anticipación, flexibilidad y adaptación continua (Floridi y otros., 2018).

Aunado a esta complejidad, la regulación de la inteligencia artificial involucra una pluralidad de actores que operan en distintos niveles. Tradicionalmente, el Estado ha sido el principal responsable de establecer marcos normativos; sin embargo, en el caso de la IA, su capacidad regulatoria se ve limitada por la naturaleza global de las tecnologías digitales y por la concentración del desarrollo tecnológico en grandes corporaciones transnacionales. En consecuencia, la regulación se configura como un proceso multinivel en el que participan gobiernos, empresas, organismos internacionales, academia y sociedad civil.

En este entramado, las empresas tecnológicas desempeñan un papel particularmente relevante, ya que no solo desarrollan las herramientas de inteligencia artificial, sino que también establecen, de facto, reglas de uso a través

de sus plataformas y servicios. Esta situación ha llevado a algunos autores a plantear la existencia de una “regulación privada” o “gobernanza algorítmica corporativa”, en la cual las decisiones sobre el funcionamiento de la tecnología se toman fuera de los marcos democráticos tradicionales. Tal como se advierte en estudios sobre organizaciones, la inteligencia artificial introduce nuevas formas de poder que operan a través de datos, algoritmos y sistemas de decisión automatizada (Sansores Guerrero y otros., 2025, p. 12).

Por su parte, los organismos internacionales han comenzado a desempeñar un rol cada vez más activo en la construcción de marcos de referencia globales. Instituciones como la Organización para la Cooperación y el Desarrollo Económicos (OCDE), la UNESCO y la Unión Europea han impulsado principios y lineamientos orientados a promover una inteligencia artificial “centrada en el ser humano”. Sin embargo, la efectividad de estos instrumentos depende en gran medida de su adopción por parte de los Estados y de su articulación con contextos locales específicos.

En el ámbito laboral, la regulación de la inteligencia artificial enfrenta desafíos adicionales, ya que implica no solo la protección de derechos individuales, sino también la preservación de estructuras sociales fundamentales como el empleo, la seguridad social y la estabilidad económica. En este sentido, uno de los principales dilemas regulatorios radica en determinar hasta qué punto es posible —y deseable— limitar el uso de tecnologías que incrementan la productividad, pero que pueden generar desplazamiento laboral.

Diversos enfoques han intentado responder a esta problemática. Algunos autores han propuesto la implementación de impuestos a la automatización, con el objetivo de redistribuir los beneficios económicos generados por la IA; otros han planteado la necesidad de establecer evaluaciones de impacto laboral previas a la adopción de tecnologías automatizadas; asimismo, se ha discutido la posibilidad de reconocer nuevos derechos laborales en entornos digitales, como el derecho a la

desconexión, la transparencia algorítmica o la supervisión humana de decisiones automatizadas.

No obstante, estas propuestas enfrentan limitaciones significativas. En primer lugar, la naturaleza global de la economía digital dificulta la aplicación de regulaciones nacionales, ya que las empresas pueden trasladar sus operaciones a jurisdicciones con menores restricciones. En segundo lugar, la velocidad del desarrollo tecnológico supera con frecuencia la capacidad de respuesta de los sistemas legislativos, lo que genera un desfase constante entre innovación y regulación. Finalmente, la complejidad técnica de la inteligencia artificial limita la capacidad de los reguladores para comprender plenamente los sistemas que buscan supervisar.

En este contexto, resulta fundamental reconocer que la regulación de la inteligencia artificial no puede basarse exclusivamente en instrumentos jurídicos tradicionales, sino que requiere enfoques más amplios que integren dimensiones éticas, sociales y organizacionales. Como se ha señalado desde la perspectiva de la intervención organizacional, las transformaciones tecnológicas no solo implican cambios técnicos, sino también reconfiguraciones en los significados, las prácticas y las relaciones de poder que estructuran la vida social (Lozano Carrillo y otros., 2024, p. 2). En consecuencia, la regulación debe considerar no solo los efectos visibles de la IA, sino también sus impactos en la cultura organizacional, la identidad laboral y la cohesión social.

Asimismo, la irrupción de la inteligencia artificial generativa intensifica estos desafíos regulatorios, ya que introduce nuevos problemas relacionados con la autoría, la propiedad intelectual, la veracidad de la información y la responsabilidad por decisiones automatizadas. A diferencia de tecnologías anteriores, la IA generativa puede ser utilizada de manera individual por millones de usuarios, lo que dificulta su control a través de mecanismos tradicionales de regulación. Esta característica plantea la necesidad de desarrollar enfoques regulatorios que combinen supervisión institucional con responsabilidad individual y organizacional.

En suma, la regulación tecnológica en la era de la inteligencia artificial se configura como un campo complejo y en constante evolución, en el que convergen múltiples actores, intereses y enfoques. Lejos de constituir un mecanismo de control absoluto, la regulación debe entenderse como un proceso dinámico de negociación social orientado a equilibrar innovación y protección, eficiencia y equidad, desarrollo tecnológico y dignidad humana. En este sentido, el desafío no radica únicamente en regular la inteligencia artificial, sino en definir colectivamente los principios que deben guiar su integración en el mundo del trabajo.

3. Evolución histórica: tecnología y desplazamiento laboral

3.1 Revoluciones industriales y empleo: antecedentes del cambio tecnológico

El análisis del impacto de la inteligencia artificial en el trabajo no puede comprenderse de manera aislada, sino que debe situarse dentro de una trayectoria histórica más amplia en la que la tecnología ha sido un factor determinante en la transformación de las estructuras productivas y laborales. A lo largo de las distintas revoluciones industriales, el desarrollo tecnológico ha generado procesos simultáneos de desplazamiento, transformación y creación de empleo, configurando dinámicas complejas que desafían interpretaciones simplistas sobre la relación entre tecnología y trabajo.

La Primera Revolución Industrial, iniciada a finales del siglo XVIII, marcó el tránsito de economías agrarias a sistemas industriales basados en la mecanización. La introducción de máquinas impulsadas por energía hidráulica y vapor transformó radicalmente los procesos productivos, desplazando formas tradicionales de trabajo artesanal y generando nuevas dinámicas laborales en las fábricas. Este proceso, si bien incrementó la productividad, también provocó tensiones sociales significativas, como lo evidencian los movimientos luditas, que expresaban el temor al reemplazo del trabajo humano por máquinas.

Posteriormente, la Segunda Revolución Industrial profundizó esta transformación mediante la introducción de la electricidad, la producción en masa y la estandarización de procesos. En esta etapa, el trabajo se reorganizó bajo principios tayloristas y fordistas, caracterizados por la división del trabajo en tareas simples y repetitivas. Si bien esta lógica permitió una expansión significativa del empleo industrial, también implicó una descalificación relativa del trabajo, en la medida en que las habilidades artesanales fueron sustituidas por funciones altamente especializadas y fragmentadas.

La Tercera Revolución Industrial, asociada con la digitalización y la automatización a partir de la segunda mitad del siglo XX, introdujo un cambio cualitativo en la relación entre tecnología y trabajo. La incorporación de computadoras, sistemas de información y robots industriales permitió automatizar procesos complejos, reduciendo la necesidad de mano de obra en sectores manufactureros y promoviendo la expansión del sector servicios. En este contexto, el trabajo comenzó a desplazarse hacia actividades intensivas en conocimiento, lo que dio origen a la denominada “economía del conocimiento”.

Sin embargo, es en la Cuarta Revolución Industrial —caracterizada por la convergencia de tecnologías digitales, inteligencia artificial, big data y sistemas ciberfísicos— donde se intensifican las tensiones en torno al futuro del trabajo. A diferencia de las revoluciones anteriores, en las que la tecnología afectaba principalmente tareas físicas o rutinarias, la actual transformación tecnológica incide directamente en funciones cognitivas, creativas y decisionales. Esta diferencia resulta fundamental, ya que amplía significativamente el espectro de actividades susceptibles de ser automatizadas.

En este sentido, diversos autores han señalado que la inteligencia artificial representa una ruptura histórica en la relación entre tecnología y trabajo, en la medida en que introduce la posibilidad de automatizar no solo la fuerza de trabajo, sino también aspectos del conocimiento humano. Como advierte Brynjolfsson y McAfee (2014), la digitalización avanzada permite que las máquinas realicen tareas

que anteriormente requerían juicio humano, lo que plantea interrogantes sobre la sostenibilidad de los modelos tradicionales de empleo.

No obstante, la evidencia histórica muestra que cada revolución tecnológica ha generado tanto procesos de desplazamiento como de creación de empleo. En este sentido, la relación entre tecnología y trabajo no es lineal ni determinista, sino que depende de múltiples factores, incluyendo políticas públicas, estructuras económicas y capacidades institucionales. Acemoglu y Restrepo (2019) destacan que la introducción de nuevas tecnologías genera un proceso dual en el que, por un lado, se automatizan ciertas tareas (efecto de desplazamiento), y por otro, se crean nuevas actividades y ocupaciones (efecto de reinstauración), lo que permite explicar la persistencia del empleo a pesar del avance tecnológico.

Sin embargo, es importante subrayar que, aunque el empleo total puede mantenerse o incluso crecer, los procesos de transición no son neutrales ni inmediatos. La historia muestra que las transformaciones tecnológicas suelen generar periodos de ajuste caracterizados por desempleo estructural, precarización laboral y aumento de la desigualdad. En este sentido, la actual revolución tecnológica impulsada por la inteligencia artificial podría reproducir, e incluso intensificar, estas dinámicas si no se acompaña de mecanismos adecuados de regulación y adaptación.

Además, la velocidad del cambio tecnológico en la actualidad introduce un elemento adicional de complejidad. Mientras que en revoluciones anteriores la difusión de nuevas tecnologías podía tomar décadas, en el contexto contemporáneo la adopción de innovaciones ocurre en cuestión de años o incluso meses. Como se ha documentado en estudios recientes, la aceleración tecnológica ha reducido significativamente los tiempos de difusión, lo que incrementa la presión sobre los sistemas laborales y educativos para adaptarse a nuevas condiciones (Lozano Carrillo, 2024, p. 12).

Esta aceleración no solo intensifica los procesos de transformación laboral, sino que también dificulta la capacidad de las instituciones para responder de manera

efectiva. En este contexto, la inteligencia artificial no debe entenderse únicamente como una continuidad de las revoluciones industriales anteriores, sino como una etapa que amplifica y acelera tendencias históricas, generando desafíos inéditos en términos de regulación, adaptación y equidad.

En suma, el análisis histórico permite comprender que el desplazamiento laboral asociado a la inteligencia artificial no constituye un fenómeno completamente nuevo, sino una manifestación contemporánea de una dinámica estructural que ha acompañado al desarrollo tecnológico. Sin embargo, las características específicas de la inteligencia artificial —particularmente su capacidad para automatizar tareas cognitivas y su rápida difusión— sugieren que nos encontramos ante una transformación de mayor alcance y profundidad, lo que exige repensar los marcos analíticos y regulatorios tradicionales.

3.2 Automatización previa a la inteligencia artificial: de la mecanización al software inteligente

Para comprender con mayor precisión el carácter disruptivo de la inteligencia artificial en el ámbito laboral, resulta indispensable analizar las etapas previas de automatización que han configurado el terreno sobre el cual emerge la actual transformación tecnológica. Si bien la automatización no es un fenómeno reciente, su evolución desde sistemas mecánicos hasta entornos digitales avanzados permite identificar continuidades y rupturas que son fundamentales para dimensionar el impacto contemporáneo de la IA.

En términos generales, la automatización puede definirse como el proceso mediante el cual determinadas tareas son ejecutadas por sistemas tecnológicos con mínima o nula intervención humana directa, particularmente en contextos donde las actividades pueden ser estandarizadas y programadas. Esta conceptualización ha sido ampliamente abordada en la literatura sobre cambio tecnológico y organización del trabajo, donde se reconoce que la automatización implica la sustitución parcial o total de funciones humanas mediante sistemas técnicos diseñados para operar de manera autónoma (Autor, Levy y Murnane,

2003, p. 1280). No obstante, esta definición no es estática, sino que encierra una diversidad de formas y niveles de automatización que han evolucionado significativamente a lo largo del tiempo en función del desarrollo tecnológico y de las transformaciones en los sistemas productivos.

En sus primeras etapas, la automatización se vinculó estrechamente con los procesos de mecanización característicos de la industrialización, particularmente en actividades físicas repetitivas. Tal como lo documenta Braverman (1974), la introducción de maquinaria en los sistemas productivos no solo transformó la organización del trabajo, sino que permitió sustituir la fuerza humana en tareas manuales, incrementando la eficiencia y la productividad, aunque al mismo tiempo generó procesos de desplazamiento laboral y descalificación del trabajo. En esta lógica, la automatización inicial se centró en la transferencia de capacidades físicas hacia sistemas mecánicos, lo que marcó un cambio estructural en la relación entre trabajo y tecnología.

Este proceso se profundizó con el avance de la industrialización, en la medida en que la incorporación de maquinaria permitió reorganizar la producción bajo esquemas más eficientes, pero también más fragmentados. Como señalan Acemoglu y Restrepo (2019), estos cambios tecnológicos generaron efectos simultáneos de sustitución y transformación del trabajo, evidenciando que la automatización no solo elimina tareas, sino que redefine la estructura ocupacional en su conjunto (p. 4).

Posteriormente, con el avance de la electrónica y la informática durante la segunda mitad del siglo XX, la automatización adquirió una dimensión distinta al incorporar sistemas programables capaces de ejecutar secuencias complejas de operaciones. En este contexto, los robots industriales se convirtieron en un elemento central de la transformación productiva, particularmente en sectores como la manufactura automotriz. Estos sistemas no solo realizaban tareas repetitivas con mayor precisión, sino que también reducían la variabilidad en los procesos, lo que contribuía a la estandarización de la producción.

No obstante, es importante señalar que esta forma de automatización, aunque significativa, se encontraba limitada por su rigidez. Los sistemas automatizados tradicionales operaban bajo lógicas deterministas, es decir, ejecutaban instrucciones predefinidas sin capacidad de adaptación a contextos cambiantes. En consecuencia, su impacto en el trabajo, si bien relevante, se concentraba principalmente en tareas rutinarias y estructuradas.

El verdadero punto de inflexión en la automatización previa a la inteligencia artificial se produce con la expansión del software y los sistemas de información, particularmente a partir de la década de 1990. La digitalización de procesos permitió automatizar no solo tareas físicas, sino también actividades administrativas, contables y de gestión. Sistemas de planificación de recursos empresariales (ERP), bases de datos avanzadas y herramientas de análisis de información comenzaron a sustituir funciones que anteriormente requerían intervención humana constante.

En este contexto, el trabajo se transformó progresivamente hacia actividades de supervisión, control y gestión de sistemas automatizados. Como se ha señalado en estudios organizacionales, la incorporación de tecnologías digitales no elimina necesariamente el trabajo, sino que modifica su naturaleza, desplazando al trabajador hacia funciones de mayor complejidad cognitiva (Sansores Guerrero y otros., 2025, p. 11). Esta transición marcó el inicio de una reconfiguración del trabajo que anticipa, en cierta medida, las dinámicas actuales asociadas a la inteligencia artificial.

Asimismo, la automatización basada en software introdujo nuevas formas de organización del trabajo, caracterizadas por la descentralización, la flexibilidad y la integración de procesos a través de plataformas digitales. Este fenómeno se intensificó con el desarrollo de internet y la economía digital, que permitió la externalización de tareas, la aparición de modelos de trabajo remoto y la consolidación de plataformas que intermedian relaciones laborales.

Sin embargo, a pesar de estos avances, la automatización previa a la inteligencia artificial presentaba limitaciones claras. En particular, los sistemas automatizados

carecían de la capacidad de aprendizaje autónomo, lo que restringía su aplicabilidad a entornos estructurados y predecibles. En otras palabras, la automatización tradicional era eficaz en contextos donde las tareas podían ser completamente especificadas, pero resultaba insuficiente frente a situaciones que requerían interpretación, juicio o creatividad.

Esta limitación resulta fundamental para entender la diferencia cualitativa que introduce la inteligencia artificial. Mientras que la automatización tradicional se basa en la ejecución de reglas predefinidas, la IA —y especialmente la IA contemporánea basada en aprendizaje automático— permite identificar patrones, adaptarse a nuevas condiciones y generar respuestas no programadas explícitamente. En este sentido, la transición de la automatización al uso de inteligencia artificial implica un desplazamiento desde sistemas deterministas hacia sistemas probabilísticos y adaptativos.

Además, la automatización previa a la IA ya había comenzado a generar debates en torno al desplazamiento laboral y la precarización del trabajo. Diversos estudios documentaron que la digitalización podía conducir a la polarización del empleo, reduciendo la demanda de trabajos intermedios y aumentando la brecha entre empleos altamente calificados y trabajos de baja remuneración (Autor, Levy & Murnane, 2003). Este fenómeno anticipa algunas de las preocupaciones actuales en torno a la inteligencia artificial, particularmente en lo que respecta a la desigualdad laboral.

En el contexto latinoamericano, estos procesos adquirieron características específicas debido a la coexistencia de sectores altamente tecnificados con amplios segmentos de economía informal. Como se observa en experiencias de intervención organizacional, la incorporación de tecnologías digitales en micro y pequeñas empresas ha sido desigual, lo que refleja no solo limitaciones tecnológicas, sino también barreras estructurales relacionadas con el acceso al conocimiento, la capacitación y los recursos (Lozano Carrillo y otros., 2025, p. 11). Esta situación resulta particularmente relevante para el caso de México, donde la

informalidad laboral y la brecha digital condicionan la forma en que la automatización impacta el trabajo.

En consecuencia, la automatización previa a la inteligencia artificial puede entenderse como una etapa de transición que sentó las bases para la transformación actual del trabajo. Si bien permitió avances significativos en términos de eficiencia y productividad, también evidenció las limitaciones de los sistemas tecnológicos para sustituir completamente las capacidades humanas. Estas limitaciones, precisamente, son las que la inteligencia artificial busca superar, ampliando el alcance de la automatización hacia dominios previamente inaccesibles.

En suma, el análisis de la automatización previa a la IA permite identificar tanto continuidades como rupturas en la relación entre tecnología y trabajo. Por un lado, se observa una continuidad en los procesos de sustitución y transformación laboral; por otro, se evidencia una ruptura en la medida en que la inteligencia artificial introduce capacidades adaptativas y cognitivas que redefinen los límites de la automatización. Esta doble dimensión resulta fundamental para comprender el carácter estructural de la transformación actual y sus implicaciones para el futuro del trabajo.

4. ¿La inteligencia artificial está desplazando empleo?

4.1 Datos globales sobre el impacto laboral de la inteligencia artificial

El debate en torno al impacto de la inteligencia artificial en el empleo ha estado marcado por una tensión constante entre posturas que anticipan una destrucción masiva de puestos de trabajo y aquellas que sostienen que la tecnología, lejos de eliminar el empleo, lo transforma y eventualmente lo expande. No obstante, más allá de estas posiciones, en los últimos años se ha acumulado un cuerpo significativo de evidencia empírica que permite matizar este debate y comprender con mayor precisión la naturaleza del fenómeno.

En primer lugar, es importante señalar que el impacto de la inteligencia artificial en el trabajo no puede analizarse como un proceso homogéneo ni lineal. Diversos estudios han demostrado que la automatización impulsada por tecnologías digitales y, más recientemente, por sistemas de inteligencia artificial, afecta de manera diferenciada a sectores, ocupaciones y regiones. En este sentido, el informe *The Future of Jobs Report 2023* del World Economic Forum estima que, en el periodo 2023–2027, aproximadamente 83 millones de empleos podrían desaparecer a nivel global como resultado de la automatización y la reestructuración económica, mientras que alrededor de 69 millones de nuevos empleos podrían generarse en el mismo periodo, lo que refleja un saldo neto negativo, pero también un proceso de transformación más que de eliminación absoluta del trabajo (World Economic Forum, 2023, p. 15).

Esta evidencia sugiere que la inteligencia artificial no actúa como un mecanismo de sustitución total del trabajo humano, sino como un factor de reconfiguración del mercado laboral. En esta línea, el estudio de McKinsey Global Institute (2023) señala que entre el 50% y el 60% de las actividades laborales actuales tienen el potencial de ser automatizadas en algún grado mediante tecnologías existentes, particularmente en lo que respecta a tareas rutinarias y predecibles. Sin embargo, el mismo informe subraya que la mayoría de los empleos no desaparecerán por completo, sino que experimentarán una transformación significativa en su contenido y en las habilidades requeridas.

Un elemento particularmente relevante en el contexto actual es el impacto de la inteligencia artificial generativa. A diferencia de tecnologías previas, la IA generativa tiene la capacidad de automatizar tareas cognitivas no rutinarias, lo que amplía considerablemente el espectro de ocupaciones potencialmente afectadas. En este sentido, el informe de Goldman Sachs (2023) estima que hasta el 18% del trabajo global podría ser automatizado mediante inteligencia artificial, con efectos más pronunciados en economías avanzadas debido a la mayor presencia de empleos intensivos en conocimiento.

Asimismo, la Organización Internacional del Trabajo (OIT) ha señalado que la IA generativa tiene una mayor probabilidad de transformar empleos que de eliminarlos completamente. En su análisis global, la OIT (2023) concluye que aproximadamente el 24% de los empleos están expuestos a algún tipo de automatización, pero solo un porcentaje relativamente menor enfrenta un riesgo alto de sustitución total. Este hallazgo refuerza la idea de que el impacto de la IA se manifiesta principalmente en la transformación de tareas más que en la desaparición de ocupaciones completas.

En términos sectoriales, la evidencia empírica muestra que los efectos de la inteligencia artificial se concentran en determinadas áreas. Los sectores más expuestos incluyen actividades administrativas, servicios financieros, atención al cliente, marketing y programación básica, donde la automatización de tareas cognitivas es más viable. En contraste, sectores que requieren interacción humana compleja, habilidades manuales no rutinarias o juicio contextual —como la salud, la educación o los servicios sociales— presentan una menor vulnerabilidad relativa.

No obstante, incluso en estos sectores, la inteligencia artificial está generando cambios significativos en la forma en que se realiza el trabajo. Como se ha documentado en estudios organizacionales, la incorporación de sistemas inteligentes no elimina necesariamente la participación humana, sino que redefine su rol dentro de los procesos productivos, desplazando al trabajador hacia funciones de supervisión, interpretación y toma de decisiones (Sansores Guerrero y otros., 2025, p. 11).

Otro aspecto relevante es la distribución desigual del impacto de la inteligencia artificial en función de factores sociodemográficos. Diversos estudios han señalado que los trabajadores con menor nivel educativo y aquellos que desempeñan tareas rutinarias son más susceptibles al desplazamiento, mientras que los trabajadores altamente calificados tienden a beneficiarse de la complementariedad entre sus habilidades y las tecnologías digitales. Este fenómeno contribuye a la polarización

del mercado laboral, incrementando las brechas de ingreso y las desigualdades estructurales.

En el contexto latinoamericano, estas dinámicas adquieren particular relevancia debido a la alta proporción de empleo informal y a las limitaciones en el acceso a educación y tecnología. Como se ha observado en experiencias de intervención organizacional en microempresas, la adopción de herramientas de inteligencia artificial puede generar mejoras significativas en la productividad, pero también evidencia profundas desigualdades en la capacidad de acceso y apropiación tecnológica (Lozano Carrillo y otros., 2025, p. 11) . Esta situación sugiere que el impacto de la IA en la región no solo dependerá de la evolución tecnológica, sino también de las condiciones estructurales que configuran el mercado laboral.

En suma, la evidencia empírica disponible permite concluir que la inteligencia artificial está transformando el trabajo a escala global, pero no de manera uniforme ni necesariamente a través de la eliminación masiva de empleos. Más bien, se trata de un proceso complejo de reconfiguración que implica la redistribución de tareas, la transformación de competencias y la redefinición de los roles laborales. Esta constatación resulta fundamental para el análisis regulatorio, ya que desplaza el foco del debate desde la simple pregunta sobre la pérdida de empleos hacia una discusión más amplia sobre la calidad del trabajo, la distribución de beneficios y la equidad en la transición tecnológica.

5. ¿Se puede regular la inteligencia artificial para limitar el desplazamiento laboral?

5.1 Límites reales de la regulación: entre innovación y protección del trabajo

La posibilidad de regular la inteligencia artificial con el objetivo de limitar el desplazamiento de la mano de obra humana constituye uno de los debates más complejos y controvertidos en la agenda contemporánea. A diferencia de otros procesos tecnológicos históricos, la velocidad, escala y naturaleza global de la inteligencia artificial plantean desafíos sustantivos a los marcos regulatorios

tradicionales, cuestionando la capacidad de los Estados para intervenir de manera efectiva en la relación entre tecnología y empleo.

En primer lugar, es necesario reconocer que la regulación de la inteligencia artificial enfrenta un límite estructural fundamental: la tensión entre innovación y control. Por un lado, la IA representa un motor clave de crecimiento económico, competitividad y eficiencia productiva; por otro, su implementación puede generar efectos adversos en el empleo, particularmente en términos de desplazamiento laboral y precarización. En este sentido, una regulación excesivamente restrictiva podría inhibir la innovación tecnológica y reducir la competitividad de las economías, mientras que una regulación insuficiente podría profundizar desigualdades y debilitar las condiciones laborales.

Esta tensión se ve acentuada por el carácter global de la economía digital. A diferencia de las industrias tradicionales, las tecnologías basadas en inteligencia artificial pueden desarrollarse, implementarse y escalarse a través de fronteras con relativa facilidad. En consecuencia, las regulaciones nacionales enfrentan el riesgo de ser eludidas mediante la relocalización de operaciones hacia jurisdicciones con menores restricciones. Este fenómeno, conocido como “arbitraje regulatorio”, limita la efectividad de las políticas públicas y obliga a pensar en esquemas de gobernanza internacional.

Asimismo, la velocidad del desarrollo tecnológico introduce un desfase constante entre innovación y regulación. Mientras que los ciclos legislativos suelen ser largos y requieren procesos deliberativos complejos, las tecnologías de inteligencia artificial evolucionan a un ritmo exponencial. Como se ha documentado en estudios recientes, la adopción de innovaciones tecnológicas que anteriormente tomaba años o décadas ahora ocurre en cuestión de meses, lo que dificulta la capacidad de los marcos regulatorios para mantenerse actualizados (Lozano Carrillo, 2024, p. 12). Este desajuste temporal genera un vacío regulatorio en el que las tecnologías se implementan sin supervisión adecuada.

Aunado a lo anterior, la complejidad técnica de la inteligencia artificial representa un obstáculo adicional para su regulación. Los sistemas basados en aprendizaje automático, particularmente aquellos de tipo generativo, operan a través de modelos altamente complejos y, en muchos casos, opacos incluso para sus propios desarrolladores. Esta “caja negra” algorítmica dificulta la supervisión, la rendición de cuentas y la atribución de responsabilidades, lo que plantea desafíos significativos para los marcos jurídicos tradicionales.

En el ámbito laboral, estos límites se traducen en una dificultad concreta: la imposibilidad de regular de manera directa la sustitución tecnológica sin afectar los incentivos económicos de las empresas. A diferencia de otras áreas donde la regulación puede establecer prohibiciones claras, en el caso del empleo la automatización forma parte de la lógica misma del desarrollo económico. Como señalan Acemoglu y Restrepo (2020), las empresas adoptan tecnologías automatizadas no solo para reducir costos, sino también para incrementar productividad, lo que genera presiones estructurales hacia la sustitución de ciertas tareas (p. 2190).

En este contexto, la pregunta no es únicamente si se puede regular la inteligencia artificial, sino hasta qué punto es viable establecer límites a su uso sin comprometer el desarrollo económico. Esta interrogante remite a una tensión más profunda entre eficiencia y equidad, que ha estado presente en todas las revoluciones tecnológicas, pero que adquiere una intensidad particular en el caso de la IA debido a su capacidad para transformar tanto el trabajo físico como el cognitivo.

No obstante, reconocer estos límites no implica asumir una postura determinista o resignada frente al avance tecnológico. Por el contrario, implica replantear el enfoque de la regulación, desplazándolo desde la idea de control absoluto hacia modelos de gobernanza adaptativa que permitan gestionar los impactos de la inteligencia artificial de manera flexible y contextualizada. En este sentido, la regulación no debe entenderse como un mecanismo para detener la innovación,

sino como una herramienta para orientar su desarrollo en función de objetivos sociales más amplios.

Desde esta perspectiva, resulta fundamental considerar que la regulación de la inteligencia artificial en el trabajo no puede centrarse exclusivamente en la tecnología en sí misma, sino en sus efectos sobre las personas y las estructuras sociales. Como se ha señalado en estudios organizacionales, las transformaciones tecnológicas implican también una reconfiguración de las relaciones de poder, los significados y las prácticas que estructuran la vida laboral (Lozano Carrillo y otros., 2024, p. 2). En consecuencia, la regulación debe abordar no solo los aspectos técnicos, sino también las dimensiones sociales y organizacionales del cambio tecnológico.

En suma, los límites de la regulación de la inteligencia artificial no derivan únicamente de la complejidad técnica de la tecnología, sino de su inserción en dinámicas económicas globales, su velocidad de evolución y su carácter estructural en los procesos productivos. Estos factores obligan a repensar los enfoques regulatorios tradicionales y a desarrollar marcos más flexibles, coordinados y orientados a la gestión de impactos, particularmente en lo que respecta al desplazamiento laboral.

5.2 Instrumentos regulatorios posibles: más allá de la prohibición tecnológica

Si bien la regulación de la inteligencia artificial enfrenta límites estructurales significativos, ello no implica que los Estados y las instituciones carezcan de herramientas para intervenir en sus efectos sobre el trabajo. Por el contrario, en los últimos años han surgido diversas propuestas orientadas a mitigar los impactos negativos de la automatización, particularmente en lo que respecta al desplazamiento laboral.

Uno de los instrumentos más debatidos es la implementación de impuestos a la automatización, comúnmente asociados con la idea de un “impuesto a los robots”. Esta propuesta, impulsada por diversos economistas y figuras públicas, parte del supuesto de que la automatización genera beneficios económicos que deberían ser

redistribuidos para compensar la pérdida de empleos. Sin embargo, su viabilidad ha sido ampliamente cuestionada, ya que resulta difícil definir qué constituye exactamente un “robot” o una tecnología automatizada, así como medir su impacto específico en el empleo.

Otra herramienta relevante es la implementación de evaluaciones de impacto laboral previas a la adopción de tecnologías de inteligencia artificial. Este enfoque, inspirado en modelos de evaluación ambiental, busca obligar a las empresas a analizar las posibles consecuencias de la automatización en el empleo antes de implementar nuevas tecnologías. Aunque esta medida podría contribuir a una mayor transparencia, su efectividad depende de la capacidad de las instituciones para supervisar y hacer cumplir estos procesos.

Asimismo, ha cobrado relevancia la idea de fortalecer los derechos laborales en el contexto digital. Esto incluye propuestas como el derecho a la transparencia algorítmica, que permitiría a los trabajadores conocer cómo los sistemas de inteligencia artificial afectan decisiones relacionadas con su empleo; el derecho a la supervisión humana, que busca evitar que decisiones críticas sean tomadas exclusivamente por sistemas automatizados; y el derecho a la reconversión laboral, que implica garantizar acceso a capacitación y educación continua.

En este sentido, la regulación no se limita a restringir el uso de la tecnología, sino que puede orientarse a fortalecer la capacidad de los trabajadores para adaptarse a los cambios. Como se ha observado en experiencias de intervención organizacional, la incorporación de la inteligencia artificial puede generar beneficios significativos cuando se acompaña de procesos de capacitación y apropiación tecnológica (Lozano Carrillo y otros., 2025, p. 12). Esto sugiere que la regulación debe articularse con políticas educativas y laborales que faciliten la transición hacia nuevos modelos de trabajo.

Por otra parte, algunos enfoques han planteado la necesidad de repensar los sistemas de protección social, incluyendo propuestas como la renta básica universal o la reducción de la jornada laboral. Estas medidas buscan desvincular

parcialmente el bienestar económico del empleo tradicional, reconociendo que la automatización podría reducir la disponibilidad de trabajo en ciertos sectores. Sin embargo, estas propuestas implican transformaciones profundas en los modelos económicos y sociales, lo que dificulta su implementación en el corto plazo.

En consecuencia, la regulación de la inteligencia artificial en el trabajo no puede reducirse a un único instrumento, sino que requiere un enfoque integral que combine medidas económicas, jurídicas, educativas y sociales. Más que limitar directamente el uso de la tecnología, el objetivo de la regulación debería ser gestionar sus efectos, garantizando que los beneficios de la innovación se distribuyan de manera equitativa y que los costos del cambio no recaigan de manera desproporcionada en los trabajadores más vulnerables.

6. Panorama internacional de la regulación de la inteligencia artificial en el trabajo

6.1 Unión Europea: regulación proactiva basada en riesgos

En el escenario internacional, la Unión Europea se ha consolidado como uno de los actores más avanzados en la construcción de un marco regulatorio integral para la inteligencia artificial. A diferencia de otros enfoques más flexibles o fragmentados, el modelo europeo se caracteriza por una orientación preventiva, sustentada en el principio de protección de derechos fundamentales y en la necesidad de anticipar los riesgos asociados al uso de tecnologías emergentes.

El instrumento más representativo de esta estrategia es el AI Act (Ley de Inteligencia Artificial), aprobado en 2024, el cual establece un sistema de regulación basado en niveles de riesgo. Bajo este enfoque, los sistemas de IA se clasifican en categorías que van desde riesgo mínimo hasta riesgo inaceptable, siendo estos últimos sujetos a prohibiciones específicas. En el ámbito laboral, esta legislación adquiere especial relevancia, ya que considera de “alto riesgo” aquellos sistemas utilizados en procesos de contratación, evaluación de desempeño y gestión del personal.

Este enfoque implica que las empresas que implementan inteligencia artificial en el trabajo deben cumplir con una serie de requisitos estrictos, incluyendo evaluaciones de impacto, transparencia en los algoritmos y supervisión humana. De esta manera, la regulación europea no busca limitar el desarrollo tecnológico, sino establecer condiciones para su uso responsable, particularmente en contextos donde pueden afectarse derechos laborales.

No obstante, este modelo también enfrenta críticas. Algunos sectores argumentan que una regulación demasiado estricta podría afectar la competitividad de las empresas europeas frente a economías con marcos más flexibles. Sin embargo, desde una perspectiva social, el enfoque europeo representa uno de los intentos más sólidos por equilibrar innovación tecnológica y protección del trabajo.

6.2 Estados Unidos: regulación fragmentada y orientada al mercado

En contraste con el modelo europeo, Estados Unidos ha adoptado un enfoque más descentralizado y orientado al mercado en la regulación de la inteligencia artificial. En lugar de una legislación integral, la gobernanza de la IA se articula a través de una combinación de lineamientos federales, regulaciones sectoriales y autorregulación empresarial.

Uno de los documentos más relevantes en este contexto es el Blueprint for an AI Bill of Rights (2022), que establece principios orientadores para el desarrollo y uso de la inteligencia artificial, incluyendo la protección contra discriminación algorítmica, la privacidad de datos y la transparencia. Sin embargo, estos principios no tienen carácter vinculante, lo que limita su capacidad de implementación efectiva.

En el ámbito laboral, la regulación en Estados Unidos se ha centrado principalmente en casos específicos, como el uso de algoritmos en procesos de contratación y evaluación de empleados. Algunas ciudades, como Nueva York, han implementado normativas que exigen auditorías de sesgo en sistemas de inteligencia artificial utilizados en recursos humanos. No obstante, la ausencia de un marco federal integral genera un entorno regulatorio fragmentado.

Este enfoque responde, en gran medida, a la lógica de promover la innovación y mantener el liderazgo tecnológico global. Sin embargo, también implica riesgos, particularmente en términos de protección laboral, ya que la responsabilidad de regular recae en gran medida en las propias empresas. En este sentido, la regulación en Estados Unidos tiende a ser reactiva, respondiendo a problemas una vez que estos se manifiestan, más que anticipando sus efectos.

6.3 China: desarrollo acelerado con control estatal

El modelo chino representa un enfoque distinto, caracterizado por una fuerte intervención estatal combinada con una estrategia de desarrollo tecnológico acelerado. En este contexto, la inteligencia artificial es concebida como un elemento central para el crecimiento económico y la competitividad global, lo que ha llevado al gobierno a impulsar su adopción en múltiples sectores, incluido el ámbito laboral.

A diferencia de Europa, donde la regulación se centra en la protección de derechos, o de Estados Unidos, donde predomina la lógica de mercado, China ha desarrollado un modelo en el que el Estado desempeña un papel central en la supervisión y control de la tecnología. Esto se traduce en regulaciones específicas sobre el uso de algoritmos, particularmente en plataformas digitales, así como en mecanismos de monitoreo que buscan garantizar la estabilidad social.

En el ámbito laboral, la inteligencia artificial se ha integrado ampliamente en sectores como la manufactura, la logística y los servicios digitales. Sin embargo, la regulación no se enfoca tanto en limitar el desplazamiento laboral, sino en gestionar sus efectos dentro de un marco de control estatal. Esto incluye políticas de reconversión laboral y capacitación, aunque subordinadas a objetivos de crecimiento económico.

Este modelo ha permitido una rápida adopción de tecnologías de inteligencia artificial, pero también ha generado preocupaciones en torno a la vigilancia, la privacidad y los derechos laborales. En este sentido, el caso chino ilustra las tensiones entre desarrollo tecnológico y control social en contextos donde el Estado tiene una capacidad de intervención más amplia.

6.4 América Latina: rezago regulatorio y desafíos estructurales

En América Latina, la regulación de la inteligencia artificial se encuentra en una etapa incipiente, caracterizada por avances fragmentados y una limitada capacidad institucional para enfrentar los desafíos asociados al desarrollo tecnológico. A diferencia de regiones como Europa, donde existe un marco normativo consolidado, en América Latina predominan iniciativas aisladas, muchas de ellas de carácter estratégico más que regulatorio.

En el ámbito laboral, esta situación se ve agravada por condiciones estructurales como la alta informalidad, la desigualdad económica y la brecha digital. En países como México, una proporción significativa de la población trabaja en el sector informal, lo que limita el alcance de cualquier regulación laboral relacionada con la inteligencia artificial. En este contexto, la automatización y la adopción de nuevas tecnologías no necesariamente se traducen en procesos formales de sustitución laboral, sino en transformaciones más difusas y difíciles de medir.

Asimismo, como se ha documentado en experiencias de intervención organizacional, la incorporación de inteligencia artificial en micro y pequeñas empresas enfrenta barreras significativas relacionadas con el acceso a tecnología, la capacitación y la infraestructura (Lozano Carrillo y otros., 2025, p. 11). Esta situación refleja que, en la región, el desafío no es únicamente regular la IA, sino también garantizar condiciones básicas para su adopción equitativa.

A pesar de estas limitaciones, algunos países han comenzado a desarrollar estrategias nacionales de inteligencia artificial que incluyen componentes regulatorios. Sin embargo, estas iniciativas suelen centrarse en la promoción de la innovación más que en la protección del trabajo, lo que evidencia un desequilibrio en las prioridades de política pública.

Síntesis del panorama internacional

El análisis comparado permite identificar que no existe un modelo único de regulación de la inteligencia artificial en el trabajo, sino una diversidad de enfoques

que reflejan diferencias en contextos económicos, políticos y sociales. Mientras que la Unión Europea apuesta por una regulación preventiva centrada en derechos, Estados Unidos privilegia la innovación y la autorregulación, China combina desarrollo tecnológico con control estatal, y América Latina enfrenta un escenario de rezago y fragmentación.

Esta diversidad pone de manifiesto que la regulación de la inteligencia artificial no es únicamente un problema técnico, sino un reflejo de decisiones políticas sobre el tipo de sociedad que se busca construir. En este sentido, el desafío para países como México no radica en replicar modelos externos, sino en desarrollar marcos regulatorios que respondan a sus propias condiciones estructurales y necesidades sociales.

7. Caso México: avances, retos y vacíos regulatorios en la inteligencia artificial y el trabajo

7.1 Marco normativo actual: entre regulación laboral tradicional y vacío tecnológico

El análisis de la regulación de la inteligencia artificial en México revela una situación caracterizada por la coexistencia de marcos normativos laborales consolidados con una notable ausencia de regulación específica en materia de tecnologías emergentes. En efecto, el país cuenta con una estructura jurídica robusta en lo que respecta al derecho laboral, sustentada principalmente en la Ley Federal del Trabajo, la cual establece principios fundamentales como la protección del empleo, la estabilidad laboral y los derechos de los trabajadores. Sin embargo, estos instrumentos fueron diseñados en un contexto previo al desarrollo de la inteligencia artificial, lo que limita su capacidad para responder a los desafíos actuales.

En este sentido, el marco normativo mexicano no contempla de manera explícita problemáticas derivadas del uso de inteligencia artificial en el trabajo, tales como la toma de decisiones automatizada, la supervisión algorítmica o el desplazamiento laboral inducido por tecnologías digitales. Como resultado, existe un vacío

regulatorio que dificulta la protección efectiva de los trabajadores frente a nuevas formas de organización del trabajo mediadas por sistemas inteligentes.

Asimismo, la legislación en materia de protección de datos personales — particularmente la Ley Federal de Protección de Datos Personales en Posesión de los Particulares— establece ciertos principios relevantes, como el consentimiento, la transparencia y la finalidad del tratamiento de datos. No obstante, estos instrumentos resultan insuficientes para abordar los desafíos específicos de la inteligencia artificial, especialmente en lo que respecta a la opacidad algorítmica y la toma de decisiones automatizadas.

En consecuencia, el marco normativo mexicano se encuentra en una situación de rezago frente a la velocidad del desarrollo tecnológico, lo que genera un entorno en el que la adopción de inteligencia artificial en el ámbito laboral ocurre sin una regulación clara que delimite sus alcances y responsabilidades.

7.2 Políticas públicas en inteligencia artificial: avances incipientes y falta de continuidad

En el ámbito de las políticas públicas, México ha mostrado avances limitados en la construcción de una estrategia nacional de inteligencia artificial. Si bien en años recientes se han impulsado iniciativas orientadas a promover el desarrollo tecnológico, estas han carecido de continuidad y de una articulación integral que vincule innovación con regulación.

Uno de los esfuerzos más relevantes fue la elaboración de la Estrategia Nacional de Inteligencia Artificial, impulsada en 2018, la cual posicionó a México como uno de los primeros países de América Latina en abordar este tema a nivel gubernamental. Sin embargo, esta estrategia tuvo un alcance limitado y no logró consolidarse como un marco operativo con impacto sostenido en políticas públicas.

Más recientemente, se han desarrollado iniciativas desde distintos sectores, incluyendo la academia y organismos internacionales, orientadas a promover el uso responsable de la inteligencia artificial. No obstante, estas iniciativas han sido

fragmentadas y, en muchos casos, carecen de mecanismos de implementación y evaluación.

En este contexto, la ausencia de una política pública integral en materia de inteligencia artificial limita la capacidad del Estado mexicano para anticipar y gestionar los impactos de estas tecnologías en el mercado laboral. A diferencia de regiones como la Unión Europea, donde la regulación forma parte de una estrategia más amplia de gobernanza tecnológica, en México persiste una desconexión entre el impulso a la innovación y la protección del trabajo.

7.3 Impacto laboral de la inteligencia artificial en México: entre informalidad y transformación tecnológica

El impacto de la inteligencia artificial en el mercado laboral mexicano se encuentra profundamente condicionado por las características estructurales de la economía nacional, particularmente por la alta prevalencia de la informalidad laboral. De acuerdo con datos del Instituto Nacional de Estadística y Geografía (INEGI), más del 50% de la población ocupada en México se encuentra en el sector informal, lo que implica la ausencia de protección social y de mecanismos formales de regulación laboral.

En este contexto, el desplazamiento laboral asociado a la inteligencia artificial no necesariamente se manifiesta en términos de despidos masivos, como podría ocurrir en economías más formalizadas, sino a través de procesos más difusos, como la reducción de ingresos, la precarización del trabajo o la sustitución gradual de actividades. Esto dificulta la medición del impacto real de la IA en el empleo, así como la implementación de políticas públicas efectivas.

Por otro lado, la adopción de inteligencia artificial en México se ha concentrado principalmente en sectores específicos, como servicios financieros, comercio electrónico, marketing digital y ciertas áreas de manufactura. En estos sectores, la IA ha permitido mejorar la eficiencia y reducir costos, pero también ha generado transformaciones en la estructura del empleo, particularmente en lo que respecta a la demanda de habilidades.

Como se ha documentado en experiencias de intervención organizacional, la incorporación de inteligencia artificial en micro y pequeñas empresas puede generar mejoras significativas en la productividad y en la gestión empresarial, pero también evidencia profundas brechas en el acceso a tecnología y capacitación (Lozano Carrillo y otros., 2025, p. 11) . Esta situación refleja que el impacto de la IA en México no es homogéneo, sino que reproduce y, en algunos casos, amplifica desigualdades preexistentes.

7.4 Retos estructurales: educación, desigualdad y brecha digital

Uno de los principales desafíos para la regulación de la inteligencia artificial en México radica en la existencia de limitaciones estructurales que condicionan tanto su adopción como su impacto en el trabajo. Entre estos desafíos, destacan la desigualdad educativa, la brecha digital y la limitada capacidad institucional.

En primer lugar, la transformación del trabajo impulsada por la inteligencia artificial exige el desarrollo de nuevas competencias laborales, particularmente en áreas relacionadas con el manejo de tecnologías digitales, el análisis de datos y la interacción con sistemas inteligentes. Sin embargo, el sistema educativo mexicano enfrenta dificultades para adaptarse a estas demandas, lo que genera un desajuste entre las habilidades disponibles y las requeridas en el mercado laboral.

En segundo lugar, la brecha digital constituye un obstáculo significativo para la adopción equitativa de la inteligencia artificial. El acceso desigual a infraestructura tecnológica, conectividad y herramientas digitales limita la capacidad de amplios sectores de la población para beneficiarse de las oportunidades que ofrece la IA, lo que contribuye a la reproducción de desigualdades.

Finalmente, la capacidad institucional del Estado para regular la inteligencia artificial se ve limitada por la falta de especialización técnica y por la ausencia de marcos normativos específicos. Como se ha señalado en estudios organizacionales, la incorporación de tecnologías emergentes requiere no solo de regulación, sino de procesos de apropiación social que permitan integrar la tecnología de manera contextualizada (Lozano Carrillo y otros., 2024, p. 2).

Síntesis del caso México

El análisis del caso mexicano permite identificar una paradoja central: mientras que la inteligencia artificial avanza de manera acelerada en ciertos sectores, el marco institucional y regulatorio permanece rezagado, lo que genera un entorno de incertidumbre y desigualdad. En este contexto, el desafío no consiste únicamente en desarrollar regulaciones específicas, sino en construir una estrategia integral que articule innovación tecnológica, protección laboral y desarrollo social.

En consecuencia, México se enfrenta a una oportunidad crítica para definir su modelo de gobernanza de la inteligencia artificial. La ausencia de regulación no implica neutralidad, sino la prevalencia de dinámicas de mercado que pueden profundizar desigualdades. Por ello, resulta fundamental avanzar hacia un enfoque que no solo promueva la adopción de tecnologías, sino que también garantice que sus beneficios se distribuyan de manera equitativa.

8. Propuesta de marco regulatorio para la inteligencia artificial y el trabajo

8.1 Principios rectores: hacia una gobernanza centrada en el trabajo humano

La construcción de un marco regulatorio para la inteligencia artificial en el ámbito laboral no puede limitarse a la adaptación de normativas existentes, sino que requiere una reconfiguración conceptual que reconozca el carácter estructural de la transformación tecnológica en curso. En este sentido, cualquier propuesta regulatoria debe partir de un conjunto de principios rectores que orienten tanto la formulación de políticas públicas como la implementación de tecnologías en los entornos laborales.

En primer lugar, resulta fundamental establecer el principio de centralidad del trabajo humano, el cual implica reconocer que la inteligencia artificial debe desarrollarse y aplicarse como un medio para potenciar las capacidades humanas, y no como un mecanismo orientado exclusivamente a su sustitución. Este principio no supone la negación del progreso tecnológico, sino la subordinación de dicho

progreso a objetivos sociales más amplios, particularmente en lo que respecta a la dignidad, la estabilidad y el bienestar de los trabajadores.

En segundo lugar, se propone el principio de justicia distributiva tecnológica, que plantea la necesidad de garantizar una distribución equitativa de los beneficios generados por la inteligencia artificial. Como se ha observado en el análisis empírico, la automatización tiende a concentrar beneficios en determinados sectores y grupos sociales, lo que puede profundizar desigualdades existentes. En consecuencia, la regulación debe incorporar mecanismos que permitan redistribuir estos beneficios, ya sea a través de políticas fiscales, programas de capacitación o sistemas de protección social.

Un tercer principio clave es el de transparencia y explicabilidad algorítmica. Dado que la inteligencia artificial puede influir de manera significativa en decisiones laborales —como la contratación, la evaluación de desempeño o la asignación de tareas— resulta indispensable que estos procesos sean comprensibles y auditables. Este principio no solo responde a una exigencia técnica, sino también a una necesidad democrática, en la medida en que permite a los trabajadores conocer y cuestionar los mecanismos que afectan su vida laboral.

Asimismo, se plantea el principio de adaptabilidad regulatoria, que reconoce la imposibilidad de establecer marcos normativos rígidos en un contexto de cambio tecnológico acelerado. En lugar de ello, se propone desarrollar esquemas regulatorios flexibles, capaces de ajustarse a nuevas condiciones y de incorporar aprendizajes derivados de la práctica. Este enfoque implica concebir la regulación como un proceso dinámico, más que como un conjunto estático de normas.

Finalmente, resulta indispensable incorporar el principio de corresponsabilidad multinivel, que reconoce que la regulación de la inteligencia artificial no es responsabilidad exclusiva del Estado, sino que involucra a empresas, instituciones educativas, organismos internacionales y sociedad civil. En este sentido, la gobernanza de la IA debe entenderse como un proceso colectivo en el que múltiples actores participan en la definición de reglas, prácticas y estándares.

8.2 Propuestas regulatorias concretas: instrumentos para la gestión del impacto laboral

A partir de los principios anteriores, es posible delinear un conjunto de propuestas concretas orientadas a gestionar el impacto de la inteligencia artificial en el trabajo, particularmente en lo que respecta al desplazamiento laboral.

En primer lugar, se propone la implementación de evaluaciones de impacto laboral obligatorias para sistemas de inteligencia artificial utilizados en entornos organizacionales. Estas evaluaciones deberían analizar no solo los efectos económicos de la automatización, sino también sus implicaciones en términos de empleo, condiciones laborales y distribución de tareas. Este instrumento permitiría anticipar riesgos y generar información relevante para la toma de decisiones tanto a nivel empresarial como gubernamental.

En segundo lugar, se plantea la creación de un marco de derechos laborales digitales, que incluya, entre otros elementos, el derecho a la transparencia algorítmica, el derecho a la supervisión humana en decisiones automatizadas y el derecho a la protección frente a sesgos y discriminación algorítmica. Este enfoque reconoce que la transformación del trabajo requiere una actualización de los derechos laborales tradicionales para adaptarse a entornos mediados por tecnología.

Asimismo, resulta fundamental fortalecer los sistemas de reconversión laboral, mediante políticas públicas orientadas a la capacitación continua y al desarrollo de nuevas competencias. En este sentido, la regulación debe articularse con el sistema educativo y con programas de formación que permitan a los trabajadores adaptarse a las nuevas demandas del mercado laboral. Como se ha observado en experiencias de intervención organizacional, la apropiación tecnológica es un factor clave para que la inteligencia artificial funcione como un complemento y no como un sustituto del trabajo humano (Lozano Carrillo y otros., 2025, p. 12).

Otra propuesta relevante es el establecimiento de incentivos regulatorios para modelos de implementación de inteligencia artificial centrados en la augmentación

del trabajo, es decir, en el uso de la tecnología como herramienta de apoyo a los trabajadores. Esto podría incluir beneficios fiscales o programas de certificación para empresas que adopten prácticas responsables en el uso de IA.

Adicionalmente, se propone explorar mecanismos de redistribución económica asociados a la automatización, como impuestos a la productividad tecnológica o contribuciones específicas destinadas a financiar sistemas de protección social. Si bien estas medidas presentan desafíos técnicos y políticos, su discusión resulta necesaria en un contexto donde la inteligencia artificial puede concentrar beneficios en determinados actores económicos.

Finalmente, se plantea la creación de organismos especializados en la supervisión de la inteligencia artificial, con capacidades técnicas y autonomía suficiente para evaluar, auditar y regular el uso de estas tecnologías en el ámbito laboral. Estos organismos podrían desempeñar un papel clave en la articulación entre regulación, innovación y protección social.

8.3 Rol de los actores clave: gobernanza colaborativa de la inteligencia artificial

La implementación de un marco regulatorio efectivo para la inteligencia artificial en el trabajo requiere la participación coordinada de múltiples actores, cada uno con responsabilidades específicas en el proceso de gobernanza tecnológica.

En primer lugar, el Estado debe asumir un rol central en la definición de marcos normativos, la protección de derechos laborales y la implementación de políticas públicas orientadas a la gestión del cambio tecnológico. Sin embargo, su papel no debe limitarse a la regulación, sino que debe incluir también la promoción de la innovación y el desarrollo de capacidades institucionales.

Por su parte, las empresas tienen la responsabilidad de adoptar prácticas éticas en el desarrollo y uso de la inteligencia artificial, garantizando que sus decisiones no generen impactos negativos desproporcionados en los trabajadores. Esto implica

no solo el cumplimiento de normativas, sino la incorporación de principios de responsabilidad social en la gestión tecnológica.

Las instituciones educativas, en tanto, desempeñan un papel fundamental en la formación de competencias necesarias para el futuro del trabajo. La adaptación de los sistemas educativos a las demandas de la inteligencia artificial constituye un elemento clave para reducir las brechas entre oferta y demanda de habilidades.

Finalmente, la sociedad civil y los organismos internacionales pueden contribuir a la construcción de marcos de gobernanza más inclusivos, promoviendo el debate público y la cooperación entre países. En un contexto globalizado, la regulación de la inteligencia artificial requiere no solo de acciones nacionales, sino de esfuerzos coordinados a nivel internacional.

Síntesis de la propuesta

La regulación de la inteligencia artificial en el trabajo no debe concebirse como un intento de frenar el avance tecnológico, sino como un proceso orientado a gestionar sus impactos de manera equitativa y sostenible. En este sentido, el desafío no radica en decidir si la inteligencia artificial debe ser regulada, sino en definir cómo hacerlo de manera que se garantice la coexistencia entre innovación y justicia social.

La propuesta aquí planteada busca contribuir a este objetivo mediante la articulación de principios, instrumentos y actores que permitan construir un modelo de gobernanza de la inteligencia artificial centrado en el trabajo humano. Este enfoque reconoce que la tecnología no es un destino inevitable, sino un proceso social que puede —y debe— ser orientado en función de los valores y objetivos de la sociedad.

9. Conclusiones

El análisis desarrollado a lo largo de este capítulo permite afirmar que la relación entre inteligencia artificial y trabajo no puede comprenderse desde una lógica simplista de sustitución tecnológica, sino como un proceso complejo de

transformación estructural que redefine las bases mismas del empleo, la organización productiva y la regulación social. En este sentido, la pregunta inicial —¿quién regula la inteligencia artificial ante el desplazamiento de la mano de obra humana? — no admite una respuesta unívoca, sino que remite a un entramado de actores, intereses y capacidades institucionales que operan en distintos niveles.

En primer lugar, la evidencia histórica y empírica sugiere que, si bien la tecnología ha generado tradicionalmente procesos de desplazamiento laboral, también ha abierto nuevas oportunidades de empleo y transformación productiva. Sin embargo, la inteligencia artificial introduce una diferencia cualitativa relevante: su capacidad para automatizar no solo tareas físicas, sino también funciones cognitivas, lo que amplía significativamente el espectro de ocupaciones potencialmente afectadas. Esta característica, sumada a la velocidad de su desarrollo y adopción, plantea desafíos inéditos para los sistemas laborales contemporáneos.

A partir del análisis empírico, se observa que la inteligencia artificial no está provocando una desaparición masiva del empleo en términos absolutos, sino una reconfiguración profunda del mercado laboral. Este proceso se manifiesta en la transformación de tareas, la redefinición de competencias y la emergencia de nuevas formas de organización del trabajo. No obstante, esta transformación no es neutral, ya que tiende a beneficiar a trabajadores con mayores niveles de capacitación y acceso a tecnología, mientras que incrementa la vulnerabilidad de aquellos que desempeñan tareas rutinarias o carecen de habilidades digitales.

En este contexto, la regulación de la inteligencia artificial emerge como un desafío central, pero también como un campo de posibilidades. Como se ha argumentado, no es viable establecer límites absolutos al uso de tecnologías que forman parte del núcleo del desarrollo económico contemporáneo. Sin embargo, ello no implica la imposibilidad de regular sus efectos. Por el contrario, la regulación debe orientarse a gestionar el impacto de la inteligencia artificial en el trabajo, garantizando condiciones de equidad, protección y adaptación.

El análisis comparado muestra que no existe un modelo único de regulación, sino una diversidad de enfoques que reflejan distintas prioridades y contextos. Mientras que la Unión Europea apuesta por una regulación preventiva centrada en derechos, Estados Unidos privilegia la innovación y la autorregulación, China combina desarrollo tecnológico con control estatal, y América Latina enfrenta un escenario de rezago estructural. Este panorama evidencia que la regulación de la inteligencia artificial no es únicamente un problema técnico, sino una decisión política sobre el tipo de sociedad que se desea construir.

En el caso de México, el análisis revela una brecha significativa entre el avance tecnológico y la capacidad institucional para regular sus efectos. La ausencia de un marco normativo específico en materia de inteligencia artificial, sumada a condiciones estructurales como la informalidad laboral y la desigualdad educativa, limita la capacidad del país para gestionar de manera efectiva la transformación del trabajo. En este sentido, el principal desafío no radica únicamente en diseñar nuevas regulaciones, sino en construir una estrategia integral que articule innovación, educación, protección social y desarrollo económico.

A partir de este diagnóstico, el capítulo propone un enfoque de gobernanza de la inteligencia artificial basado en principios como la centralidad del trabajo humano, la justicia distributiva, la transparencia algorítmica y la adaptabilidad regulatoria. Asimismo, se plantean instrumentos concretos, como evaluaciones de impacto laboral, fortalecimiento de derechos laborales digitales y sistemas de reconversión profesional, que permiten gestionar los efectos de la automatización sin inhibir la innovación.

En este sentido, la regulación de la inteligencia artificial no debe concebirse como un mecanismo de contención tecnológica, sino como una herramienta para orientar su desarrollo en función de objetivos sociales. La pregunta no es si la inteligencia artificial debe ser regulada, sino cómo hacerlo de manera que sus beneficios se distribuyan de forma equitativa y sus costos no recaigan de manera desproporcionada en los sectores más vulnerables.

Finalmente, es importante subrayar que el futuro del trabajo en la era de la inteligencia artificial no está predeterminado. Si bien las tendencias tecnológicas imponen ciertas condiciones, las decisiones políticas, institucionales y sociales desempeñan un papel fundamental en la configuración de los resultados. En este sentido, la inteligencia artificial no debe entenderse como una fuerza autónoma e inevitable, sino como un proceso social que puede —y debe— ser orientado a través de mecanismos de gobernanza adecuados.

En consecuencia, el desafío central para las sociedades contemporáneas, y particularmente para países como México, consiste en transitar hacia modelos de regulación que no solo respondan a los cambios tecnológicos, sino que anticipen sus efectos y promuevan una integración más justa, inclusiva y sostenible de la inteligencia artificial en el mundo del trabajo. Solo a través de este enfoque será posible construir un futuro en el que la innovación tecnológica y la dignidad humana no sean fuerzas opuestas, sino dimensiones complementarias de un mismo proceso de desarrollo.

Referencias

1. Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), 3–30. <https://doi.org/10.1257/jep.33.2.3>
2. Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, 128(6), 2188–2244. <https://doi.org/10.1086/705716>
3. Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4), 1279–1333. <https://doi.org/10.1162/003355303322552801>
4. Braverman, H. (1974). *Labor and monopoly capital: The degradation of work in the twentieth century*. Monthly Review Press.
5. Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.
6. European Commission. (2024). *Artificial Intelligence Act*. <https://artificialintelligenceact.eu/>
7. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
8. Goldman Sachs. (2023). The potentially large effects of artificial intelligence on economic growth. <https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html>

9. Instituto Nacional de Estadística y Geografía (INEGI). (2023). Encuesta Nacional de Ocupación y Empleo. <https://www.inegi.org.mx>
10. International Labour Organization (OIT). (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality. <https://www.ilo.org>
11. Lozano Carrillo, O. (2024). El futuro de la inteligencia artificial. REMINEO.
12. Lozano Carrillo, O., Zavala Osorio, Y., & Couturier Bañuelos, D. P. (2024). Reflexiones sobre la intervención en las organizaciones en la UAM. REMINEO.
13. Lozano Carrillo, O., Zavala Osorio, Y., & Couturier Bañuelos, D. P. (2025). El PIAESS: Una IA para la vida. REMINEO.
14. McKinsey Global Institute. (2023). Generative AI and the future of work in America. <https://www.mckinsey.com>
15. Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2019). OECD Principles on Artificial Intelligence. <https://oecd.ai>
16. Sansores Guerrero, E. A., Lozano Carrillo, O., & Garibay Suárez, A. (2025). Inteligencia artificial en las organizaciones: Innovación, poder y ética en la era digital. REMINEO.
17. The White House. (2022). Blueprint for an AI Bill of Rights. <https://www.whitehouse.gov>
18. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org>
19. World Economic Forum. (2023). The Future of Jobs Report 2023. <https://www.weforum.org/reports/the-future-of-jobs-report-2023>

CAPÍTULO V

REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL Y SOSTENIBILIDAD COMPETITIVA DE LAS MIPYMES DIGITALES EN GUAYAQUIL

Regulation of Artificial Intelligence and Competitive Sustainability of Digital
MSMEs in Guayaquil

Econ. María Paulina Brito Ochoa, PhD.

maria.britooch@ug.edu.ec

Universidad de Guayaquil, Ecuador

<https://orcid.org/0000-0002-0051-2220>

Econ. Fabricio Alfonso Vasco Mora, Mgs.

favascomo@uide.edu.ec

Universidad Internacional del Ecuador

<https://orcid.org/0000-0002-4832-4982>

Resumen

La adopción de inteligencia artificial (IA) en las micro, pequeñas y medianas empresas (MIPYMES) ha dejado de ser un fenómeno periférico para convertirse en un componente creciente de la transformación digital empresarial. En particular, herramientas de automatización, analítica, segmentación, asistentes conversacionales y generación de contenido están ingresando en áreas como marketing, atención al cliente, logística, ventas y gestión administrativa. Sin embargo, la literatura reciente advierte que la incorporación de IA en pequeñas empresas ocurre en un entorno tensionado por brechas de capacidades, limitaciones financieras, dependencia de proveedores tecnológicos y crecientes exigencias regulatorias y éticas (Ayinaddis, 2025; Oldemeyer y otros., 2025; OECD, 2025a; OECD, 2025b). En Ecuador, esta discusión adquiere especial relevancia debido a la vigencia de la Ley Orgánica de Protección de Datos Personales, la política pública de transformación digital y la Estrategia para el Fomento del

Desarrollo y Uso Ético y Responsable de la Inteligencia Artificial en el Ecuador, que configuran una base institucional aún en consolidación (Ley Orgánica de Protección de Datos Personales, 2021; MINTEL, 2025, 2026).

Este capítulo analiza cómo la regulación de la IA impacta y puede fortalecer a las MIPYMES digitales de Guayaquil. A partir de una revisión crítica de literatura científica reciente y de un enfoque aplicado al ecosistema empresarial local, se sostiene que la regulación no debe concebirse como un obstáculo exógeno a la innovación, sino como una infraestructura institucional para ordenar el uso de datos, reducir riesgos, promover confianza y sostener ventajas competitivas. El texto articula tres planos: la transformación digital de las MIPYMES, la gobernanza algorítmica y el entorno normativo ecuatoriano. Asimismo, incorpora un estudio aplicado con escenarios plausibles en comercio, servicios, educación, salud y logística, evitando inventar evidencia empírica donde no existe. El argumento central es que las MiPymes guayaquileñas requieren una regulación proporcional, pedagógica y operativamente viable, capaz de proteger derechos sin bloquear la adopción de tecnologías de IA en contextos de recursos limitados.

Palabras clave: MIPYMES digitales; inteligencia artificial; regulación tecnológica; gobernanza algorítmica; protección de datos; transformación digital; Guayaquil.

Abstract

The adoption of artificial intelligence (AI) in micro, small and medium-sized enterprises (MSMEs) has ceased to be a peripheral phenomenon and has become a growing component of business digital transformation. In particular, automation, analytics, segmentation, conversational assistants and content generation tools are entering areas such as marketing, customer service, logistics, sales and administrative management. However, recent literature warns that the incorporation of AI in small businesses occurs in an environment stressed by capacity gaps, financial constraints, dependence on technology providers, and growing regulatory and ethical demands (Ayinaddis, 2025; Oldemeyer y otros., 2025; OECD, 2025a;

OECD, 2025b). In Ecuador, this discussion acquires special relevance due to the validity of the Organic Law on the Protection of Personal Data, the public policy of digital transformation and the Strategy for the Promotion of the Development and Ethical and Responsible Use of Artificial Intelligence in Ecuador, which make up an institutional base still in consolidation (Organic Law on the Protection of Personal Data, 2021; MINTEL, 2025, 2026).

This chapter analyzes how AI regulation impacts and can strengthen digital MSMEs in Guayaquil. Based on a critical review of recent scientific literature and an approach applied to the local business ecosystem, it is argued that regulation should not be conceived as an exogenous obstacle to innovation, but as an institutional infrastructure to organize the use of data, reduce risks, promote trust and sustain competitive advantages. The text articulates three levels: the digital transformation of MSMEs, algorithmic governance and the Ecuadorian regulatory environment. It also incorporates an applied study with plausible scenarios in trade, services, education, health and logistics, avoiding inventing empirical evidence where it does not exist. The central argument is that Guayaquil's MSMEs require proportional, pedagogical and operationally viable regulation, capable of protecting rights without blocking the adoption of AI technologies in contexts of limited resources.

Keywords: digital MSMEs; artificial intelligence; technological regulation; algorithmic governance; data protection; digital transformation; Guayaquil.

Introducción

La transformación digital ha modificado de manera estructural la forma en que las empresas producen valor, organizan sus procesos y se relacionan con sus clientes. En el caso de las MIPYMES, este proceso ha sido desigual y condicionado por factores como madurez digital, infraestructura tecnológica, disponibilidad de talento, presión competitiva y apoyo institucional (González-Tamayo y otros., 2023; Pozo-Benites, 2025; Valdez-Juárez y otros., 2024). A esta dinámica se suma hoy la expansión de la inteligencia artificial, cuya accesibilidad a través de plataformas

comerciales ha permitido que empresas pequeñas utilicen soluciones antes reservadas a organizaciones de gran escala, especialmente en marketing, predicción de demanda, atención automatizada y análisis de datos (Arroyabe y otros., 2024; Le Dinh y otros., 2025; Ardito y otros., 2024).

Sin embargo, el avance de la IA en pequeñas empresas no puede entenderse únicamente en clave de eficiencia. La literatura reciente muestra que la adopción de IA en MIPYMES está marcada por restricciones de recursos, problemas de integración organizacional, riesgos de opacidad algorítmica, dificultades de gobernanza y desafíos éticos relacionados con sesgos, privacidad, explicabilidad y supervisión humana (Soudi, 2024; Oldemeyer y otros., 2025; Batool y otros., 2025; Gunasekara y otros., 2025). En otras palabras, la IA abre oportunidades productivas, pero también crea nuevas exigencias de regulación y de gestión empresarial responsable.

En Ecuador, esta discusión es especialmente pertinente. La vigencia de la Ley Orgánica de Protección de Datos Personales y su reglamento, junto con la Ley Orgánica para la Transformación Digital y Audiovisual, la Política Pública para la Transformación Digital 2025-2030 y la estrategia nacional de IA de 2026, muestran que el país se encuentra en una etapa de construcción institucional para gobernar tecnologías emergentes (Ley Orgánica de Protección de Datos Personales, 2021; Reglamento General a la Ley Orgánica de Protección de Datos Personales, 2023; MINTEL, 2025, 2026). No obstante, el propio entorno nacional aún presenta vacíos en capacidades, estandarización y apropiación empresarial, especialmente en unidades económicas pequeñas.

Guayaquil constituye un espacio particularmente revelador para este análisis. Su centralidad comercial y de servicios, su dinamismo logístico y su expansión en comercio digital la convierten en un territorio donde la IA puede insertarse con rapidez en procesos cotidianos de negocio. A la vez, la ciudad evidencia las limitaciones típicas de economías urbanas latinoamericanas con fuerte presencia de MIPYMES: heterogeneidad tecnológica, informalidad relativa, formación

desigual y dependencia creciente de plataformas digitales externas (INEC, 2024; CECE, 2025; GEM Ecuador, 2025; Brito Ochoa, 2022).

Planteamiento del problema

Las MIPYMES digitales enfrentan una tensión crítica entre la urgencia de adoptar IA para mantener su competitividad y la dificultad de hacerlo dentro de marcos de gobernanza y cumplimiento suficientemente robustos. La evidencia comparada sugiere que la adopción de IA en pequeñas empresas no depende solo de la disponibilidad tecnológica, sino también de capacidades dinámicas, cultura organizacional, preparación digital, liderazgo y soporte institucional (Arroyabe y otros., 2024; Badghish y otros., 2024; Sánchez y otros., 2025). En contextos de recursos restringidos, la empresa pequeña suele adoptar aplicaciones de IA de forma incremental, centradas en tareas de impacto inmediato, sin desarrollar necesariamente estructuras formales de evaluación de riesgos o gobierno del dato.

Este patrón es problemático por al menos cuatro razones. Primero, porque la IA empresarial se alimenta de datos personales, comerciales y operativos cuya gestión inadecuada puede generar vulneraciones normativas y reputacionales. Segundo, porque muchas MIPYMES consumen herramientas empaquetadas por terceros y, por tanto, operan con escaso control sobre los mecanismos de inferencia, almacenamiento o reutilización de la información. Tercero, porque la automatización puede reproducir sesgos y errores en decisiones de segmentación, priorización o trato al cliente. Cuarto, porque una regulación mal calibrada puede profundizar la brecha competitiva entre empresas pequeñas y grandes, trasladando costos de cumplimiento difíciles de absorber para las primeras (Wirtz y otros., 2022; Papagiannidis y otros., 2025; OECD, 2025b; OECD, 2026).

En Guayaquil, el problema adquiere un perfil operativo claro. Las MIPYMES ya emplean o pueden emplear IA en publicidad digital, ventas online, soporte posventa, logística urbana, análisis de comportamiento de clientes, diseño de contenidos, administración documental y gestión de relaciones comerciales. Sin embargo, la literatura local disponible aún es escasa y más concentrada en digitalización que

en gobernanza específica de IA, lo que obliga a construir un análisis aplicado a partir de estudios sobre transformación digital, madurez empresarial y adopción tecnológica en Ecuador y América Latina (Brito Ochoa, 2022; Mariscal y otros., 2025; Dávila y otros., 2025; Dicao y otros., 2025).

Objetivos

Objetivo general

Analizar cómo la regulación de la inteligencia artificial impacta, condiciona y puede fortalecer el desarrollo de las MIPYMES digitales de Guayaquil, Ecuador, considerando capacidades tecnológicas, protección de datos, riesgos éticos, gobernanza algorítmica, cumplimiento normativo, competitividad e innovación.

Objetivos específicos

1. Examinar la literatura científica reciente sobre adopción de IA en MIPYMES y sus principales barreras y facilitadores.
2. Analizar los fundamentos conceptuales de regulación de IA, ética algorítmica y gobernanza organizacional.
3. Identificar el marco normativo ecuatoriano pertinente para el uso de IA en entornos empresariales.
4. Desarrollar un análisis aplicado al contexto de las mipymes guayaquileñas en sectores representativos.
5. Formular recomendaciones académicamente fundamentadas para política pública, universidades, gremios e iniciativas de apoyo empresarial.

Marco teórico

MIPYMES digitales y transformación digital

La literatura sobre transformación digital de MIPYMES coincide en que digitalizar no equivale únicamente a incorporar software o abrir canales en redes sociales. Se trata, más bien, de una reconfiguración organizacional que involucra procesos,

cultura, capacidades de aprendizaje e innovación, así como nuevas formas de interacción con clientes y mercados (Pozo-Benites, 2025; Yesuf, 2026; Valdez-Juárez y otros., 2024). Estudios en América Latina muestran que la madurez digital de las pequeñas empresas depende de factores internos —liderazgo, habilidades, gestión del cambio— y externos —entorno institucional, acceso a financiamiento, presión competitiva—, lo que explica por qué muchas unidades empresariales avanzan de forma fragmentaria y con baja integración estratégica (González-Tamayo y otros., 2023; Afan Torres y otros., 2025).

Adopción de IA en pequeñas empresas

El campo específico de adopción de IA en MIPYMES ha crecido con rapidez. Revisiones sistemáticas recientes destacan que las aplicaciones más frecuentes se concentran en marketing, atención al cliente, gestión de riesgos, analítica comercial, logística y desarrollo de productos, mientras que persisten obstáculos vinculados con infraestructura, calidad de datos, costos, cultura organizacional, gobernanza y escasez de talento especializado (Oldemeyer y otros., 2025; Ayinaddis, 2025; Chotisarn y otros., 2025). También se ha encontrado que la adopción de IA puede impactar positivamente en crecimiento de ingresos, reducción de riesgos, desempeño e innovación, aunque esos efectos están mediados por preparación tecnológica, capacidades humanas y contexto de mercado (Drydakakis, 2022; Ardito y otros., 2024; Haq y otros., 2025).

Gobernanza algorítmica y regulación de IA

La gobernanza algorítmica se refiere al conjunto de principios, reglas, procesos y mecanismos mediante los cuales una organización diseña, supervisa, documenta y corrige el uso de sistemas algorítmicos. En el debate contemporáneo, este concepto se ha ampliado desde el sector público hacia el ámbito organizacional y corporativo, donde la gobernanza de IA se relaciona con transparencia, supervisión humana, accountability, gestión del riesgo y alineación con valores éticos y regulatorios (Wirtz y otros., 2022; Batool y otros., 2025; de Almeida y otros., 2025). En esa línea, revisiones recientes subrayan que la gobernanza responsable de IA

exige prácticas estructurales, relacionales y procedimentales, no simples declaraciones normativas (Papagiannidis y otros., 2025; Sule, 2026).

Ética, sesgos y responsabilidad

Los estudios sobre ética de la IA han mostrado que la automatización puede reproducir desigualdades, ocultar sesgos en los datos, reforzar discriminaciones y generar opacidad decisonal, especialmente cuando las organizaciones confían excesivamente en salidas algorítmicas sin revisión humana significativa (UNESCO, 2021; Rodgers y otros., 2023; Voutyrakou, 2025). A ello se añade un problema central para las pequeñas empresas: la asimetría cognitiva frente a proveedores tecnológicos, que limita la posibilidad de auditar de forma realista los sistemas que usan. Por ello, la ética aplicada en MIPYMES no puede descansar únicamente en sofisticadas auditorías técnicas, sino en criterios de proporcionalidad, trazabilidad básica, claridad sobre finalidades de uso y mecanismos de intervención humana (Soudi, 2024; Gunasekara y otros., 2025; Liapis, 2026).

IA, innovación y competitividad

La relación entre IA e innovación empresarial no es lineal. La literatura muestra que la IA puede fortalecer productividad, capacidad de decisión, innovación en productos y eficiencia comercial, pero estos beneficios no se distribuyen automáticamente. Dependen de absorción organizacional, gobernanza del conocimiento, integración estratégica y capacidades complementarias (Cooper, 2025; Jobstreibizer, 2025; Štrukelj, 2025). En pequeñas empresas, la IA es más útil cuando se integra como soporte a decisiones y no como sustitución acrítica del juicio estratégico, idea que dialoga con el enfoque de “estrategia amplificada” desarrollado en el documento adjunto para el caso guayaquileño.

Contexto de Ecuador y Guayaquil

Ecuador presenta un tejido empresarial ampliamente compuesto por unidades de menor tamaño. El Registro Estadístico de Empresas de 2024 confirma el peso estructural de las micro y pequeñas empresas en la economía nacional, mientras

que estudios recientes sobre emprendimiento y comercio digital muestran un ecosistema dinámico, aunque todavía marcado por brechas de sofisticación tecnológica y de capacidades organizativas (INEC, 2024; GEM Ecuador, 2025; CECE, 2025). Esta estructura importa porque la regulación de IA en el país no puede pensarse solo para grandes plataformas, sino para empresas pequeñas que consumen soluciones tecnológicas externas en su operación cotidiana.

En materia normativa, el país dispone de una base relevante. La Ley Orgánica de Protección de Datos Personales establece principios, derechos y obligaciones aplicables al tratamiento de datos personales. La Ley Orgánica para la Transformación Digital y Audiovisual impulsa el ecosistema digital, y la política pública de transformación digital 2025-2030 plantea lineamientos sobre modernización, servicios digitales e innovación. Más recientemente, la estrategia ecuatoriana de IA de 2026 propone un marco de desarrollo y uso ético y responsable de la IA, alineado con referencias multilaterales como la UNESCO y la OCDE (MINTEL, 2025, 2026; OECD, 2024; UNESCO, 2021).

Guayaquil, por su parte, concentra condiciones especialmente favorables y desafiantes a la vez. Es un espacio urbano donde convergen comercio, servicios, logística, educación privada, salud y emprendimiento, sectores en los que la IA puede aplicarse con relativa rapidez. Pero también es un entorno donde muchas MIPYMES adoptan tecnología de forma reactiva, impulsadas por competencia y necesidad de supervivencia, más que por una planificación robusta de gobernanza digital. La literatura local sobre digitalización de mipymes en Guayas y Guayaquil sugiere que el salto tecnológico existe, pero no siempre va acompañado de estructuras de gestión, estandarización de procesos o marcos internos claros de seguridad y cumplimiento (Brito Ochoa, 2022; Mariscal y otros., 2025; Bustos y otros., 2025).

Desarrollo analítico

La primera implicación regulatoria para las MIPYMES digitales es la protección de datos personales. La IA empresarial requiere datos y, en muchos casos, datos de

clientes, empleados o usuarios. En un contexto como el ecuatoriano, esto implica obligaciones jurídicas relacionadas con licitud, finalidad, seguridad, minimización y ejercicio de derechos. Para una pequeña empresa que usa CRM, campañas automatizadas o asistentes conversacionales, el problema no es solo jurídico, sino también organizacional: necesita saber qué datos trata, dónde se almacenan, quién accede y bajo qué base se reutilizan (Ley Orgánica de Protección de Datos Personales, 2021; Pérez-Campdesuñer y otros., 2025; Aguilar, 2026).

La segunda implicación es la dependencia de proveedores y la opacidad algorítmica. Muchas MIPYMES no desarrollan modelos propios, sino que usan plataformas de terceros integradas a servicios de marketing, ventas o gestión. Esto crea asimetrías técnicas y contractuales: la empresa usuaria asume responsabilidades frente a sus clientes, pero sin control real sobre ciertos procesos internos del proveedor. Desde la literatura de gobernanza, esta situación exige mecanismos mínimos de diligencia, trazabilidad y evaluación de riesgos antes de adoptar una solución de IA, incluso cuando la empresa no posea capacidades avanzadas de auditoría técnica (Papagiannidis y otros., 2025; Beulen, 2026; Albulayhi, 2026).

La tercera implicación es la transparencia algorítmica. No se trata de exigir a toda MIPYME explicaciones matemáticas sofisticadas, sino de establecer una transparencia razonable y proporcional: informar cuándo existe automatización relevante, cuándo se usan datos para personalización y cuándo es necesario que intervenga un criterio humano. El debate comparado sobre regulación de IA insiste precisamente en esa lógica de proporcionalidad basada en riesgo, visible tanto en estudios de gobernanza como en análisis del modelo europeo de regulación (Hulok, 2025; García-Llorente, 2026; Haidar, 2026).

La cuarta implicación se refiere a la gestión de riesgos y confianza del consumidor. La evidencia reciente muestra que la gobernanza responsable de IA no solo reduce exposición a daños, sino que puede incidir favorablemente en desempeño organizacional, crecimiento sostenible y legitimidad empresarial (Xia, 2026;

Mansour, 2025). Para una MIPYME guayaquileña, esto significa que prácticas como control de accesos, autenticación reforzada, protocolos de incidentes, revisión humana de respuestas automatizadas y límites al tratamiento de información sensible no son cargas accesorias, sino condiciones de confianza en mercados digitales competitivos. En este punto, el documento adjunto ofrece un marco operativo útil al proponer “micro-regulación” de seguridad y uso aceptable como base mínima de gobernanza.

La quinta implicación es la competitividad. La literatura no respalda una visión simplista según la cual toda regulación frena innovación. Más bien, sugiere que la ausencia de reglas claras puede generar incertidumbre, concentración de poder en grandes plataformas y desprotección de empresas pequeñas. Una regulación bien calibrada puede nivelar parcialmente el terreno competitivo al ofrecer certidumbre, proteger datos, aumentar confianza y facilitar adopción responsable. El verdadero riesgo aparece cuando la regulación se diseña sin considerar capacidades reales de las MIPYMES, trasladando costos de cumplimiento incompatibles con su escala (OECD, 2025b; OECD, 2026; Poveda-Valverde y otros., 2026).

Estudio aplicado a las MIPYMES guayaquileñas

No se identifican todavía suficientes estudios públicos, robustos y recientes que permitan afirmar con precisión estadística cómo usan IA, sector por sector, las MIPYMES guayaquileñas. Por ello, el presente capítulo propone un análisis aplicado de carácter contextual y propositivo, basado en evidencia científica sobre adopción de IA en pequeñas empresas y en estudios locales sobre digitalización, comercio electrónico y transformación empresarial en Guayas y Ecuador. Esta delimitación metodológica es importante para evitar inferencias empíricas indebidas.

En el retail digital, una pequeña empresa de ropa, alimentos o accesorios en Guayaquil puede usar IA para redactar fichas de producto, automatizar respuestas, segmentar campañas y anticipar momentos de mayor conversión. La literatura muestra que estos usos son frecuentes y potencialmente rentables en MIPYMES,

especialmente cuando se articulan con analítica de clientes y procesos de marketing digital (Le Dinh y otros., 2025; Ham, 2026). Pero también generan riesgos si la empresa reutiliza bases de datos sin claridad de finalidad, conserva datos más allá de lo necesario o deja decisiones de trato comercial enteramente automatizadas.

En servicios profesionales, por ejemplo, consultorías, estudios jurídicos o servicios migratorios, la IA puede asistir en síntesis normativa, clasificación documental, generación de borradores y respuesta inicial a consultas. Aquí el beneficio de eficiencia es claro, pero el tratamiento de información sensible obliga a un estándar reforzado de diligencia. Estudios sobre IA responsable y gobernanza centrada en datos subrayan precisamente que el valor organizacional de la IA depende de marcos de control sobre calidad del dato, seguridad y accountability (Aguilar, 2026; Sule, 2026). En el caso guayaquileño, el enfoque del documento adjunto resulta especialmente pertinente para este tipo de negocios intensivos en confianza y reputación.

En educación y capacitación, centros de formación, tutorías y emprendimientos edtech pueden incorporar IA para apoyo tutorial, diseño de materiales y seguimiento personalizado. Sin embargo, la literatura sobre ética y gobernanza advierte que el uso intensivo de IA generativa puede afectar calidad, autoría, transparencia y no discriminación, especialmente cuando no existen reglas claras sobre revisión humana y comunicación al usuario (UNESCO, 2021; Liapis, 2026; Zahn-Muñoz, 2026). Por tanto, una adopción responsable en este sector exige no solo cumplir con datos personales, sino también establecer límites pedagógicos y mecanismos de verificación.

En logística y cadenas de suministro, la IA puede ayudar a optimizar rutas, asignar entregas, proyectar demanda y gestionar inventarios. La literatura reciente sobre IA en supply chains de PYMES confirma el potencial de estas aplicaciones, aunque también destaca barreras ligadas a interoperabilidad, costos, habilidades y adopción gradual (Bigliardi, 2026; Fatorachian, 2026). En Guayaquil, donde la

logística urbana es un componente decisivo de competitividad, estas herramientas pueden resultar especialmente valiosas; sin embargo, si las decisiones automatizadas afectan tiempos, cargas o evaluación de trabajadores, la gobernanza debe incorporar también una dimensión laboral y ética.

Discusión

La principal tensión identificada en la literatura es que la IA promete agilidad, pero la regulación exige cautela. Esta tensión, sin embargo, no debe resolverse en términos binarios. Los estudios sobre gobernanza responsable muestran que la regulación puede ser un habilitador de innovación cuando provee criterios claros, incentivos adecuados, mecanismos de aprendizaje institucional y obligaciones proporcionales al riesgo (Papagiannidis y otros., 2025; Poti, 2025; Aarab, 2025). En sentido contrario, la ausencia de gobernanza no equivale a libertad innovadora, sino a incertidumbre, exposición a fallos y mayor dependencia de actores tecnológicos dominantes.

Para las MIPYMES de Guayaquil, la clave no está en replicar esquemas complejos de cumplimiento pensados para grandes corporaciones o entornos regulatorios altamente sofisticados, sino en construir una gobernanza simplificada, escalable y pedagógica. La revisión científica disponible sugiere que las pequeñas empresas adoptan IA de forma más incremental, pragmática y funcional, por lo que las políticas y marcos internos deben acompañar esa lógica sin renunciar a los principios de transparencia, supervisión humana, seguridad y protección de derechos (Steiner, 2025; Poveda-Valverde y otros., 2026; Yesuf, 2026). Este criterio coincide con la orientación práctica del documento adjunto, donde la seguridad básica, la autenticidad de marca y el uso aceptable de sistemas se proponen como condiciones mínimas de una integración sostenible de IA.

Recomendaciones

La evidencia revisada permite sostener que Ecuador necesita avanzar hacia lineamientos de IA aplicables a contextos empresariales de pequeña escala, con enfoque de riesgo y criterios diferenciados. Para la política pública, esto implica

traducir principios generales a guías simples de uso empresarial de IA, incluyendo tratamiento de datos, selección de proveedores, automatización responsable y manejo de incidentes. Esta orientación sería coherente con la experiencia comparada regional, que muestra avances en políticas de IA en América Latina, aunque aún con fragmentación y asimetrías importantes (Benedikter, 2026; Correa, 2025; Burkhardt, 2025).

Para universidades, incubadoras y centros de emprendimiento, la prioridad debería ser la formación de perfiles híbridos capaces de conectar negocio, datos, ética y regulación. La literatura sobre adopción de IA en MIPYMES muestra que la falta de preparación organizacional y de capacidades humanas es una de las barreras más persistentes (Oldemeyer y otros., 2025; Sánchez y otros., 2025; Popa, 2025). En Guayaquil, esto podría traducirse en laboratorios aplicados, clínicas de gobernanza digital y programas de alfabetización en IA responsable para gerentes y equipos de pequeñas empresas.

Para propietarios y gerentes de MIPYMES, la recomendación más importante es evitar la adopción irreflexiva. Antes de implementar una herramienta de IA, deberían poder responder al menos cinco preguntas: qué problema resuelve, qué datos utiliza, qué proveedor interviene, qué error sería más costoso y qué decisión debe seguir siendo humana. Esta lógica coincide con la literatura sobre responsable AI y con la idea de “estrategia amplificada” trabajada en el documento adjunto: la IA es más valiosa cuando refuerza el juicio empresarial, no cuando lo sustituye mecánicamente.

Conclusiones

La regulación de la inteligencia artificial en las MIPYMES digitales de Guayaquil no puede ser abordada como una cuestión marginal ni exclusivamente jurídica. Se trata de un problema de competitividad, organización, confianza y sostenibilidad empresarial. La literatura científica revisada muestra con claridad que la adopción de IA en pequeñas empresas puede generar beneficios importantes en eficiencia, innovación y capacidad de respuesta, pero también revela que esos beneficios

están condicionados por preparación organizacional, gobernanza del dato, supervisión humana y marcos regulatorios apropiados (Ayinaddis, 2025; Oldemeyer y otros., 2025; OECD, 2025b; OECD, 2026).

En el caso ecuatoriano, la existencia de normas sobre protección de datos, políticas de transformación digital y una estrategia nacional de IA configura una base institucional valiosa, aunque todavía insuficiente en términos de aterrizaje operativo para pequeñas empresas. Guayaquil, por su densidad comercial y su dinamismo digital, constituye un escenario especialmente relevante para pensar una gobernanza algorítmica proporcional, realista y contextualizada. La evidencia local disponible, junto con el marco operativo propuesto en el documento adjunto, sugiere que una micro-regulación empresarial basada en seguridad básica, trazabilidad, autenticidad y revisión humana puede ser una vía adecuada para avanzar sin asfixiar la innovación.

En definitiva, las mipymes guayaquileñas no necesitan elegir entre innovar o cumplir. Lo que necesitan es una arquitectura de adopción responsable que les permita innovar con criterios, usar IA con legitimidad y convertir la confianza digital en un activo competitivo duradero.

Referencias

Aarab, A. (2025). *Integrating AI in Public Governance: A Systematic Review*.

Afan Torres, J. M., y otros. (2025). *Factors influencing the digitization process of Peruvian SMEs*.

Aguilar, R. M. (2026). *Data-centric AI governance for responsible organizational value*.

Albulayhi, O. (2026). *AI Governance Risk Tiering for Sustainable Digital Transformation*.

Ardito, L., y otros. (2024). *Artificial intelligence adoption and revenue growth in SMEs*.

Arroyabe, M. F., y otros. (2024). *Analyzing AI adoption in European SMEs: A study of digital capabilities, innovation capabilities and environmental support*.

Badghish, S., y otros. (2024). *Artificial Intelligence Adoption by SMEs to Achieve Sustainability and Innovation*.

Batool, A., y otros. (2025). *AI governance: A systematic literature review*.

Beulen, E. (2026). *Artificial Intelligence Governance Mechanisms—The Chief AI Officer and steering structures*.

Benedikter, R. (2026). *AI in Latin America: Attempts of regulating Artificial Intelligence within the geopolitical paradigm of Active Non-Alignment*.

Bigliardi, B. (2026). *Artificial intelligence in SME supply chains: A bibliometric review*.

Brito Ochoa, M. P. (2022). *Potencial digitalización en MiPymes Guayas*.

Burkhardt, A. (2025). *Artificial intelligence in the work and view of Latin American societies*.

Bustos, C. A. Q., y otros. (2025). *Análisis del impacto de la digitalización en la cadena de suministro de las MIPYMES ecuatorianas.*

CECE. (2025). *Estudio de comercio electrónico en Ecuador: VII medición.*

Chotisarn, N., y otros. (2025). *A bibliometric analysis insights into the intellectual structure of AI adoption in MSMEs.*

Cooper, R. G. (2025). *SMEs' use of AI for new product development: Adoption, readiness and applications.*

Correa, T. (2025). *Reimagining AI in Latin America: Situated narratives of users and governance.*

Dávila, N. V. M., y otros. (2025). *Transformación digital sostenible en las pymes ecuatorianas.*

De Almeida, P. G. R., y otros. (2025). *Artificial intelligence governance: Understanding how organizations incorporate principles into practice.*

Dicao, G. N. B., y otros. (2025). *Transformación digital como motor de crecimiento en la gestión empresarial del sector textil ecuatoriano.*

Drydakis, N. (2022). *Artificial Intelligence and Reduced SMEs' Business Risks.*

Fatorachian, H. (2026). *Overcoming barriers to digital transformation for SMEs in cold chain management.*

García-Llorente, C. (2026). *Algorithmic governance in banking: A comparative analysis of supervisory models.*

GEM Ecuador. (2025). *Global Entrepreneurship Monitor Ecuador 2024/2025.*

González-Tamayo, L. A., y otros. (2023). *Factors influencing small and medium size enterprises development in Latin America.*

Gunasekara, L., y otros. (2025). *A Systematic Review of Responsible Artificial Intelligence.*

Haidar, A. (2026). *Ethics Without Teeth? Challenges and Opportunities in AI Governance*.

Haq, F. ul, y otros. (2025). *Exploring AI Adoption and SME Performance in Resource-Constrained Environments*.

Ham, M. (2026). *Rethinking targeting strategies for SMEs: How artificial intelligence transforms digital marketing*.

Hulok, M. (2025). *The EU model of AI governance: Regulating artificial intelligence under the AI Act*.

INEC. (2024). *Registro Estadístico de Empresas 2024*. Informe adjunto.
(2026). *Gobernanza algorítmica y estrategia amplificada: Un marco regulatorio y operativo para la integración de inteligencia artificial en las MIPYMES de Guayaquil*. Documento de trabajo.

Jobstreibizer, J. (2025). *The impact of artificial intelligence on business models*.

Le Dinh, T., y otros. (2025). *Artificial Intelligence in SMEs: Enhancing Business Operations and Competitiveness*.

Ley Orgánica de Protección de Datos Personales. (2021). Registro Oficial Suplemento 459.

Liapis, C. M. (2026). *Ethics in Artificial Intelligence: A Cross-Sectoral Review*.

Mansour, M. (2025). *Artificial Intelligence, ESG Governance, and Green Competitiveness*.

Mariscal, E. S. S., y otros. (2025). *Digitalización de las Pymes en Ecuador: Desafíos y oportunidades*.

MINTEL. (2025). *Política Pública para la Transformación Digital del Ecuador 2025–2030*.

MINTEL. (2026). *Estrategia para el Fomento del Desarrollo y Uso Ético y Responsable de la Inteligencia Artificial en el Ecuador*.

OECD. (2024). *Recommendation of the Council on Artificial Intelligence*.

OECD. (2025a). *The adoption of artificial intelligence in firms*.

OECD. (2025b). *AI adoption by small and medium-sized enterprises*.

OECD. (2026). *Empowering SMEs in the age of AI*.

Oldemeyer, L., y otros. (2025). *Investigation of artificial intelligence in SMEs: A systematic literature review*.

Papagiannidis, E., y otros. (2025). *Responsible artificial intelligence governance: A review and research agenda*.

Pérez-Campdesuñer, R., y otros. (2025). *Artificial Intelligence in Ecuadorian SMEs: Drivers and Obstacles to Adoption*.

Pozo-Benites, K. B. (2025). *Transformación digital de las PYMES en América Latina*.

Poti, S. P. (2025). *Enabling ethics mechanisms in the governance of artificial intelligence*.

Reglamento General a la Ley Orgánica de Protección de Datos Personales. (2023).

Rodgers, W., y otros. (2023). *An artificial intelligence algorithmic approach to ethical decision-making*.

Sánchez, E., y otros. (2025). *Artificial Intelligence Adoption in SMEs: Survey Based on TOE and Adoption Models*.

Soudi, M. S. (2024). *AI Guidelines and Ethical Readiness Inside SMEs: A Review*.

Steiner, A. (2025). *How and Why Social Enterprises Adopt Artificial Intelligence*.

Štrukelj, T., y otros. (2025). *The Role of AI in Decision-Making and Regulatory Compliance in SMEs*.

Sule, D. F. (2026). *Towards Responsible Artificial Intelligence Adoption*.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.

Valdez-Juárez, L. E., y otros. (2024). *Digital transformation and innovation, dynamic capabilities and performance*.

Vásquez-Vásquez, L. M. (2026). *Artificial Intelligence for Business Decision-Making in Latin America: A systematic review*.

Voutyrakou, D. A. (2025). *Algorithmic Governance: Gender Bias in AI-Generated Systems*.

Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). *Governance of artificial intelligence: A risk and guideline-based integrative framework*.

World Bank. (2025). *Digital Progress and Trends Report 2025: Strengthening AI Foundations*.

Xia, H. (2026). *Exploring the impact of responsible AI governance on business performance*.

Yesuf, Y. M. (2026). *SMEs, digital transformation and sustainable development*.

Zahn-Muñoz, C. (2026). *Policies and Guidelines for the Use of Artificial Intelligence in Latin American journals*.

CAPÍTULO VI

ÉTICA EN EL USO DE LA INTELIGENCIA ARTIFICIAL: ENTRE LA INNOVACIÓN Y LA RESPONSABILIDAD.

Ethics in the use of Artificial Intelligence: Between innovation and
responsibility

Sergio Cruz Hernández

spqr_Checo@hotmail.com

Universidad Autónoma Metropolitana, Unidad Iztapalapa

<https://orcid.org/0009-0006-2016-9809>

Didre Estefani Morales Rodríguez

estefi.mor28@gmail.com

Universidad Autónoma Metropolitana, Unidad Iztapalapa

<https://orcid.org/0009-0007-6958-9772>

Resumen

El presente capítulo analiza la inteligencia artificial desde una perspectiva ética, social y política, destacando que su impacto trasciende el ámbito tecnológico para incidir directamente en la vida cotidiana, el trabajo, la educación, la justicia y los sistemas de gobernanza. A partir de un recorrido histórico sobre el desarrollo de la IA, el texto examina los principales dilemas éticos asociados a su implementación, entre ellos la falsa neutralidad algorítmica, la privacidad, la vigilancia masiva, la opacidad de las “cajas negras”, la automatización del trabajo y la reproducción de desigualdades estructurales. Asimismo, se aborda cómo la inteligencia artificial reconfigura las relaciones de poder y transforma las dinámicas laborales mediante sistemas de gestión algorítmica, automatización de tareas y nuevas formas de control digital.

El análisis evidencia que la IA no constituye una tecnología neutral ni homogénea, sino un fenómeno condicionado por intereses económicos, estructuras sociales y decisiones políticas. En este sentido, se estudian casos concretos relacionados con reconocimiento facial, vigilancia, sistemas judiciales automatizados, desinformación, deepfakes y plataformas digitales, mostrando cómo estas tecnologías pueden amplificar sesgos, afectar derechos fundamentales y generar nuevas formas de exclusión social. Del mismo modo, el capítulo reflexiona sobre los efectos de la IA en ámbitos sensibles como la salud y la educación, donde sus beneficios potenciales coexisten con riesgos asociados a la dependencia tecnológica, la pérdida de autonomía y la desigualdad en el acceso.

Finalmente, el texto analiza los principales marcos éticos y regulatorios impulsados por organismos internacionales, gobiernos y empresas tecnológicas, enfatizando la necesidad de construir modelos de gobernanza flexibles, transparentes y orientados al bienestar colectivo. Se concluye que el futuro de la inteligencia artificial no depende únicamente de los avances técnicos, sino de las decisiones éticas, sociales y políticas que orienten su desarrollo. En consecuencia, la regulación, la responsabilidad compartida y el pensamiento crítico se presentan como elementos indispensables para asegurar que la innovación tecnológica contribuya a sociedades más justas, inclusivas y democráticas.

Palabras clave: Inteligencia artificial, ética digital, algoritmos, vigilancia, automatización, trabajo, desigualdad tecnológica, gobernanza, transparencia algorítmica, regulación de la IA.

Abstract

This chapter analyzes artificial intelligence from an ethical, social and political perspective, highlighting that its impact transcends the technological field to directly affect daily life, work, education, justice and governance systems. Based on a historical overview of the development of AI, the text examines the main ethical dilemmas associated with its implementation, including false algorithmic neutrality,

privacy, mass surveillance, the opacity of "black boxes", the automation of work and the reproduction of structural inequalities. It also addresses how artificial intelligence reconfigures power relations and transforms work dynamics through algorithmic management systems, task automation and new forms of digital control.

The analysis shows that AI is not a neutral or homogeneous technology, but a phenomenon conditioned by economic interests, social structures and political decisions. In this sense, specific cases related to facial recognition, surveillance, automated judicial systems, disinformation, deepfakes and digital platforms are studied, showing how these technologies can amplify biases, affect fundamental rights and generate new forms of social exclusion. Similarly, the chapter reflects on the effects of AI in sensitive areas such as health and education, where its potential benefits coexist with risks associated with technological dependence, loss of autonomy and inequality in access.

Finally, the text analyzes the main ethical and regulatory frameworks promoted by international organizations, governments, and technology companies, emphasizing the need to build flexible, transparent, and collective well-being-oriented governance models. It is concluded that the future of artificial intelligence does not depend only on technical advances, but also on the ethical, social and political decisions that guide its development. Consequently, regulation, shared responsibility and critical thinking are presented as indispensable elements to ensure that technological innovation contributes to fairer, more inclusive and democratic societies.

Keywords: Artificial intelligence, digital ethics, algorithms, surveillance, automation, work, inequality tecnológica, gobernanza, transparencia algorítmica, regulación de la IA.

Parte I: Fundamentos, historia y primeros dilemas

Introducción

La inteligencia artificial (IA) se ha consolidado como una de las tecnologías más influyentes del mundo contemporáneo. Su presencia atraviesa múltiples dimensiones de la vida cotidiana: desde los sistemas de recomendación que sugieren qué contenido consumir, hasta algoritmos que apoyan diagnósticos médicos, decisiones financieras o procesos de contratación laboral.

Sin embargo, el avance acelerado de la IA ha superado, en muchos casos, la capacidad de las sociedades para reflexionar críticamente sobre sus implicaciones. La discusión no se limita a lo que la IA puede hacer, sino a lo que debería hacer, en qué condiciones y con qué límites.

Este capítulo parte de una premisa central: la inteligencia artificial no es únicamente una herramienta técnica, sino un fenómeno social, político y ético. Su diseño, implementación y uso están atravesados por valores, intereses y estructuras de poder.

A lo largo de este texto se abordarán los principales dilemas éticos asociados a la IA, integrando ejemplos concretos que ilustran cómo estos problemas se manifiestan en la práctica.

Orígenes de la inteligencia artificial: entre ciencia y ambición

La inteligencia artificial no surgió de manera espontánea, sino como resultado de una serie de desarrollos científicos que buscaban comprender y replicar la inteligencia humana.

El punto de partida formal suele ubicarse en la conferencia de Dartmouth de 1956, donde John McCarthy propuso el término “inteligencia artificial”. La idea era ambiciosa: construir máquinas capaces de simular procesos cognitivos humanos como el razonamiento, el aprendizaje y la resolución de problemas.

Durante sus primeras décadas, la IA se centró en sistemas basados en reglas. Un ejemplo clásico son los programas capaces de jugar ajedrez o resolver problemas matemáticos mediante instrucciones predefinidas. Sin embargo, estos sistemas tenían limitaciones importantes: funcionaban bien en entornos controlados, pero fallaban en situaciones complejas o ambiguas.

Un momento clave en la historia de la IA ocurrió en 1997, cuando la computadora **Deep Blue** de IBM derrotó al campeón mundial de ajedrez Garry Kasparov. Este evento fue simbólicamente poderoso, ya que mostró que una máquina podía superar a un humano en una tarea considerada altamente intelectual.

Años después, en 2016, el sistema **AlphaGo** de DeepMind venció al campeón mundial de Go, un juego mucho más complejo que el ajedrez. Este logro fue posible gracias a técnicas de aprendizaje profundo, que permiten a los sistemas aprender a partir de grandes cantidades de datos.

Estos hitos reflejan un cambio fundamental: la IA dejó de depender exclusivamente de reglas explícitas y comenzó a aprender de la experiencia.

El auge contemporáneo de la IA

El desarrollo reciente de la IA está estrechamente ligado a tres factores:

1. **Disponibilidad masiva de datos**
2. **Incremento en la capacidad de cómputo**
3. **Avances en algoritmos de aprendizaje automático**

Empresas como Google, Amazon y Microsoft han integrado sistemas de IA en productos y servicios utilizados por millones de personas.

Por ejemplo, los sistemas de recomendación de plataformas de streaming analizan patrones de consumo para sugerir contenido personalizado. Aunque esto puede mejorar la experiencia del usuario, también plantea preguntas sobre la creación de burbujas informativas y la limitación de la diversidad cultural.

Otro ejemplo es el uso de IA en redes sociales, donde algoritmos priorizan ciertos contenidos en función de su potencial para generar interacción. Este tipo de sistemas ha sido señalado por amplificar información polarizante o incluso desinformación.

Primer dilema ético: la falsa neutralidad

Uno de los errores más comunes al analizar la inteligencia artificial es asumir que es neutral. Esta idea se basa en la creencia de que los algoritmos simplemente procesan datos de manera objetiva.

Sin embargo, los sistemas de IA están profundamente condicionados por:

- Los datos con los que se entrenan
- Las decisiones de diseño de sus desarrolladores
- Los objetivos que se les asignan

Un caso emblemático es el de sistemas de reconocimiento facial que presentan mayores tasas de error al identificar a personas con tonos de piel más oscuros. Esto se debe, en gran medida, a que fueron entrenados con bases de datos poco diversas.

Otro ejemplo relevante ocurrió con herramientas de reclutamiento automatizado que penalizaban a candidatas mujeres debido a patrones históricos en los datos de contratación. En este caso, el sistema no “decidió discriminar”, sino que reprodujo desigualdades existentes, estos ejemplos muestran que la IA no solo refleja la realidad social, sino que puede reforzarla.

Segundo dilema: privacidad y vigilancia

La inteligencia artificial depende de grandes volúmenes de datos, muchos de los cuales son personales. Esto ha dado lugar a lo que algunos autores denominan una “economía de la vigilancia”.

Empresas tecnológicas recopilan información sobre hábitos de consumo, ubicaciones, interacciones sociales y preferencias individuales. Esta información se utiliza para entrenar algoritmos que predicen comportamientos futuros.

Un ejemplo claro es la publicidad personalizada. Plataformas digitales pueden mostrar anuncios altamente específicos basados en el historial de navegación de un usuario. Si bien esto puede parecer útil, también implica una forma de seguimiento constante.

En contextos más extremos, algunos gobiernos han implementado sistemas de vigilancia masiva apoyados en IA, como el uso de cámaras con reconocimiento facial para monitorear a la población. Esto plantea serias preocupaciones sobre libertades individuales y derechos humanos.

Tercer dilema: decisiones automatizadas y responsabilidad

A medida que la IA se integra en procesos críticos, la cuestión de la responsabilidad se vuelve central. Por ejemplo, en el ámbito financiero, los algoritmos pueden decidir si una persona es apta para recibir un crédito. En el sector salud, pueden sugerir diagnósticos o tratamientos. En el transporte, los vehículos autónomos deben tomar decisiones en situaciones de riesgo.

En todos estos casos surge una pregunta fundamental: ¿quién responde cuando algo sale mal?

Un caso relevante es el de accidentes involucrando vehículos autónomos. Aunque estos sistemas están diseñados para reducir errores humanos, no son infalibles. Cuando ocurre un accidente, es difícil determinar si la responsabilidad recae en el fabricante, el programador o el usuario. Además, muchos sistemas de IA funcionan como “cajas negras”, lo que significa que sus procesos internos no son fácilmente comprensibles. Esto dificulta la rendición de cuentas.

Hasta este punto, es evidente que la inteligencia artificial no puede analizarse únicamente desde una perspectiva técnica. Sus implicaciones éticas atraviesan dimensiones fundamentales de la vida social: la justicia, la privacidad, la autonomía

y la responsabilidad. La historia de la IA muestra que su desarrollo ha estado impulsado por una combinación de ambición científica y oportunidades económicas. Sin embargo, su impacto actual exige una reflexión más profunda sobre sus consecuencias.

Parte II: Impacto social, trabajo y desigualdad

La inteligencia artificial en la vida cotidiana

La inteligencia artificial ya no es una tecnología lejana o exclusiva de laboratorios especializados; se ha integrado de manera casi invisible en la vida cotidiana. Cada interacción digital —desde desbloquear un teléfono con reconocimiento facial hasta recibir recomendaciones en plataformas digitales— está mediada por sistemas algorítmicos.

Por ejemplo, aplicaciones como TikTok o Instagram utilizan algoritmos altamente sofisticados para determinar qué contenido mostrar a cada usuario. Estos sistemas analizan patrones de comportamiento, tiempo de visualización y preferencias implícitas para maximizar la atención.

El problema ético no radica únicamente en el uso de estos algoritmos, sino en su opacidad. La mayoría de los usuarios no comprende cómo se construyen estas recomendaciones ni qué criterios se utilizan. Esto genera una relación asimétrica donde la tecnología influye en las decisiones individuales sin que exista plena conciencia de ello.

IA y manipulación del comportamiento

Uno de los debates más relevantes en torno a la IA es su capacidad para influir en el comportamiento humano. Un caso ampliamente discutido es el escándalo de Cambridge Analytica, donde datos de millones de usuarios de Facebook fueron utilizados para diseñar campañas políticas altamente segmentadas. A través del análisis de datos, se construyeron perfiles psicológicos que permitían dirigir mensajes específicos a distintos grupos.

Este tipo de prácticas plantea preguntas profundas sobre la autonomía individual. Si las decisiones pueden ser influenciadas mediante algoritmos que conocen nuestras preferencias mejor que nosotros mismos, ¿hasta qué punto seguimos siendo completamente libres en nuestras elecciones?

Otro ejemplo se encuentra en plataformas de comercio electrónico como Amazon, donde los sistemas de recomendación no solo sugieren productos, sino que también pueden incentivar compras impulsivas mediante estrategias como la urgencia artificial (“últimas unidades disponibles”).

El impacto de la IA en el trabajo

Uno de los temas más debatidos en la ética de la inteligencia artificial es su impacto en el empleo.

Por un lado, la IA tiene el potencial de automatizar tareas repetitivas, lo que puede aumentar la eficiencia y reducir costos. Por otro lado, esta automatización puede desplazar a trabajadores, especialmente en sectores donde las tareas son altamente estructuradas.

Por ejemplo:

- En la industria manufacturera, robots inteligentes han reemplazado procesos manuales.
- En el sector servicios, chatbots y asistentes virtuales han reducido la necesidad de atención humana en centros de contacto.
- En el ámbito administrativo, sistemas automatizados pueden procesar grandes volúmenes de información en segundos.

Sin embargo, el impacto no es uniforme. Mientras algunos empleos desaparecen, otros se transforman o surgen nuevos roles relacionados con el desarrollo y supervisión de tecnologías.

El dilema ético aquí no es simplemente la automatización, sino la **distribución de sus efectos**. ¿Quién se beneficia de la eficiencia generada por la IA? ¿Qué ocurre

con quienes pierden su empleo? ¿Existen mecanismos para su reconversión laboral?

Algoritmos en el ámbito laboral: control y vigilancia

Más allá del desplazamiento laboral, la IA también está transformando las condiciones de trabajo. En empresas de logística como Amazon, los trabajadores son monitoreados mediante sistemas que evalúan su productividad en tiempo real. Estos sistemas pueden medir tiempos de entrega, pausas y eficiencia operativa, generando métricas constantes sobre el desempeño.

Este tipo de control algorítmico introduce nuevas formas de presión laboral. A diferencia de la supervisión tradicional, que era visible y directa, el control basado en IA es continuo y muchas veces invisible.

Un caso interesante es el de plataformas de transporte como Uber, donde los conductores están sujetos a sistemas de evaluación automática basados en calificaciones de usuarios y métricas de desempeño. Aunque estos sistemas buscan garantizar calidad, también pueden generar incertidumbre y dependencia de criterios poco transparentes.

Desigualdad tecnológica

La inteligencia artificial no afecta a todos por igual. Su desarrollo y acceso están concentrados en ciertos países, empresas y grupos sociales. Las grandes corporaciones tecnológicas como Google o Microsoft poseen los recursos necesarios para desarrollar sistemas avanzados de IA. Esto les otorga una ventaja significativa en términos económicos y de influencia global.

A nivel internacional, existe una brecha entre países que lideran el desarrollo tecnológico y aquellos que dependen de tecnologías importadas. Esto puede generar nuevas formas de dependencia y desigualdad. Finalmente, a nivel social, el acceso a tecnologías basadas en IA también puede estar condicionado por factores como el nivel socioeconómico, la educación o la infraestructura digital disponible.

Transformación del trabajo en la era de la inteligencia artificial

Más allá de los efectos que la inteligencia artificial genera en la forma en que las personas acceden a la información o toman decisiones cotidianas, su impacto se extiende hacia uno de los ejes fundamentales de la organización social: el trabajo. En este sentido, la IA no solo influye en cómo se produce conocimiento o se gestionan datos, sino que comienza a redefinir las condiciones mismas bajo las cuales se crea valor en las economías contemporáneas.

A diferencia de procesos tecnológicos anteriores, en los que la automatización se concentraba principalmente en tareas físicas o repetitivas, la inteligencia artificial contemporánea introduce la posibilidad de intervenir en actividades cognitivas, analíticas e incluso creativas. Este cambio marca una ruptura significativa, ya que desplaza el foco de la automatización hacia ámbitos que históricamente habían sido considerados exclusivamente humanos.

Sin embargo, este proceso no implica necesariamente la desaparición inmediata del trabajo humano. Más bien, se manifiesta como una reconfiguración progresiva de tareas, en la que ciertas actividades son automatizadas, mientras otras se transforman o emergen. En este contexto, el trabajo deja de ser una categoría estable para convertirse en un espacio dinámico, sujeto a constantes redefiniciones.

Ética del desplazamiento laboral: entre eficiencia y justicia social

En este escenario, la discusión sobre la inteligencia artificial en el trabajo no puede limitarse a un análisis técnico o económico, sino que requiere incorporar una dimensión ética que permita problematizar sus implicaciones sociales. En particular, el desplazamiento laboral deja de ser un fenómeno meramente funcional para convertirse en un dilema que involucra decisiones sobre qué tipo de desarrollo tecnológico se considera legítimo.

Uno de los principales problemas éticos radica en la distribución de los beneficios y costos asociados a la automatización. Mientras que las ganancias derivadas de

la inteligencia artificial tienden a concentrarse en actores con capacidad tecnológica, los efectos negativos —como la pérdida de empleo o la precarización— recaen con mayor frecuencia en sectores vulnerables. Esta asimetría pone en evidencia que el desarrollo tecnológico no es neutral, sino que se inserta en estructuras sociales preexistentes.

En este sentido, la inteligencia artificial no solo transforma el trabajo, sino también los mecanismos mediante los cuales se asignan oportunidades. Las decisiones automatizadas, que en apariencia responden a criterios objetivos, pueden reproducir sesgos estructurales si no se diseñan con una perspectiva ética. De esta manera, la cuestión no es únicamente si la tecnología es eficiente, sino si su implementación es socialmente justa.

Reconfiguración del trabajo: transformación más que sustitución

A pesar de las preocupaciones que genera la automatización, la evidencia sugiere que la inteligencia artificial no está eliminando el trabajo en términos absolutos, sino transformándolo. Este proceso se caracteriza por la fragmentación de las actividades laborales en tareas específicas, algunas de las cuales pueden ser automatizadas, mientras que otras requieren intervención humana.

En consecuencia, el trabajo contemporáneo se configura cada vez más como una interacción entre personas y sistemas inteligentes. Esta relación no implica necesariamente una sustitución total, sino una forma de colaboración que redefine las competencias necesarias. Sin embargo, esta transformación también introduce tensiones, particularmente en lo que respecta a la calidad del empleo y la estabilidad laboral.

El desafío, por tanto, no radica únicamente en la cantidad de trabajo disponible, sino en las condiciones bajo las cuales este se desarrolla. La creación de nuevos empleos no garantiza necesariamente la mejora en las condiciones laborales, lo que obliga a replantear la discusión más allá de una lógica cuantitativa.

Control algorítmico y nuevas formas de poder en el trabajo

Otro aspecto relevante de esta transformación es la manera en que la inteligencia artificial reconfigura las relaciones de poder dentro de los entornos laborales. La incorporación de sistemas algorítmicos en la gestión del trabajo ha dado lugar a nuevas formas de control que operan de manera menos visible, pero más constante.

En este contexto, la supervisión del trabajo ya no depende exclusivamente de relaciones jerárquicas tradicionales, sino de sistemas que registran y evalúan el desempeño de forma continua. Este tipo de gestión permite optimizar procesos, pero también introduce una vigilancia permanente que puede afectar la autonomía de los trabajadores.

Aunado a ello, la opacidad de los algoritmos dificulta la comprensión de los criterios utilizados en la toma de decisiones. Esto genera una asimetría significativa entre quienes diseñan los sistemas y quienes son evaluados por ellos, lo que plantea interrogantes sobre la transparencia y la posibilidad de cuestionar dichas decisiones.

Desigualdad tecnológica y trabajo: una ética situada

El impacto de la inteligencia artificial en el trabajo no es homogéneo, sino que depende de las condiciones específicas en las que se implementa. En contextos como el de México, caracterizados por altos niveles de informalidad laboral y desigualdad estructural, la incorporación de tecnologías puede amplificar brechas existentes.

En este sentido, la inteligencia artificial no actúa en un vacío, sino sobre estructuras sociales que condicionan su impacto. Mientras que en algunos sectores puede generar mejoras en productividad, en otros puede traducirse en precarización o exclusión. Por ello, cualquier reflexión ética sobre la IA debe considerar el contexto en el que se desarrolla.

Responsabilidad, gobernanza y trabajo en la era de la IA

Frente a estos desafíos, resulta necesario avanzar hacia modelos de gobernanza que permitan gestionar la relación entre inteligencia artificial y trabajo de manera más equilibrada. La responsabilidad en el uso de estas tecnologías no puede recaer en un solo actor, sino que debe distribuirse entre gobiernos, empresas y sociedad.

En este sentido, la regulación no debe entenderse únicamente como un mecanismo de control, sino como una herramienta para orientar el desarrollo tecnológico hacia objetivos sociales. Esto implica garantizar transparencia en los sistemas, así como la posibilidad de supervisar y cuestionar sus efectos.

En última instancia, el futuro del trabajo en la era de la inteligencia artificial no está determinado exclusivamente por la tecnología, sino por las decisiones que se tomen en torno a su uso. La ética y la gobernanza, por tanto, se convierten en elementos centrales para asegurar que la innovación tecnológica contribuya al bienestar colectivo y no a la profundización de desigualdades.

IA en la educación: oportunidades y riesgos

El uso de inteligencia artificial en la educación ha crecido significativamente. Plataformas educativas utilizan algoritmos para personalizar el aprendizaje, identificar áreas de mejora y ofrecer contenidos adaptados a cada estudiante, sin embargo, esto también plantea desafíos éticos.

Por ejemplo, los sistemas de evaluación automatizada pueden reproducir sesgos o limitar la creatividad al priorizar respuestas estandarizadas. Asimismo, el uso intensivo de datos estudiantiles genera preocupaciones sobre privacidad. Otro riesgo es la dependencia excesiva de la tecnología. Si bien la IA puede apoyar el aprendizaje, no puede sustituir completamente la interacción humana, que es fundamental en los procesos educativos.

IA en la salud: decisiones críticas

Uno de los campos donde la IA tiene mayor potencial y riesgo es la salud. Sistemas basados en IA pueden analizar imágenes médicas, detectar patrones en historiales clínicos y apoyar diagnósticos. Esto puede mejorar la precisión y rapidez en la atención médica.

Sin embargo, los errores en este contexto pueden tener consecuencias graves. Además, existe el riesgo de confiar excesivamente en los sistemas automatizados sin cuestionar sus resultados.

Un problema adicional es el acceso desigual. Mientras algunos hospitales cuentan con tecnologías avanzadas, otros carecen de recursos básicos, lo que puede ampliar las brechas en la atención médica.

Hasta este punto, los ejemplos analizados muestran que la inteligencia artificial no es una tecnología neutral ni homogénea en sus efectos. Su impacto depende del contexto en el que se implementa, así como de las decisiones políticas, económicas y sociales que la rodean. La ética en la IA no puede reducirse a principios abstractos; debe considerar situaciones concretas donde se ponen en juego derechos, oportunidades y condiciones de vida.

Parte III: Justicia, vigilancia y dilemas críticos

IA en el sistema de justicia

Uno de los usos más controversiales de la inteligencia artificial se encuentra en el ámbito judicial. En algunos países, se han implementado sistemas algorítmicos para evaluar el riesgo de reincidencia de personas acusadas o condenadas por delitos.

Un caso ampliamente debatido es el sistema **COMPAS** (Correctional Offender Management Profiling for Alternative Sanctions), utilizado en Estados Unidos para apoyar decisiones judiciales. Este sistema analiza distintos factores como historial criminal, entorno social y comportamiento para generar una puntuación de riesgo.

El problema surge cuando estos sistemas presentan sesgos. Investigaciones han señalado que COMPAS tendía a asignar mayor riesgo a personas afroamericanas en comparación con personas blancas, incluso en condiciones similares.

Este ejemplo ilustra un dilema ético central: ¿es legítimo delegar decisiones que afectan la libertad de las personas a sistemas que pueden reproducir desigualdades estructurales? Además, estos sistemas suelen ser opacos. Las personas afectadas no siempre tienen acceso a la lógica detrás de las decisiones, lo que dificulta su derecho a la defensa.

Reconocimiento facial y vigilancia masiva

El uso de tecnologías de reconocimiento facial ha crecido de manera significativa en los últimos años. Estas herramientas permiten identificar personas a partir de imágenes o video en tiempo real.

Empresas como Clearview AI han desarrollado sistemas capaces de rastrear rostros en grandes bases de datos obtenidas de internet. Aunque estas tecnologías pueden ser útiles en investigaciones criminales, también plantean riesgos importantes.

En algunos países, el reconocimiento facial se ha utilizado para monitorear a la población en espacios públicos. Esto puede contribuir a la seguridad, pero también abre la puerta a formas de vigilancia masiva que afectan la privacidad y la libertad individual. El problema ético no es solo el uso de la tecnología, sino su posible abuso. Sin regulaciones claras, estas herramientas pueden emplearse para controlar, discriminar o reprimir.

La lógica de la “caja negra”

Uno de los mayores desafíos de la inteligencia artificial contemporánea es su falta de transparencia.

Muchos sistemas avanzados, especialmente aquellos basados en aprendizaje profundo, funcionan como “cajas negras”. Esto significa que, aunque se conocen

los datos de entrada y los resultados, el proceso interno que conecta ambos no es fácilmente comprensibles.

Este problema es crítico en contextos sensibles. Por ejemplo:

- Un sistema que rechaza una solicitud de crédito
- Un algoritmo que prioriza pacientes en un hospital
- Un modelo que determina la elegibilidad para beneficios sociales

En estos casos, la falta de explicabilidad limita la posibilidad de cuestionar decisiones y detectar errores o injusticias.

Desde una perspectiva ética, esto plantea una tensión entre eficiencia y transparencia. Los sistemas más precisos no siempre son los más comprensibles.

IA y seguridad: decisiones en situaciones límite

La inteligencia artificial también ha sido incorporada en sistemas de seguridad y defensa, incluyendo el desarrollo de armas autónomas. Estas tecnologías pueden identificar objetivos y tomar decisiones sin intervención humana directa. Esto ha generado un debate global sobre los llamados “robots asesinos”.

Organizaciones como la ONU han discutido la necesidad de regular o incluso prohibir este tipo de sistemas. El principal argumento es que delegar decisiones de vida o muerte a máquinas plantea problemas éticos fundamentales.

Un dilema clásico en este contexto es el equivalente al “problema del tranvía”: ¿cómo debe actuar un sistema autónomo en una situación donde cualquier decisión implica daño? Aunque, este tipo de escenarios puede parecer extremo, refleja una cuestión profunda: la dificultad de traducir valores humanos en reglas algorítmicas.

Desinformación y generación de contenido

La inteligencia artificial también ha transformado la producción de información.

Herramientas avanzadas pueden generar texto, imágenes, audio y video con un alto grado de realismo. Esto tiene aplicaciones positivas —como la creación de contenido educativo—, pero también riesgos importantes.

Uno de los más preocupantes es la creación de **deepfakes**, es decir, contenidos manipulados que pueden hacer parecer que una persona dijo o hizo algo que nunca ocurrió.

Este tipo de tecnología puede utilizarse para:

- Difundir desinformación
- Manipular opiniones públicas
- Dañar reputaciones

En contextos políticos, esto puede afectar procesos democráticos. La capacidad de producir contenido falso creíble desafía la confianza en la información.

IA y toma de decisiones públicas

Los gobiernos también han comenzado a utilizar inteligencia artificial para diseñar políticas públicas, asignar recursos y gestionar servicios.

Por ejemplo, algunos sistemas analizan datos para predecir zonas con mayor probabilidad de criminalidad (policing predictivo). Aunque esto puede mejorar la eficiencia, también puede reforzar sesgos históricos.

Si un algoritmo identifica ciertos barrios como “de alto riesgo” basándose en datos pasados, puede contribuir a una mayor vigilancia en esas zonas, perpetuando dinámicas de desigualdad.

Este fenómeno muestra cómo la IA puede institucionalizar decisiones que, aunque aparentemente objetivas, están cargadas de implicaciones sociales.

Dilema central: ¿puede la IA ser justa?

Todos los casos anteriores apuntan a una pregunta fundamental: ¿es posible construir sistemas de inteligencia artificial completamente justos? La respuesta no

es sencilla. La justicia no es un concepto único; depende de valores, contextos y perspectivas. Lo que se considera justo en un caso puede no serlo en otro.

Por ejemplo:

- ¿Es justo maximizar la eficiencia si esto afecta a ciertos grupos?
- ¿Debe priorizarse la igualdad de resultados o de oportunidades?
- ¿Cómo se equilibran derechos individuales y beneficios colectivos?

Estas preguntas no pueden resolverse únicamente mediante programación. Requieren deliberación ética, participación social y reflexión crítica.

A medida que la inteligencia artificial se integra en ámbitos más sensibles, sus implicaciones éticas se vuelven más profundas y complejas. Los casos analizados muestran que los riesgos no son hipotéticos: ya están presentes en distintas partes del mundo. La IA no solo amplifica capacidades humanas, sino también sus desigualdades, sesgos y conflictos.

Esto refuerza la necesidad de abordar la tecnología desde una perspectiva ética integral, que no se limite a corregir errores técnicos, sino que cuestione los fines y valores que orientan su desarrollo.

Parte IV: Principios éticos, regulación y gobernanza

La necesidad de marcos éticos en la IA

A medida que los riesgos de la inteligencia artificial se han vuelto más evidentes, distintos actores gobiernos, empresas, organizaciones internacionales y academia han impulsado la creación de marcos éticos para orientar su desarrollo y uso.

Estos marcos buscan responder a una pregunta fundamental: ¿cómo garantizar que la IA beneficie a la sociedad sin vulnerar derechos fundamentales?

Sin embargo, establecer principios éticos no es suficiente. El verdadero desafío radica en su implementación, ya que implica traducir valores abstractos en prácticas concretas dentro de contextos organizacionales, políticos y económicos diversos.

Principios fundamentales de la ética en IA

Aunque existen múltiples propuestas, la mayoría coincide en un conjunto de principios básicos que orientan el desarrollo responsable de la inteligencia artificial.

1. Transparencia

Los sistemas de IA deben ser comprensibles y explicables, especialmente cuando afectan decisiones importantes.

Esto implica que los usuarios puedan saber:

- Cómo se toman las decisiones
- Qué datos se utilizan
- Qué criterios influyen en los resultados

Sin transparencia, se limita la capacidad de cuestionar y corregir posibles errores.

2. Justicia y no discriminación

La IA debe evitar reproducir o amplificar desigualdades existentes.

Esto implica:

- Evaluar los datos de entrenamiento
- Detectar sesgos
- Diseñar sistemas que consideren la diversidad social

Sin este principio, la IA puede legitimar prácticas injustas bajo la apariencia de objetividad.

3. Responsabilidad

Debe existir claridad sobre quién es responsable de las decisiones tomadas por sistemas de IA.

Esto incluye:

- Desarrolladores
- Empresas
- Instituciones que implementan la tecnología

La ausencia de responsabilidad genera vacíos éticos y legales.

4. Privacidad

El uso de datos debe respetar los derechos de las personas.

Esto implica:

- Consentimiento informado
- Protección de información personal
- Limitación en la recolección de datos

La privacidad no es solo un aspecto técnico, sino un derecho fundamental.

5. Beneficio social

La IA debe orientarse hacia el bienestar colectivo.

Esto implica cuestionar:

- ¿A quién beneficia la tecnología?
- ¿Quién queda excluido?
- ¿Qué impactos genera a largo plazo?

Iniciativas internacionales

Diversas organizaciones han trabajado en la construcción de marcos éticos globales.

La UNESCO, por ejemplo, ha desarrollado recomendaciones sobre la ética de la inteligencia artificial que buscan guiar a los Estados en la creación de políticas públicas responsables.

Por su parte, la Unión Europea ha impulsado regulaciones más concretas, como la Ley de Inteligencia Artificial (AI Act), que clasifica los sistemas de IA según su nivel de riesgo y establece obligaciones específicas para cada categoría. Estas iniciativas reflejan un consenso creciente: la IA no puede desarrollarse sin regulación.

El papel de las empresas tecnológicas

Las empresas tienen un rol central en el desarrollo de la inteligencia artificial, ya que son responsables de gran parte de la innovación en este campo. Compañías como Google, Microsoft y OpenAI han publicado principios éticos y lineamientos internos para el uso de IA.

Sin embargo, existe una tensión importante: las empresas operan bajo incentivos económicos, lo que puede entrar en conflicto con consideraciones éticas.

Por ejemplo:

- Priorizar crecimiento sobre privacidad
- Implementar sistemas sin evaluación suficiente
- Minimizar riesgos para acelerar innovación

Esto plantea la necesidad de mecanismos externos de supervisión.

Gobernanza de la IA

La gobernanza de la inteligencia artificial implica la coordinación entre distintos actores:

- Gobiernos
- Empresas

- Academia
- Sociedad civil

No existe un modelo único, pero sí un consenso en que la regulación debe ser:

- Flexible (para adaptarse a la innovación)
- Clara (para evitar ambigüedades)
- Aplicable (para no quedarse en el plano teórico)

Un desafío importante es que la IA es una tecnología global, mientras que las regulaciones suelen ser nacionales. Esto genera vacíos legales y dificultades para su control.

Ética aplicada: de los principios a la práctica

Uno de los mayores retos es pasar de principios generales a acciones concretas.

Algunas estrategias incluyen:

- Auditorías algorítmicas
- Evaluaciones de impacto ético
- Equipos interdisciplinarios (tecnología + ética + ciencias sociales)
- Mecanismos de rendición de cuentas

Por ejemplo, antes de implementar un sistema de IA en una institución pública, podría realizarse un análisis que evalúe posibles riesgos de discriminación o impacto social.

Educación y cultura ética

La ética en la IA no depende únicamente de regulaciones externas. También requiere una transformación cultural.

Esto implica:

- Formar desarrolladores con sensibilidad ética

- Promover el pensamiento crítico en usuarios
- Fomentar debates públicos informados

La educación juega un papel clave en este proceso. No se trata solo de enseñar cómo funciona la IA, sino de reflexionar sobre sus implicaciones.

Hasta el este punto, los marcos éticos y las regulaciones representan un avance importante, pero no son soluciones definitivas. La inteligencia artificial evoluciona constantemente, lo que implica que los dilemas éticos también cambian. Por ello, la ética en la IA debe entenderse como un proceso dinámico, no como un conjunto fijo de reglas.

Parte V: Reflexiones finales, retos futuros y conclusiones

La IA como espejo de la sociedad

A lo largo de este capítulo se ha argumentado que la inteligencia artificial no es una tecnología neutral. Más bien, funciona como un espejo que refleja y en muchos casos amplifica las estructuras, valores y desigualdades de la sociedad en la que se desarrolla.

Los sesgos en los algoritmos, las desigualdades en el acceso a la tecnología y las tensiones en torno a la privacidad no son problemas exclusivamente técnicos. Son manifestaciones de dinámicas sociales más amplias.

Esto implica que la ética en la IA no puede limitarse a corregir errores en los sistemas, sino que debe cuestionar los contextos en los que estos sistemas emergen. En otras palabras, no basta con hacer algoritmos más precisos; es necesario preguntarse para qué y para quién se diseñan.

El desafío de anticipar el futuro

Uno de los mayores retos en la ética de la inteligencia artificial es su carácter anticipatorio. Muchas de las decisiones que se toman hoy tendrán efectos a largo plazo, en contextos que aún no pueden preverse completamente.

Por ejemplo:

- ¿Cómo afectará la IA a las futuras generaciones en términos de empleo?
- ¿Qué tipo de relaciones sociales se configurarán en entornos altamente automatizados?
- ¿Cómo evolucionará la noción de autonomía en un mundo mediado por algoritmos?

La dificultad radica en que la tecnología avanza más rápido que la capacidad de regulación y reflexión ética. Esto genera un desfase que puede tener consecuencias significativas si no se aborda a tiempo.

Entre la innovación y la precaución

Existe una tensión constante entre dos posturas:

- **Impulsar la innovación tecnológica**
- **Regular para prevenir riesgos**

Un exceso de regulación puede frenar el desarrollo tecnológico, mientras que una falta de regulación puede permitir abusos y consecuencias negativas.

El desafío consiste en encontrar un equilibrio. Esto implica no solo establecer normas, sino también crear espacios de diálogo entre distintos actores: desarrolladores, legisladores, académicos y ciudadanos.

La responsabilidad compartida

Uno de los puntos centrales que atraviesa todo el capítulo es que la ética en la inteligencia artificial no recae en un solo actor.

La responsabilidad es compartida:

- **Desarrolladores**, que diseñan los sistemas
- **Empresas**, que los implementan y comercializan
- **Gobiernos**, que regulan su uso

- **Usuarios**, que interactúan con la tecnología

Cada uno de estos actores tiene un papel en la construcción de un ecosistema tecnológico más justo y responsable. Sin embargo, esta distribución de responsabilidades también puede diluir la rendición de cuentas. Por ello, es fundamental establecer mecanismos claros que definan obligaciones y límites.

La importancia del pensamiento crítico

En un contexto donde la inteligencia artificial influye cada vez más en la vida cotidiana, el pensamiento crítico se vuelve una herramienta indispensable.

Esto implica:

- Cuestionar los resultados de los sistemas automatizados
- No asumir que la tecnología es infalible
- Comprender sus limitaciones

El riesgo de una aceptación acrítica de la IA es que se legitimen decisiones injustas bajo la apariencia de objetividad.

Hacia una ética situada

Uno de los errores más comunes en la discusión sobre la IA es buscar soluciones universales para problemas que son profundamente contextuales.

La ética en la inteligencia artificial debe ser **situada**, es decir, adaptada a contextos específicos:

- No es lo mismo implementar IA en salud que en publicidad
- No es igual en países con distintos niveles de desarrollo
- No afecta de la misma manera a todos los grupos sociales

Esto implica reconocer la diversidad de realidades y evitar enfoques simplistas.

Retos futuros

De cara al futuro, pueden identificarse algunos desafíos clave:

1. Regulación efectiva

Diseñar marcos legales que sean capaces de adaptarse a la velocidad del cambio tecnológico.

2. Transparencia real

Superar la opacidad de los sistemas y garantizar que las decisiones puedan ser comprendidas.

3. Reducción de desigualdades

Evitar que la IA amplíe brechas sociales, económicas y tecnológicas.

4. Control democrático

Hay que asegurar que el desarrollo de la IA no quede concentrado en unos pocos actores.

5. Integración ética desde el diseño

Incorporar la ética no como un elemento posterior, sino como parte del proceso de desarrollo tecnológico.

Conclusión

La inteligencia artificial representa una de las transformaciones más significativas de nuestra época. Su capacidad para procesar información, automatizar tareas y apoyar decisiones la convierte en una herramienta poderosa, pero también en un fenómeno complejo desde el punto de vista ético.

Este capítulo ha mostrado que los dilemas asociados a la IA no son marginales, sino centrales. La cuestión no es si debemos usar inteligencia artificial, sino cómo hacerlo de manera responsable. La ética en la IA no puede entenderse como un

conjunto de reglas estáticas, sino como un proceso continuo de reflexión, debate y adaptación. Implica reconocer que cada decisión tecnológica tiene consecuencias sociales.

En última instancia, el desafío no es solo desarrollar sistemas más avanzados, sino construir una relación más consciente y crítica con la tecnología. Una relación en la que la innovación no esté desligada de la responsabilidad, y en la que el progreso tecnológico contribuya realmente al bienestar colectivo.

Ahora siguiendo una conclusión sobre la inteligencia artificial, trabajo y desafíos éticos emergentes, el análisis desarrollado nos permite comprender que la inteligencia artificial no solo transforma procesos técnicos o productivos, sino que incide directamente en la configuración del trabajo como eje estructurante de la vida social. Lejos de representar un fenómeno aislado, la IA se inserta en dinámicas económicas, políticas y culturales que redefinen tanto las formas de producción como las condiciones en las que las personas participan en ellas.

En este sentido, uno de los principales hallazgos es que la inteligencia artificial no está provocando una desaparición absoluta del trabajo, sino una reconfiguración profunda de sus contenidos, sus formas de organización y sus relaciones de poder. Sin embargo, esta transformación no es neutral. Por el contrario, se encuentra atravesada por desigualdades estructurales que tienden a reproducirse —e incluso intensificarse— en contextos de adopción tecnológica acelerada. La automatización de tareas, la gestión algorítmica y la opacidad de los sistemas no solo modifican la naturaleza del empleo, sino que plantean interrogantes sobre la autonomía, la dignidad y la capacidad de los trabajadores para incidir en decisiones que afectan directamente su vida laboral.

A partir de ello, la ética de la inteligencia artificial adquiere una dimensión ampliada. Ya no se trata únicamente de garantizar que los sistemas sean transparentes o libres de sesgos, sino de cuestionar las implicaciones sociales de su implementación, particularmente en lo que respecta a la distribución de oportunidades, la calidad del empleo y la concentración del poder tecnológico. En

este marco, el trabajo deja de ser un ámbito secundario dentro del debate ético para convertirse en uno de sus núcleos fundamentales.

No obstante, reconocer estos desafíos implica también identificar los límites de los enfoques actuales. La velocidad del desarrollo tecnológico, la globalización de los sistemas digitales y la complejidad de los algoritmos dificultan la construcción de marcos regulatorios capaces de responder de manera efectiva a los cambios en curso. En consecuencia, la gobernanza de la inteligencia artificial en el ámbito laboral requiere avanzar hacia modelos más flexibles, adaptativos y, sobre todo, contextualizados.

De cara al futuro, se identifican al menos cuatro retos fundamentales. En primer lugar, la necesidad de desarrollar mecanismos que permitan anticipar los efectos de la inteligencia artificial en el trabajo, superando enfoques reactivos que actúan únicamente cuando los impactos ya se han materializado. En segundo lugar, resulta imprescindible fortalecer la capacidad de los sistemas educativos y de formación para responder a la transformación de las competencias laborales, evitando que amplios sectores de la población queden excluidos de las nuevas dinámicas productivas.

En tercer lugar, se presenta el desafío de construir marcos de gobernanza que equilibren innovación y protección social, lo que implica no solo regular la tecnología, sino también redefinir los sistemas de protección laboral en contextos de creciente automatización. Finalmente, emerge la necesidad de garantizar que la inteligencia artificial no profundice las desigualdades existentes, particularmente en países con estructuras laborales frágiles, donde los efectos de la automatización pueden ser más severos.

En suma, el futuro del trabajo en la era de la inteligencia artificial no está determinado únicamente por el avance tecnológico, sino por las decisiones sociales, políticas e institucionales que orienten su desarrollo. En este sentido, la ética y la gobernanza no deben entenderse como elementos complementarios, sino

como condiciones indispensables para asegurar que la inteligencia artificial contribuya a la construcción de sociedades más justas, inclusivas y sostenibles.

Referencias

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), 3–30. <https://doi.org/10.1257/jep.33.2.3>

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, 128(6), 2188–2244. <https://doi.org/10.1086/705716>

AI Now Institute. (2018). AI Now Report 2018. New York University. <https://ainowinstitute.org/>

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4), 1279–1333. <https://doi.org/10.1162/003355303322552801>

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Braverman, H. (1974). *Labor and monopoly capital: The degradation of work in the twentieth century*. Monthly Review Press.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). <https://eur-lex.europa.eu/>

European Commission. (2024). Artificial Intelligence Act. <https://artificialintelligenceact.eu/>

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>

Goldman Sachs. (2023). The potentially large effects of artificial intelligence on economic growth. <https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html>

Instituto Nacional de Estadística y Geografía (INEGI). (2023). Encuesta Nacional de Ocupación y Empleo. <https://www.inegi.org.mx>

International Labour Organization (OIT). (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality. <https://www.ilo.org>

Lozano Carrillo, O. (2024). El futuro de la inteligencia artificial. REMINEO.

Lozano Carrillo, O., Zavala Osorio, Y., & Couturier Bañuelos, D. P. (2024). Reflexiones sobre la intervención en las organizaciones en la UAM. REMINEO.

Lozano Carrillo, O., Zavala Osorio, Y., & Couturier Bañuelos, D. P. (2025). El PIAESS: Una IA para la vida. REMINEO.

McKinsey Global Institute. (2023). Generative AI and the future of work in America. <https://www.mckinsey.com>

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing Group.

OECD. (2019). OECD principles on artificial intelligence. <https://oecd.ai/>

Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2019). OECD Principles on Artificial Intelligence. <https://oecd.ai>

Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking.

Sansores Guerrero, E. A., Lozano Carrillo, O., & Garibay Suárez, A. (2025). Inteligencia artificial en las organizaciones: Innovación, poder y ética en la era digital. REMINEO.

The White House. (2022). Blueprint for an AI Bill of Rights. <https://www.whitehouse.gov>

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. <https://unesdoc.unesco.org/>

World Economic Forum. (2023). The Future of Jobs Report 2023. <https://www.weforum.org/reports/the-future-of-jobs-report-2023>

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.

CAPÍTULO VII

LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA ADMINISTRACION PUBLICA: RETOS JURÍDICOS, CRITERIOS JURISPRUDENCIALES, Y PERSPECTIVAS DE GOBERNANZA

The Regulation of Artificial Intelligence in Public Administration: Legal
Challenges, Jurisprudential Criteria, and Governance Perspectives

Miguel Francisco Barrera Aldama

migfraba@gmail.com

Servicio de Administración Tributaria

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0001-9407-6548>

RESUMEN

El presente ensayo aborda la urgente necesidad de establecer un marco regulatorio robusto para la Inteligencia Artificial (IA) desde una perspectiva jurídica y administrativa. Ante la creciente implementación de sistemas algorítmicos y de aprendizaje automático en las funciones del Estado, particularmente en la administración pública y la justicia administrativa, surgen interrogantes fundamentales sobre la transparencia, la rendición de cuentas y la salvaguarda del debido proceso. El análisis se centra en la tensión existente entre la eficiencia administrativa y el respeto a los derechos humanos, examinando el deber de fundamentación y motivación de los actos administrativos automatizados a la luz del artículo 16 de la Constitución Política de los Estados Unidos Mexicanos. Asimismo, se exploran diversos criterios jurisprudenciales a nivel nacional e internacional que delinear los límites del actuar estatal frente a la "caja negra" algorítmica. Finalmente, se proponen directrices de gobernanza orientadas a una regulación ética y socialmente responsable, asegurando que el avance tecnológico se subordine al principio de legalidad y al Estado Constitucional de Derecho.

Palabras clave: Inteligencia artificial, derecho administrativo, fundamentación y motivación, algoritmos, justicia administrativa, regulación.

Abstract

This essay addresses the urgent need to establish a robust regulatory framework for Artificial Intelligence (AI) from a legal and administrative perspective. In the face of the increasing implementation of algorithmic and machine learning systems in state functions, particularly in public administration and administrative justice, fundamental questions arise about transparency, accountability, and safeguarding due process. The analysis focuses on the tension between administrative efficiency and respect for human rights, examining the duty to substantiate and motivate automated administrative acts in the light of Article 16 of the Political Constitution of the United Mexican States. Likewise, various jurisprudential criteria are explored at the national and international level that delineate the limits of state action in the face of the algorithmic "black box". Finally, governance guidelines aimed at ethical and socially responsible regulation are proposed, ensuring that technological progress is subordinated to the principle of legality and the Constitutional Rule of Law.

Keywords: Artificial intelligence, administrative law, rationale and motivation, algorithms, administrative justice, regulation.

DECLARACIÓN DE USO ÉTICO DE INTELIGENCIA ARTIFICIAL

En congruencia con los principios de gobernanza, transparencia y explicabilidad que se defienden a lo largo de este ensayo, el autor declara que en las fases de investigación, estructuración documental y revisión de estilo de la presente obra se utilizaron herramientas de Inteligencia Artificial generativa (Modelos de Lenguaje de Gran Escala).

El uso de estas tecnologías se realizó bajo un estricto marco de ética académica e integridad investigativa, observando las siguientes directrices:

ASISTENCIA, NO SUSTITUCIÓN: La inteligencia artificial fue empleada exclusivamente como un mecanismo de asistencia técnica para la sistematización de información, la optimización del proceso de redacción y la estructuración del formato APA (7ª edición). En ningún momento la herramienta sustituyó el análisis crítico, la valoración cualitativa ni la capacidad deductiva inherente al trabajo académico.

SUPERVISIÓN HUMANA SIGNIFICATIVA (*HUMAN-IN-THE-LOOP*): Como docente e investigador en la materia jurídico-administrativa, el autor mantuvo en todo momento el control intelectual y metodológico del documento. La selección de las premisas, la formulación de los argumentos sobre la fundamentación y motivación de los actos administrativos automatizados, así como la interpretación hermenéutica de los criterios jurisprudenciales de la Suprema Corte de Justicia de la Nación y del Tribunal Federal de Justicia Administrativa, son producto exclusivo del raciocinio humano.

VERIFICACIÓN Y VALIDACIÓN DE FUENTES: Toda la información proporcionada o sugerida por la herramienta algorítmica —incluyendo las citas bibliográficas, referencias hemerográficas y sentencias referidas— fue sometida a un escrutinio riguroso y contrastada directamente con las fuentes primarias oficiales y las bases de datos de los tribunales correspondientes, garantizando su absoluta veracidad, actualidad y pertinencia.

RESPONSABILIDAD AUTORAL E IMPUTABILIDAD: En apego al principio de responsabilidad jurídica abordado en el cuerpo de este texto, el autor asume la responsabilidad total, patrimonial e intelectual sobre el contenido, las conclusiones y las propuestas vertidas en este ensayo.

La integración de herramientas tecnológicas en la labor dogmática y de investigación no debe ser motivo de opacidad, sino una oportunidad para ejercer una transparencia proactiva. Declarar su uso es el primer paso hacia la

normalización de una inteligencia artificial supeditada a la ética y al Estado Constitucional de Derecho.

INTRODUCCIÓN

La disrupción tecnológica que representa la Inteligencia Artificial (IA) ha trascendido el ámbito de la ciencia ficción y la especulación académica para instalarse como una realidad ineludible en el núcleo de las interacciones sociales, económicas y gubernamentales.

En la actualidad, las capacidades de procesamiento masivo de datos (*Big Data*) combinadas con modelos de aprendizaje automático (*Machine Learning*) otorgan a las máquinas la facultad de identificar patrones, realizar predicciones y tomar decisiones con una autonomía sin precedentes.

Este fenómeno impone un desafío colosal para las estructuras normativas tradicionales, las cuales fueron diseñadas en una época donde la agencia y la toma de decisiones eran prerrogativas exclusivas del ser humano.

Dentro de las esferas del derecho público y la administración, la adopción de estas tecnologías promete una optimización sin igual. Las instituciones estatales, en su constante búsqueda por la eficiencia y la reducción de la burocracia, ven en la IA una herramienta invaluable para la prestación de servicios, la fiscalización tributaria, la evaluación de políticas públicas y, en última instancia, el ejercicio de la potestad sancionadora.

Sin embargo, esta automatización de las funciones gubernamentales trae consigo un replanteamiento profundo del derecho administrativo. La opacidad inherente a ciertos sistemas algorítmicos —conocida como el problema de la "caja negra"— amenaza los principios más elementales del debido proceso, la seguridad jurídica y el derecho a una buena administración.

El presente ensayo tiene como objetivo principal reflexionar sobre los riesgos y las oportunidades que subyacen en la regulación de la Inteligencia Artificial, con un enfoque particular en el aspecto jurídico-administrativo en México. Se parte de la

premisa de que la regulación no debe concebirse como un freno al desarrollo tecnológico, sino como el cauce institucional necesario para garantizar que dicha tecnología sirva al interés público sin menoscabar los derechos fundamentales.

Para ello, el documento se estructura en cinco apartados principales: el primero delimita la naturaleza y el impacto de la IA en el ámbito público; el segundo aborda la tensión entre los actos administrativos automatizados y el mandato constitucional de fundamentación y motivación; el tercero analiza los riesgos de sesgos y la responsabilidad patrimonial del Estado; el cuarto explora los criterios jurisprudenciales más relevantes a nivel comparado y nacional; y, finalmente, el quinto esboza propuestas concretas para la gobernanza y regulación de la IA en la administración pública mexicana.

LA NATURALEZA DE LA INTELIGENCIA ARTIFICIAL Y SU IRRUPCIÓN EN LA ADMINISTRACIÓN PÚBLICA

La Inteligencia Artificial podría definirse, en términos generales, como la capacidad de un sistema informático para interpretar correctamente datos externos, aprender de dichos datos y emplear esos conocimientos para lograr tareas y metas específicas a través de una adaptación flexible.

Desde el punto de vista jurídico, la dificultad para regular la IA radica en su naturaleza evolutiva. A diferencia del software tradicional, cuyo código obedece a instrucciones deterministas programadas por un operador humano (reglas de "*si ocurre A, entonces B*"), los sistemas de IA basados en redes neuronales desarrollan sus propias lógicas internas para resolver problemas.

En el ámbito de la administración pública, esta capacidad adaptativa transforma radicalmente la actuación del Estado. Históricamente, el derecho administrativo se ha construido sobre la figura del funcionario público, sujeto a un régimen estricto de competencias, procedimientos y responsabilidades. La delegación de la toma de

decisiones (o la asistencia sustancial en la misma) a sistemas algorítmicos desdibuja la línea de imputabilidad.

Ejemplos de esto ya son visibles en la gestión tributaria contemporánea, donde autoridades fiscales utilizan algoritmos predictivos para seleccionar contribuyentes sujetos a auditorías, detectar patrones de evasión o emitir requerimientos prellenados. Estas acciones, que agilizan de manera impresionante la recaudación y la revisión, constituyen verdaderos actos de molestia o privación que deben estar sujetos al escrutinio constitucional.

La transición del "gobierno electrónico" (digitalización de trámites) a la "administración algorítmica" (toma de decisiones automatizadas o semi-automatizadas) requiere una recalibración dogmática. El derecho administrativo no puede operar bajo la ficción de que la tecnología es una simple herramienta neutral.

Los algoritmos reflejan, consciente o inconscientemente, las prioridades, los sesgos y los valores de quienes los diseñan y de los datos con los que son entrenados. Por lo tanto, el uso de la IA por parte del Estado no es una mera cuestión técnica, sino un asunto profundamente político y jurídico que incide directamente en el ejercicio del poder público y la tutela de los derechos de los gobernados.

ACTOS ADMINISTRATIVOS AUTOMATIZADOS: EL RETO DE LA FUNDAMENTACIÓN Y MOTIVACIÓN

Uno de los pilares del Estado Constitucional de Derecho en México se encuentra consagrado en el artículo 16 de la Constitución Federal, el cual establece el principio de legalidad en su vertiente de fundamentación y motivación. Ningún acto de autoridad puede emitirse sin que exprese con precisión el precepto legal aplicable al caso (fundamentación) y señale con claridad las circunstancias especiales, razones particulares o causas inmediatas que se hayan tenido en consideración para su emisión (motivación), debiendo existir adecuación entre los motivos aducidos y las normas aplicables.

Cuando un acto administrativo —por ejemplo, la determinación de un crédito fiscal, la denegación de un permiso, o la imposición de una sanción— es generado total o parcialmente por un sistema de IA, surge la interrogante crítica: ¿Cómo motiva la autoridad un acto cuya lógica deductiva o inductiva se encuentra encriptada en una red neuronal inescrutable? La opacidad algorítmica entra en conflicto directo con el derecho de defensa del particular. Si el administrado desconoce cómo el sistema ponderó los elementos de hecho para llegar a una conclusión jurídica, se encuentra en un estado de indefensión material, imposibilitado para rebatir adecuadamente la resolución ante los tribunales, como el Tribunal Federal de Justicia Administrativa (TFJA).

La doctrina administrativista moderna ha comenzado a exigir el "derecho a una explicación" (*right to explanation*). No basta con que la autoridad cite la base de datos utilizada o el nombre del software empleado. La debida motivación de un acto administrativo automatizado exige que la administración pública sea capaz de traducir a un lenguaje comprensible y lógico-jurídico el iter iterativo del algoritmo.

Esto significa que la utilización de IA en funciones de imperio debe limitarse, o al menos condicionarse, a la existencia de sistemas de "IA explicable" (*Explainable AI* o *XAI*). Si la administración no puede explicar las razones de una decisión de forma que permita el control jurisdiccional, el acto deviene inconstitucional por vicios en su motivación.

Además, la automatización plantea desafíos sobre la competencia material y la firma de los actos. La Suprema Corte y los Tribunales Colegiados han analizado exhaustivamente la validez de las firmas electrónicas y los actos emitidos por sistemas informáticos, concluyendo que la validez del acto recae en el funcionario competente que asume la titularidad de dicho sistema.

Sin embargo, en la era de la IA, asumir la titularidad técnica no resuelve el problema sustantivo de la discrecionalidad administrativa. ¿Puede un algoritmo ejercer facultades discrecionales? La doctrina se inclina a sostener que los algoritmos solo pueden operar válidamente en el marco de facultades regladas, donde la

subsunción del hecho a la norma es mecánica. Delegar el arbitrio o la valoración cualitativa (facultades discrecionales) a una máquina vulnera la naturaleza de la potestad administrativa.

ÉTICA, SEGOS ALGORÍTMICOS Y RESPONSABILIDAD SOCIAL Y PATRIMONIAL DEL ESTADO

Más allá del cumplimiento formal de los requisitos constitucionales, la regulación de la IA debe abordar los imperativos éticos y de derechos humanos. Los sistemas de aprendizaje automático aprenden del pasado empírico a través de inmensos volúmenes de datos. Si estos datos históricos contienen asimetrías, prejuicios o prácticas discriminatorias estructurales, el algoritmo no solo las reproducirá, sino que las invisibilizará bajo una falsa pátina de objetividad matemática y neutralidad científica. A esto se le conoce como el problema del sesgo algorítmico o "*mathwashing*" (el blanqueo a través de las matemáticas).

En sectores como la salud, la educación, la prevención del delito o la seguridad social, un algoritmo sesgado puede perpetuar la exclusión de grupos vulnerables. Por ejemplo, un sistema automatizado para la asignación de becas universitarias o para la determinación de riesgo de reincidencia penal que asigne puntajes desfavorables basados implícitamente en el código postal, el origen étnico o el género de las personas, viola frontalmente el principio de igualdad y no discriminación consagrado en el Artículo Primero Constitucional y en diversos Tratados Internacionales en materia de Derechos Humanos.

Frente a estas vulneraciones, emerge el complejo tema de la Responsabilidad Patrimonial del Estado. De acuerdo con el marco jurídico mexicano, el Estado tiene la obligación de responder por los daños que cause su actividad administrativa irregular.

La actuación de un algoritmo defectuoso, sesgado o mal implementado que genere un daño en la esfera de derechos de un particular constituye una actividad administrativa irregular. El reto procesal consiste en probar el nexo causal en un entorno de opacidad algorítmica. Para que la responsabilidad patrimonial sea un mecanismo efectivo de reparación en la era de la IA, los tribunales deberán adoptar mayor cuidado respecto a la carga de la prueba, imponiendo a la Administración la obligación de demostrar la fiabilidad técnica y jurídica del sistema utilizado (carga dinámica de la prueba).

ANÁLISIS JURISPRUDENCIAL Y PERSPECTIVAS COMPARADAS

El desarrollo jurisprudencial sobre la IA y el control de algoritmos se encuentra en sus fases incipientes en México, pero la experiencia comparada ofrece directrices valiosas. En el contexto europeo, el Reglamento General de Protección de Datos (RGPD) estableció en su artículo 22 el derecho de toda persona a no ser objeto de una decisión basada únicamente en el tratamiento automatizado de datos, incluida la elaboración de perfiles, que produzca efectos jurídicos o que la afecte significativamente de modo similar.

A nivel jurisprudencial, destaca el caso "SyRI" en los Países Bajos (2020), donde un Tribunal de Distrito de La Haya ordenó detener un sistema de evaluación de riesgos impulsado por IA utilizado por el gobierno para detectar fraudes en prestaciones sociales. El tribunal determinó que la legislación que amparaba a SyRI no lograba un equilibrio justo entre los beneficios del uso de nuevas tecnologías para prevenir fraudes y la injerencia en el derecho a la vida privada (Artículo 8 del Convenio Europeo de Derechos Humanos). La falta de transparencia sobre cómo el algoritmo ponderaba los factores de riesgo impidió el control democrático y vulneró los derechos fundamentales.

En Italia, el Consejo de Estado emitió sentencias emblemáticas en 2019 relacionadas con el algoritmo "*Buona Scuola*", utilizado para la movilidad del personal docente. El tribunal italiano dictaminó que, aunque el uso de algoritmos en la administración pública es lícito e incluso deseable por su eficiencia, debe

someterse a tres principios rectores: 1) El principio de cognoscibilidad (todo sujeto tiene derecho a conocer las reglas del algoritmo), 2) El principio de no exclusividad de la decisión algorítmica (debe haber supervisión humana significativa), y 3) El principio de no discriminación algorítmica.

En México, si bien no existen sentencias de la Suprema Corte de Justicia de la Nación (SCJN) que aborden directamente la palabra "Inteligencia Artificial" desde la óptica del Derecho Administrativo de fondo, existe una vasta línea jurisprudencial aplicable por analogía.

Tratándose de los procedimientos de revisión electrónica (buzón tributario y auditorías digitales), la SCJN ha validado el uso de la tecnología siempre y cuando se respete el derecho de audiencia y la posibilidad de desvirtuar las presunciones de la autoridad (Amparos en Revisión 1287/2015 y acumulados).

El principio extraído de la jurisprudencia nacional es claro: la sistematización informática y los cruces de información no eximen a la autoridad de probar los hechos constitutivos de su acción y de fundar y motivar la resolución final. Este criterio jurisprudencial será la piedra angular para el control jurisdiccional de la IA en México: la máquina puede sugerir, calcular o presumir, pero la imputación jurídica y la motivación exhaustiva siguen siendo cargas ineludibles del ente público.

Actualmente, la Suprema Corte de Justicia de la Nación (SCJN) se encuentra en una fase de transición hacia el análisis sustantivo de la Inteligencia Artificial. Si bien no existe todavía un cuerpo robusto de jurisprudencia que utilice la "Inteligencia Artificial" como término central de una *ratio decidendi* (razón de la decisión), se han emitido criterios fundamentales en materia de derecho digital, justicia administrativa y fiscalización electrónica que sientan las bases para el control de algoritmos.

A continuación, se presentan los criterios y precedentes más relevantes que la SCJN ha establecido y que son aplicables por analogía o extensión al fenómeno de la IA:

CONSTITUCIONALIDAD DE LAS REVISIONES ELECTRÓNICAS (ASUNTO LÍDER)

El precedente más importante para entender la automatización administrativa es el **Amparo en Revisión 1287/2015**. En este caso, la Segunda Sala analizó la validez de las facultades de fiscalización a través de medios electrónicos (Buzón Tributario).

Criterio: La Corte determinó que el uso de tecnologías para la revisión de la situación fiscal de los contribuyentes no vulnera los principios de seguridad jurídica ni de audiencia.

Aplicación a la IA: Establece que la autoridad puede valerse de cruces de información automatizados y sistemas informáticos para emitir actos de molestia, siempre que el sistema permita al particular conocer las irregularidades detectadas y le brinde la oportunidad de desvirtuarlas (Derecho de Audiencia). Este es el precedente fundamental para la "Administración Algorítmica". Al analizar las revisiones electrónicas, la entonces Segunda Sala determinó que el uso de sistemas informáticos para realizar cruces de información masiva y emitir resoluciones provisionales es constitucional. El punto clave es la distinción entre el "acto de molestia" y el "derecho de defensa". La Corte estableció que la automatización no es arbitraria siempre que el particular tenga una vía clara (como el buzón tributario) para conocer las inconsistencias detectadas por el sistema y pueda aportar pruebas para desvirtuarlas. Para la IA, esto significa que el Estado puede usar algoritmos predictivos o de detección de riesgo, pero el resultado algorítmico nunca puede ser definitivo sin una etapa de intervención humana o de réplica por parte del gobernado.

FUNDAMENTACIÓN Y MOTIVACIÓN EN ACTOS DIGITALES

La SCJN ha sostenido que la digitalización no exime a la autoridad de los requisitos del artículo 16 constitucional.

Tesis de Jurisprudencia 2a./J. 153/2017 (10a.): "ACTOS ADMINISTRATIVOS. REQUISITOS QUE DEBEN OBSERVARSE PARA QUE SE TENGAN POR CUMPLIDOS LOS DE FUNDAMENTACIÓN Y MOTIVACIÓN CUANDO SE EMITAN A TRAVÉS DE MEDIOS ELECTRÓNICOS."

Criterio: Para que un acto emitido por un sistema informático sea válido, la autoridad debe precisar con claridad los datos procesados y el fundamento legal. Esto es el precursor del "**Derecho a la Explicabilidad**" de los algoritmos en México: si un sistema informático genera una liquidación fiscal, la autoridad debe ser capaz de "explicar" el cálculo, no solo arrojar el resultado.

VALIDEZ DE LA FIRMA ELECTRÓNICA Y CERTIFICADOS DIGITALES

Diversas tesis (como la **2a./J. 13/2016**) han validado el uso de la firma electrónica avanzada (FIEL) para la emisión de actos de autoridad.

Implicación: Esto reconoce la validez jurídica de la voluntad de la administración pública expresada a través de medios digitales, lo cual es esencial para la imputación de responsabilidad legal cuando un sistema de IA firma o valida una decisión administrativa.

PRIVACIDAD Y DATOS PERSONALES EN ENTORNOS DIGITALES

En asuntos relacionados con el acceso a comunicaciones y la protección de datos (como el **Amparo en Revisión 550/2018** sobre neutralidad de la red), la Corte ha enfatizado la protección del "Entorno Digital" como una extensión de la vida privada.

Aplicación a la IA: Estos criterios limitan el tipo de datos que los algoritmos estatales pueden procesar sin consentimiento o sin una orden judicial, especialmente en tareas de vigilancia o perfilamiento. En este asunto, la Primera Sala reconoció que el entorno digital es un espacio donde se ejercen derechos fundamentales que el Estado debe respetar. Al analizar la neutralidad y la privacidad, la Corte determinó que la recolección y el procesamiento masivo de datos (esenciales para entrenar modelos de IA) deben estar limitados por el principio de proporcionalidad y finalidad. Este criterio impide que el Estado realice

"minería de datos" indiscriminada para alimentar algoritmos de perfilamiento sin un marco legal robusto que lo justifique. Establece que la dignidad humana y la intimidad son barreras infranqueables para la eficiencia tecnológica.

ACCESO A LA JUSTICIA Y EL EXPEDIENTE ELECTRÓNICO

A través de los Acuerdos Generales de la SCJN y el Consejo de la Judicatura Federal (como el 12/2020), se ha formalizado el juicio en línea. Aunque son normas administrativas, han sido convalidadas implícitamente por la Corte al resolver recursos sobre la validez de notificaciones electrónicas.

Perspectiva: La Corte ya reconoce al entorno digital como un espacio válido para el ejercicio de derechos procesales, lo que abre la puerta a la impugnación de sesgos algorítmicos en la justicia administrativa.

SITUACIÓN ACTUAL Y PROSPECTIVA

Es importante mencionar que la SCJN ha creado recientemente la **Unidad de Inteligencia Artificial y Tecnologías Emergentes**. Aunque esta unidad tiene funciones internas (como el sistema "*Sor Juana*" para la búsqueda de precedentes), su existencia marca la pauta de que la Corte está reconociendo la IA como una herramienta auxiliar de la justicia, pero siempre bajo el principio de "**Supervisión Humana**".

GOBERNANZA, REGULACIONES SECTORIALES Y PROPUESTAS NORMATIVAS EN MÉXICO

La regulación de la inteligencia artificial no puede dejarse a las fuerzas del mercado tecnológico, ni puede consistir en una prohibición total que inhiba la innovación. La vía óptima es una regulación basada en riesgos, tal como lo propone la histórica "*AI Act*" (Ley de Inteligencia Artificial) de la Unión Europea.

Bajo este enfoque, las obligaciones legales son proporcionales al nivel de riesgo que el sistema de IA representa para los derechos humanos y la seguridad.

Para México, la conformación de un marco regulatorio de la IA exige acciones a nivel legislativo y administrativo. En primer lugar, es menester la promulgación de una Ley General de Inteligencia Artificial y Algoritmos que establezca principios transversales: transparencia, explicabilidad, rendición de cuentas, supervisión humana, equidad y ciberseguridad. En segundo lugar, tratándose de la Administración Pública Federal, la Ley Federal de Procedimiento Administrativo debe ser reformada para incluir un capítulo específico sobre la "Actuación Administrativa Automatizada". Dicha reforma debe incluir:

- a) Creación de Registros Públicos de Algoritmos: La administración debe mantener un catálogo público de los sistemas algorítmicos que utiliza, explicando en lenguaje ciudadano su propósito, los datos que procesa y la lógica de su funcionamiento.
- b) Evaluaciones de Impacto Algorítmico (EIA): Previo al despliegue de cualquier sistema de IA en el sector público, la entidad debe realizar una EIA exhaustiva, analizando los posibles impactos en los derechos humanos, los riesgos de sesgo y las medidas de mitigación adoptadas.
- c) Supervisión Humana Significativa (*Human in the loop*): Queda prohibida la delegación absoluta de decisiones administrativas discrecionales o sancionatorias a sistemas de IA. La última decisión, así como la asunción de la responsabilidad legal de la misma, debe recaer invariable y comprobablemente en un servidor público facultado.
- d) *Sandboxes* Regulatorios: Fomentar el uso de entornos de prueba controlados para evaluar nuevas aplicaciones de IA en la gestión pública o en sectores regulados (salud, educación, medio ambiente) sin que la rigidez normativa inicial ahogue el desarrollo de soluciones sostenibles, promoviendo así la innovación con salvaguardas.

CONCLUSIONES

La inteligencia artificial representa uno de los cambios paradigmáticos más significativos para el derecho contemporáneo; su incorporación a las tareas del Estado no es una opción, sino una inercia inevitable de la modernidad administrativa. Sin embargo, como se ha sostenido a lo largo de este ensayo, la eficiencia tecnocrática nunca puede erigirse como un valor superior a la dignidad humana, la legalidad y la seguridad jurídica.

1. El Derecho Administrativo mexicano debe evolucionar para enfrentar la "caja negra". La exigencia de fundamentación y motivación del artículo 16 constitucional obliga a que las decisiones públicas asistidas por IA sean transparentes, explicables y auditables. Un algoritmo que el Estado no puede explicar es un algoritmo que el Estado no debe utilizar.

2. Los sesgos ocultos en los datos de entrenamiento representan una amenaza silenciosa a los derechos de igualdad y no discriminación. La mitigación de estos riesgos requiere equipos interdisciplinarios y evaluaciones rigurosas de impacto algorítmico previas a la implementación de estas herramientas en la gestión pública.

3. La jurisprudencia comparada, especialmente en Europa, está trazando la ruta de defensa de los ciudadanos frente a la automatización estatal. En México, los tribunales, en particular el Tribunal Federal de Justicia Administrativa y el Poder Judicial de la Federación, están llamados a ser la última barrera de contención frente a la arbitrariedad algorítmica, expandiendo la interpretación de los derechos procesales para ajustarse a los litigios tecnológicos.

4. Se propone la imperiosa necesidad de emitir un marco legislativo integral (basado en niveles de riesgo) y reformar las leyes de procedimiento administrativo. Dicha regulación debe establecer registros de algoritmos gubernamentales, exigir auditorías de código en casos de afectación a derechos, y garantizar la responsabilidad patrimonial del Estado por daños generados a través de sistemas de inteligencia artificial deficientes.

En síntesis, el gran reto del siglo XXI para los juristas, legisladores y administradores no es frenar el avance de las máquinas, sino garantizar que el Derecho mantenga el control sobre ellas. La Inteligencia Artificial debe someterse a la Constitución, y la administración algorítmica debe ser, ante todo, una buena administración.

REFERENCIAS

Alarie, B., Niblett, A., & Yoon, A. (2018). How artificial intelligence will affect the practice of law. *University of Toronto Law Journal*, 68*(1), 106-124. <https://doi.org/10.3138/utlj.2017-0052>

Barrio Andrés, M. (Dir.). (2018). *Derecho Público y de la Artificial*. Tirant lo Blanch.

Boix Palop, A. (2020). Los algoritmos en la Administración Pública. *Revista de Administración Pública*, (211), 63-102.

Cerrillo i Martínez, A. (2019). El impacto de la inteligencia artificial en el derecho administrativo ¿nuevos conceptos para nuevas realidades? *Revista General de Derecho Administrativo*, (50).

Coglianesi, C., & Lehr, D. (2017). Regulating by Robot: Administrative Decision Making in the Machine-Learning Era. *The Georgetown Law Journal*, 105*, 1147-1223.

Comisión Europea. (2021). *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial)*. COM(2021) 206 final.

Constitución Política de los Estados Unidos Mexicanos [CPEUM]. (1917). Diario Oficial de la Federación, 5 de febrero de 1917 (México).

Cotino Hueso, L. (2019). Riesgos e impactos del big data, la inteligencia artificial y la robótica. En *Derecho Público, Inteligencia Artificial y Nuevas Tecnologías* (pp. 53-84). Tirant lo Blanch.

Díaz de León, H. (2021). La justicia administrativa digital y el impacto de las tecnologías en los procedimientos de revisión. *Revista del Tribunal Federal de Justicia Administrativa*, 8*(74), 45-68.

Echeverría, J. (2018). *Ética de la Inteligencia Artificial*. Ediciones Cátedra.

Esteva, E. (2022). Algoritmos, derechos fundamentales y administración pública en México. **Cuestiones Constitucionales**, (46), 115-140.

Gamero Casado, E. (2021). **Inteligencia Artificial y Derecho Administrativo**. Aranzadi.

Haba, E. P. (2020). Teoría del Estado y transformación tecnológica: El paradigma de los datos algorítmicos. **Revista de Derecho y Ciencia Política**, 32*(2), 89-112.

Lessig, L. (2006). **Code: And Other Laws of Cyberspace** (Versión 2.0). Basic Books.

Ley Federal de Procedimiento Administrativo [LFPA]. (1994). Diario Oficial de la Federación, 4 de agosto de 1994 (México).

Mayer-Schönberger, V., & Cukier, K. (2013). **Big Data: A Revolution That Will Transform How We Live, Work, and Think**. Houghton Mifflin Harcourt.

Naser, A., Ramírez-Alujas, A., & Rosales, D. (Eds.). (2017). **Desde el gobierno abierto al Estado abierto en América Latina y el Caribe**. CEPAL.

O'Neil, C. (2016). **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. Crown.

Pasquale, F. (2015). **The Black Box Society: The Secret Algorithms That Control Money and Information**. Harvard University Press.

Ponce Solé, J. (2019). Inteligencia artificial, derecho público y buena administración. **Revista de la Escuela Jacobea de Posgrado**, (16), 1-28.

Suprema Corte de Justicia de la Nación [SCJN]. (2016). Revisiones electrónicas. Los artículos 42, fracción IX y 53-B del Código Fiscal de la Federación que prevén el procedimiento relativo, no violan el derecho de audiencia. [Amparo en revisión 1287/2015]. **Semanario Judicial de la Federación**.

Suprema Corte de Justicia de la Nación. (2016). *Revisiones electrónicas. El procedimiento previsto en los artículos 42, fracción IX y 53-B del Código Fiscal de la Federación, no vulnera el derecho de audiencia ni el principio de seguridad jurídica*. [Amparo en revisión 1287/2015]. Semanario Judicial de la Federación.

Suprema Corte de Justicia de la Nación. (2018). *Derecho a la privacidad. Su protección se extiende al entorno digital y a los datos personales que el Estado procesa con fines administrativos o de seguridad*. [Amparo en revisión 550/2018]. Semanario Judicial de la Federación.

Susskind, R. (2019). **Online Courts and the Future of Justice**. Oxford University Press.

Tribunal de Distrito de La Haya. (2020). Sentencia sobre el Sistema de Indicación de Riesgos (SyRI). Caso C/09/554458 / HA ZA 18-388. Países Bajos.

Valero Torrijos, J. (2020). Las garantías jurídicas de la inteligencia artificial en la actuación administrativa. En **Tratado de Derecho Administrativo y Tecnología** (pp. 301-332). Thomson Reuters.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. **Science Robotics*, 2*(6).

Zalnieriute, M., Bennett Moses, L., & Williams, G. (2019). The rule of law and automation of government decision-making. **The Modern Law Review*, 82*(3), 425-455. <https://doi.org/10.1111/1468-2230.12412>

CAPÍTULO VIII

LA IMPLEMENTACIÓN DE LA IA EN PRÁCTICAS DE GOBERNANZA PARA LA DEMOCRATIZACIÓN DEL CONOCIMIENTO

The implementation of AI in governance practices for the democratization of knowledge

Erandi Nina Velarde Torres

eninavelt@gmail.com

Universidad Autónoma Metropolitana, Unidad Iztapalapa

<https://orcid.org/0009-0003-4194-9933>

Resumen

La gobernanza contemporánea ha transitado de modelos jerárquicos tradicionales a redes de interdependencia entre actores públicos y privados, donde la Inteligencia Artificial (IA) actúa como un catalizador para transformar el saber en un bien público regido por la libertad y la igualdad. Sus alcances incluyen la optimización de la gestión documental, la personalización del aprendizaje y la creación de una "racionalidad artificial" capaz de procesar volúmenes masivos de datos para una toma de decisiones más ágil y precisa. Además, potencia el Acceso Abierto y la participación ciudadana mediante plataformas digitales que fomentan un "pueblo decisor" activo. Sin embargo, la tecnología enfrenta límites significativos: la opacidad algorítmica de las "cajas negras" dificulta la rendición de cuentas, mientras que los sesgos inherentes pueden amplificar discriminaciones sociales y raciales.

En América Latina, la brecha digital —con un 40% de la población rural sin conexión— y la dependencia de conocimientos importados restringen su potencial, evidenciando que la IA no es una solución mágica para problemas estructurales como la pobreza o la corrupción. Una democratización real exige una meta

gobernanza ética y soberana que priorice la dignidad humana y el bienestar colectivo sobre los intereses mercantiles.

Palabras clave: Inteligencia artificial, gobernanza, democratización del conocimiento, desigualdad, políticas públicas.

Abstract

Contemporary governance has moved from traditional hierarchical models to networks of interdependence between public and private actors, where Artificial Intelligence (AI) acts as a catalyst to transform knowledge into a public good governed by freedom and equality. Its scope includes the optimization of document management, the personalization of learning and the creation of an "artificial rationality" capable of processing massive volumes of data for more agile and accurate decision-making. In addition, it promotes Open Access and citizen participation through digital platforms that promote an active "decision-making people". However, technology faces significant limits: the algorithmic opacity of "black boxes" makes accountability difficult, while inherent biases can amplify social and racial discriminations.

In Latin America, the digital divide – with 40% of the rural population offline – and dependence on imported knowledge restrict its potential, showing that AI is not a magic solution to structural problems such as poverty or corruption. A real democratization requires an ethical and sovereign meta-governance that prioritizes human dignity and collective well-being over commercial interests.

Keywords: Artificial intelligence, governance, democratization of knowledge, inequality, public policies.

Introducción

En 1635, Pedro Calderón de la Barca escribió "La vida es sueño", como Segismundo en su torre barroca, la humanidad navega hoy entre los espejismos de una realidad difusa dentro de las nuevas prácticas de gobernanza. Bajo la densa sombra de cajas negras e indescifrables algoritmos, el ciudadano se debate entre el sueño y la vigilia de las decisiones del poder.

Mas la inteligencia artificial no ha de ser nuestra celda informativa, sino una posible herramienta de alcance que democratice el saber cómo un bien público y una forma de vida. Esto no puede darse per se, pues solo al derribar las barreras estructurales que sumen a la sociedad en la pasividad del consumo, florecerá el sujeto cognoscente activo capaz de transformar su realidad. En una gobernanza ética reside el instrumento de nuestra liberación: convertir la ciencia en un derecho humano palpable, cimiento de la justicia social en nuestra tierra.

La sociedad contemporánea atraviesa una transformación profunda en sus formas de organización política, marcada por la erosión de las bases tradicionales del poder estatal y el surgimiento de la gobernanza como un modelo de decisión compartida entre actores públicos y privados. Este cambio de paradigma, donde el Estado deja de ser el centro indiscutible de la acción colectiva, se ve potenciado por la irrupción masiva de la Inteligencia Artificial (IA), la cual actúa como un catalizador para reconfigurar la gestión de lo público.

El presente capítulo analiza cómo esta integración tecnológica busca no solo optimizar procesos administrativos, sino alcanzar la democratización del conocimiento, transformando el saber científico en un bien público regido por los principios de libertad e igualdad.

En el primer apartado, se define la gobernanza como una "realidad difusa" cuyo éxito reside en su capacidad para transitar desde el mando jerárquico tradicional hacia redes de interdependencia horizontal (Gaudin, 2002; Martínez Brouchoud, 2010). Etimológicamente derivada del griego *kybernan* (dirigir una nave), la gobernanza moderna integra la denominada Racionalidad Artificial, donde los

algoritmos procesan volúmenes masivos de datos para influir en la conducta humana y la planificación estratégica de las organizaciones.

No obstante, este alcance técnico encuentra sus límites en la persistente debilidad institucional y en la falta de estrategias nacionales integrales, especialmente en regiones que aún se encuentran en niveles de gobernanza básica. (Torres Ramírez, 2024)

El segundo eje de este estudio profundiza en lo que implica la democratización del conocimiento, entendiéndola no como una mera estructura jurídica, sino como una forma de vida donde la ciencia es para todos (Ordóñez, 2023). Este proceso exige que la población deje de ser un consumidor pasivo de información y se convierta en un ente cognocente activo, involucrado en la tríada de producción, difusión y consumo del saber. Sin embargo, el panorama de ejercicio de poder revela asimetrías críticas: el control histórico del conocimiento por parte de élites y corporaciones, sumado a la opacidad algorítmica de las "cajas negras", genera nuevas brechas que desafían el ideal democrático de transparencia.

Al abordar la implementación tecnológica en la gobernanza, es imperativo mantener la objetividad y reconocer que la IA no es una solución mágica para los problemas sistémicos (Norman-Acevedo, 2023). En América Latina, este despliegue ocurre sobre un terreno minado por prácticas estructurales injustas, donde la arbitrariedad y la desigualdad social condicionan cualquier avance. La tecnología, por sí sola, carece de la capacidad para resolver crisis educativas de fondo o conflictos políticos de larga data, debiendo ser entendida siempre como un complemento supervisado de la acción humana.

Uno de los obstáculos más severos en la región es la brecha digital, que actúa como una antinomia de la democracia al dejar a cerca del 40% de la población rural sin acceso a internet de calidad (Ordóñez, 2023; Zacatenco Arellano, 2024; Silva Escamilla, 2024). Esta exclusión limita la soberanía tecnológica y profundiza la dependencia de conocimientos importados, agravada por una inversión regional en investigación y desarrollo que apenas alcanza el 0.6% del PIB. Sin una

infraestructura digital pública robusta, los beneficios de la inteligencia artificial seguirán concentrados en las élites metropolitanas, reforzando ciclos históricos de pobreza.

Asimismo, la implementación tecnológica enfrenta riesgos éticos prominentes, como el sesgo algorítmico, que puede perpetuar discriminaciones raciales y socioeconómicas presentes en los datos de entrenamiento. A esto se suma la ansiedad social por el desplazamiento laboral y las preocupaciones sobre la vigilancia masiva en contextos de debilidad democrática.

Por ello, se propone una meta gobernanza basada en principios normativos internacionales, como los de la UNESCO, que asegure que el uso de la IA esté alineado con la dignidad humana y la rendición de cuentas.

La ruta hacia una gobernanza que democratice efectivamente el conocimiento requiere abatir la pasividad ciudadana y el consumismo capitalista en favor de la creatividad y la crítica reflexiva. La integración de la inteligencia natural con la artificial es la clave para resolver dilemas laborales y sociales, siempre que exista la voluntad política para retirar las barreras financieras y legales al saber. El futuro de América Latina dependerá de su capacidad para construir sistemas de gobernanza éticos y soberanos que pongan la tecnología al servicio del progreso humano y la justicia social.

¿Cómo entender la gobernanza?

La gobernanza se define actualmente como una "realidad difusa", cuyo éxito político radica en un encanto consensual y reconfortante que permite transitar de modelos de autoridad tradicionales a esquemas de decisión compartida. Etimológicamente, el término proviene del griego *kybernan* (dirigir o conducir una nave), y aunque se rastrea hasta la Edad Media, su uso analítico contemporáneo emergió en la década de los ochenta para promover la transparencia y la participación de la sociedad civil en la gestión de lo público.

Este concepto marca un cambio de paradigma: se ha pasado de un escenario donde el Estado era el centro indiscutible del poder político a uno donde las decisiones son el producto de la interacción y las interdependencias entre actores públicos y privados (Pierre & Peters, 2000). La gobernanza no implica necesariamente el "vaciamiento" de las funciones estatales, sino la adaptación de las instituciones a una sociedad más compleja, dinámica y globalizada, donde la fuerza legal formal del Estado cede espacio a la coordinación horizontal.

El surgimiento de la gobernanza moderna se explica por factores críticos como la crisis financiera del Estado, que obligó a involucrar a actores privados en la prestación de servicios, y el cambio ideológico hacia el mercado que cuestionó el rol exclusivo de la burocracia centralizada. Asimismo, la globalización ha erosionado el centralismo estatal, forzando la creación de niveles de gobierno multinivel que van desde lo local hasta organismos sobre estatales como la Unión Europea.

En un sentido económico, la gobernanza industrial puede entenderse como el conjunto de instituciones que regulan la actividad económica a través de mercados, jerarquías, asociaciones o redes informáticas (Linding y otros., 1991). Esta última dimensión cobra especial relevancia con la inteligencia artificial (IA), ya que la gobernanza contemporánea debe integrar redes de información para movilizar recursos dispersos y articular la acción colectiva frente a problemas de interés público.

El alcance de la gobernanza de la IA es vasto e incluye dimensiones técnicas como el establecimiento de estándares de precisión, ciberseguridad y equidad estadística. No obstante, también aborda preocupaciones existenciales sobre la supervivencia de la humanidad ante el desarrollo de máquinas autónomas conscientes, lo que exige marcos regulatorios que equilibren la innovación con la protección de valores fundamentales.

La visión tradicional de la gobernanza se fundamentaba en el ejercicio de la coerción y el mando jerárquico "top-down". En contraste, la perspectiva moderna

asume la gobernanza como una propiedad emergente de las interacciones sociales, donde los ciudadanos a menudo "saben mejor" que el Estado lo que necesitan, favoreciendo procesos de autoorganización y dirección a distancia (Kickert, 1994).

Un elemento ineludible en este debate son las redes de políticas, entendidas como conjuntos estables de relaciones no jerárquicas donde diversos actores intercambian recursos para alcanzar intereses comunes. Las redes permiten a los gobiernos compensar sus capacidades limitadas al incorporar fuentes externas de conocimiento y participación, mejorando así la eficacia gubernamental en la resolución de problemas complejos.

Sin embargo, la gobernanza tiene límites claros en su ejecución; por ejemplo, el riesgo de caer en modelos corporativistas donde los grupos más poderosos y homogéneos cuentan con ventajas insoslayables en la negociación plural. En América Latina, la mayoría de las naciones aún carece de estrategias nacionales integrales de IA, lo que sitúa a la región en un nivel de "gobernanza básica" con debilidades institucionales significativas y falta de visión a largo plazo.

La legitimidad democrática es otra de las tensiones principales, ya que los actores privados en las redes a menudo carecen de validación popular directa. Por ello, se propone la meta gobernanza, que consiste en aplicar principios normativos para guiar la conducta de los implicados, asegurando que el proceso de toma de decisiones sea explícito y evaluable ante la ciudadanía.

El "juego político" dentro de la gobernanza requiere una respuesta colaborativa e interdisciplinaria entre gobiernos, universidades y empresas privadas. Esta interacción permite diseñar soluciones adaptadas a los desafíos locales, fomentar la innovación y establecer mejores prácticas compartidas, especialmente en sectores críticos como la ciberseguridad y la salud pública.

En este contexto, las organizaciones mismas pueden ser vistas como espacios de gobernanza que varían según el nivel de participación de sus miembros. El espectro de la gobernanza democrática en las organizaciones se mide por la fuerza de la

participación de los trabajadores, que puede ir desde lo meramente consultivo conocido como participación débil hasta derechos de decisión estructuralmente garantizados, aquellos denominados como participación fuerte.

El uso de nuevas tecnologías por parte de los actores está transformando estas prácticas a través de lo que se denomina Racionalidad Artificial. Los tomadores de decisiones ahora se apoyan en algoritmos de IA para procesar volúmenes masivos de datos en tiempo real, lo que otorga una "agencia" a las máquinas capaz de influir directamente en la conducta humana y en la planificación estratégica organizacional.

Actores gubernamentales y privados utilizan la IA Generativa para automatizar auditorías, detectar amenazas de seguridad de forma predictiva y generar reportes detallados que facilitan la comunicación técnica hacia niveles directivos. Estas herramientas permiten a las organizaciones con recursos limitados acceder a capacidades avanzadas de análisis, optimizando los tiempos de respuesta ante incidentes críticos.

No obstante, un límite tecnológico crítico es la opacidad algorítmica o "caja negra", donde los procesos internos de los modelos de IA son ininteligibles para el usuario común. Esta falta de transparencia puede generar desconfianza y dificultar la rendición de cuentas, exigiendo el desarrollo de tecnologías de Inteligencia Artificial Explicable (XAI) para que las decisiones automatizadas sean justificables.

Otro uso innovador es la implementación de plataformas digitales y asambleas ciudadanas mediadas por tecnología, que permiten a las comunidades participar directamente en la elaboración de leyes y presupuestos. Estas herramientas de "tecnología cívica" demuestran que es posible escalar la deliberación democrática y reducir la distancia entre gobernantes y gobernados.

Los límites de la gobernanza de la IA deben estar marcados por los derechos humanos y la privacidad. Es imperativo regular el uso de datos biométricos y evitar la vigilancia masiva intrusiva, asegurando que la recolección de información

mediante Big Data no derive en prácticas discriminatorias o en un "capitalismo de vigilancia" que vulnere la autonomía individual.

Sobre la democratización del conocimiento

La democratización del conocimiento trasciende la mera acumulación de datos para definirse como un proceso donde la ciencia y el saber se transforman en un bien público regido por la libertad y la igualdad. No se limita a una estructura jurídica, sino que se entiende como una forma de vida que permite la discusión y transmisión del saber hacia el mayor número de personas sin obstáculos previos. Este concepto busca responder a la interrogante de para quién es la ciencia, afirmando que el conocimiento debe ser accesible para todos los estratos de la sociedad.

Este apartado se interrelaciona profundamente con lo que se entiende por gobernanza, ya que esta última proporciona el marco institucional de redes e interdependencias necesario para movilizar recursos dispersos. Mientras la gobernanza gestiona la interacción entre actores públicos y privados, la democratización asegura que el contenido de dichas interacciones el conocimiento, no sea monopolizado por una élite. Así, una gobernanza democrática es aquella que garantiza el acceso a la educación y los instrumentos para actuar en la vida política.

El alcance de esta democratización abarca la tríada de producción, difusión y consumo del saber científico. No basta con generar investigación de excelencia si esta permanece en secreto o bajo el control de sectores privilegiados; es imperativo que el producto de la ciencia cumpla su cometido social. En este sentido, la población debe dejar de ser un ente pasivo de consumo para convertirse en un sujeto cognoscente activo de la realidad.

Un panorama crítico del ejercicio de poder revela que el conocimiento ha sido históricamente controlado por grandes grupos editoriales y corporaciones que imponen barreras financieras. La democratización busca desplazar este poder hacia la mayoría, asegurando que los beneficios del progreso científico recaigan en la sociedad y no solo en quienes ostentan el capital. La gobernanza de la IA debe,

por tanto, vigilar que los intereses comerciales no capturen los modos de comunicación electrónica (Ordóñez, 2023; Silva Escamilla, 2025).

Las asimetrías políticas se manifiestan con dureza en América Latina a través de la brecha digital, considerada una antinomia de la democracia. Mientras que las grandes metrópolis actúan como hubs tecnológicos, el 40% de la población rural en la región carece de acceso estable a internet. Esta desigualdad impide una participación equitativa en la sociedad del conocimiento, reforzando ciclos de pobreza y exclusión social. (Torres Ramírez, 2024; Silva Escamilla, 2025).

El uso de la Inteligencia Artificial introduce nuevos desafíos en el ejercicio del poder, como la opacidad algorítmica. Muchos modelos operan como "cajas negras", donde los procesos internos son ininteligibles para el ciudadano común, dificultando la rendición de cuentas. Esta falta de transparencia crea una asimetría de información entre los desarrolladores de la tecnología y quienes son objeto de sus decisiones automatizadas.

Los límites de la democratización también están marcados por la concentración de poder en manos de unas pocas empresas de Estados Unidos y China. Esta centralización geográfica limita la soberanía tecnológica de las naciones latinoamericanas, que a menudo dependen de conocimientos e infraestructuras importadas. Para democratizar el saber, es esencial fomentar capacidades locales e investigación propia adaptada a las necesidades de la región.

Otro límite fundamental reside en los sesgos algorítmicos, que pueden perpetuar discriminaciones históricas de género, etnia o estatus socioeconómico. Si la IA se entrena con datos sesgados, las decisiones en áreas como salud o contratación laboral no serán democráticas, sino que amplificarán las injusticias preexistentes. Una gobernanza ética exige auditorías algorítmicas constantes para detectar y corregir estos sesgos.

La implementación de la IA en la gobernanza permite potenciar la participación ciudadana mediante plataformas de deliberación y gobierno abierto. Sin embargo, la eficacia de esta participación depende de una educación STEM inclusiva que

reduzca la brecha de talento especializado. Sin ciudadanos formados en habilidades digitales, las herramientas tecnológicas pueden convertirse en instrumentos de manipulación en lugar de empoderamiento.

El instrumento práctico más potente para esta causa es el Acceso Abierto (Open Access), que busca retirar las barreras financieras y legales a la literatura científica (Ordóñez, 2023). El conocimiento generado con recursos públicos debe ser un bien de carácter universal, permitiendo que cualquier persona lea, descargue y distribuya el saber sin restricciones técnicas. Esto acelera la investigación y une a la humanidad en una conversación intelectual común.

No obstante, existen límites legales impuestos por marcos de propiedad intelectual que a menudo favorecen el lucro privado sobre el interés social. La democratización exige que los autores y depositarios de derechos garanticen al público el derecho libre e irrevocable de acceso a sus obras eruditas. La gobernanza debe mediar en esta tensión, protegiendo la integridad del trabajo del autor mientras asegura su difusión masiva.

En América Latina, la fuga de talento es una asimetría política crítica que debilita la producción de conocimiento local. Muchos investigadores migran hacia polos de desarrollo fuera de la región, dejando a sus países de origen con una infraestructura científica precaria. Democratizar el conocimiento implica crear incentivos para que los especialistas permanezcan en sus comunidades y contribuyan al desarrollo regional.

La libertad académica es un pilar indispensable, permitiendo que las ideas sean discutidas, contrastadas y refutadas. La IA puede facilitar este proceso al automatizar la revisión de literatura y detectar patrones, pero no debe sustituir la reflexión crítica humana. El límite de la tecnología es la preservación de la autonomía del investigador y el docente ante la estandarización educativa.

El panorama del poder en la era de la IA también plantea riesgos de vigilancia masiva y control social. (Norman-Acevedo, 2023) La recolección de grandes volúmenes de datos personales puede llevar a un monitoreo intrusivo que vulnere

la privacidad mental. Una gobernanza responsable debe establecer "barandillas" éticas que impidan el uso de la IA para la represión o la manipulación electoral.

Las asimetrías socioeconómicas entre los actores del mundo laboral dificultan el diálogo social en torno a las nuevas tecnologías. Quienes poseen las habilidades digitales obtienen ventajas competitivas, mientras que los trabajadores con baja calificación enfrentan el riesgo de desplazamiento laboral. La democratización del conocimiento debe incluir programas de reciclaje profesional para asegurar una transición justa hacia la economía digital.

Es vital reconocer que la IA no es una solución mágica que pueda resolver por sí sola los problemas educativos o políticos de la región. Las prácticas estructurales injustas, como la pobreza y la falta de infraestructura básica, requieren voluntad política y reformas sociales profundas. La tecnología es solo un multiplicador de las capacidades que la sociedad decida fomentar.

La interrelación con la gobernanza se hace evidente al proponer una meta gobernanza, que aplique principios normativos para guiar la conducta de todos los actores. Esto asegura que la toma de decisiones sea explícita, evaluable y fundamentada en la transparencia y la rendición de cuentas. Solo así se puede garantizar que la IA se utilice para el bien común y no para reforzar monopolios de información.

La democratización del conocimiento implica también abatir el consumismo capitalista en favor de la creatividad y la crítica reflexiva. Se busca transformar al ciudadano en un agente activo que utilice la tecnología para cuestionar la realidad y proponer soluciones a los desafíos locales. El acceso al saber debe ser un derecho fundamental para enfrentar la incertidumbre del futuro.

Finalmente, este proceso requiere una colaboración regional robusta entre México, Ecuador y el resto de América Latina para crear redes de investigación soberanas. Compartir conocimientos y estándares éticos permitirá a la región incidir en la gobernanza global de la IA y proteger sus intereses específicos. La unión de

esfuerzos entre universidades, gobiernos y empresas es el camino para cerrar las brechas existentes.

Sobre la implementación tecnológica y la gobernanza

La implementación de la IA en prácticas de gobernanza para la democratización del conocimiento se presenta hoy como una "realidad difusa", cuyo éxito político depende de un encanto consensual que permite transitar de modelos jerárquicos tradicionales a esquemas de decisión compartida. Este proceso no debe entenderse como una transición meramente técnica, sino como una transformación profunda de las bases del poder político, donde la interacción y la interdependencia entre actores públicos y privados sustituyen al monopolio estatal del interés colectivo.

Sin embargo, es imperativo mantener la objetividad: la Inteligencia Artificial no es una solución mágica para los grandes problemas educativos, sociales ni políticos que aquejan a la humanidad (Velázquez Sagahón y otros., 2024). En el contexto específico de América Latina, la tecnología se despliega sobre un terreno minado por prácticas estructurales injustas, arbitrariedad y una desigualdad social persistente que condiciona cualquier intento de progreso.

América Latina es una región marcada por desafíos estructurales donde más del 30% de la población vive en condiciones de pobreza, lo que limita severamente las oportunidades de desarrollo. En este escenario, la implementación de la IA corre el riesgo de convertirse en un instrumento que profundice las brechas existentes en lugar de cerrarlas, si no se acompaña de políticas inclusivas que reconozcan estas realidades materiales.

La brecha digital constituye el obstáculo más crítico para la democratización del conocimiento en la región, funcionando como una antinomia de la democracia. Cerca del 40% de la población latinoamericana carece de acceso a internet de calidad, una exclusión que atenta contra los valores de igualdad y libertad al privar a millones de personas de los beneficios de la sociedad de la información.

Sin una infraestructura digital pública robusta, las áreas rurales y las comunidades de bajos ingresos permanecen rezagadas, perpetuando un ciclo de desigualdad. La inversión desigual en tecnologías de seguridad y conectividad entre las grandes metrópolis y las zonas periféricas crea asimetrías políticas que la IA, por sí sola, es incapaz de revertir sin una voluntad gubernamental decidida.

El "juego político" de la gobernanza tecnológica implica que los gobiernos se vuelven dependientes de fuentes externas de conocimiento y de la intervención de diversos actores en la definición de políticas. No obstante, en América Latina existe una fuerte dependencia tecnológica externa y una inversión limitada en investigación y desarrollo, que apenas alcanza el 0.6% del PIB regional en comparación con el 2-3% de las potencias mundiales.

Esta falta de soberanía tecnológica se agrava con la fuga de talento especializado, donde los investigadores migran hacia polos de desarrollo fuera de la región, debilitando la base científica local. Por tanto, la implementación tecnológica en la gobernanza debe priorizar la formación de talento local y la creación de ecosistemas de innovación propios que respondan a las necesidades específicas de nuestras sociedades.

Desde una perspectiva de gobernanza moderna, la IA puede actuar como un catalizador para redes de políticas que compensen las capacidades limitadas del Estado. Herramientas como la IA Generativa ya están transformando áreas como la ciberseguridad, permitiendo automatizar auditorías y detectar amenazas de forma predictiva en sectores críticos como las finanzas y la salud.

En el ámbito educativo, las plataformas adaptativas tienen el potencial de personalizar el aprendizaje y reducir las brechas de acceso en contextos rurales. Sin embargo, la tecnología debe ser siempre un complemento supervisado que no sustituya el pensamiento crítico humano ni la interacción docente-discente, esenciales para un desarrollo integral.

Uno de los mayores riesgos de la implementación tecnológica es la opacidad algorítmica o el efecto "caja negra". La falta de transparencia en los procesos

internos de los modelos de IA impide que los ciudadanos comprendan cómo se llegan a decisiones que afectan sus vidas, lo que socava la rendición de cuentas, un pilar fundamental de la buena gobernanza.

A esto se suma el desafío del sesgo algorítmico, donde los sistemas reflejan y amplifican prejuicios humanos preexistentes. Si los datos de entrenamiento no son representativos de la diversidad étnica y socioeconómica de América Latina, las decisiones automatizadas en áreas como el crédito o la justicia penal pueden derivar en prácticas discriminatorias injustas.

La democratización del conocimiento a través de la IA exige que el saber científico se transforme en un bien público universal mediante el Acceso Abierto. Es injusto que la literatura generada con fondos públicos sea monopolizada por intereses comerciales que imponen barreras financieras, privando a la mayoría de las herramientas para resolver sus problemas cotidianos.

La implementación tecnológica debe fomentar una ciudadanía activa, transformando al habitante de un consumidor pasivo de información en un sujeto cognoscente capaz de participar en el control social. La IA puede facilitar el gobierno abierto mediante plataformas digitales que involucren a la población en la planificación urbana y la gestión de recursos.

Sin embargo, en contextos de debilidad institucional, existe el peligro de que la tecnología se utilice para la vigilancia masiva y el control social invasivo. La recolección masiva de datos personales sin marcos regulatorios robustos puede vulnerar la privacidad mental y la dignidad humana, amenazando las libertades civiles en la región (Zavala Osorio & Gómez-Quintero, 2025)

Las asimetrías socioeconómicas también se manifiestan en el mercado laboral, donde la automatización genera una ansiedad justificada o "AI-nxiety" ante el posible desplazamiento de empleos de baja calificación. Las políticas de gobernanza deben incluir programas de reciclaje profesional para evitar que la transición tecnológica se traduzca en una mayor precariedad laboral.

Es fundamental reconocer que la mayoría de los países latinoamericanos aún se encuentran en una fase de "gobernanza básica" o "exploradora" respecto a la IA, careciendo de estrategias nacionales integrales y marcos regulatorios actualizados. Sin estas guías éticas, la implementación tecnológica queda sujeta a la arbitrariedad y al beneficio de unos pocos.

La cooperación regional entre naciones como México, Ecuador, Brasil y Chile es indispensable para fortalecer las capacidades tecnológicas y negociar estándares éticos comunes frente a los gigantes tecnológicos globales. Solo mediante una voz unificada, América Latina podrá asegurar que sus intereses y diversidad cultural sean protegidos en la era de la IA.

La implementación de IA en la gestión de servicios básicos, como la electricidad, ha mostrado éxitos en proyectos piloto que reducen fallas en comunidades rurales, demostrando que la tecnología puede ser un catalizador de justicia social cuando se aplica con objetivos de equidad. Estos casos de éxito deben escalarse mediante alianzas público-privadas transparentes.

No obstante, la sostenibilidad debe ser un eje transversal: la infraestructura necesaria para la IA consume vastos recursos energéticos y hídricos, un coste ambiental que a menudo se ignora en los discursos de progreso. Una gobernanza responsable debe equilibrar la innovación con el cuidado del planeta y los límites territoriales.

La ética no puede ser un añadido posterior, sino la raíz del diseño de cualquier sistema sociotécnico. Es necesario involucrar a la sociedad civil y a las comunidades afectadas en la elaboración de directrices, asegurando que la IA hable sus lenguas y respete sus conocimientos ancestrales sin irrumpir violentamente en sus culturas.

En conclusión, la tecnología debe servir para abatir la pasividad ciudadana y el consumismo capitalista, fomentando la creatividad y la crítica reflexiva. La implementación de la IA en la gobernanza solo será exitosa si logra que los

beneficios del progreso técnico recaigan verdaderamente en todos los estratos de la sociedad.

Debemos alejarnos de un tecno-optimismo ingenuo: los problemas de corrupción, arbitrariedad e injusticia estructural en América Latina requieren voluntad política y reformas sociales profundas que ningún algoritmo puede generar por sí solo. La IA es una herramienta de apoyo, no un sustituto de la responsabilidad humana.

La democratización del conocimiento se alcanza cuando se retiran todas las barreras financieras, legales y técnicas que impiden que la ciencia sea un derecho para todos. La gobernanza, mediada por la IA, debe ser el vehículo que garantice el acceso a la educación y proporcione los instrumentos necesarios para actuar en la vida política.

El futuro de la región dependerá de nuestra capacidad para diseñar sistemas de "Racionalidad Artificial" que integren la precisión de la máquina con el juicio ético y la empatía humana. Si los sistemas tecnológicos fallan, la mente humana y su capacidad de organización democrática siguen siendo nuestro soporte más confiable.

La implementación de la Inteligencia Artificial en las prácticas de gobernanza no puede ser interpretada como un proceso puramente técnico o neutral, sino como una batalla política por la democratización del conocimiento en una región fracturada por la arbitrariedad. Es imperativo alejarse de un tecno-optimismo ingenuo: la IA no es una solución mágica para las crisis educativas o la corrupción sistémica. En América Latina, este despliegue tecnológico ocurre sobre un terreno marcado por prácticas estructurales injustas, donde más del 30% de la población vive en la pobreza y el 40% carece de un acceso a internet que le permita ser parte de la sociedad del saber. Esta brecha digital constituye una verdadera "antinomía de la democracia", ya que mientras las élites metropolitanas se consolidan como centros tecnológicos, las comunidades rurales e indígenas son condenadas a un "acceso prohibido" que perpetúa ciclos históricos de exclusión y pobreza (Martínez

Brouchoud, 2010; Norman-Acevedo, 2023; Silva Escamilla, 2025; Sansores Guerrero & Zavala Osorio, 2025).

El panorama actual revela una asimetría de poder alarmante, donde el desarrollo y control de la IA están concentrados en manos de unas pocas corporaciones gigantes y potencias extranjeras, lo que debilita la soberanía tecnológica de nuestras naciones. Esta centralización facilita la mercantilización del saber y el surgimiento de un "capitalismo de vigilancia", donde los datos personales son explotados sin transparencia. En este "juego político", los actores no son iguales: los grupos más poderosos cuentan con ventajas insoslayables en la negociación, mientras que el ciudadano común queda atrapado ante la opacidad algorítmica de las "cajas negras", incapaz de comprender las decisiones que afectan su vida. La gobernanza de la IA debe, por tanto, vigilar rigurosamente que los intereses comerciales no capturen los modos de comunicación ni monopolicen el conocimiento generado con recursos públicos (Zavala Osorio & Gómez-Quintero, 2025; Avilés Chacón & Cruz Rodríguez, 2025; Nova & Rivera-Aguilera, 2024; Ordóñez, 2023).

Frente a estos riesgos, el papel del Estado y la sociedad civil en la gobernanza es irrenunciable. La gobernanza no debe significar el "vaciamiento" de las funciones estatales ni el abandono del interés general al libre juego de intereses grupales o mercantiles. Es necesaria una metagobernanza ética que garantice la participación ciudadana real, permitiendo que investigadores y comunidades afectadas tengan voz en la elaboración de las directrices políticas. El rol de la IA debe ser estrictamente el de una herramienta de apoyo que fomente la creatividad y la crítica reflexiva, retirando las barreras financieras y legales al saber para que la ciencia sea, finalmente, un bien público universal (Martínez Brouchoud, 2010; Santos Hortelano, 2025; Guerrero Carrasco y otros., 2025; Ordóñez, 2023).

En última instancia, el futuro de la región depende de nuestra capacidad para construir una "Racionalidad Artificial" que no sustituya, sino que potencie el juicio ético y la empatía humana. Debemos transformar al habitante de un consumidor

pasivo de mercancías digitales en un sujeto cognocente activo con capacidad de incidencia política. Si los algoritmos fallan, si los sesgos perpetúan la discriminación o si las infraestructuras colapsan, la mente humana y su capacidad de organización democrática siguen siendo nuestro soporte más confiable y la única defensa contra la deshumanización. La tecnología solo será un aliado del progreso si se pone al servicio de la justicia social y la dignidad de todos los estratos de la sociedad (Velázquez Sagahón y otros., 2024; Ordóñez, 2023; Silva Escamilla, 2025).

Conclusiones: Hacia una gobernanza ética frente a la ambivalencia tecnológica

La implementación de la IA en prácticas de gobernanza para la democratización del conocimiento aún da para trazar un vasto campo de oportunidades, que no debe interpretarse como un destino final exento de peligros, pues aún se trata de un proceso inacabado y profundamente ambivalente.

Si bien esta tecnología promete optimizar la gestión de lo público y expandir las fronteras del saber, su despliegue en América Latina ocurre en un terreno minado por prácticas estructurales injustas, arbitrariedad y una desigualdad social que condicionan cualquier intento de progreso real. La Inteligencia Artificial no es una solución mágica que pueda resolver, por sí sola, las crisis educativas, políticas o de pobreza que aquejan a la región.

Uno de los riesgos más prominentes es el sesgo algorítmico, el cual tiene el potencial de perpetuar y amplificar prejuicios históricos presentes en los datos de entrenamiento. Al ser sistemas que aprenden de patrones pasados, la IA puede derivar en decisiones discriminatorias en áreas críticas como la contratación laboral, la concesión de créditos o la justicia penal, afectando desproporcionadamente a los grupos más vulnerables. Esta situación representa una amenaza directa a la equidad y la justicia social, pilares fundamentales de cualquier sistema democrático.

La opacidad algorítmica o el efecto de "caja negra" constituye otro desafío técnico y ético insoslayable para la gobernanza. La falta de transparencia en los procesos

internos de los modelos avanzados impide que el ciudadano común comprenda la lógica detrás de las decisiones automatizadas que afectan su vida cotidiana. Sin mecanismos claros de rendición de cuentas y explicación en el uso de la IA en el gobierno corre el riesgo de volverse autoritario y de erosionar la confianza pública en las instituciones.

En materia de privacidad y seguridad, la recolección masiva de datos personales para alimentar estos sistemas plantea riesgos de monitoreo intrusivo y vigilancia estatal o corporativa. El manejo de información biométrica y el desarrollo de tecnologías de reconocimiento facial en contextos de debilidad institucional pueden derivar en un "capitalismo de vigilancia" que atente contra los derechos civiles y la autonomía individual. La protección de la privacidad mental emerge, así, como una nueva frontera de lucha por los derechos humanos en la era digital.

El impacto económico de la automatización genera una ansiedad social justificada, conocida como "AI-nxiety", ante el posible desplazamiento laboral masivo. Existe el riesgo de que la inversión en capital constante (maquinaria e IA) sustituya de manera agresiva al trabajo vivo, profundizando la desigualdad económica y la precariedad. En una economía totalmente automatizada, la base de la generación de valor y la redistribución de la riqueza entrarían en una crisis sistémica sin precedentes

En América Latina, la brecha digital actúa como una antinomia de la democracia, excluyendo a cerca del 40% de la población del acceso a internet de calidad. Esta exclusión tecnológica no solo es una barrera para la información, sino que representa una forma de injusticia social que coloca a las comunidades rurales y de bajos ingresos en una desventaja competitiva y cognitiva insalvable. Sin una infraestructura digital pública robusta, los beneficios de la IA seguirán concentrados en las élites metropolitanas.

La fuerte dependencia tecnológica externa es otro factor que debilita la soberanía de la región. América Latina apenas invierte alrededor del 0.6% de su PIB en investigación y desarrollo, lo que la obliga a importar soluciones diseñadas bajo

visiones culturales ajenas. Esta falta de inversión propia se agrava con la fuga de talento especializado, donde los mejores investigadores migran hacia potencias como Estados Unidos o China, dejando a los países locales sin la capacidad crítica para auditar o adaptar estas tecnologías.

Las asimetrías de poder se reflejan también en la concentración de la industria de la IA en manos de unas pocas corporaciones gigantes. Este monopolio sobre los modos de comunicación electrónica y el procesamiento de datos subordina el interés colectivo a las necesidades financieras del mercado. Para democratizar el conocimiento, es imperativo que la gobernanza promueva el Acceso Abierto como un bien público universal que retire las barreras financieras impuestas por los grandes grupos editoriales.

El riesgo de deshumanización de los servicios esenciales, como la educación y la salud, es una preocupación latente. La sustitución de la interacción humana por sistemas automáticos puede eliminar la empatía, la compasión y la reflexión crítica necesarias para un desarrollo social integral. La IA debe ser vista como un complemento supervisado y no como un reemplazo de la conexión humana que da sentido a la vida comunitaria.

En el ámbito político, el uso de IA para la manipulación social y electoral representa un peligro para la estabilidad democrática. La creación de "deepfakes" y la difusión de desinformación a través de algoritmos pueden erosionar la veracidad de la información y polarizar a la sociedad. Una gobernanza responsable exige marcos regulatorios que sancionen el uso malicioso de la tecnología y garanticen la integridad del debate público.

La inclusión de poblaciones vulnerables, como comunidades indígenas y minorías étnicas, es una tarea pendiente en la agenda tecnológica. La IA a menudo ignora la diversidad lingüística y cultural, imponiendo estándares homogéneos que invisibilizan los saberes ancestrales. Democratizar el conocimiento implica asegurar que estas tecnologías sean culturalmente adaptativas y que estas comunidades tengan voz en el diseño de las herramientas que las afectan.

El coste medioambiental del entrenamiento de grandes modelos de IA es una injusticia que a menudo se omite en los discursos de progreso. El consumo masivo de energía y agua en los centros de datos presiona los límites del planeta y afecta la sostenibilidad de los recursos locales. Una gobernanza ética debe integrar la sustentabilidad ambiental como un criterio irrenunciable para la aprobación de proyectos tecnológicos.

Es fundamental reconocer que los problemas de corrupción y falta de transparencia en la política tradicional no se eliminan con la tecnología. Si la IA se implementa sobre estructuras institucionales corruptas, simplemente se convertirá en un medio más eficiente para perpetuar la arbitrariedad. La reforma social profunda y la voluntad política siguen siendo los motores primordiales del cambio; el algoritmo es solo un instrumento.

Para enfrentar estos riesgos, se requiere una meta gobernanza fundamentada en principios normativos internacionales, como las recomendaciones de la UNESCO. Es necesario establecer "barandillas" éticas que aseguren que la IA se utilice para el bienestar colectivo y el respeto a la dignidad humana. La colaboración interdisciplinaria entre gobiernos, universidades y sociedad civil es la única vía para diseñar soluciones que cierren las brechas en lugar de profundizarlas.

La educación en competencias STEM inclusivas y la alfabetización digital avanzada son esenciales para que la población deje de ser un consumidor pasivo y se convierta en un sujeto cognoscente activo. Sin ciudadanos formados en el uso crítico de la tecnología, la democratización del conocimiento será solo un espejismo. El fortalecimiento de las habilidades blandas, como el pensamiento crítico y la ética, es más importante hoy que nunca.

A nivel regional, la cooperación entre países latinoamericanos es indispensable para negociar estándares comunes frente a los gigantes tecnológicos y crear redes de investigación soberanas. Iniciativas como el "Gran Modelo de Lenguaje Latino" demuestran que es posible unir esfuerzos para proteger la diversidad cultural y los

intereses económicos de la región. La unión de esfuerzos es la clave para no quedar rezagados en la carrera global por la innovación.

En conclusión, la implementación de la IA en la gobernanza debe estar guiada por una brújula humanista que priorice la equidad sobre el lucro individual. No debemos caer en un tecno-optimismo ingenuo que ignore las realidades materiales de exclusión en las que vivimos. La tecnología debe servir para abatir el consumismo capitalista y fomentar la creatividad y la crítica reflexiva en favor del progreso social.

El horizonte final es lograr que el conocimiento sea, verdaderamente, un bien público accesible para todos los estratos de la sociedad. Esto requiere retirar todas las barreras financieras, legales y técnicas que han monopolizado históricamente la ciencia y la información. Solo así la IA dejará de ser una herramienta de control para convertirse en un aliado en la construcción de una América Latina más justa y soberana.

Debemos recordar siempre que, si los sistemas tecnológicos llegan a fallar o a actuar en contra de los valores humanos, nuestra mayor defensa sigue siendo la mente humana y su capacidad de organización democrática. El futuro no está predeterminado por el código; es una construcción social que depende de las decisiones éticas que tomemos hoy. La responsabilidad de guiar este cambio paradigmático reside en nosotros, no en las máquinas.

Finalmente, la democratización del conocimiento en la era de la IA es un imperativo ético para la supervivencia de la democracia misma. Al enfrentar con valentía los riesgos de sesgo, vigilancia y exclusión, podemos aspirar a una sociedad donde el progreso técnico se traduzca en una mejora real de la calidad de vida para los más de 8 mil millones de habitantes de la tierra. La meta es, y siempre será, poner la inteligencia al servicio de la humanidad y la justicia social.

Bibliografía

Aguado-López, E., Rogel Salazar, R., Garduño Oropeza, G., Zúñiga Roca, M. F., y Baca Zapata, G. (2009). Redalyc y la democratización del conocimiento. Universidad Autónoma del Estado de México.

Battilana, J., Beckman, C. M., y Yen, J. (2025). On Democratic Organizing and Organization Theory. *Administrative Science Quarterly*, 70(2), 297–327. <https://doi.org/10.1177/00018392251322430>

De la Garza Toledo, E. (2018). Introducción: El futuro del Trabajo en América Latina. *Trabajo*, 10(15), 5-10.

De la Garza Toledo, E., y Otero, G. (2020). Trabajo y Sindicalismo en México: Desafíos de la cuarta transformación del gobierno de MORENA. *Trabajo*, (177), enero-julio 2020.

Juca-Maldonado, F. (2023). El impacto de la inteligencia artificial en los trabajos académicos y de investigación. *Revista Metropolitana de Ciencias Aplicadas*, 6(S1), 289-296.

Lozano Carrillo, O., y Garibay Suárez, A. (Coords.). (2024). El futuro de la inteligencia artificial. *Red Mexicana de Investigadores en Estudios Organizacionales (REMINEO)*.

Lozano Carrillo, O., y Garibay Suárez, A. (Coords.). (2025). Inteligencia artificial para América Latina. Reflexiones desde México y Ecuador. *Red Mexicana de Investigadores en Estudios Organizacionales (REMINEO)*.

Martínez Brouchoud, M. F. (2010). Gobernanza y legitimidad democrática. *Reflexión Política*, 12(23), 96-107.

Muñoz Cruz, R. C. (2017). La Nueva Percepción de la Identidad y la Cultura en la Subcontratación [Ponencia]. Universidad Autónoma Metropolitana Unidad Iztapalapa.

Norman-Acevedo, E. (2023). La inteligencia artificial en la educación: una herramienta valiosa para los tutores virtuales universitarios y profesores universitarios. *Panorama*, 17(32). <https://doi.org/10.15765/pnrm.v17i32.3681>

Ordóñez, J. (2023). El Open Access: Un medio para la democratización del conocimiento. *Revista de la Educación Superior*, 52(205). <https://doi.org/10.36857/resu.2023.205.2371>

Ruiz-Moreno, A., Ortega-Egea, T., Haro-Domínguez, C., y Roldán-Bravo, M. (2014). El proceso de co-creación de valor y su impacto en la estrategia de innovación en empresas de servicios. *Intangible Capital*, 10(2), 266-293. <http://dx.doi.org/10.3926/ic.504>

Sánchez Vila, E. M., y Lama Penín, M. (2007). Monografía: Técnicas de la Inteligencia Artificial Aplicadas a la Educación. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 11(33), 7-12.

Urrego-Sandoval, C. (2022). [Reseña del libro: Ahora o nunca. Gobernanza, coproducción y bioeconomía contra el cambio climático, por M. I. Gómez-Lee (Ed.)]. *Oasis*, (35), 361-365. <https://doi.org/10.18601/16577558.n35.18>

CAPÍTULO IX

INTELIGENCIA ARTIFICIAL, ÉTICA Y ACCESIBILIDAD: DESAFÍOS EN CONTEXTOS DE DISCAPACIDAD Y BRECHA DIGITAL

Artificial Intelligence, Ethics and Accessibility: Challenges in Contexts of
Disability and Digital Divide

Nadia Estefania Dalcol

nadia.dalcol@udemmm.edu.ar

Universidad de la Marina Mercante

<https://orcid.org/0009-0004-3232-2925>

Resumen

El avance de la Inteligencia Artificial (IA) ha abierto nuevas posibilidades para la inclusión de personas con discapacidad, especialmente en términos de accesibilidad, autonomía, comunicación y participación social. Herramientas basadas en IA permiten adaptar contenidos, facilitar la interacción con entornos digitales y ofrecer apoyos personalizados que hace algunos años resultaban impensados. Sin embargo, este desarrollo también plantea interrogantes éticos profundos vinculados a la desigualdad en el acceso, la falta de formación para su uso y la reproducción de brechas digitales preexistentes.

La expansión acelerada de estas tecnologías no impacta de manera homogénea en la población: mientras algunos sujetos logran ampliar sus oportunidades educativas, laborales y comunicativas mediante el uso de IA, otros quedan excluidos por limitaciones económicas, tecnológicas o formativas. En este sentido, la discusión sobre IA no puede reducirse únicamente a sus capacidades técnicas, sino que debe incorporar preguntas vinculadas a la justicia social, la equidad y los derechos humanos.

Este artículo analiza el potencial inclusivo de la IA en diálogo con los riesgos de exclusión que emergen cuando su acceso y uso no están garantizados de manera

equitativa. Asimismo, reflexiona sobre la necesidad de construir enfoques éticos y políticas públicas que promuevan una IA accesible, inclusiva y socialmente responsable.

Palabras clave: Inteligencia Artificial; accesibilidad; discapacidad; inclusión digital; brecha digital; ética tecnológica; equidad educativa; alfabetización digital; tecnologías de apoyo; Diseño Universal para el Aprendizaje (DUA); inclusión social; tecnologías accesibles.

Abstract

The advancement of Artificial Intelligence (AI) has opened up new possibilities for the inclusion of people with disabilities, especially in terms of accessibility, autonomy, communication and social participation. AI-based tools make it possible to adapt content, facilitate interaction with digital environments and offer personalised support that was unthinkable a few years ago. However, this development also raises deep ethical questions linked to inequality in access, lack of training for its use and the reproduction of pre-existing digital divides.

The accelerated expansion of these technologies does not have a homogeneous impact on the population: while some individuals manage to expand their educational, work, and communication opportunities through the use of AI, others are excluded due to economic, technological, or training limitations. In this sense, the discussion on AI cannot be reduced only to its technical capabilities, but must incorporate questions related to social justice, equity and human rights.

This article discusses the inclusive potential of AI in dialogue with the risks of exclusion that emerge when its access and use are not guaranteed in an equitable manner. It also reflects on the need to build ethical approaches and public policies that promote accessible, inclusive and socially responsible AI.

Keywords: Artificial Intelligence; accessibility; disability; digital inclusion; digital divide; technological ethics; educational equity; digital literacy; assistive technologies; Universal Design for Learning (UDL); social inclusion; accessible technologies.

Introducción

La Inteligencia Artificial se ha consolidado como una de las tecnologías más influyentes del siglo XXI, transformando prácticas sociales, educativas, laborales y culturales. Desde asistentes virtuales hasta sistemas de recomendación, pasando por herramientas de generación automática de texto, reconocimiento de voz o análisis predictivo, la IA atraviesa cada vez más dimensiones de la vida cotidiana.

En el campo de la discapacidad, su potencial resulta especialmente significativo. Las tecnologías basadas en IA pueden convertirse en herramientas de apoyo fundamentales para favorecer la autonomía, reducir barreras y ampliar oportunidades de participación. Sistemas de lectura automatizada, subtítulo en tiempo real, descripción de imágenes, asistentes conversacionales o aplicaciones de comunicación aumentativa representan ejemplos concretos de cómo estas tecnologías pueden contribuir a mejorar la calidad de vida de muchas personas.

Sin embargo, el desarrollo tecnológico no ocurre en un vacío social. La incorporación de IA se produce en contextos atravesados por desigualdades económicas, educativas y digitales que condicionan profundamente quiénes pueden acceder a estas herramientas y quiénes quedan excluidos de sus beneficios. En este sentido, la IA no es neutral: sus posibilidades de inclusión dependen de las condiciones materiales, políticas y culturales en las que se implementa.

La discusión ética sobre IA suele enfocarse en cuestiones vinculadas a la privacidad de datos, los sesgos algorítmicos o la automatización del trabajo. No obstante, en relación con discapacidad y accesibilidad, emerge una preocupación adicional: la

posibilidad de que estas tecnologías profundicen desigualdades ya existentes. Cuando el acceso a IA depende del poder adquisitivo, de la conectividad o de competencias digitales específicas, las brechas tecnológicas se amplían y afectan especialmente a los grupos históricamente vulnerados.

Por ello, pensar la IA desde una perspectiva inclusiva implica reconocer tanto sus potencialidades como sus límites. La pregunta central ya no es únicamente qué puede hacer la IA, sino quiénes pueden acceder a ella, en qué condiciones y con qué apoyos.

Transformaciones sociales producidas por la IA

La expansión de la IA no solo modifica herramientas tecnológicas, sino también formas de aprender, comunicarse, trabajar y participar socialmente. En muchos sentidos, la IA está redefiniendo las dinámicas de producción de conocimiento y los modos en que las personas interactúan con la información.

En el ámbito educativo, por ejemplo, las plataformas inteligentes permiten personalizar contenidos, automatizar evaluaciones y generar apoyos inmediatos. En el ámbito laboral, numerosos procesos se realizan actualmente mediante automatización o asistencia algorítmica. En la vida cotidiana, asistentes virtuales y sistemas predictivos median decisiones de consumo, comunicación y organización personal.

Estas transformaciones generan oportunidades, pero también nuevas dependencias tecnológicas. Quienes dominan el uso de herramientas digitales avanzadas cuentan con mayores posibilidades de acceso al conocimiento y al empleo, mientras que quienes no poseen dichas competencias pueden quedar progresivamente excluidos.

En este contexto, la alfabetización digital adquiere una dimensión aún más compleja. Ya no se trata únicamente de aprender a usar dispositivos tecnológicos, sino de comprender cómo funcionan los sistemas inteligentes, qué decisiones automatizan y qué impactos producen en la vida social.

Asimismo, las tecnologías basadas en IA están modificando las formas de producir y validar conocimiento. La posibilidad de generar textos, imágenes o respuestas automatizadas plantea interrogantes sobre autoría, confiabilidad de la información y pensamiento crítico. Esto resulta especialmente relevante en contextos educativos, donde la enseñanza no puede limitarse al uso instrumental de herramientas, sino que debe promover procesos reflexivos y éticos.

IA y discapacidad: oportunidades para la accesibilidad

Desde el paradigma del Diseño Universal y la accesibilidad, la IA puede ser entendida como un facilitador clave para reducir barreras y ampliar posibilidades de participación. Su capacidad para adaptar contenidos, reconocer patrones y responder a necesidades específicas permite generar entornos más flexibles y personalizados.

Accesibilidad cognitiva

La IA puede favorecer la comprensión de la información mediante herramientas que simplifican textos complejos, generan resúmenes automáticos, explican conceptos en lenguaje claro o adaptan contenidos según diferentes niveles de lectura. Estas aplicaciones resultan especialmente valiosas para personas con discapacidad intelectual, trastornos del aprendizaje o dificultades en la comprensión lectora.

Además, los asistentes conversacionales posibilitan realizar consultas inmediatas, recibir explicaciones paso a paso y acceder a apoyos personalizados. Esto puede fortalecer procesos de autonomía y aprendizaje autorregulado.

Accesibilidad sensorial

Las tecnologías basadas en IA han mejorado significativamente las posibilidades de accesibilidad para personas con discapacidad visual o auditiva. Entre ellas se encuentran:

- Lectores de pantalla más precisos.
- Subtitulado automático en tiempo real.

- Reconocimiento de voz.
- Descripción automática de imágenes.
- Traducción instantánea.
- Conversión de texto a voz y de voz a texto.

Estas herramientas favorecen el acceso a contenidos educativos, culturales y comunicacionales que históricamente resultaban inaccesibles.

Autonomía comunicativa

La IA también potencia sistemas de comunicación aumentativa y alternativa (CAA), permitiendo predecir palabras, interpretar patrones de uso y adaptar respuestas según las necesidades del usuario. Esto favorece la participación social y comunicativa de personas con dificultades en el lenguaje oral o escrito.

Personalización educativa

En contextos educativos, la IA permite adaptar actividades, formatos y ritmos de aprendizaje. Esto resulta especialmente relevante desde una perspectiva inclusiva, ya que posibilita diversificar estrategias pedagógicas y ofrecer múltiples formas de representación y participación.

Sin embargo, estas potencialidades solo se concretan cuando existen condiciones reales de acceso, formación y acompañamiento. De lo contrario, las mismas tecnologías que prometen inclusión pueden transformarse en nuevos factores de desigualdad.

La brecha digital en la era de la IA

La expansión de la IA no elimina la brecha digital; por el contrario, en muchos casos la redefine y profundiza. Las desigualdades ya no se limitan únicamente al acceso a internet o a dispositivos tecnológicos, sino también a la capacidad de comprender, utilizar y apropiarse críticamente de estas herramientas.

Brecha de acceso

No todas las personas cuentan con dispositivos adecuados, conectividad estable o acceso a plataformas de IA. Muchas herramientas requieren computadoras actualizadas, teléfonos de alta gama o servicios pagos, lo que limita significativamente su disponibilidad.

En contextos de vulnerabilidad social y económica, esta situación implica una exclusión directa de los beneficios asociados a la IA.

Brecha de uso

Tener acceso a una herramienta no garantiza un uso significativo. Muchas personas no han recibido formación para interactuar con sistemas de IA, comprender sus funciones o aprovechar sus posibilidades.

Esto resulta particularmente significativo en adultos mayores, estudiantes en contextos vulnerables y personas con discapacidad que no cuentan con apoyos tecnológicos adecuados.

Brecha de apropiación crítica

Existe además una diferencia entre usar tecnología y comprender críticamente su funcionamiento. La apropiación crítica implica poder interpretar resultados, identificar errores o sesgos y utilizar la IA de manera estratégica y ética.

Quienes desarrollan estas competencias cuentan con mayores oportunidades educativas y laborales, mientras que quienes no acceden a ellas quedan en una situación de dependencia tecnológica.

Sesgos algorítmicos y discriminación

Uno de los principales desafíos éticos vinculados a la IA es la reproducción de sesgos presentes en los datos utilizados para entrenar los sistemas. Los algoritmos aprenden a partir de grandes volúmenes de información producida por sociedades atravesadas por desigualdades históricas. Como consecuencia, muchas

herramientas replican discriminaciones relacionadas con género, etnia, clase social o discapacidad.

En relación con discapacidad, esto puede manifestarse de distintas maneras:

- Sistemas de reconocimiento facial que no identifican correctamente ciertos rasgos.
- Plataformas de empleo que excluyen perfiles no normativos.
- Herramientas educativas que no contemplan diferentes formas de aprendizaje.
- Asistentes virtuales que interpretan incorrectamente patrones de habla diversos.

La discriminación algorítmica no siempre es visible. Muchas veces opera de forma automática e invisible, dificultando su identificación y cuestionamiento.

Por ello, resulta fundamental incorporar perspectivas de diversidad y accesibilidad en el diseño tecnológico. La inclusión no puede pensarse únicamente como adaptación posterior, sino como criterio constitutivo del desarrollo de sistemas inteligentes.

Equidad educativa e inclusión digital

La incorporación de tecnologías emergentes en educación requiere analizar no solo las posibilidades de innovación, sino también las condiciones de equidad en las que dichas herramientas se implementan. Como señala Dalcol (2025), pensar la equidad implica reconocer que los estudiantes no parten de las mismas condiciones materiales, sociales y tecnológicas.

En este marco, la IA introduce nuevas tensiones vinculadas a la inclusión digital. La disponibilidad desigual de dispositivos, conectividad y formación específica puede profundizar las brechas existentes, especialmente en estudiantes con discapacidad o pertenecientes a contextos de vulnerabilidad.

Además, muchas instituciones educativas enfrentan dificultades para integrar estas tecnologías de manera crítica e inclusiva. La incorporación acelerada de IA suele producirse sin procesos suficientes de capacitación docente o análisis pedagógico.

Esto genera desafíos importantes:

- Cómo garantizar accesibilidad digital.
- Cómo evitar dependencias tecnológicas.
- Cómo promover el pensamiento crítico.
- Cómo acompañar usos éticos de la IA.
- Cómo reducir desigualdades de acceso.

Desde una perspectiva inclusiva, la innovación tecnológica debe ir acompañada de políticas institucionales que garanticen participación real y accesibilidad para todos los estudiantes.

Personas sin acceso a IA: nuevas formas de exclusión

Uno de los aspectos menos visibilizados del debate actual es la situación de aquellas personas que no acceden a herramientas de IA por razones económicas, tecnológicas o educativas.

La falta de acceso puede producir consecuencias concretas:

- Menores oportunidades educativas.
- Desigualdad en el acceso a información.
- Dificultades para insertarse laboralmente.
- Dependencia de terceros para tareas digitales.
- Exclusión de espacios sociales y comunicativos.

Esto afecta especialmente a personas con discapacidad, ya que muchas herramientas accesibles requieren dispositivos costosos o conocimientos específicos para su utilización.

En este sentido, la exclusión tecnológica no implica solamente quedar fuera de determinadas plataformas, sino también perder oportunidades de autonomía y participación social.

Formación y alfabetización en IA: un desafío urgente

Frente a este escenario, la alfabetización en IA aparece como una necesidad urgente. No basta con garantizar acceso tecnológico; es necesario promover competencias que permitan comprender, utilizar y evaluar críticamente estas herramientas.

La alfabetización en IA debería contemplar:

- Uso funcional de herramientas.
- Comprensión de límites y sesgos.
- Evaluación crítica de resultados.
- Uso ético y responsable.
- Aplicación en contextos específicos.
- Enfoque accesible e inclusivo.

En el caso de las personas con discapacidad, esta formación debe diseñarse desde el paradigma de accesibilidad universal, contemplando apoyos, adaptaciones y múltiples formas de interacción.

Asimismo, resulta fundamental formar a docentes, profesionales y equipos de apoyo para que puedan acompañar estos procesos y generar propuestas inclusivas.

Políticas públicas y responsabilidad social

La construcción de una IA inclusiva no puede quedar únicamente en manos del mercado tecnológico. Los Estados y las instituciones públicas tienen un rol central en la regulación, promoción y democratización del acceso a estas herramientas.

Entre las políticas necesarias se destacan:

- Garantizar conectividad universal.
- Promover acceso gratuito o subsidiado a tecnologías accesibles.
- Desarrollar programas de alfabetización digital inclusiva.
- Regular criterios éticos en el diseño de IA.
- Exigir accesibilidad digital en plataformas públicas y privadas.
- Financiar investigaciones sobre tecnología y discapacidad.

Asimismo, las empresas tecnológicas deben asumir responsabilidades vinculadas a la accesibilidad y la diversidad. Diseñar productos accesibles no debe entenderse como una acción opcional o complementaria, sino como parte de los derechos digitales contemporáneos.

Hacia una IA inclusiva y ética

Avanzar hacia una IA verdaderamente inclusiva requiere un compromiso conjunto entre Estados, instituciones educativas, empresas tecnológicas y sociedad civil.

Algunas acciones fundamentales incluyen:

- Diseñar tecnologías accesibles desde el inicio.
- Garantizar acceso equitativo a dispositivos y conectividad.
- Promover políticas públicas inclusivas.
- Fomentar programas de alfabetización digital e IA.
- Incorporar la voz de las personas con discapacidad en el diseño tecnológico.
- Desarrollar marcos éticos situados y contextualizados.

La accesibilidad no debe entenderse como un agregado opcional, sino como un principio fundamental del desarrollo tecnológico contemporáneo.

Conclusiones

La Inteligencia Artificial representa una oportunidad inédita para avanzar en accesibilidad, autonomía e inclusión social. Sus aplicaciones pueden contribuir significativamente a reducir barreras históricas y ampliar oportunidades para personas con discapacidad.

Sin embargo, el desarrollo tecnológico no garantiza por sí mismo procesos de inclusión. Sin políticas activas, formación accesible y enfoques éticos, la IA puede reforzar desigualdades preexistentes y generar nuevas formas de exclusión.

El desafío actual no es únicamente tecnológico, sino profundamente social, político y educativo. Pensar una IA inclusiva implica reconocer que el acceso a la tecnología, la conectividad y la alfabetización digital forman parte de los derechos fundamentales en las sociedades contemporáneas.

Garantizar que nadie quede excluido de los beneficios de la IA supone construir políticas y prácticas orientadas a la equidad, la accesibilidad y la participación. Solo así la Inteligencia Artificial podrá convertirse verdaderamente en una herramienta de inclusión y no en un nuevo privilegio reservado para unos pocos.

Referencias

CAST. (2018). *Universal Design for Learning Guidelines version 2.2*.
<http://udlguidelines.cast.org>

Dalcol, N. E. (2025). *Equidad en la educación. Voces del Deseo*, 1(1).
Universidad de la Marina Mercante.
<https://publicaciones.udemm.edu.ar/index.php/vde/article/view/370/320>

Dussel, I., & Southwell, M. (2024). La escuela y la igualdad: Renovar la apuesta. *El Monitor*, octubre.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Gardner, H. (1991). *La mente no escolarizada: Cómo piensan los niños y cómo deberían enseñar las escuelas*. Paidós.

Gardner, H. (1993). *Estructuras de la mente: La teoría de las inteligencias múltiples*. Fondo de Cultura Económica.

Gardner, H. (2021). *La inteligencia reformulada*. Paidós.

Meirieu, P. (1992). *Aprender, sí. Pero ¿cómo?*. Octaedro.

Meirieu, P. (1996). *Frankenstein educador*. Laertes.

Meirieu, P. (2007). Es responsabilidad del educador provocar el deseo de aprender. *Cuadernos de Pedagogía*, 373, 42–48.

Meirieu, P. (2016). *Recuperar la pedagogía: De lugares comunes a conceptos claves*. Paidós.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.
UNESCO. <https://unesdoc.unesco.org>

Van Dijk, J. (2020). *The digital divide*. Polity Press.

World Health Organization. (2011). *World report on disability*. WHO Press.

CAPÍTULO X

INTELIGENCIA ARTIFICIAL, TRABAJO E INNOVACIÓN EN EL CONTEXTO DE LA EMPRESA FAMILIAR EN MÉXICO

Artificial Intelligence, Work and Innovation in the Context of Family Business
in Mexico

Juan Cayetano Niebla Zatarain⁵

juan.niebla@uadeo.mx. México
Universidad Autónoma de Occidente
<https://orcid.org/0000-0002-7978-6573>

Xiomara Penélope Zaldívar Colado⁶

xiomara.zadivar@uas.edu.mx
Universidad Autónoma de Sinaloa
<https://orcid.org/0009-0005-0137-5909>

José Guadalupe Robles Hernández⁷

jose.robles@uadeo.mx
Universidad Autónoma de Occidente
<https://orcid.org/0009-0000-0849-8323>

José Efrén Leyva Duarte⁸

joseleyva@uas.edu.mx México
Universidad Autónoma de Sinaloa
<https://orcid.org/0000-0001-8855-5737>

⁵ Doctor en Estudios Organizacionales. PITC UAdeO Unidad Regional Culiacán. Email: juan.niebla@uadeo.mx. <https://orcid.org/0000-0002-7978-6573>.

⁶ Doctora en Educación. PITC UAS Facultad de Informática Mazatlán. xiomara.zadivar@uas.edu.mx
<https://orcid.org/0009-0005-0137-5909>

⁷ Doctor en Derecho a la Información. PITC de la UAdeO Culiacán. Email: jose.robles@uadeo.mx.
<https://orcid.org/0009-0000-0849-8323>

⁸ Doctor en Ciencias Administrativas. PITC UAS Facultad de Química. Email: joseleyva@uas.edu.mx
<https://orcid.org/0000-0001-8855-5737>

Resumen

La inteligencia artificial (IA) se ha consolidado como una de las tecnologías más influyentes en la reconfiguración de la economía global, los mercados de trabajo y los modelos organizacionales. Su incorporación en las empresas familiares mexicanas no implica solo un cambio técnico, sino también una transformación cultural y relacional, dada la estrecha vinculación entre familia, patrimonio y empresa. Este capítulo examina la relación entre IA, trabajo, economía del conocimiento e innovación en la empresa familiar, con énfasis en el contexto mexicano.

Se desarrolla un marco teórico que concibe la empresa familiar como sistema socio-técnico, articulado con las nociones de destrucción creativa y economía del conocimiento, así como con los enfoques de aprendizaje organizacional y liderazgo intergeneracional. Con base en un enfoque cualitativo-descriptivo, se utiliza un estudio de caso único instrumental de una empresa familiar mexicana del sector alimentario que ha avanzado en la adopción de herramientas de IA para la gestión de demanda, logística, percepción del consumidor y mantenimiento industrial.

Los hallazgos evidencian que la IA puede convertirse en palanca para la permanencia y profesionalización de la empresa familiar en la economía del conocimiento, siempre que se integre en una estrategia de aprendizaje, liderazgo intergeneracional y gobernanza ética. Asimismo, se advierten riesgos vinculados a la opacidad algorítmica, la brecha de competencias digitales y la posible erosión del capital relacional cuando la tecnología se adopta sin reflexión crítica sobre los valores organizacionales.

Palabras clave: Inteligencia artificial, empresa familiar, trabajo, innovación, economía del conocimiento.

Abstract

Artificial intelligence (AI) has established itself as one of the most influential technologies in the reconfiguration of the global economy, labor markets, and organizational models. Their incorporation into Mexican family businesses does not only imply a technical change, but also a cultural and relational transformation, given the close link between family, heritage and business. This chapter examines the relationship between AI, work, knowledge economy, and innovation in family businesses, with an emphasis on the Mexican context.

A theoretical framework is developed that conceives the family business as a sociotechnical system, articulated with the notions of creative destruction and knowledge economy, as well as with the approaches of organizational learning and intergenerational leadership. Based on a qualitative-descriptive approach, a unique instrumental case study of a Mexican family business in the food sector that has advanced in the adoption of AI tools for demand management, logistics, consumer perception and industrial maintenance is used.

The findings show that AI can become a lever for the permanence and professionalization of the family business in the knowledge economy, as long as it is integrated into a learning strategy, intergenerational leadership and ethical governance. Likewise, risks linked to algorithmic opacity, the digital skills gap and the possible erosion of relational capital are warned when technology is adopted without critical reflection on organizational values.

Keywords: artificial intelligence, family business, work, innovation, knowledge economy.

Introducción

La aceleración tecnológica de las últimas décadas ha colocado a la inteligencia artificial (IA) en el centro de los debates sobre el futuro del trabajo, la competitividad empresarial y el desarrollo económico. Sistemas de aprendizaje automático, analítica predictiva, visión computacional y procesamiento de lenguaje natural forman parte ya de la infraestructura informacional de múltiples sectores productivos. La IA permite automatizar decisiones, procesar grandes volúmenes de datos y aprender de la experiencia, produciendo cambios profundos en los procesos organizativos y en los perfiles de empleo (Brynjolfsson & McAfee, 2014).

En México, las empresas familiares constituyen la base del tejido productivo, al representar la mayoría de las unidades económicas privadas y una proporción significativa del empleo formal (INEGI, 2024). Estas organizaciones combinan lógicas económicas y afectivas: la rentabilidad convive con la preservación del legado, la cohesión familiar y la reputación construida a lo largo de generaciones (Chrisman y otros., 2005). La irrupción de la IA en este contexto plantea preguntas de gran calado: ¿cómo se integra una tecnología disruptiva en organizaciones orientadas a la continuidad?; ¿qué ocurre con las formas tradicionales de autoridad cuando los datos y los algoritmos se vuelven centrales en la toma de decisiones?; ¿qué implicaciones tiene todo ello para la organización del trabajo y la innovación?

Este capítulo parte de la premisa de que la empresa familiar no es un receptáculo pasivo de la tecnología, sino una entidad que interpreta y resignifica la IA desde sus propias historias, valores y relaciones de poder. Analizar la interacción entre IA, trabajo, economía e innovación implica atender, de forma simultánea, las dimensiones técnicas de la transformación digital y las dimensiones culturales y simbólicas propias de la familia empresarial. Con este propósito, se articula un marco teórico sólido con un estudio de caso en profundidad, orientado a comprender procesos más que a generalizar resultados estadísticos.

Para comprender en toda su complejidad la irrupción de la IA en las organizaciones, es necesario contextualizar su desarrollo histórico. Desde los primeros trabajos de Turing (1950) y los intentos de la IA simbólica en los años cincuenta y sesenta, hasta el resurgimiento del aprendizaje automático impulsado por el aumento exponencial de la capacidad computacional y la disponibilidad de datos masivos a partir de la década de 2010, la IA ha atravesado ciclos de expectativas, desencanto y renovación (Russell & Norvig, 2020). La irrupción del aprendizaje profundo (deep learning), las redes neuronales convolucionales y, más recientemente, los modelos de lenguaje de gran escala, ha producido avances sin precedentes en reconocimiento de imágenes, procesamiento de lenguaje natural, predicción de demanda y optimización logística. Estas capacidades, antes reservadas a grandes corporaciones con infraestructura tecnológica sofisticada, son hoy accesibles para empresas medianas e incluso pequeñas mediante soluciones en la nube y plataformas de software como servicio (SaaS), lo que democratiza parcialmente el acceso a la innovación digital (Schwab, 2016).

No obstante, la adopción de IA en América Latina y, en particular, en México presenta particularidades que la distinguen del patrón observado en economías avanzadas. Según datos de la CEPAL (2022), la región enfrenta una brecha digital estructural que limita el acceso equitativo a infraestructura de telecomunicaciones, conectividad de banda ancha y talento especializado en tecnologías de la información. En México, si bien existe un ecosistema tecnológico dinámico concentrado en grandes centros urbanos como la Ciudad de México, Monterrey y Guadalajara, la mayoría de las empresas, en especial las de tamaño mediano y pequeño, aún operan con sistemas de gestión básicos y carecen de las competencias internas necesarias para implementar soluciones de inteligencia artificial (IMCO, 2023). A ello se suma un entorno regulatorio en construcción: a diferencia de la Unión Europea, que ha avanzado en marcos normativos específicos para la IA (European Commission, 2021), México no cuenta aún con una regulación integral que oriente el uso ético y responsable de estas tecnologías en el ámbito empresarial.

El estudio de la empresa familiar ha adquirido creciente relevancia académica en las últimas tres décadas, consolidando un campo de investigación propio con revistas especializadas, asociaciones internacionales y líneas de trabajo bien establecidas (Sharma y otros., 2012). En el contexto latinoamericano, investigadores como Belausteguigoitia (2010) y Ogliastri (2016) han documentado las características singulares de la empresa familiar en la región: mayor concentración de la propiedad en manos de la familia, menor separación entre gestión y propiedad, alta informalidad en los mecanismos de gobernanza y un peso significativo de las redes de confianza personal en las relaciones comerciales. Estas características configuran un terreno especialmente sensible para la introducción de tecnologías como la IA, que por su naturaleza tiende a formalizar, sistematizar y objetivar procesos que antes descansaban en el juicio personal y en las relaciones informales de autoridad. La escasez de investigaciones empíricas sobre la interacción entre IA y empresa familiar en México y América Latina representa una brecha relevante que este trabajo pretende contribuir a cerrar.

Desde el punto de vista académico, la investigación sobre transformación digital en empresas familiares presenta un campo todavía fragmentado: abundan estudios sobre adopción tecnológica en pequeñas y medianas empresas en general, pero escasean aquellos que integran las dinámicas familiares, generacionales y culturales propias de este tipo de organizaciones (De Massis y otros., 2018). Los trabajos existentes tienden a enfatizar las barreras para la innovación, resistencia al cambio del fundador, escasez de recursos, aversión al riesgo, sin profundizar en los mecanismos a través de los cuales ciertas empresas familiares logran integrar exitosamente tecnologías disruptivas, manteniendo su identidad y cohesión. Este capítulo se propone explorar precisamente esos mecanismos, prestando especial atención a los factores institucionales, culturales y de liderazgo que median entre la disponibilidad de herramientas de IA y su efectiva incorporación en la lógica organizacional de la empresa familiar.

En términos de estructura, el capítulo se organiza en torno a cuatro grandes bloques. En primer lugar, se presenta el marco teórico que articula las nociones de sistema socio-técnico, destrucción creativa, economía del conocimiento, capacidad dinámica y gobernanza ética aplicadas a la empresa familiar. En segundo lugar, se describe la metodología cualitativa adoptada y los criterios de selección del caso. En tercer lugar, se presentan y discuten los hallazgos, organizados en torno a los impactos de la IA sobre el trabajo, la innovación, la gobernanza y la sucesión. Finalmente, se elaboran conclusiones y se proponen líneas de investigación futura que permitan avanzar en la comprensión de este fenómeno en el contexto mexicano y latinoamericano.

1. Marco teórico

1.1. La empresa familiar como sistema socio-técnico complejo

La empresa familiar puede comprenderse como un sistema socio-técnico complejo en el que interactúan, de forma dinámica, al menos tres subsistemas: el familiar, el empresarial y el patrimonial (Gersick y otros., 1997). El subsistema familiar está atravesado por vínculos afectivos, relaciones de parentesco, narrativas identitarias y expectativas respecto al legado. El subsistema empresarial responde a las exigencias del mercado, la eficiencia, la productividad y la competitividad. El subsistema patrimonial se ocupa de la propiedad, la sucesión y la gestión de la riqueza familiar.

La introducción de IA afecta simultáneamente a estos tres niveles. A nivel empresarial, modifica procesos productivos, logísticos, comerciales y administrativos, al automatizar tareas y mejorar la capacidad de análisis. A nivel familiar, incide en la distribución de roles entre generaciones: la generación joven suele liderar la adopción tecnológica, mientras la generación fundadora aporta experiencia y prudencia (Zellweger, 2017). A nivel patrimonial, la IA suele requerir inversiones en infraestructura digital, aliados tecnológicos y nuevas formas de

financiamiento, lo que puede alterar decisiones sobre reparto de utilidades y control de la propiedad.

Concebir a la empresa familiar como sistema socio-técnico permite entender que la tecnología nunca entra “en vacío”, sino que se inserta en una trama de relaciones, normas e identidades preexistentes. De ahí que una misma herramienta de IA pueda reforzar la cohesión y profesionalización en una empresa, mientras genera conflicto y resistencia en otra con configuraciones culturales distintas.

1.2. Destrucción creativa, continuidad y cambio en contextos familiares

Desde la perspectiva de la destrucción creativa, la innovación tecnológica desestabiliza estructuras productivas y abre paso a nuevas combinaciones económicas (Schumpeter, 1942). La IA, al automatizar procesos cognitivos y decisionales, constituye un ejemplo paradigmático de este tipo de innovación. En la empresa familiar, sin embargo, el cambio tecnológico no solo afecta rutinas productivas, sino también símbolos y figuras de autoridad. Las prácticas y criterios que sostuvieron el éxito de la primera generación pueden verse cuestionados cuando los algoritmos proponen decisiones basadas en patrones de datos que no siempre coinciden con la experiencia acumulada. La tensión no es solo técnica, sino identitaria: se confrontan formas de saber encarnadas en personas con formas de saber codificadas en modelos.

El reto consiste en transformar la destrucción creativa en continuidad transformadora. Esto ocurre cuando la familia empresaria logra reinterpretar la adopción de IA no como sustitución de la experiencia, sino como una herramienta que potencia esa experiencia para afrontar entornos más complejos. Así, el fundador puede validar su intuición al contrastarla con indicadores generados por IA o enriquecer sus decisiones al descubrir patrones que antes permanecían invisibles.

1.3. Economía del conocimiento y capital intelectual familiar

En la economía del conocimiento, la principal fuente de valor es la capacidad de producir, procesar y aplicar conocimiento de manera sistemática (Drucker, 1993), las empresas familiares poseen un importante capital intelectual acumulado en forma de saber técnico, redes de confianza, reputación y conocimiento del entorno local pero gran parte de ese capital permanece en estado tácito.

La IA ofrece oportunidades para transformar ese conocimiento tácito en conocimiento explícito, al sistematizar información que antes quedaba dispersa. Sistemas de gestión del conocimiento, bases de datos de experiencias comerciales, registros detallados de interacciones con clientes y proveedores, y herramientas de analítica avanzada permiten convertir la experiencia en datos reutilizables.

No obstante, este proceso genera ambivalencias. Algunos actores pueden percibir que su valor diferenciador, basado en un “saber hacer” personal, se diluye al ser objetivado en sistemas accesibles a otros. Por ello, es crucial que la implementación de IA reconozca explícitamente la contribución de las personas, articulando el capital intelectual individual con el capital estructural de la empresa familiar.

1.4. Aprendizaje organizacional, cultura y liderazgo intergeneracional

El aprendizaje organizacional se refiere a la capacidad de una organización para revisar y modificar sus modelos mentales, rutinas y prácticas mediante la reflexión sobre la experiencia (Senge, 1990). En la empresa familiar, esta capacidad está condicionada por su cultura y su estructura de poder.

La IA puede convertirse en un catalizador del aprendizaje organizacional al proporcionar información más precisa y oportuna sobre el desempeño, los mercados y los procesos internos. Sin embargo, también puede tensionar creencias arraigadas, al evidenciar ineficiencias o mostrar que ciertas decisiones tradicionales ya no resultan óptimas. La reacción ante estos datos depende del tipo de liderazgo predominante.

En contextos de liderazgo autoritario, los sistemas de IA pueden utilizarse principalmente para reforzar controles y validar decisiones preestablecidas. En cambio, bajo un liderazgo transformacional, la información generada por la IA se discute, se interpreta colectivamente y alimenta procesos de innovación (Bass & Riggio, 2006).

La dimensión intergeneracional es central: la generación fundadora aporta memoria histórica y valores; las generaciones sucesoras suelen aportar competencias digitales y sensibilidad hacia la economía de datos. Estructuras como consejos de familia, comités de innovación o foros intergeneracionales permiten articular estas perspectivas y construir visiones compartidas sobre el papel de la IA en el proyecto familiar.

1.5. Capacidades dinámicas y ventaja competitiva en la era de la IA

La teoría de las capacidades dinámicas, desarrollada por Teece, Pisano y Shuen (1997), propone que la ventaja competitiva sostenible en entornos turbulentos no reside en los recursos estáticos de la empresa, sino en su habilidad para integrar, construir y reconfigurar competencias internas y externas en respuesta a cambios acelerados del entorno. Esta perspectiva resulta particularmente relevante para comprender cómo las empresas familiares mexicanas pueden afrontar la irrupción de la IA: no se trata simplemente de adquirir tecnología, sino de desarrollar la capacidad organizacional para aprender de ella, adaptarla a sus procesos y transformar sus rutinas productivas y directivas.

En la empresa familiar, el desarrollo de capacidades dinámicas vinculadas a la IA implica tres dimensiones interrelacionadas. La primera es la capacidad de detección (sensing): la habilidad para identificar oportunidades y amenazas tecnológicas antes de que se vuelvan críticas, lo que exige acceso a redes de información sobre tendencias digitales, relaciones con proveedores tecnológicos y participación en ecosistemas de innovación sectoriales. La segunda es la capacidad de aprovechamiento (seizing): la disposición y competencia para tomar decisiones de inversión, adaptar procesos y gestionar el cambio organizacional que

supone la introducción de nuevas herramientas de IA, incluyendo la formación del personal, el rediseño de flujos de trabajo y la gestión de resistencias internas. La tercera es la capacidad de transformación (transforming): la renovación continua de los activos y rutinas de la empresa a medida que las tecnologías evolucionan y los entornos competitivos se reconfiguran (Teece, 2007).

Para la empresa familiar, el desarrollo de estas capacidades dinámicas vinculadas a la IA presenta ventajas y limitaciones específicas. Entre las ventajas se destaca la agilidad decisional: la concentración de la propiedad y el mando en pocas personas puede facilitar decisiones rápidas de adopción tecnológica cuando el líder está convencido de su valor estratégico. Además, el capital social de la familia empresaria, sus redes de proveedores, clientes y socios construidas a lo largo de generaciones, puede funcionar como puente hacia ecosistemas de innovación a los que de otro modo sería difícil acceder. Entre las limitaciones, destaca la propensión al paternalismo tecnológico: el riesgo de que el fundador o líder familiar concentre las decisiones sobre IA sin involucrar adecuadamente a equipos técnicos, lo que puede derivar en implementaciones sub idóneas o en resistencias no gestionadas (De Massis y otros., 2018).

1.6. Gobernanza tecnológica y ética algorítmica en la empresa familiar

La gobernanza tecnológica se refiere al conjunto de estructuras, procesos y normas mediante los cuales una organización toma decisiones sobre la adopción, el uso y la supervisión de tecnologías digitales, asegurando que estén alineadas con sus objetivos estratégicos y sus valores (Van Alstyne y otros., 2016). En el contexto de la empresa familiar, la gobernanza tecnológica adquiere una dimensión adicional: no solo se trata de gestionar el riesgo técnico u operativo, sino de preservar la coherencia entre la lógica de la familia y la lógica del algoritmo. Esta tensión se vuelve especialmente aguda cuando los sistemas de IA producen recomendaciones que contradicen la experiencia acumulada o los valores del fundador.

La ética algorítmica se ocupa de las implicaciones morales del diseño, entrenamiento y uso de sistemas de IA: quién define los objetivos del algoritmo, qué datos se utilizan para entrenarlo, cómo se manejan los sesgos intrínsecos y cuáles son los efectos sobre los trabajadores, los clientes y la comunidad (Floridi, 2019; O'Neil, 2016). Para la empresa familiar, estas preguntas tienen connotaciones específicas: una empresa que ha construido su reputación durante décadas sobre la base del trato cercano, la lealtad y el compromiso con su comunidad local puede ver esa imagen amenazada si adopta prácticas algorítmicas que sus grupos de interés perciben como opacas, injustas o deshumanizadas. La gestión ética de la IA no es, por tanto, una cuestión meramente técnica, sino una dimensión estratégica vinculada directamente a la reputación y la legitimidad de la organización familiar.

Un enfoque prometedor para articular gobernanza tecnológica y ética algorítmica en la empresa familiar es el concepto de IA responsable (Responsible AI), que propone principios como la transparencia, la explicabilidad, la equidad, la privacidad, la seguridad y la rendición de cuentas como guías para el diseño e implementación de sistemas de IA (European Commission, 2021). Trasladar estos principios al contexto de la empresa familiar implica crear mecanismos concretos de supervisión: comités o consejos de gobernanza digital que revisen periódicamente el uso de los algoritmos, establecer criterios claros sobre qué decisiones pueden ser delegadas a sistemas automáticos y cuáles deben permanecer bajo control humano, y diseñar protocolos de comunicación que expliquen a trabajadores y clientes cómo se utilizan sus datos y qué salvaguardas existen para proteger su privacidad.

1.7. Identidad organizacional, capital socioemocional e innovación en la empresa familiar

La identidad organizacional hace referencia a las creencias compartidas sobre lo que una organización es: sus valores centrales, su misión, la narrativa sobre su origen y su trayectoria histórica (Albert & Whetten, 1985). En la empresa familiar, la

identidad organizacional y la identidad familiar se entrelazan de manera especialmente estrecha: el nombre de la familia, la historia del fundador y los valores transmitidos de generación en generación conforman el núcleo identitario de la organización. Esta fusión puede constituir una fortaleza, proporciona cohesión, lealtad y reputación, pero también puede convertirse en un obstáculo para la innovación cuando los cambios tecnológicos son percibidos como una amenaza a esa identidad consolidada (Berrone y otros., 2012).

El concepto de capital socioemocional (socioemotional Wealth, SEW), propuesto por Gómez-Mejía y otros. (2007), resulta especialmente útil para explicar las decisiones de innovación en empresas familiares. El SEW engloba el conjunto de beneficios no económicos que la familia obtiene de la empresa: identidad, estatus, influencia, perpetuación del legado y satisfacción emocional derivada del control de la organización. Según esta teoría, las familias empresarias tomarán decisiones orientadas primariamente a preservar su capital socioemocional, incluso cuando ello implique sacrificar rentabilidad o crecimiento económico. En el contexto de la adopción de IA, esto significa que una familia puede rechazar determinadas herramientas tecnológicas si las percibe como una amenaza a su control sobre la empresa o a los valores que definen su identidad, aunque esas herramientas ofrezcan claras ventajas competitivas.

Sin embargo, la relación entre capital socioemocional e innovación no es necesariamente negativa. Investigaciones recientes sugieren que las empresas familiares con un SEW elevado pueden ser más innovadoras en áreas donde la innovación refuerza la reputación familiar o la posición de liderazgo en el sector (Carney, 2005). En el caso de la IA, esto se traduciría en una disposición mayor a invertir en tecnologías que se presenten como instrumentos para perpetuar el legado familiar, mejorar la calidad del producto o servicio históricamente asociado al nombre de la familia, o fortalecer la posición de liderazgo en el mercado local o regional. La narrativa sobre la adopción de IA, si se enmarca como modernización al servicio de los valores familiares o como ruptura con ellos, resulta, por tanto, un factor determinante en la velocidad y profundidad de la transformación digital.

1.8. Síntesis del marco teórico: un modelo integrador

Los enfoques teóricos revisados en las secciones precedentes configuran, en conjunto, un modelo integrador para comprender la adopción de IA en la empresa familiar mexicana. Dicho modelo articula tres niveles de análisis interdependientes. El nivel macro sitúa el fenómeno en el marco de la cuarta revolución industrial, la economía del conocimiento y la destrucción creativa, que configuran las presiones del entorno que impulsan a las organizaciones hacia la transformación digital. El nivel meso se ocupa de la empresa familiar como sistema socio-técnico, con sus capacidades dinámicas, sus estructuras de gobernanza y sus mecanismos de aprendizaje organizacional. El nivel micro atiende a los actores, sus identidades, sus patrimonios socioemocionales y las tensiones intergeneracionales que modulan la interpretación y la apropiación concreta de las herramientas de IA.

La articulación de estos tres niveles permite superar tanto el determinismo tecnológico (que concibe la IA como una fuerza externa que impacta de forma uniforme sobre todas las organizaciones) como el voluntarismo estratégico, que reduce la adopción tecnológica a una decisión racional del líder empresarial. En cambio, el modelo propuesto enfatiza la mediación de factores institucionales, culturales e intergeneracionales que producen trayectorias de adopción de IA singulares e irreductibles en cada empresa familiar. Esta perspectiva abre espacios para la agencia: aunque el entorno impone presiones hacia la digitalización, cada familia empresaria puede y debe construir su propio camino, alineando la transformación tecnológica con sus valores, su historia y su proyecto de continuidad generacional.

2. Metodología

2.1. Enfoque y propósito

El estudio se inscribe en un enfoque cualitativo-descriptivo e interpretativo, adecuado para abordar fenómenos complejos donde confluyen dimensiones

técnicas, culturales y relacionales (Yin, 2018). Más que medir el grado de adopción de IA en términos estadísticos, el propósito es comprender cómo se integra esta tecnología en la lógica institucional de la empresa familiar mexicana y qué efectos produce en el trabajo, la innovación y la continuidad generacional.

Este enfoque permite analizar significados, tensiones y procesos a partir de un diálogo constante entre teoría y evidencia. En lugar de formular hipótesis verificables en términos cuantitativos, se construyen categorías interpretativas que ayudan a comprender la experiencia de la empresa familiar frente a la transformación digital.

2.2. Diseño de estudio de caso único instrumental

Se adopta un diseño de estudio de caso único instrumental (Yin, 2018). El caso seleccionado no busca ser estadísticamente representativo del conjunto de empresas familiares mexicanas, sino ofrecer una ventana analítica privilegiada para observar la interacción entre IA, trabajo e innovación en un contexto real.

El estudio de caso instrumental permite utilizar la experiencia de una organización concreta como medio para profundizar en la comprensión de categorías teóricas, tales como aprendizaje organizacional, liderazgo intergeneracional y gobernanza tecnológica. Este enfoque resulta adecuado cuando el fenómeno es reciente, poco explorado y fuertemente situado, como ocurre con la adopción de IA en familias empresarias latinoamericanas.

2.3. Criterios de selección del caso

La selección del caso se realizó mediante muestreo intencional, siguiendo criterios de pertinencia teórica y riqueza informativa (Denzin, 2017). Los criterios fueron:

1. Empresa familiar mexicana con trayectoria de varias décadas, en la que la familia fundadora mantiene un papel relevante en la propiedad y el gobierno (Gersick y otros., 1997).

2. Relevancia sectorial, preferentemente en un sector clave para la economía nacional, como el alimentario.
3. Evidencia pública de una estrategia de transformación digital que incluya herramientas de IA o analítica avanzada en áreas como logística, producción, mercadotecnia o relación con clientes.
4. Disponibilidad de información documental suficiente (informes, reportes, entrevistas, notas de prensa) para reconstruir la trayectoria tecnológica y su articulación con la lógica familiar.

El caso seleccionado cumple con estas condiciones, lo que permite explorar de manera detallada cómo una empresa familiar consolidada afronta el reto de integrar IA sin perder su identidad.

2.4. Fuentes de información y procedimiento analítico

Se utilizaron exclusivamente fuentes secundarias:

1. Informes anuales y reportes de sostenibilidad de la empresa.
2. Comunicados institucionales y presentaciones corporativas sobre innovación y digitalización.
3. Entrevistas a directivos y miembros de la familia publicadas en medios especializados.
4. Artículos y reportajes en prensa económica y de negocios sobre la trayectoria del grupo.
5. Literatura especializada sobre empresa familiar, IA, economía del conocimiento y ética tecnológica (p. ej., Brynjolfsson & McAfee, 2014; Floridi, 2019; Le Breton-Miller & Miller, 2018).

El análisis se desarrolló mediante técnicas de análisis de contenido cualitativo (Denzin, 2017). Se identificaron unidades de significado vinculadas con categorías

teóricas (trabajo, innovación, continuidad, gobernanza, cultura) y se elaboraron mapas temáticos que facilitaron la articulación entre evidencia y marco conceptual.

La relación entre teoría y datos fue iterativa: las categorías orientaron la lectura de documentos, y los hallazgos empíricos llevaron a matizar o enriquecer las categorías iniciales. Esta dinámica de ida y vuelta permitió construir una interpretación coherente sobre la adopción de IA en el caso estudiado.

2.5. Validez, fiabilidad y limitaciones

La validez se abordó mediante triangulación de fuentes, contrastando información procedente de documentos corporativos con reportes de prensa y literatura especializada (Denzin, 2017). Aunque no se pretende una validación estadística, esta triangulación contribuye a construir una narrativa plausible y bien sustentada.

La fiabilidad se apoyó en un registro sistemático de decisiones analíticas: criterios de selección, categorías utilizadas, pasos del análisis de contenido. Este registro permite que otros investigadores puedan seguir el razonamiento y evaluar la consistencia del proceso, incluso si trabajan con marcos teóricos distintos.

Entre las principales limitaciones se encuentran la dependencia de fuentes públicas, el posible sesgo positivo de los discursos corporativos y la ausencia de entrevistas directas con actores internos. No obstante, la riqueza descriptiva de los materiales disponibles y la relevancia del caso lo convierten en un insumo valioso para la reflexión académica sobre empresas familiares y IA.

3. Transformación digital en las empresas familiares mexicanas

La transformación digital en las empresas familiares mexicanas avanza de forma heterogénea, condicionada por el tamaño, el sector, el grado de profesionalización y la generación al mando (INEGI, 2024). En empresas de segunda o tercera generación, con sucesores formados en administración, ingeniería o tecnologías de

la información, suele observarse una mayor apertura hacia la digitalización y el uso de datos para la toma de decisiones.

La IA tiende a incorporarse inicialmente en procesos “duros” donde el retorno es más evidenciable, como la gestión de inventarios, la planificación de la producción o la optimización de rutas de distribución. Herramientas de pronóstico de demanda y sistemas de planificación avanzados permiten reducir costos y mejorar el nivel de servicio, algo especialmente crítico en sectores con productos perecederos, como el alimentario.

A medida que la organización gana confianza, la IA se expande a áreas más sensibles como el análisis de clientes, la fijación de precios o la segmentación de mercados. Esta expansión exige revisar prácticas de gobierno corporativo, dado que la toma de decisiones deja de depender exclusivamente de la intuición de uno o dos líderes y se apoya en información compartida por diversos niveles de la organización (Le Breton-Miller & Miller, 2018).

En este contexto, la transformación digital se convierte también en transformación institucional: se fortalecen los órganos de gobierno, se clarifican roles entre familia y empresa, y se profesionalizan procesos antes gestionados de manera informal. La IA, así, no solo cambia tecnologías, sino que reordena la estructura de poder y los mecanismos de coordinación en la empresa familiar. Con el fin de sintetizar los principales efectos organizativos de la inteligencia artificial en la empresa familiar, la Tabla 1 presenta una comparación esquemática entre organizaciones que aún no han integrado IA en su gestión y aquellas que ya la utilizan de manera sistemática en sus procesos clave.

Tabla 1.

Diferencias entre empresa familiar sin IA y con IA

Dimensión	Empresa familiar sin IA	Empresa familiar con IA
Base de decisiones	Predominio de intuición y experiencia personal del fundador	Decisiones apoyadas en datos, indicadores y modelos predictivos
Organización del trabajo	Tareas manuales y rutinarias, fuerte dependencia de personas clave	Automatización de tareas repetitivas, foco en análisis, supervisión y resolución de problemas
Gestión del conocimiento	Conocimiento tácito, poco documentado, concentrado en algunos miembros	Conocimiento sistematizado en bases de datos y sistemas de información
Sucesión generacional	Transición basada en carisma y trayectoria personal	Transición apoyada en resultados, proyectos de innovación y liderazgo digital
Innovación	Cambios incrementales, poco formalizados	Proyectos de innovación apoyados en analítica de datos y seguimiento sistemático

Dimensión	Empresa familiar sin IA	Empresa familiar con IA
Relación con el mercado	Conocimiento del cliente principalmente informal y local	Segmentación y comprensión del cliente basada en analítica avanzada y monitoreo digital
Gestión de riesgos	Evaluación cualitativa, escasa simulación de escenarios	Modelos de escenarios y pronósticos que anticipan variaciones de demanda y operación
Gobernanza y transparencia	Información fragmentada, reportes básicos	Tableros de control, indicadores compartidos y órganos de gobierno mejor informados
Cultura organizacional	Fuerte tradición, posible resistencia al cambio tecnológico	Cultura de aprendizaje continuo, mayor apertura a la experimentación y al cambio
Reputación y ética tecnológica	Reputación basada en historia y trato cercano	Reputación basada en historia más prácticas visibles de uso responsable de datos y algoritmos

Fuente: elaboración propia

Como se aprecia en la Tabla 1, la incorporación de IA no se limita a la automatización de tareas, sino que transforma la base informacional de la toma de decisiones, los perfiles de trabajo, la forma de gestionar el conocimiento y los mecanismos de gobernanza. Esta reconfiguración explica por qué la adopción de

IA en la empresa familiar debe entenderse como un proceso estratégico y cultural, más que como una simple actualización tecnológica puntual.

4. Organización del trabajo y perfiles profesionales

En primer lugar, se produce una redistribución de tareas. Actividades rutinarias como captura de datos, elaboración de reportes estándar, conteo de inventarios o programación básica de producción son asumidas por sistemas automatizados o por módulos inteligentes integrados en plataformas de gestión empresarial. El personal deja de dedicar tiempo a estas tareas repetitivas y pasa a concentrarse en actividades de supervisión, análisis de información, resolución de problemas y atención personalizada a clientes o proveedores.

En segundo lugar, emergen nuevos perfiles laborales vinculados a competencias digitales. Las empresas familiares empiezan a requerir analistas de datos, responsables de comercio electrónico, especialistas en sistemas y coordinadores de proyectos tecnológicos (World Economic Forum, 2023). En algunos casos, estos roles son asumidos por miembros jóvenes de la familia, en otros por profesionales externos que se integran en posiciones clave.

En tercer lugar, cambian las relaciones jerárquicas. Cuando la información está disponible en paneles y sistemas accesibles a diversos niveles, la autoridad basada exclusivamente en el control de la información pierde peso. Mandos medios y equipos de trabajo disponen de más datos para argumentar sus posiciones, lo que puede favorecer formas de liderazgo más participativas, pero también generar conflictos si no existen espacios claros de coordinación.

Finalmente, la IA plantea desafíos relacionados con la calidad del empleo y la equidad. La automatización puede generar miedo entre trabajadores con baja cualificación digital, especialmente en contextos donde el empleo familiar se ha basado en la lealtad más que en la especialización técnica. La manera en que la empresa familiar gestiona esta transición —recapacitación, reubicación, despidos— revela el grado en que sus decisiones tecnológicas son coherentes con

los valores de responsabilidad y cuidado que suele enarbolar (Chrisman y otros., 2005).

Cuando la IA se acompaña de estrategias de formación continua y de una comunicación transparente sobre sus objetivos y efectos, puede convertirse en una herramienta para dignificar el trabajo, ampliar las oportunidades de desarrollo profesional y fortalecer el sentido de pertenencia dentro de la empresa familiar.

5. La economía del conocimiento aplicada a la empresa familiar

La economía del conocimiento demanda transformar la experiencia acumulada en conocimiento sistemático, y la información dispersa en activos estratégicos (Drucker, 1993). En la empresa familiar mexicana, este tránsito implica un cambio profundo desde la gestión “por costumbre” hacia la gestión basada en datos y aprendizaje organizacional.

La IA aporta capacidades para capturar, almacenar, procesar y analizar datos de manera masiva y veloz. Por ejemplo, registros de ventas, hábitos de compra, tiempos de entrega, incidencias de calidad o niveles de merma pueden integrarse en sistemas de inteligencia de negocios que alimentan decisiones sobre portafolio de productos, segmentación de clientes, diseño de promociones y configuración de la cadena de suministro.

No obstante, la simple disponibilidad de datos no garantiza la generación de conocimiento. Es necesario construir rutinas que favorezcan el análisis colectivo de la información, la reflexión sobre los resultados y la incorporación de aprendizajes en las prácticas cotidianas (Senge, 1990). Reuniones intergeneracionales para revisar indicadores, comités de innovación que discutan hallazgos de la analítica y mecanismos para documentar decisiones y sus resultados son ejemplos de cómo la empresa familiar puede institucionalizar el aprendizaje.

Además, la IA interactúa con distintas dimensiones del capital intelectual familiar:

Con el capital humano, al requerir nuevas competencias y facilitar procesos formativos apoyados en simulaciones y análisis de escenarios. Con el capital estructural, al fortalecer sistemas, procedimientos y bases de datos que permanecen en la organización más allá de las personas. Con el capital relacional, al mejorar la comprensión de las necesidades de clientes y proveedores y permitir interacciones más personalizadas. La empresa familiar que logra articular estos elementos puede transformar su experiencia histórica en ventaja competitiva sostenible en la economía del conocimiento.

6. Innovación y continuidad generacional

En la empresa familiar, la innovación no se limita a la introducción de productos nuevos, sino que se vincula con la redefinición del legado a la luz de las condiciones cambiantes del entorno (Le Breton-Miller & Miller, 2018). La continuidad generacional implica, por tanto, una combinación de preservación y cambio: preservar valores, propósito y vínculos; cambiar estrategias, procesos y modelos de negocio.

La IA se convierte en una palanca clave para esta innovación orientada a la continuidad. La generación joven suele ver en la tecnología un espacio para aportar valor y legitimarse como sucesora, mientras la generación fundadora la percibe con una mezcla de curiosidad y cautela. Cuando se construyen espacios de diálogo, la IA puede incorporarse a una narrativa de futuro compartido en la que el proyecto familiar se proyecta hacia nuevas formas de competitividad.

Desde la perspectiva del liderazgo transformacional, las y los líderes que facilitan este proceso son aquellos capaces de articular una visión que integra tradición e innovación, y de inspirar confianza en que la adopción de IA fortalecerá, y no debilitará, la identidad familiar (Bass & Riggio, 2006). Esta visión se traduce en decisiones concretas: invertir en formación digital de la familia y del personal,

institucionalizar órganos de gobierno que incluyan a varias generaciones y vincular los proyectos tecnológicos con metas de largo plazo (p. ej., expansión internacional, sostenibilidad, impacto social).

De este modo, la innovación impulsada por IA no rompe con el legado, sino que lo actualiza, permitiendo que la empresa familiar siga siendo relevante y competitiva para las siguientes generaciones.

7. Estudio de caso: empresa familiar mexicana del sector alimentario

7.1. Origen y evolución familiar

La empresa analizada se originó a inicios del siglo XX como un pequeño negocio familiar dedicado a la producción y comercialización de alimentos procesados. A lo largo de varias décadas, se expandió territorialmente, diversificó su portafolio de productos y consolidó alianzas con marcas internacionales, manteniendo siempre un núcleo de control familiar en la propiedad y en el consejo de administración. La narrativa fundacional destaca valores como esfuerzo, austeridad, calidad y compromiso con el consumidor. Estos valores se han transmitido intergeneracionalmente mediante relatos familiares, prácticas cotidianas y decisiones de negocio, configurando una identidad organizacional sólida que combina orientación al mercado con responsabilidad social.

Con la apertura comercial, la entrada de nuevos competidores y la creciente complejidad regulatoria, la empresa enfrentó la necesidad de modernizar sus procesos, profesionalizar su estructura y fortalecer su capacidad de innovación. La tercera generación familiar, con formación en administración e ingeniería, asumió un rol activo en este proceso, impulsando la transformación digital como estrategia para asegurar la continuidad del proyecto en un entorno más competitivo.

7.2. Estrategia de transformación digital

A partir de la segunda década del siglo XXI, la empresa diseñó una estrategia integral de transformación digital que incluyó:

- Implementación de un sistema de planificación de recursos empresariales (ERP) que integrara información de producción, logística, finanzas y ventas.
- Desarrollo de capacidades de analítica avanzada para pronosticar demanda, optimizar inventarios y planificar la producción con base en escenarios.
- Introducción de herramientas de IA para el análisis de datos de mercado, comportamiento del consumidor y gestión de la cadena de suministro.
- Expansión de canales digitales de venta y comunicación, incluidos comercio electrónico y presencia activa en redes sociales.

Esta estrategia se articuló con cambios en la gobernanza: se creó un comité de innovación y tecnología con participación de miembros de la familia y directivos profesionales; se redefinieron funciones en la alta dirección para incorporar perfiles con experiencia en transformación digital; y se establecieron indicadores específicos para evaluar el avance de los proyectos tecnológicos.

7.3. Aplicaciones de IA en la operación

La empresa desplegó varias aplicaciones concretas de IA:

- Pronóstico de demanda: se construyeron modelos de aprendizaje automático que integran datos históricos de ventas, estacionalidad, promociones y variables macroeconómicas para prever la demanda por producto y región. Esta información permite ajustar la producción, reducir quiebres de stock y minimizar mermas.
- Optimización logística: mediante algoritmos de optimización, se diseñan rutas de distribución que consideran costos de transporte, tiempos de entrega,

restricciones de capacidad y condiciones geográficas. La IA genera escenarios alternativos y sugiere configuraciones óptimas que luego son revisadas por el equipo logístico.

- **Análisis de percepción del consumidor:** herramientas de procesamiento de lenguaje natural monitorean menciones de las marcas de la empresa en redes sociales y plataformas digitales, clasificando comentarios por sentimiento y tema. Esta información alimenta decisiones de marketing, ajustes en el portafolio de productos y diseño de campañas.
- **Mantenimiento predictivo:** sensores instalados en maquinaria crítica envían datos sobre vibración, temperatura y otros parámetros a sistemas de análisis que detectan patrones asociados a fallas. La IA genera alertas tempranas y recomienda intervenciones preventivas, reduciendo tiempos muertos y costos de reparación.

7.4. Impactos en el trabajo y la cultura organizacional

La implementación de estas herramientas ha tenido efectos significativos en la organización del trabajo y en la cultura de la empresa. En logística, el personal pasó de diseñar rutas manualmente a supervisar y ajustar propuestas generadas por los algoritmos, concentrándose más en la coordinación con clientes y en la solución de contingencias. En planta, el acceso en tiempo real a indicadores de producción y mantenimiento incrementó la responsabilidad compartida sobre el desempeño y fomentó una cultura de mejora continua.

Al inicio se registraron resistencias, especialmente entre mandos intermedios habituados a tomar decisiones con base en intuición y experiencia. La dirección respondió con programas de capacitación y espacios de diálogo en los que se explicó el funcionamiento de los sistemas de IA, se aclaró que las decisiones seguirían siendo responsabilidad humana y se enfatizó que la tecnología estaba al servicio de los valores de calidad y servicio al cliente.

Este enfoque contribuyó a que la IA se percibiera gradualmente como aliada más que como amenaza, y a que se fortaleciera una cultura de aprendizaje en la que los errores se interpretan como oportunidades de mejora y no como fracasos individuales.

7.5. Continuidad generacional y proyecto

La familia empresaria integró la transformación digital en su narrativa de continuidad. En documentos estratégicos y reuniones de consejo se explicitó que la adopción de IA formaba parte del compromiso con garantizar que la empresa siguiera siendo competitiva y relevante para las generaciones futuras.

Se establecieron mecanismos para involucrar a miembros jóvenes de la familia en proyectos tecnológicos específicos, ya fuera como coordinadores, miembros de equipos de innovación o participantes en comités de seguimiento. De este modo, la IA se convirtió en un espacio de ejercicio de liderazgo para la nueva generación, bajo la tutela y el acompañamiento de la generación anterior.

Este enfoque permitió que la transformación digital fuera percibida como continuidad evolutiva del legado, y no como ruptura con él. La empresa familiar se concibe a sí misma como una organización centenaria que se renueva de manera permanente, y la IA emerge como pieza clave de esa renovación, al servicio de los valores fundacionales que la distinguen.

8. Desafíos éticos y gobernanza tecnológica

La adopción de IA en la empresa familiar plantea desafíos éticos que se entrelazan con su reputación y su capital relacional. Floridi (2019) destaca que la ética de la información debe abordar problemas como la opacidad algorítmica, el sesgo en los datos, la vigilancia excesiva y la concentración de poder en quienes controlan los sistemas.

En el caso analizado, la empresa ha debido afrontar, entre otros, los siguientes retos:

1. Privacidad de datos: la recopilación de información sobre consumidores y empleados exige políticas claras sobre uso, almacenamiento y acceso a los datos, respetando derechos fundamentales y normativa vigente.
2. Sesgos algorítmicos: los modelos de IA pueden reproducir patrones discriminatorios presentes en los datos de entrenamiento. La empresa ha identificado la necesidad de revisar periódicamente sus algoritmos y de incorporar criterios de equidad en el diseño de modelos.
3. Responsabilidad en decisiones automatizadas: aunque los sistemas generen recomendaciones, la última responsabilidad sobre decisiones de impacto significativo, por ejemplo, asignación de créditos, evaluaciones de desempeño, debe recaer en personas identificables.

Para gestionar estos desafíos, la empresa ha avanzado en la construcción de una gobernanza tecnológica que incluye códigos de conducta, protocolos para el desarrollo y uso de IA, y mecanismos de supervisión y auditoría interna. Esta gobernanza se vincula explícitamente con los valores familiares, reforzando la idea de que la innovación tecnológica debe ser congruente con la ética que ha sustentado la reputación de la organización.

En la medida en que estas prácticas se consoliden, la empresa familiar puede convertir la gestión ética de la IA en un elemento diferenciador de su ventaja competitiva y en un fundamento de su legitimidad social y reputación de largo plazo.

9. Conclusiones

Este capítulo analizó cómo la adopción de inteligencia artificial reconfigura el trabajo, la innovación y la continuidad generacional en la empresa familiar mexicana, articulando un marco teórico que combina la perspectiva de los sistemas

socio-técnicos, la economía del conocimiento, las capacidades dinámicas y el capital socioemocional con la evidencia proporcionada por un estudio de caso del sector alimentario. Los resultados muestran que la IA no opera como un factor meramente técnico, sino como un vector de cambio institucional que modifica la distribución de poder, las rutinas de trabajo y los mecanismos de aprendizaje organizacional, al tiempo que abre oportunidades para la profesionalización y la permanencia de la empresa familiar cuando se integra en una estrategia deliberada de liderazgo intergeneracional y gobernanza ética.

En síntesis, la adopción de inteligencia artificial en la empresa familiar mexicana solo se convierte en una verdadera palanca de continuidad e innovación cuando se articula con el aprendizaje organizacional, el liderazgo intergeneracional y una gobernanza ética coherente con los valores de la familia empresaria. A partir de estos hallazgos, es posible traducir las implicaciones teóricas y empíricas del estudio en orientaciones concretas de acción para las empresas familiares, las cuales se presentan en la sección siguiente como recomendaciones prácticas sobre preparación organizacional, gestión del talento, implementación gradual de la IA y fortalecimiento de la gobernanza y de la responsabilidad socio-tecnológica.

10. Recomendaciones prácticas para empresas familiares mexicanas

Con base en el análisis teórico y empírico desarrollado en este capítulo, se proponen a continuación algunas recomendaciones prácticas para que las empresas familiares mexicanas integren la inteligencia artificial de manera gradual, alineada con su identidad y con su proyecto de continuidad generacional

10.1. Preparar el terreno organizacional

- Definir una visión digital compartida. Formular, desde el consejo de familia y la dirección, una declaración explícita sobre para qué se quiere la IA (mejorar servicio, reducir mermas, profesionalizar la gestión) y cómo se vincula con el legado familiar.

- Diagnosticar capacidades actuales. Identificar procesos con problemas recurrentes (errores, tiempos muertos, sobrecostos) donde la IA pueda aportar valor, así como brechas de competencias digitales en el personal.
- Empezar con proyectos piloto. Iniciar con iniciativas acotadas (por ejemplo, pronóstico de demanda o monitoreo de mantenimiento) que permitan aprender con riesgo controlado antes de escalar la transformación.

10.2. Gestionar el talento y las competencias digitales

- Diseñar un plan de formación intergeneracional. Ofrecer capacitación básica en datos y herramientas digitales tanto a miembros de la familia como al personal clave, evitando que la brecha generacional se convierta en brecha de poder.
- Crear equipos mixtos. Integrar equipos de proyecto donde convivan experiencia operativa de larga data y perfiles jóvenes con formación tecnológica, de modo que la IA se enraíce en el conocimiento real del negocio.
- Reconocer el valor del conocimiento tácito. Asegurar que los expertos “de oficio” participen en el diseño de sistemas, para que los modelos incorporen la realidad del terreno y ellos se perciban como coautores del cambio.

10.3. Implementar IA de forma gradual y alineada al negocio

- Priorizar casos de uso con impacto claro. Concentrarse primero en aplicaciones que mejoren indicadores visibles (mermas, tiempos de entrega, rotación de inventarios), lo que facilita construir credibilidad interna.
- Asegurar que la IA apoye, no reemplace, el juicio humano. Establecer como principio que los modelos sugieren y las personas deciden, especialmente en cuestiones estratégicas o sensibles para el personal.
- Documentar aprendizajes. Registrar resultados, errores y lecciones de cada proyecto de IA para alimentar un repositorio interno de experiencias que oriente futuros proyectos y facilite la transferencia de conocimiento a nuevas generaciones.

10.4. Fortalecer la gobernanza familiar y corporativa

- Separar espacios de decisión. Utilizar consejos de familia para discutir los impactos de la tecnología en el legado y el patrimonio, y órganos de gobierno corporativo para decidir sobre inversiones y prioridades técnicas.
- Incorporar la IA a los órganos de gobierno. Incluir en el consejo o en comités especializados al menos a una persona con comprensión sólida de datos e IA, capaz de traducir sus implicaciones estratégicas y éticas a la familia y a la dirección.
- Establecer políticas tecnológicas claras. Definir por escrito criterios para adquirir, desarrollar y evaluar sistemas de IA, incluyendo responsables, objetivos, indicadores y límites de uso.

10.5. Cuidar la dimensión ética y la reputación

- Elaborar un código de ética digital. Incluir principios sobre protección de datos, transparencia en decisiones automatizadas y uso responsable de algoritmos, vinculados explícitamente con los valores familiares.
- Comunicar con honestidad. Explicar a empleados y clientes qué datos se recaban, con qué propósitos y qué beneficios se esperan, evitando prácticas invasivas o poco claras.
- Revisar impactos sociales y laborales. Evaluar periódicamente cómo los cambios tecnológicos afectan condiciones de trabajo, oportunidades de desarrollo y relaciones con la comunidad, ajustando la estrategia cuando sea necesario.

10.6. Vinculación con el entorno y apoyo externo

- Colaborar con universidades y centros de investigación. Establecer convenios para proyectos conjuntos de analítica o IA aplicada, que permitan acceder a conocimiento especializado sin asumir todos los costos internamente.

- Participar en redes empresariales. Integrarse a cámaras y asociaciones que impulsen la digitalización de pymes y empresas familiares, compartiendo experiencias y buenas prácticas.
- Utilizar apoyos públicos y financieros. Explorar programas gubernamentales y líneas de crédito orientadas a innovación y transformación digital, procurando siempre conservar el control familiar y la coherencia con la estrategia de largo plazo.

Referencias

Bass, B. M., & Riggio, R. E. (2006). *Transformational leadership* (2nd ed.). Psychology Press.

Belausteguigoitia, I. (2010). *Gobierno corporativo en la empresa familiar*. Panorama Editorial.

Berrone, P., Cruz, C., & Gómez-Mejía, L. R. (2012). Socioemotional wealth in family firms: Theoretical dimensions, assessment approaches, and agenda for future research. *Family Business Review*, 25(3), 258–279.

Carney, M. (2005). Corporate governance and competitive advantage in family-controlled firms. *Entrepreneurship Theory and Practice*, 29(3), 249–265.

Chrisman, J. J., Chua, J. H., & Sharma, P. (2005). Trends and directions in the development of a strategic management theory of the family firm. *Entrepreneurship Theory and Practice*, 29(5), 555–576.

Comisión Económica para América Latina y el Caribe. (2022). *Impacto económico de la inteligencia artificial en América Latina: Transformación digital para el desarrollo*. CEPAL.

De Massis, A., Frattini, F., & Lichtenthaler, U. (2018). Research on technological innovation in family firms: Present debates and future directions. *Family Business Review*, 31(2), 117–170.

Denzin, N. K. (2017). *The research act: A theoretical introduction to sociological methods*. Routledge.

Drucker, P. F. (1993). *Post-capitalist society*. HarperBusiness.

European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Publications Office of the European Union.

Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Gersick, K. E., Davis, J. A., McCollom Hampton, M., & Lansberg, I. (1997). *Generation to generation: Life cycles of the family business*. Harvard Business School Press.

Gómez-Mejía, L. R., Haynes, K. T., Núñez-Nickel, M., Jacobson, K. J., & Moyano-Fuentes, J. (2007). Socioemotional wealth and business risks in family-controlled firms: Evidence from Spanish olive oil mills. *Administrative Science Quarterly*, 52(1), 106–137.

Instituto Mexicano para la Competitividad. (2023). *La inteligencia artificial (IA) revolucionará la educación. México no puede quedarse atrás*. IMCO.

Instituto Nacional de Estadística y Geografía. (2024). *Estadísticas de empresas familiares en México*. INEGI.

Le Breton-Miller, I., & Miller, D. (2018). Family businesses and the institutional environment: Theoretical perspectives. *Entrepreneurship Theory and Practice*, 42(2), 231–241.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy*. Harper & Brothers.

Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum.

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. Doubleday.

Sharma, P., Melin, L., & Nordqvist, M. (2012). Theoretical perspectives on family business. Edward Elgar.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.

Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. (2016). Pipelines, platforms, and the new rules of strategy. *Harvard Business Review*, 94(4), 54–62.

World Economic Forum. (2023). The future of jobs report 2023. World Economic Forum.

Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). Sage Publications.

Zellweger, T. M. (2017). Managing the family business: Theory and practice. Edward Elgar.

Albert, S., & Whetten, D. A. (1985). Organizational identity. En L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 7, pp. 263–295). JAI Press.

CAPÍTULO XI

TRANSPARENCIA ALGORITMICA: POR QUÉ EXPLICAR LA INTELIGENCIA ARTIFICIAL ES MÁS IMPORTANTE QUE REGULARLA.

Algorithmic transparency: Why explaining artificial intelligence is more important than regulating it.

Rafael Angeles García

agrafael17@icloud.com

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

0000-0002-9357-4761

Resumen

Este texto propone una reconfiguración de la transparencia algorítmica, superando su enfoque técnico para entenderla como un ecosistema integral de rendición de cuentas. El autor sostiene que la regulación legal tradicional suele nacer obsoleta ante la velocidad del avance tecnológico. Por ello, plantea la necesidad de transitar hacia un Tercer Nivel de Transparencia, centrado en la figura del sujeto-implementador (docentes, funcionarios o directivos).

Bajo este paradigma, la transparencia deja de ser un mero cumplimiento burocrático para convertirse en una obligación ética de usuario a usuario. El implementador debe actuar como un gestor de nodos —según la teoría del conectivismo— encargado de traducir y desmitificar la lógica de la IA ante su comunidad. Esta labor de curaduría y supervisión activa es vital para evitar la "desprofesionalización" y garantizar que la tecnología no actúe como un "oráculo inescrutable" que oculte sesgos o errores bajo una falsa infalibilidad matemática.

Finalmente, el análisis aboga por institucionalizar una "cultura de la inteligibilidad" basada en la costumbre y en políticas transversales internas. Al

convertir la explicación del algoritmo en un hábito cotidiano, las organizaciones aseguran que el juicio humano prevalezca sobre la automatización, protegiendo así la dignidad de las personas y la soberanía del conocimiento en la era digital.

Palabras clave: Transparencia algorítmica, rendición de cuentas, inteligencia artificial, gobernanza digital, sujeto-implementador, conectivismo, inteligibilidad algorítmica, supervisión humana, ética tecnológica, soberanía del conocimiento.

Abstract

This text proposes a reconfiguration of algorithmic transparency, going beyond its technical approach to understand it as a comprehensive ecosystem of accountability. The author argues that traditional legal regulation is usually obsolete in the face of the speed of technological advance. For this reason, it raises the need to move towards a Third Level of Transparency, focused on the figure of the subject-implementer (teachers, officials or directors).

Under this paradigm, transparency ceases to be a mere bureaucratic compliance to become an ethical obligation from user to user. The implementer must act as a node manager – according to the theory of connectivism – in charge of translating and demystifying the logic of AI to its community. This work of curatorship and active supervision is vital to avoid "deprofessionalization" and ensure that technology does not act as an "inscrutable oracle" that hides biases or errors under false mathematical infallibility.

Finally, the analysis advocates institutionalizing a "culture of intelligibility" based on custom and internal transversal policies. By making the algorithm's explanation an everyday habit, organizations ensure that human judgment prevails over automation, thus protecting people's dignity and knowledge sovereignty in the digital age.

Keywords: Algorithmic transparency, accountability, artificial intelligence, digital governance, subject-implementer, connectivism, algorithmic intelligibility, human supervision, technological ethics, knowledge sovereignty.

Para desarticular la profunda complejidad que reviste la integración de la Inteligencia Artificial (IA) en los pilares de la esfera pública y el ámbito educativo contemporáneo, resulta fundamental realizar una disección de la naturaleza misma de la transparencia algorítmica. Este concepto no debe ser interpretado de manera restrictiva o simplista como un atributo puramente técnico del software o una característica aislada de la programación; por el contrario, debe concebirse como un ecosistema integral de rendición de cuentas. Dicho sistema es extenso y complejo, pues abarca desde la estructuración primaria y la selección de los datos que alimentan al sistema, pasando por la lógica interna de sus procesos, hasta llegar al impacto final y tangible que estas herramientas ejercen sobre la dignidad de las personas y el ejercicio de sus derechos. En este sentido, la transparencia deja de ser una meta estática para convertirse en una estructura dinámica y multinivel que debe ser capaz de transitar orgánicamente desde la responsabilidad técnica de los desarrolladores hacia una necesaria traducción social en forma de divulgación científica. El objetivo de este enfoque es garantizar que la tecnología no opere en un vacío ético, sino que sus desarrolladores pero sobre todo sus implementadores sean capaces de comunicar su funcionamiento de forma que la sociedad pueda comprender, auditar y, en última instancia, validar las decisiones que afectan su realidad cotidiana.

El primer nivel: La trazabilidad como Imperativo ético y normativo

En esta dimensión primaria, la transparencia algorítmica se erige como el principio rector fundamental que mandata a los desarrolladores, proveedores tecnológicos y entidades gubernamentales a garantizar una apertura absoluta sobre el marco estructural, las lógicas internas y los criterios de ponderación que rigen sus sistemas. Este nivel inicial se fundamenta en la trazabilidad, entendida no como un

simple registro de eventos, sino como la capacidad metodológica y operativa de auditar el rastro preciso que sigue un algoritmo para transformar una entrada de información en una resolución o predicción con impacto en la vida de las personas. Bajo esta visión, la trazabilidad deja de ser un accesorio de ingeniería o un requisito técnico secundario para convertirse en un límite ético infranqueable que define la validez de cualquier implementación tecnológica.

Como lo establecen con rigor los Principios de Chapultepec, es preciso reconocer que la tecnología bajo ninguna circunstancia puede ser considerada un agente neutro u objetivo; por el contrario, su diseño está profundamente impregnado por las visiones de mundo, los valores y las prioridades de sus creadores. Esta falta de neutralidad obliga a que la transparencia sea la única vía para evitar que los prejuicios humanos se automaticen y se oculten tras una fachada de infalibilidad digital. En este contexto, si un proceso de decisión asistido por inteligencia artificial resulta ininteligible para quienes lo operan o para quienes reciben sus efectos, su automatización debe considerarse ilegítima. El derecho a la explicación surge entonces como una garantía ciudadana inalienable que permite a cualquier individuo afectado reconstruir la lógica del proceso, asegurando que el ejercicio de los derechos fundamentales prevalezca siempre frente a la posible arbitrariedad o el hermetismo tecnológico.

A pesar de los esfuerzos por consolidar marcos normativos que garanticen la rendición de cuentas, la sofisticación alcanzada por los modelos contemporáneos —particularmente aquellos fundamentados en el aprendizaje profundo y en arquitecturas de redes neuronales de alta densidad— plantea un desafío que la regulación tradicional no ha logrado cubrir con suficiencia: la incapacidad de conocer plenamente el proceso interno de razonamiento de la máquina. Para abordar este fenómeno a profundidad, resulta indispensable analizar la transparencia desde una doble perspectiva que armonice la construcción técnica con su impacto social. Este enfoque integral busca minimizar el riesgo del hermetismo sistémico, frecuentemente descrito bajo la metáfora de la "caja negra", en el cual la escala masiva y la complejidad de las operaciones estadísticas superan

la capacidad de comprensión incluso de los propios diseñadores del sistema. El peligro sustancial de esta opacidad radica en que la inteligencia artificial deje de funcionar como una herramienta de apoyo a la decisión humana para transformarse en un oráculo inescrutable, cuyas determinaciones se aceptan sin cuestionamiento bajo una falsa premisa de infalibilidad tecnológica (Rentería & Ordellin Font, 2026).

Desde esta postura analítica, la transparencia de segundo nivel no puede limitarse únicamente a la apertura formal del código, sino que exige a los expertos mantener una gobernanza activa del modelo. Esto implica la responsabilidad técnica y ética de identificar no solo los errores visibles en los resultados finales, sino también los sesgos de representatividad y las distorsiones narrativas que suelen anidarse de forma invisible en los conjuntos de datos utilizados para el entrenamiento de los sistemas. Combatir el hermetismo desde la fase de diseño significa, en última instancia, asegurar que la autonomía y el control sobre la decisión final permanezcan firmemente bajo una supervisión humana crítica. Como se advierte en las investigaciones sobre la adopción de estas tecnologías en el ámbito público mexicano, esta vigilancia es la única defensa efectiva para evitar que la inteligencia artificial actúe como un amplificador de inequidades históricas, disfrazando prejuicios sociales bajo un barniz de objetividad matemática y neutralidad estadística (Rentería & Ordellin Font, 2026).

La transparencia técnica, por muy exhaustiva que se presente desde la óptica del diseño de sistemas, corre el riesgo inminente de perder su valor democrático fundamental si no logra transformarse en una verdadera accesibilidad general. En este punto crítico, el análisis debe romper con el formalismo burocrático y legalista para abrazar una lógica basada en el sentido común, entendida como el puente necesario entre la abstracción matemática y la comprensión humana. No resulta suficiente con el cumplimiento de requisitos mínimos que se limitan a la apertura de repositorios de código o a la publicación de arquitecturas algorítmicas que, a menudo, resultan ininteligibles para el ciudadano promedio; por el contrario, la transparencia auténtica exige una "traducción pedagógica" de la lógica operativa. Esta traducción debe ser capaz de explicar, de forma clara y accesible, cómo las

decisiones de la máquina se entrelazan con los derechos y las expectativas de las personas en el mundo real.

Traducir la transparencia al terreno del sentido común implica el compromiso de proporcionar información que sea no solo técnica, sino significativamente contextual. Si un sistema de inteligencia artificial se despliega en la administración pública mexicana o en organismos académicos con el fin de asignar recursos, evaluar desempeños o gestionar trayectorias de aprendizaje, la explicación del sistema debe ser capaz de responder a la pregunta fundamental del "por qué" en términos humanos y profesionales estrictos. El sentido común dicta que la transparencia no puede ser reducida a una entrega pasiva de datos en un portal digital de difícil acceso; debe constituirse como un diálogo horizontal y constante. Este modelo de apertura es el que permite a la sociedad civil y a las comunidades académicas auditar, cuestionar y, en su caso, impugnar las premisas sobre las cuales se construye y moldea su realidad digital, garantizando que el uso de la tecnología sea un reflejo de los valores sociales y no una imposición del hermetismo tecnológico (Rentería & Ordellin Font, 2026).

El segundo nivel: Hacia una reconceptualización de la opacidad algorítmica

La insuficiencia detectada en los dos niveles previos de transparencia revela una limitación estructural que la regulación contemporánea no ha logrado resolver: aunque el desarrollador comprenda profundamente su código y el ciudadano reciba una explicación técnica simplificada, persiste una brecha crítica en el momento exacto de la implementación cotidiana. Este vacío operativo sugiere que la transparencia algorítmica, cuando se limita exclusivamente a ser un requisito exigido a la empresa proveedora o al fabricante del software, corre el riesgo de degradarse hasta convertirse en lo que podemos denominar un "escudo de cumplimiento". Bajo esta figura, las instituciones y corporaciones pueden legitimar sistemas opacos al ampararse en el cumplimiento formal de la ley, creando una ficción de apertura que no ofrece garantías reales de justicia ni de entendimiento en la práctica laboral o académica. En el contexto de la administración pública

mexicana, este fenómeno es particularmente preocupante, ya que la adopción de herramientas automatizadas sin una supervisión situada puede derivar en una rendición de cuentas fragmentada donde nadie asume la responsabilidad final sobre los efectos del sistema (Rentería & Ordellin Font, 2026).

Este escenario también plantea la necesidad de explorar una dimensión hasta ahora ignorada por la literatura académica tradicional y por los marcos legales estáticos: el papel determinante del usuario que decide implementar la herramienta en entornos específicos y sensibles, como el aula universitaria o la oficina gubernamental y su correlación con un nuevo nivel de transparencia algorítmica aún más importante. Es precisamente en este punto donde el análisis debe evolucionar hacia la propuesta central de este artículo: el establecimiento de un Tercer Nivel de Transparencia. A diferencia de los niveles anteriores, este enfoque no busca su validez únicamente en la norma rígida o en la auditoría externa, sino en la reestructuración de la costumbre como fuente natural del Derecho y en el desarrollo de una ética de la implementación situada. Se trata de reconocer que la transparencia no se agota en el diseño del objeto tecnológico, sino que debe completarse en la interacción humana. Esta propuesta sostiene que la verdadera seguridad tecnológica y la protección de la dignidad de las personas dependen de que el acto de explicar y justificar el uso de un algoritmo se convierta en una práctica social y profesional arraigada, transformando la opacidad técnica en un compromiso de inteligibilidad compartida (Secretaría de Ciencia, Humanidades, Tecnología e Innovación & Agencia de Transformación Digital y Telecomunicaciones, 2026).

El tercer nivel de transparencia: El sujeto-implementador y la reconfiguración conectivista de la responsabilidad

A diferencia de los niveles iniciales de transparencia, que suelen encontrar su límite en el cumplimiento de normas técnicas y requisitos de diseño por parte de las empresas fabricantes —lo que podríamos considerar una transparencia de origen—, el surgimiento de un tercer nivel se vuelve una exigencia ineludible debido a la

forma masiva en que la inteligencia artificial se ha entretelado en la estructura de la vida diaria. Este planteamiento propone un cambio profundo y necesario en el eje de la responsabilidad, desplazando el foco analítico desde la entidad que construye el código hacia la figura determinante del sujeto-implementador. Dicho cambio reconoce que la tecnología no posee un impacto social real mientras permanece en un entorno de laboratorio o como una simple oferta de mercado; su verdadero efecto ético y práctico se manifiesta únicamente en el instante en que un ser humano decide utilizarla para mediar, influir o decidir sobre la realidad de otros. Así, el sujeto-implementador se convierte en el mediador crítico: es el docente frente a su grupo, el directivo que coordina un equipo de trabajo o el servidor público que gestiona recursos desde el Estado. En manos de estos actores, la transparencia deja de ser un expediente técnico o un manual de uso para transformarse en un proceso de acompañamiento humano y pedagógico, donde la validez de la herramienta no reside en su sofisticación matemática, sino en la claridad con la que se pone al servicio de la comunidad.

Bajo este nuevo paradigma, la transparencia se consolida como una obligación directa y ética de usuario a usuario. No resulta aceptable bajo este modelo que quien posee la autoridad para introducir una inteligencia artificial en un proceso productivo, administrativo o educativo, ignore el deber ineludible de explicar las lógicas de funcionamiento a quienes ejecutan las tareas o reciben las consecuencias de dichas decisiones automatizadas. Omitir esta labor de comunicación y guía constituye una falta de responsabilidad que vulnera la autonomía de las personas y convierte la interacción profesional en un entorno de opacidad e incertidumbre. Por lo tanto, el tercer nivel exige que el implementador asuma el papel de un facilitador que desmitifica la tecnología, despojándola de cualquier aura de infalibilidad y exponiendo sin ambigüedades cómo el sistema filtra la información, cuáles son sus posibles sesgos de origen y cuáles son los márgenes de error de su razonamiento estadístico. Solo a través de esta labor constante de traducción y acompañamiento situado, se puede asegurar que el uso de la tecnología avance en concordancia con la protección de la dignidad humana y la

búsqueda de entornos más equitativos, evitando que la innovación se convierta en una herramienta de exclusión oculta tras un velo de eficiencia tecnológica.

El Conectivismo y la Gestión de Nodos

Para profundizar en esta transición fundamental hacia un modelo de responsabilidad compartida, resulta indispensable acudir a los fundamentos del conectivismo desarrollados por George Siemens (2004), los cuales ofrecen el marco teórico necesario para comprender cómo se construye el aprendizaje y cómo se gestiona el conocimiento en la complejidad de la era digital. En una realidad donde el flujo de información crece de manera exponencial y la vigencia del conocimiento se ha reducido drásticamente a lapsos de apenas meses o incluso semanas, la productividad y el desarrollo intelectual ya no pueden depender de la simple acumulación individual de datos. Por el contrario, dependen de la capacidad crítica para conectar, supervisar y validar diversos puntos de información especializados dentro de una red. Bajo esta mirada, la inteligencia artificial deja de ser percibida como un simple instrumento de oficina o un accesorio técnico para ser reconocida como un nodo no humano de procesamiento autónomo que se integra profundamente en la ecología de las organizaciones y las instituciones educativas. Por lo tanto, si el sujeto encargado de la implementación —ya sea un docente frente a su grupo, un directivo en una empresa o un servidor público— decide incorporar este nodo de inteligencia artificial a su red de trabajo, asume de inmediato la responsabilidad ética de garantizar que la arquitectura de dicha conexión sea transparente. De lo contrario, se permite de forma negligente que el flujo de conocimiento se contamine con lógicas inexplicables que comprometen la integridad de los resultados y la veracidad de la información que circula en el entorno institucional.

En esta nueva ecología informativa, la metáfora de Siemens sobre la primacía de la infraestructura adquiere una relevancia jurídica y social definitiva: en el entorno actual, la tubería es abismalmente más importante que el contenido que transporta. Traducido al ámbito de la transparencia algorítmica de tercer nivel, esto significa

que el acto de comprender bajo qué criterios y a través de qué procesos se filtra la información mediante la inteligencia artificial es mucho más determinante para la autonomía profesional que el resultado final que entrega la máquina. El sujeto que implementa la herramienta asume, por tanto, el rol fundamental de gestor de nodos, cuya obligación principal no se limita a proveer tecnología a sus subordinados o alumnos, sino a velar por la salud intelectual y la claridad operativa de la red que dirige. Si un líder de equipo o un profesor introduce un sistema tecnológico de naturaleza hermética sin ofrecer la explicación pedagógica u organizacional correspondiente, está sabotando activamente la capacidad de su entorno para aprender y actuar de forma crítica. Esta omisión no debe considerarse un error menor, sino que constituye una forma de opacidad que convierte el proceso laboral o educativo en una caja cerrada desde su base, donde los participantes pierden el control y el entendimiento sobre las premisas que guían sus propias actividades profesionales.

Desde una perspectiva de análisis profundo, el abandono de este tercer nivel de transparencia provoca una erosión preocupante de la capacidad humana para generar juicios propios. Como ha advertido Marc Prensky (2010), la delegación ciega de tareas intelectuales complejas a sistemas automatizados, sin una comprensión mínima de sus parámetros de funcionamiento, deriva inevitablemente en un fenómeno de desprofesionalización o pérdida de pericia. Nos enfrentamos a un escenario crítico donde los funcionarios, académicos o estudiantes corren el riesgo de perder la facultad de justificar las bases lógicas de sus propias resoluciones, volviéndose dependientes de una herramienta cuyos sesgos y márgenes de error desconocen por completo. Sin el ejercicio metódico de la transparencia de usuario a usuario, el entorno de trabajo se transforma en un espacio de asimetría radical donde la información fluye sin sentido. En este contexto, la inteligencia artificial no actúa como un facilitador de la innovación, sino como una máscara que oculta vacíos éticos y técnicos bajo el disfraz de entregables aparentemente perfectos, pero que carecen de un verdadero sustento humano. Como señalan los Principios de Chapultepec (2026), el uso ético de la tecnología

exige que la gestión de los datos y los algoritmos sea un bien público cuidado con responsabilidad, asegurando que la innovación no se convierta en una herramienta de exclusión oculta tras un velo de eficiencia estadística.

Para contrarrestar esta deriva hacia el hermetismo, este análisis propone que el tercer nivel de transparencia exige que el sujeto que implementa la tecnología defina y comunique de manera explícita el grado de autonomía compartida que ha otorgado al sistema. No todas las integraciones tecnológicas tienen el mismo impacto social, y por ello la transparencia debe ajustarse a la función específica que desempeña el algoritmo en cada proceso. En una primera instancia, cuando la inteligencia artificial opera únicamente como un gestor de rutinas mecánicas, la explicación debe centrarse en la fidelidad y la seguridad de la automatización. En un segundo plano, si el sistema actúa como un colaborador que asiste en procesos creativos o de lluvia de ideas, la transparencia debe exponer cómo la máquina influye en la narrativa y en la construcción del pensamiento del usuario. Finalmente, si la herramienta se utiliza como un consultor avanzado cuyas recomendaciones sustentan decisiones críticas para la vida pública o académica, el implementador tiene la obligación inalienable de desglosar exhaustivamente las variables y ponderaciones que utiliza el sistema. Solo a través de esta clasificación clara y este acompañamiento situado se puede garantizar que la tecnología funcione verdaderamente como un puente hacia la innovación ética y no como un obstáculo que nuble el juicio profesional o la dignidad de las personas involucradas en la red.

El análisis tradicional sobre la gobernanza de la inteligencia artificial ha quedado, con frecuencia, atrapado en una visión puramente legalista y estática que resulta insuficiente ante la complejidad del fenómeno. Existe una creencia generalizada de que, para controlar una tecnología de vanguardia, basta con que los órganos legislativos nacionales e internacionales redacten marcos normativos extensos o que los gobiernos emitan reglamentos oficiales cargados de sanciones punitivas. Sin embargo, esta postura ignora una realidad fundamental de la era digital: la ley escrita, por su propia naturaleza burocrática y pausada, es estructuralmente incapaz de seguirle el ritmo a una tecnología que experimenta transformaciones

gigantes cada pocos meses. El proceso de creación de una ley —desde su propuesta hasta su publicación y entrada en vigor— suele tomar años, lo que garantiza que, para el momento en que la norma está lista, la herramienta que pretendía regular ya opera bajo una lógica técnica completamente distinta. Esta brecha insalvable entre la velocidad de la política pública y la velocidad del avance tecnológico genera lo que podemos definir como una parálisis por regulación, donde la norma no solo llega tarde, sino que nace obsoleta, perdiendo su capacidad de guiar el comportamiento social de manera efectiva.

Cuando se intenta imponer desde una jerarquía vertical una serie de reglas rígidas y estandarizadas a un ecosistema compuesto por usuarios, académicos y funcionarios que todavía están en proceso de asimilación de estas herramientas, el resultado natural no es el aumento de la seguridad, sino el rechazo operativo. Las personas en sus frentes de trabajo, ya sea en las aulas universitarias o en las dependencias de la administración pública, perciben estas legislaciones como obstáculos burocráticos ajenos que no resuelven los dilemas éticos y técnicos que enfrentan en su interacción diaria con la máquina. Esta desconexión profunda entre la letra de la norma y la trinchera de la realidad laboral termina convirtiendo a la legislación en "letra muerta": un documento de cumplimiento formal que se archiva para satisfacer auditorías externas, pero que no altera de ninguna manera la opacidad de los procesos internos. En este sentido, la regulación que realmente impacta en la protección de los derechos no puede emanar exclusivamente de un escritorio gubernamental distante, sino que debe encontrar su legitimidad y su fuerza en la costumbre y en la repetición de prácticas saludables por parte de quienes implementan la tecnología.

La propuesta de este análisis sostiene que la regulación del tercer nivel de transparencia algorítmica debe fundamentarse en una reestructuración consciente de la costumbre dentro de las instituciones. En la teoría clásica del derecho, la costumbre es reconocida como una fuente de reglas mucho más noble, orgánica y adaptable que la ley impuesta por un soberano, ya que nace del convencimiento colectivo sobre la necesidad de actuar de una forma determinada. Al ser la

integración de la inteligencia artificial un territorio relativamente nuevo y lleno de dilemas éticos menores pero constantes, la costumbre permite corregir y moldear las conductas de forma natural, sin detener el flujo de la innovación. Reestructurar la costumbre implica sentar las bases culturales para que el acto de explicar el funcionamiento de un algoritmo deje de ser visto como una interrupción molesta del flujo de trabajo y se convierta en una conducta inherente, esperada y necesaria en cualquier relación profesional o pedagógica. En lugar de depender de una vigilancia externa que rara vez comprende el contexto específico del aula o de la oficina pública, las organizaciones deben apostar por el desarrollo de políticas transversales internas que reflejen sus propias necesidades operativas y consensos éticos.

Estas políticas transversales —manifestadas en manuales de buenas prácticas, guías de ética compartida y comités de revisión de proyectos— constituyen un cuerpo normativo vivo que se adapta en tiempo real a la evolución de los nodos de información que gestionamos. Cuando un docente o un líder de proyecto asume como un hábito cotidiano la tarea de desglosar las variables y criterios de la inteligencia artificial ante su equipo, está sentando un precedente de transparencia que tiene mucha más fuerza vinculante que una sanción legal lejana. Este enfoque se alinea con los Principios de Chapultepec (2026), los cuales defienden que el gobierno de la tecnología es más seguro cuando las decisiones sobre su uso se toman de forma colectiva y horizontal. Al promover institucionalmente que la lectura del algoritmo sea una costumbre administrativa y docente, aseguramos que la responsabilidad no se diluya en la complejidad del código. En última instancia, regular desde la práctica social garantiza que cualquier decisión automatizada pueda ser reconstruida cognitivamente por los actores involucrados, erradicando el peligro de una sociedad alienada donde el juicio profesional quede oculto tras el velo de un sistema que nadie se toma el tiempo de explicar.

Para que esta reestructuración de la costumbre no se mantenga como una simple aspiración teórica o una idea abstracta desvinculada de la práctica, su consolidación definitiva debe materializarse a través de lo que definimos como

políticas transversales internas. Este enfoque normativo no busca suplantar a las leyes nacionales o a los tratados internacionales —los cuales conservan su relevancia para establecer límites éticos macro y proteger los derechos humanos fundamentales—, sino que propone la creación de acuerdos situados dentro de cada organización que operen como guías dinámicas para el comportamiento de todos sus integrantes. Hablamos de la implementación de instrumentos concretos como manuales de buenas prácticas, decálogos de ética compartida y comités de revisión, donde los propios usuarios tengan la facultad de analizar y cuestionar la viabilidad de los proyectos tecnológicos antes de su puesta en marcha. La virtud de estas políticas radica en su capacidad para actuar en la micro-realidad de las instituciones, ofreciendo una respuesta ágil y contextualmente relevante allí donde la legislación estatal, por su propia naturaleza burocrática y generalista, suele llegar tarde o de forma ineficiente.

La ventaja estratégica de estas políticas transversales reside en que nacen directamente de la realidad operativa de los mismos individuos que gestionan los nodos de información. Cuando una universidad, un organismo público o un corporativo privado asume la tarea de diseñar sus propias reglas de transparencia algorítmica, dichas normas dejan de percibirse como una imposición coercitiva externa para ser comprendidas como un consenso ético destinado a proteger la calidad del trabajo y la integridad del aprendizaje. Al establecer formalmente que el docente o el líder de proyecto posee un deber ineludible de acompañamiento, se garantiza que la responsabilidad sobre el uso de la inteligencia artificial no se diluya en la complejidad del sistema técnico. En este sentido, la transparencia se transforma de un requisito estático a un diálogo constante y significativo; se aleja de la simple entrega de documentos técnicos o códigos ininteligibles que, con frecuencia, solo se publican para simular que se cumple con la ley. El objetivo es erradicar la "ficción de cumplimiento" y sustituirla por una cultura de la inteligibilidad, donde el uso de la tecnología sea siempre un acto consciente y justificado.

Bajo este modelo de gobernanza interna, la legitimidad de la inteligencia artificial se construye mediante la repetición de conductas que priorizan la agencia humana

sobre la automatización ciega. Al integrar estos procesos de revisión y explicación en la rutina diaria, se asegura que cualquier decisión asistida por una máquina pueda ser reconstruida cognitivamente por los actores involucrados, evitando que la pericia profesional sea reemplazada por una dependencia tecnológica opaca. Estas políticas internas funcionan, en última instancia, como el tejido que une la teoría del conectivismo con la práctica jurídica; permiten que la red de conocimiento institucional se mantenga saludable y abierta, asegurando que los nodos no humanos —los algoritmos— operen bajo la supervisión de una comunidad que entiende sus alcances y limitaciones. Solo a través de esta institucionalización de la costumbre, donde la claridad operativa sea la norma y no la excepción, podremos asegurar que el avance tecnológico en México y en el mundo sea un facilitador del progreso social que no comprometa la autonomía de los ciudadanos ni el juicio crítico de los profesionales.

El rol del implementador en la validación de la verdad (en la época de la posverdad)

Si bien he establecido que el sujeto-implementador tiene el deber ineludible de abrir la lógica del algoritmo ante su comunidad, esta labor de visibilidad resulta estéril si no se acompaña de un proceso activo de verificación de la veracidad. En la ecología del conectivismo planteada por Siemens (2004), la red de aprendizaje y trabajo solo es valiosa en la medida en que los nodos que la componen aportan información verificada y útil para el crecimiento del ecosistema. Por lo tanto, el sujeto-implementador no pueden limitarse a ser simples espectadores de los resultados que arroja la inteligencia artificial. Su función debe evolucionar necesariamente hacia la de un curador especializado, encargado de filtrar, contrastar y validar las posibles distorsiones o falsedades estadísticas que estos sistemas suelen generar debido a su propia naturaleza probabilística. Esta labor de curaduría asegura que la "tubería" informativa de la que hablaba Siemens no solo sea transparente en su estructura, sino que transporte un contenido que respete la integridad intelectual de quienes lo reciben, evitando que la fascinación por la velocidad tecnológica sacrifique el rigor de la verdad.

El sujeto-implementador, al asumir esta función de curaduría, se convierte en el último filtro de seguridad y ética entre la máquina y el ciudadano o el estudiante. Se trata de una vigilancia consciente que busca proteger la calidad del pensamiento humano frente a la estandarización informativa. Ignorar esta tarea de validación significaría aceptar una transparencia superficial: una transparencia que explica el camino pero permite la circulación de resultados contaminados o sesgados. En este sentido, la curaduría es el acto que dota de propósito real al tercer nivel de transparencia; es el momento en que el implementador deja de ser un mero facilitador técnico para transformarse en un garante de la calidad democrática y académica en el entorno digital.

Como preámbulo a las conclusiones finales, es preciso enfatizar que esta capacidad de curaduría es lo que define la verdadera pericia profesional en el siglo XXI. Al ejercer un juicio crítico sobre lo que la inteligencia artificial produce, el sujeto-implementador reafirma su agencia y autonomía frente al sistema. Este compromiso con la verdad y con la calidad del conocimiento es lo que permite que la tecnología no sea percibida como una entidad que sustituye el discernimiento humano, sino como una herramienta que, bajo la supervisión de una conciencia ética y profesional, potencia nuestra capacidad para navegar en un mundo de información sobreabundante. Solo a través de esta validación constante, integrada en la práctica diaria y convertida en una nueva costumbre organizacional, podremos asegurar que la inteligencia artificial funcione verdaderamente como un puente hacia la innovación que no comprometa, bajo ninguna circunstancia, la dignidad de las personas ni la salud intelectual de nuestras instituciones.

De hecho, las experiencias más enriquecedoras y transformadoras en la interacción entre el ser humano y la inteligencia artificial no ocurren mediante la simple entrega de resultados, sino en el espacio dinámico que se genera cuando estos ejercicios de transparencia son llevados a la práctica supervisada. Se ha observado que la integración tecnológica alcanza su máximo potencial ético y formativo cuando un líder organizacional o un docente no se limita a autorizar el uso de la herramienta, sino que motiva activamente a sus equipos o alumnos a utilizarla bajo un esquema

de vigilancia crítica y acompañamiento constante. Este modelo de supervisión activa permite trascender la mera operatividad para convertir el uso de la inteligencia artificial en un ejercicio dialógico de alta complejidad. En este entorno, la figura del implementador cobra un relieve fundamental: al estar presente durante el proceso de interacción, el docente o jefe puede guiar al usuario en el señalamiento preciso de lo que en el ámbito técnico se conoce como alucinaciones o distorsiones informativas, transformando el error estadístico en una oportunidad inigualable para el aprendizaje y el fortalecimiento del juicio crítico.

Este ejercicio de supervisión que constituye el cimiento pedagógico del uso ético de la tecnología. Al fomentar un espacio donde el líder es capaz de identificar inconsistencias, señalar los sesgos narrativos en tiempo real y, simultáneamente, aplaudir los aciertos y las síntesis logradas por la máquina, se está construyendo una nueva cultura de la responsabilidad. Este tipo de dinámicas promueve una relación de colaboración donde la inteligencia artificial es sometida al escrutinio humano de forma habitual y constructiva. De este modo, la transparencia de tercer nivel deja de ser un concepto teórico para materializarse en una conducta profesional y académica: se enseña que la validez de un nodo no humano —la IA— depende siempre de la verificación del nodo humano que lo dirige. Esta práctica cotidiana de supervisión es la forma más efectiva de evitar la desprofesionalización, pues asegura que el personal o los estudiantes desarrollen la pericia necesaria para discernir entre la información útil y el ruido estadístico, recuperando la autonomía y la dignidad en el flujo de trabajo digital.

Esta reestructuración de la conducta profesional implica un cambio en las competencias individuales del implementador a la par que exige una transformación profunda en la identidad misma de nuestras instituciones: el tránsito definitivo de la vieja "cultura del expediente" hacia una renovada "cultura de la inteligibilidad". Históricamente, en el ámbito de la administración pública y en las estructuras académicas, la transparencia ha sido interpretada de forma reduccionista como un ejercicio de cumplimiento documental, donde la apertura se agota en el acto de publicar informes, códigos fuente o manuales técnicos que, aunque cumplen

formalmente con los requisitos legales, resultan inalcanzables para el entendimiento ciudadano o estudiantil. Este modelo de transparencia basada en la simulación técnica genera una ilusión de control que, en la práctica, solo sirve para proteger la opacidad de los sistemas. Frente a esta inercia, la propuesta de un tercer nivel de transparencia algorítmica demanda que las instituciones abandonen la idea de que ser transparentes es "entregar datos" para asumir el compromiso de "generar comprensión". La inteligibilidad institucional se convierte entonces en el estándar de legitimidad del siglo XXI: una organización ya no es transparente porque permite ver su información, sino porque tiene la capacidad y la disposición de explicar los criterios humanos y estadísticos que sustentan sus decisiones automatizadas, garantizando que el uso de la tecnología no se convierta en un refugio para evitar la rendición de cuentas.

La resignificación del concepto de transparencia se presenta como una urgencia impostergable ante el avance de la inteligencia artificial, exigiendo una evolución profunda que supere las nociones tradicionales centradas casi exclusivamente en la rendición de cuentas financiera o el ejercicio de recursos públicos. Históricamente, las organizaciones han comprendido la transparencia como un mecanismo de vigilancia sobre el flujo del capital y la legalidad administrativa; sin embargo, en la era de la automatización masiva, este enfoque resulta insuficiente para salvaguardar la integridad de las estructuras sociales. El nuevo imperativo de la transparencia debe orientarse hacia la preservación de un orden humano dentro de procesos que, en un futuro cercano, serán ejecutados íntegramente por sistemas computacionales. Esta transición implica reconocer que la delegación de decisiones críticas a los algoritmos no es un simple cambio de herramienta, sino un desplazamiento de la lógica del poder que requiere nuevos puntos de anclaje. La transparencia algorítmica debe entenderse, por tanto, como la facultad de mantener la brújula ética y el control consciente sobre el sentido de nuestras instituciones, asegurando que la eficiencia técnica no termine por devorar la capacidad de justificar el porqué de cada acto que impacta en la vida ciudadana.

Bajo esta nueva arquitectura conceptual, el riesgo que conlleva mantener procesos bajo la opacidad se vuelve drásticamente mayor al que cualquier institución ha enfrentado previamente. Mientras que la falta de transparencia administrativa suele derivar en ineficiencias o actos de corrupción localizados, el hermetismo en los sistemas de inteligencia artificial amenaza con provocar una alienación sistémica, donde el orden social deje de responder a voluntades humanas para someterse a patrones estadísticos inescrutables. Estamos ante la inminencia de una autonomía tecnológica total en las líneas de producción, en la gestión del talento y en la administración del Estado, lo que convierte a la inteligibilidad del proceso en la última frontera de la soberanía del conocimiento. Si las organizaciones no asumen que transparentar el algoritmo es la única forma de garantizar que la inteligencia artificial funcione como una extensión de nuestras capacidades y no como un reemplazo de nuestra conciencia, el vacío de entendimiento generado por la máquina terminará por desarticular la confianza que sostiene el tejido social. La transparencia resignificada es el compromiso de que ninguna decisión automatizada quede desvinculada de un marco de interpretación humano, recuperando la pericia profesional y la dignidad institucional frente a la velocidad de una tecnología que pronto realizará la totalidad de las tareas operativas.

Esta visión propositiva exige que la transparencia deje de ser percibida como un informe estático de cumplimiento legal para consolidarse como una vigilancia activa de la racionalidad de los sistemas. Mantener un orden humano en procesos perfectamente automatizados requiere que el sujeto encargado de la implementación sea capaz de desglosar la lógica interna de la máquina, transformando el oscurantismo técnico en claridad pedagógica. La gravedad de la opacidad en este contexto radica en su capacidad para ocultar sesgos estructurales y errores de razonamiento bajo una apariencia de objetividad matemática, lo cual es abismalmente más peligroso que la simple falta de información sobre un presupuesto. Al redefinir la transparencia como la garantía de que el ser humano sigue siendo el arquitecto del sentido en sus propios procesos, las instituciones mexicanas y globales pueden transitar hacia una modernización que potencie la

innovación sin renunciar a la responsabilidad ética. El éxito de la convivencia entre humanos y máquinas dependerá de esta capacidad de explicar, supervisar e intervenir en la tubería informativa, asegurando que el avance tecnológico sea siempre un facilitador de la justicia y nunca un refugio para la arbitrariedad automatizada.

Este cambio de paradigma encuentra su mayor resistencia en la costumbre burocrática de ocultar la responsabilidad tras la complejidad técnica. Sin embargo, es precisamente en el entorno digital donde la reestructuración de esa costumbre cobra una urgencia democrática. Cuando una universidad o una dependencia pública integra políticas transversales internas que obligan al diálogo sobre el algoritmo, está desmantelando la jerarquía del hermetismo. Al institucionalizar la práctica de explicar el "por qué" de una recomendación de la inteligencia artificial, se fomenta un entorno donde la confianza no se deposita ciegamente en la infalibilidad de la máquina, sino en la solidez del proceso humano que la supervisa. Esta transición hacia la inteligibilidad es, en esencia, lo que permite que la tecnología sea asimilada de forma orgánica por la comunidad, transformando el miedo a la automatización en un ejercicio de soberanía del conocimiento. De este modo, la protección de los derechos de los ciudadanos y los estudiantes deja de depender exclusivamente de una legislación rígida y distante, para sostenerse sobre una red de confianza construida a través del hábito diario de la claridad operativa. En última instancia, la cultura de la inteligibilidad es la única defensa efectiva contra un futuro donde el juicio profesional quede sepultado bajo el peso de sistemas inescrutables, permitiendo recuperar el sentido humano de la justicia y la educación en la era de la inteligencia artificial.

Conclusiones

A lo largo de este análisis, se ha demostrado que la transparencia algorítmica no puede seguir siendo interpretada como un requisito técnico estático que se agota en el diseño del código o en el cumplimiento burocrático de las empresas desarrolladoras. La conclusión fundamental de este artículo es que la efectividad

ética de la inteligencia artificial reside en la capacidad humana de traducir su complejidad al terreno de lo social y lo cotidiano. Al desplazar el eje de la responsabilidad hacia el sujeto-implementador, se reconoce que el verdadero impacto de la tecnología ocurre en el punto de contacto entre quien dirige una actividad y quien la ejecuta. Esta figura, ya sea en el aula o en la oficina pública, se convierte en el garante de un conocimiento que no debe ser alienante, sino facilitador. Así, entendemos la transparencia en tres niveles distintos, y es el tercer nivel que aquí se propone una condición de posibilidad para que la innovación tecnológica avance sin sacrificar la autonomía de las personas ni la integridad de los procesos institucionales.

Con el conectivismo por su parte, la integración de la inteligencia artificial debe ser entendida como la incorporación de un nodo de procesamiento masivo a nuestra ecología de conocimiento, cuya validez depende enteramente de la salud y la claridad de la red que lo sostiene. Por ello, la labor de curaduría y supervisión activa del implementador no es una tarea secundaria, sino el ejercicio mismo de la pericia profesional en la era digital. Se concluye que las mejores experiencias de interacción entre humanos y sistemas automatizados son aquellas donde el error y la distorsión estadística se utilizan como herramientas pedagógicas para fortalecer el juicio crítico. El silencio o el hermetismo frente al funcionamiento del algoritmo no solo genera desconfianza, sino que provoca una desprofesionalización progresiva donde el ser humano pierde la facultad de defender sus propias decisiones, delegando su voluntad a un sistema inescrutable. La transparencia de tercer nivel es, por tanto, la única vía para asegurar que la inteligencia artificial potencie nuestras capacidades en lugar de reemplazarlas por una dependencia opaca.

Finalmente, este estudio sostiene que la gobernanza de las tecnologías disruptivas en México y en el mundo no debe quedar supeditada exclusivamente a una legislación rígida y pausada que, por su propia naturaleza, suele nacer obsoleta. La verdadera regulación, la más legítima y efectiva, es la que emana de la reestructuración de una fuente tan noble, antigua y flexible como la costumbre en conjunto con la creación de una cultura de la inteligibilidad institucional. Al convertir

el acto de explicar y justificar el uso de la tecnología en un hábito organizacional diario, las instituciones dejan de simular que cumplen con la ley para empezar a proteger de forma real la dignidad y los derechos de su comunidad. La soberanía del conocimiento y la seguridad tecnológica se construirán en el consenso horizontal de los usuarios que deciden dominar sus herramientas en lugar de ser dominados por ellas. Así, el futuro de la justicia y la educación en un mundo automatizado dependerá de nuestra capacidad para garantizar que, detrás de cada proceso asistido por una máquina, prevalezca siempre una conciencia humana capaz de dar cuenta de sus actos.

BIBLIOGRAFÍA

Instituto para el Futuro de la Educación. (2024). *Inteligencia Artificial Generativa en la Educación Superior: Una Visión Objetiva*. IFE Insights Reports. Tecnológico de Monterrey.

Prensky, M. (2010). *Nativos e Inmigrantes Digitales*. Cuadernos SEK 2.0. Institución Educativa SEK. (Original publicado en 2001).

Rentería Marín, C., & Ordellin Font, J. L. (2026). *Inteligencia artificial en el ámbito público mexicano: Dónde estamos y hacia dónde vamos*. Facultad Latinoamericana de Ciencias Sociales (FLACSO México).

Secretaría de Ciencia, Humanidades, Tecnología e Innovación, & Agencia de Transformación Digital y Telecomunicaciones. (2026, 29 de enero). *Principios de Chapultepec: Declaración de ética y buenas prácticas para el uso y desarrollo de la Inteligencia Artificial*. Gobierno de México.

Siemens, G. (2004). *Conectivismo: Una teoría de aprendizaje para la era digital* (D. E. Leal Fonseca, Trad., 2007). elearnspace.org.

CAPÍTULO XII

LA INTELIGENCIA ARTIFICIAL Y LA AUTORÍA PEDAGÓGICA: EL LIBRO DE TEXTO HUMANIZADO EN LA EDUCACIÓN MEDIA SUPERIOR

Artificial Intelligence and Pedagogical Authorship: The Humanized Textbook
in Upper Secondary Education

César Augusto Arellano Rodríguez

cesar1arellano@gmail.com

Estudiante de la Tercera generación del D.I.O.

<https://orcid.org/0009-0008-6811-2734>

Ivanna Arellano Reyes

ivannaarellanor@gmail.com

Universidad Nacional Autónoma de México

<https://orcid.org/0009-0005-3523-9737>

Resumen

Este texto analiza el papel de la inteligencia artificial en la elaboración de libros de texto para la Educación Media Superior. Se sostiene que la IA puede ser una herramienta útil para apoyar tareas de búsqueda, organización, revisión y contraste de información, pero no debe sustituir la autoría pedagógica humana. El argumento central plantea que el libro de texto no es un simple contenedor de información, sino una mediación curricular, pedagógica, cultural y docente que participa en la formación de los estudiantes y en la organización de la enseñanza. Desde esta perspectiva, se advierte que la producción algorítmica de materiales puede debilitar la sensibilidad didáctica, la responsabilidad autoral y la comprensión del aula real. El texto también considera las condiciones juveniles, sociales y desiguales de la Educación Media Superior, así como las presiones del mercado editorial, la racionalización productiva y la necesidad de establecer criterios éticos,

pedagógicos y editoriales sobre el uso de IA. Finalmente, se defiende el libro de texto humanizado como una construcción que requiere experiencia docente, juicio pedagógico, cuidado del lenguaje y responsabilidad frente a los efectos formativos del material escolar

Palabras clave: Inteligencia artificial; autoría pedagógica; libro de texto; Educación Media Superior; experiencia docente; mediación pedagógica; ética educativa.

Abstract

This text analyzes the role of artificial intelligence in the development of textbooks for Upper Secondary Education. It argues that AI can be a useful tool for supporting information search, organization, revision, and comparison tasks, but it should not replace human pedagogical authorship. The central argument states that a textbook is not merely a container of information, but a curricular, pedagogical, cultural, and teaching mediation that participates in students' formation and in the organization of teaching practice. From this perspective, the text warns that algorithmic production of educational materials may weaken didactic sensitivity, authorial responsibility, and the understanding of the real classroom. It also considers the youthful, social, and unequal conditions of Upper Secondary Education, as well as the pressures of the publishing market, productive rationalization, and the need to establish ethical, pedagogical, and editorial criteria for the use of AI. Finally, it defends the humanized textbook as a construction that requires teaching experience, pedagogical judgment, careful language use, and responsibility for the formative effects of school materials.

Keywords: Artificial intelligence; pedagogical authorship; textbook; Upper Secondary Education; teaching experience; pedagogical mediation; educational ethics

La educación frente a la aceleración social y tecnológica

El mundo ha cambiado. Con frecuencia se dice que la sociedad avanza a una velocidad cada vez mayor. Existe una tendencia a confundir la velocidad con la inteligencia, la productividad con la satisfacción y lo tecnológico con la legitimidad. La educación también ha sido arrastrada por esta lógica de aceleración, pues en lugar de detenerse a reflexionar sobre el rumbo que está tomando la sociedad, parece empeñada en seguirle el paso, aun cuando ese esfuerzo pueda dejar atrás procesos, aprendizajes y personas que requieren otro ritmo.

Cada vez que aparece una herramienta capaz de agilizar búsquedas, resumir información, redactar párrafos o crear materiales con aparente coherencia, el ser humano se coloca frente a artefactos que le permiten acelerar sus procesos para intentar estar a la altura de ese ritmo. No obstante, también surge la posibilidad de reducir esfuerzos y facilitar actividades mediante el uso de estas nuevas herramientas. La Inteligencia Artificial (IA) se ubica precisamente en ese punto: puede apoyar tareas humanas, pero también puede desplazar procesos que requieren criterio, experiencia y responsabilidad.

La IA es el resultado de un desarrollo tecnológico de varias décadas y aún no se sabe con certeza hasta dónde llegará. Con ella se han automatizado tareas administrativas, se gestionan datos en investigaciones, se escriben textos académicos, se crean actividades didácticas, se planean clases, se diseñan evaluaciones y se producen materiales escolares. En este contexto, la pregunta no es solamente si la IA puede crear textos correctos, sino si resulta adecuado utilizarla para producir materiales educativos sin alterar el sentido formativo de la educación. La modernidad tecnológica parece haber avanzado más rápido que la modernidad contestataria, como ya advertía Wallerstein (1995).

Para la Educación Media Superior (EMS) se requiere un análisis particular. No se trata únicamente de indagar si la IA puede crear textos gramaticalmente correctos o reunir con rapidez varios documentos para producir un escrito único. Eso ya está comprobado: puede hacerlo de manera rápida y, en muchos casos, eficiente. La

cuestión central es otra: no debería asignarse a la IA la escritura de libros de texto para la EMS, pues este nivel educativo puede ser especialmente vulnerable a la producción algorítmica de materiales.

Una herramienta no puede asumir por sí sola la redacción de materiales dirigidos a jóvenes con experiencias de vida desiguales, situados en contextos heterogéneos y atravesados por tensiones culturales, afectivas y sociales propias de su desarrollo. La IA puede desempeñar un papel útil como apoyo en la investigación, la organización de materiales y el contraste de información; sin embargo, no puede ocupar el lugar de un autor en la creación de libros de texto para la EMS. Cuando deja de ser un recurso auxiliar y se convierte en una instancia central de escritura, surgen riesgos técnicos, pedagógicos, formativos y éticos.

Para crear un libro de texto con IA basta con elaborar un *prompt*; incluso puede pedírsele a la propia IA que lo diseñe. Con ese *prompt* se pueden indicar las características del texto, las directrices que debe seguir, el tipo de lenguaje que debe emplear, la forma de escritura e incluso la sensación que se busca generar en el lector. Sin embargo, aun con todo ello, el resultado carece de humanidad, porque esa capacidad técnica no equivale a la experiencia humana. Aunque la IA pueda imitar una intención comunicativa o producir un texto aparentemente adecuado, no posee sensibilidad propia ni comprensión real del contexto formativo de los estudiantes.

Este trabajo propone defender que los libros de texto para la EMS deben seguir siendo escritos por seres humanos con sensibilidad docente, responsabilidad ética y conocimiento de lo que sucede en el aula. No basta con que estén bien escritos o sean técnicamente correctos; deben surgir de una experiencia docente real y de una comprensión concreta de las condiciones en que se forma el alumnado. La IA puede asistir, pero no puede humanizar un texto; puede acelerar, pero no puede asumir el juicio pedagógico; puede producir sintaxis, pero no puede prever si sus palabras tendrán un efecto formativo en los lectores. Esta postura no representa un rechazo al cambio, sino el resguardo de la experiencia humana, que sigue siendo

insustituible cuando lo que está en juego no es únicamente la creación de contenidos educativos, sino la formación de personas.

La IA como apoyo y no como sustitución de la autoría pedagógica

La discusión no debe formularse como un rechazo simplista a la tecnología. Se reconocen los aportes de la IA en la creación de materiales, su pertinencia en algunas revisiones, su utilidad para ordenar información y su apoyo para mantener ciertos objetivos educativos, siempre que estas acciones sean conducidas y revisadas por autores humanos. La IA puede ampliar capacidades, pero no debe suplantarlas. En este sentido, Quiñónez Salcido (2025) sostiene que la IA es una “herramienta simbólica orientada al desarrollo de capacidades humanas y no un sustituto de la inteligencia humana” (p. 446).

Quiñónez Salcido (2025) también afirma que “la IA educativa debe estar guiada por principios éticos, donde el docente mantenga su rol de mediador y la tecnología actúe como un catalizador del pensamiento reflexivo” (p. 446). No obstante, considerar al docente únicamente como mediador de la IA puede restarle importancia a su dimensión creativa, pues lo desplaza hacia un papel de juez o modificador de contenidos, en lugar de reconocerlo como autor y constructor del sentido pedagógico. Tal como se ilustra en la figura 1.

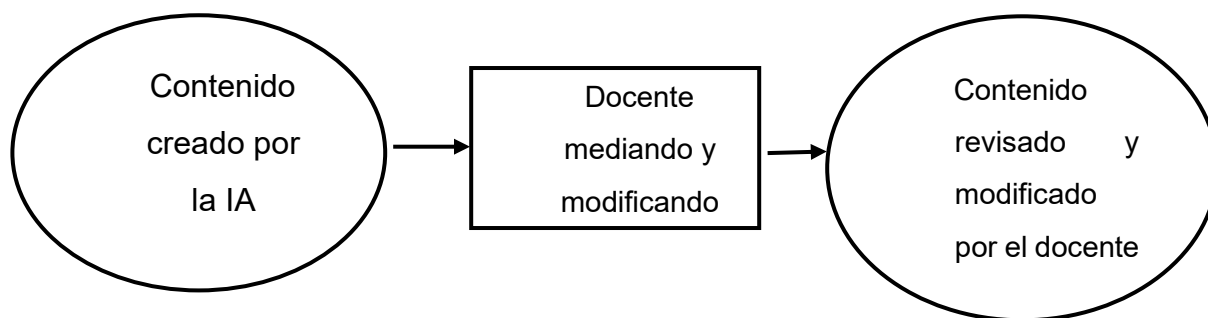


Figura 1. El papel del docente como mediador del contenido creado por la IA.

El docente es quien debe crear el contenido y, si así lo decide, puede apoyarse en la IA como herramienta. En ese sentido, debe ser mediador de sus propias creaciones, no únicamente de contenidos producidos por una tecnología. La

responsabilidad de crear, orientar y dar sentido pedagógico al contenido debe permanecer en el docente. Como señala Arellano Rodríguez (2025), el valor de la IA en la educación “no será reemplazar al docente, sino ampliar el horizonte del proceso formativo” (p. 765). Como se aprecia en la Figura 2.

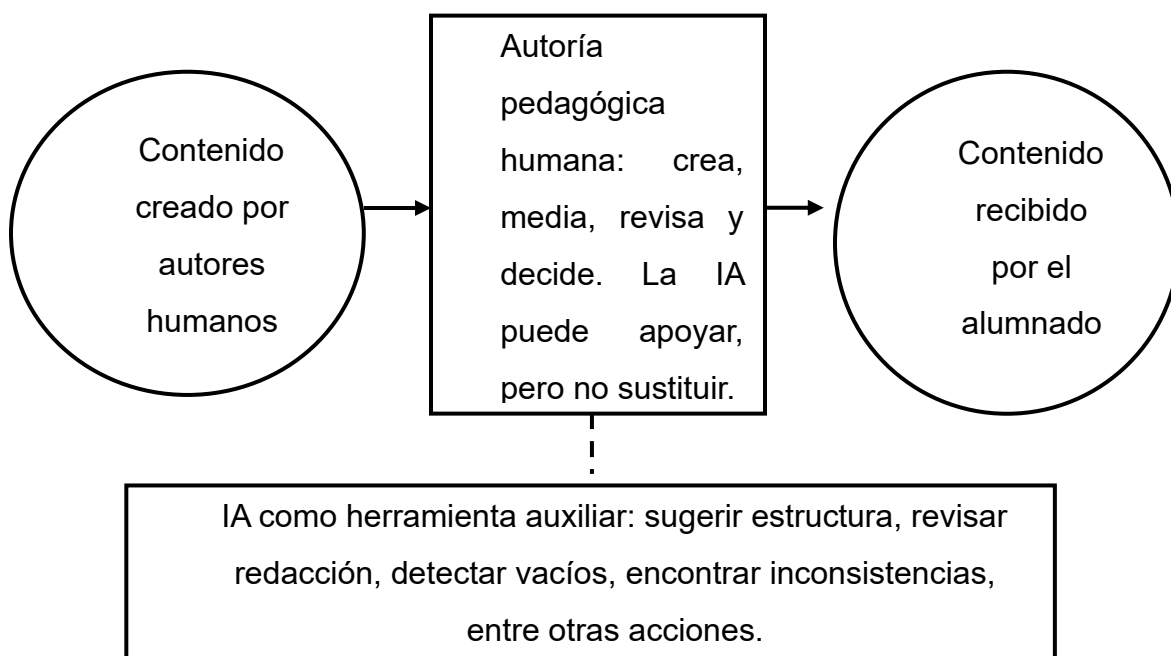


Figura 2. La IA como apoyo y no como sustitución de la autoría docente.

Por su carga pedagógica y formativa, la discusión sobre la IA en la elaboración de materiales escolares no puede realizarse por ocurrencia ni desde posiciones extremas. Una parte del debate público sobre la IA en educación se expresa en dos posturas simplificadoras: por un lado, la aceptación acrítica de sus aportaciones; por otro, el rechazo tecnófobo. Ambas posiciones reducen el problema. Por eso, resulta necesario observar el uso de la IA desde dos lógicas distintas: una lógica sustitutiva y una lógica de apoyo.

La lógica sustitutiva aparece cuando la IA se concibe como reemplazo del autor-docente y no como apoyo a su trabajo. En este caso, la herramienta deja de ser un medio para convertirse en el centro del proceso. Se utiliza para abreviar procedimientos, reducir tiempos, abaratar costos, simular escritura académica y

ensamblar materiales con una intervención humana mínima. Bajo esta lógica, el libro se concibe sobre todo como un producto de información, mientras que la autoría pedagógica humana es tratada como un obstáculo para la eficiencia o como una fase residual del proceso. El costo, aunque muchas veces no se reconozca, es la pérdida de sensibilidad didáctica, de lectura situada del destinatario y de responsabilidad intelectual sobre los efectos del texto.

La lógica de apoyo, en cambio, mantiene el centro del proceso en la autoría pedagógica humana. En este caso, la IA puede servir para documentar, ordenar, explorar variantes, sugerir secuencias, detectar vacíos, proponer reorganizaciones y acelerar tareas repetitivas, pero siempre bajo la conducción sustantiva de un autor con experiencia docente. No se piensa como fuente de criterio, sino como un recurso técnico subordinado a una racionalidad pedagógica y ética. La diferencia no es menor: en la lógica de apoyo, la IA no decide la voz del libro, no cierra la versión final ni reemplaza el juicio sobre qué es enseñable, qué conviene graduar, qué ejemplo debe retirarse o cómo debe nombrarse un problema sensible para no producir exclusión.

Negar la asistencia de la tecnología en las diversas actividades humanas, desde una postura extrema, equivale a negar una parte del desarrollo histórico de la sociedad. La tecnología ha ampliado capacidades, acelerado procesos y facilitado múltiples tareas en prácticamente todos los ámbitos de la vida humana. Sin embargo, reconocer su utilidad no implica aceptarla como centro rector de las acciones humanas. En el caso de los libros de texto, la IA no debe ocupar el lugar del autor ni desplazar la responsabilidad creativa, intelectual y pedagógica que corresponde al ser humano.

Debe rechazarse la idea de que la IA produzca el contenido y que la persona quede reducida únicamente a mediar con la editorial o, en un nivel todavía más problemático, a fungir como intermediaria entre la máquina y los alumnos que recibirán los materiales. Lo preocupante de esa lógica es que el ser humano se desplaza a sí mismo del espacio creador, reflexivo y formativo para ocupar un papel

secundario, correctivo y subordinado. Deja de ser autor para convertirse en revisor de un producto que no hizo; deja de ser constructor del sentido pedagógico para limitarse a ajustar, contener o suavizar aquello que la herramienta tecnológica ya generó.

Cuando el ser humano abandona el núcleo creativo y la IA ocupa su lugar, no solo cambia el procedimiento de elaboración del libro; también se altera la naturaleza misma de la autoría. El riesgo no consiste únicamente en que el material pueda presentar errores o limitaciones, sino en que el proceso formativo quede cada vez más alejado de la experiencia humana que debería sostenerlo. Un libro de texto elaborado bajo esa lógica puede conservar una apariencia de orden, claridad e incluso funcionalidad, pero pierde una parte esencial de su humanidad. Esto ocurre porque el criterio que lo origina ya no nace de quien enseña, escribe y comprende a los estudiantes, sino de un sistema incapaz de habitar el aula y de asumir la responsabilidad formativa de sus palabras.

La EMS en un contexto juvenil, desigual y heterogéneo

En la Educación Media Superior esta exigencia adquiere mayor relevancia, ya que se trata de un nivel marcado por trayectorias, contextos y experiencias estudiantiles profundamente diversas. Por ello, sus materiales no deben producirse desde una lógica impersonal, fría o genérica, sino desde una comprensión concreta de las condiciones reales en las que aprenden los estudiantes y de lo que ocurre en las distintas aulas.

Para Villa Lever (2010), la Educación Media Superior “constituye una construcción social que no tiene que ver solamente con los distintos planes de estudio, sino también con las trayectorias escolares que orientan a diplomas diferentes —cuyo reconocimiento es distinto y diferenciado pues tienen grados de aceptación diferentes en el mercado de trabajo—, regidos por distintas categorías pues suponen una inserción social específica de acuerdo con las diferencias de currículo, lo que conduce a la obtención de un diploma heterogéneo de ems, dando como resultado la segmentación del nivel: cada modalidad tiene su manera de acreditar

los estudios, los estudiantes que se inscriben en cada tipo de ems tienen orígenes socioculturales distintos y las historias escolares particulares y sus posibilidades de futuro están estrechamente relacionadas a la modalidad estudiada” (pp. 291-292).

El empleo de la IA en la creación de materiales educativos puede aumentar ciertos problemas ya presentes en la EMS, como la fragmentación, la falta de articulación entre programas y la producción de contenidos rígidos o desconectados. En este sentido, puede favorecer la “desarticulación, dispersión y heterogeneidad de modalidades y carreras, y el predominio de programas académicos inconexos y rígidos, así como regulaciones excesivas” (Székely Pardo, 2010, p. 314).

En la EMS, el estudiante no es una figura neutra; es un joven situado, con una relación concreta con su escuela, el lenguaje, su entorno, el cansancio, la desigualdad, sus posibilidades de futuro, sus problemas cotidianos, sus complicaciones internas y externas, sus confusiones, sus dudas, sus alegrías y sus tristezas. Por ello, el libro de texto debe ser capaz de tomar en serio esa realidad y responder de manera rigurosa no solo a los contenidos exigidos por la autoridad educativa, sino también al proceso formativo del alumnado.

La condición juvenil de los estudiantes de la EMS vuelve todavía más delicado el problema. Feixa (1999) define las culturas juveniles como las formas en que las experiencias sociales de los jóvenes “son expresadas colectivamente mediante la construcción de estilos de vida distintivos localizados fundamentalmente en el tiempo libre, o en espacios intersticiales de la vida institucional” (p. 1). Además, recuerda que los jóvenes suelen tener “escaso control sobre la mayor parte de aspectos decisivos en su vida” (p. 1), al permanecer bajo la tutela de instituciones adultas.

Un libro de texto dirigido a este sector etario no entra en un vacío; entra en una escena cargada de conflicto, negociación cultural, búsqueda de identidad y sensibilidad frente al reconocimiento o al desprecio. Escribir para jóvenes no es lo mismo que escribir para lectores adultos o para niños: implica tomar en serio sus formas de apropiación, resistencias, fragilidades, humores grupales, modas,

gustos, temporalidades y mediaciones lingüísticas. Estos aspectos difícilmente pueden ser captados por una escritura producida solo desde instrucciones técnicas.

No se trata únicamente de edad biológica, sino de una posición social vinculada con la experiencia escolar. En la EMS convergen adolescencia tardía, primera juventud, presiones de rendimiento, incertidumbres sobre el futuro, desigualdades familiares y trayectorias marcadas muchas veces por el abandono, el trabajo y la continuidad incierta de los estudios. Por eso, el libro de texto no puede ser un objeto indiferente. Debe reconocer al estudiante como un sujeto capaz de pensar, comprender y participar en su formación. No puede dirigirse a él desde un registro distante, frío, impersonal o excesivamente tecnológico, porque ello puede alejarlo simbólicamente del proceso formativo.

Tampoco puede dejarse de lado que el proceso educativo ocurre dentro de una estructura social desigual. La idea de ofrecer la misma educación a poblaciones profundamente disímiles puede resultar pedagógicamente engañosa si no se reconocen las diferencias de contexto, capital cultural y trayectorias disponibles. Incluso una educación aparentemente meritocrática puede reforzar comportamientos específicos de clase y contribuir a la reproducción de la desigualdad social. Por ello, el libro de texto de EMS necesita escribirse desde la conciencia de la pluralidad de sus lectores. Los materiales pensados para sociedades homogéneas —que, en realidad, no existen— pueden ampliar la distancia entre el diseño formal del currículum y la vida escolar que efectivamente experimentan los actores del Sistema Educativo Nacional.

La Educación Media Superior debe entenderse como uno de los niveles formativos que requiere especial cuidado, debido a los contextos y características de su alumnado. Las preguntas sobre la pertinencia de los libros de texto no pueden reducirse a si transmiten contenidos, favorecen la comprensión de los temas, problematizan o facilitan el aprendizaje mediante su uso. También es necesario preguntarse si reconocen la condición humana de quienes los leen; si logran establecer un diálogo con el lector para acompañarlo en su etapa formativa; si

fortalecen la capacidad de reconocerse a sí mismos y a los demás como sujetos sociales y humanos; y si inciden en su manera de pensar y de sentir.

Mercado editorial, reformas y racionalización de los libros de texto

Resulta conveniente realizar una distinción fundamental dentro del Sistema Educativo Nacional: no todos los libros de texto circulan bajo la misma lógica institucional ni todos son gratuitos. En la educación básica, los libros de texto gratuitos forman parte de una política pública histórica de homologación de contenidos, impulsada por el Estado a través de la Comisión Nacional de los Libros de Texto Gratuitos (CONALITEG). Su surgimiento debe entenderse en el contexto del México de la posguerra y de la Guerra Fría, particularmente durante el gobierno del entonces presidente Adolfo López Mateos.

La creación de la CONALITEG en 1959 respondió, entre otras razones, a la necesidad de unificar contenidos, ampliar el acceso a los materiales escolares, construir una identidad nacional, reforzar la orientación ideológica de ese momento histórico y afianzar la presencia del Estado en la formación de las nuevas generaciones. Desde entonces, la gratuidad de los libros en la educación básica no solo ha representado un apoyo económico para las familias, sino también una vía política, cultural e institucional para mantener cierta homogeneidad en la enseñanza desde el poder público.

La Educación Media Superior, en cambio, no ha sido incorporada de la misma manera a esa tradición de gratuidad. Sus libros de texto no forman parte de la CONALITEG. Por ello, gran parte de los materiales que circulan en este nivel proviene del mercado editorial. Esto significa que, a diferencia de lo que ocurre en la educación básica, en la EMS los materiales suelen ser ofrecidos por editoriales y seleccionados por escuelas o docentes conforme a distintos criterios: utilidad, extensión, costo y ajuste curricular. La consecuencia es que el libro de texto de media superior no solo se produce desde una lógica pedagógica, sino también desde una lógica mercantil, en la que el número de páginas, el costo final, la

competencia entre materiales, la urgencia por responder a cada reforma educativa y la viabilidad comercial condicionan directamente su construcción.

A lo anterior se suma otro problema de gran relevancia. Las reformas educativas no siempre se presentan con la claridad suficiente para orientar la elaboración de los materiales. En muchas ocasiones, los cambios curriculares, los nuevos enfoques, las categorías de análisis, las progresiones de aprendizaje y los intereses políticos exigen un trabajo de interpretación por parte de autores y editores. Bajo la premura del mercado editorial, ese trabajo muchas veces se convierte en una especie de traducción apresurada, cuyo propósito es transformar en libro aquello que la reforma enuncia de manera general, dispersa o todavía inestable.

El problema aparece cuando la velocidad de salida al mercado pesa más que el tiempo necesario para comprender pedagógicamente la reforma. Las editoriales necesitan publicar rápido para responder a los cambios curriculares, competir en el mercado y colocar sus materiales. Como resultado, el libro puede elaborarse con prisa, mediante recortes, ajustes y decisiones tomadas más por urgencia editorial que por comprensión pedagógica profunda. De esa manera, el libro de texto corre el riesgo de nacer más como una respuesta editorial inmediata que como una verdadera mediación didáctica.

El trabajo editorial no debe entenderse como algo ajeno al trabajo autoral, sino como una labor que lo complementa. El editor interviene en la organización del material, la claridad expositiva, la extensión, el ritmo de lectura y la adecuación del libro tanto a la línea de la casa editorial como a su público objetivo. Idealmente, esa relación entre edición y autoría tendría que fortalecer la posibilidad de que el texto conecte con sus lectores. Sin embargo, cuando un libro no logra establecer esa conexión, el problema no siempre radica exclusivamente en el autor. En muchos casos, la dificultad también proviene de la línea editorial, las decisiones de mercado, los recortes de contenido, el tono exigido o el tipo de lector que la editorial imagina.

Algunas editoriales han comenzado a recurrir al uso de IA para la elaboración de libros de texto de EMS. El problema no radica únicamente en el uso de la herramienta, sino en el lugar que esta ocupa dentro del proceso de producción. Cuando la IA se inserta en un entorno editorial presionado por tiempos, costos, ajustes de extensión y rapidez de salida al mercado, existe el riesgo de que deje de ser un recurso auxiliar y se convierta en un atajo de escritura. Además, su uso puede resultar atractivo para algunas editoriales porque reduce la participación autoral humana y, en consecuencia, disminuye o evita el pago de regalías a escritores especializados.

Dicho de otra forma, en la EMS la discusión sobre la IA no puede separarse de las condiciones concretas de producción de los materiales, porque en ellas intervienen no solo decisiones pedagógicas, sino también intereses editoriales y criterios de mercado. Estos factores pueden empujar la racionalización del texto escolar hasta volverlo más eficiente desde el punto de vista productivo, pero menos cuidadoso en términos pedagógicos y humanos.

La racionalización ha sido una de las principales búsquedas del ser humano a lo largo de la historia. La IA puede integrarse a esa lógica, ya que trabaja con rapidez, estandariza procesos, economiza tiempo y facilita la construcción de variantes discursivas. Por ello, los libros de texto son proclives a convertirse en productos elaborados desde criterios de eficiencia. No obstante, aquello que se gana en rapidez puede perderse en humanidad. Si el criterio principal para elaborar materiales se desplaza hacia la eficiencia formal y la reducción de costos cognitivos y económicos, la experiencia docente deja de ser vista como una fuente de criterio pedagógico y pasa a ser tratada como algo prescindible.

El principal daño no proviene del uso de la IA en sí mismo, sino de la facilidad con que puede ser absorbida por organizaciones interesadas en acelerar la producción, disminuir los tiempos de elaboración y tratar la escritura educativa como un problema de administración. Allí, la lógica sustitutiva encuentra terreno fértil. Lo que se vende como innovación puede ser, en realidad, una forma de debilitamiento de

la mediación humana. Lo que se presenta como modernización puede reducir el proceso autoral a una serie de operaciones técnicas. Por ello, el problema de la IA no es puramente tecnológico, sino también organizacional y político: en contextos de presión editorial, puede convertirse en un atajo productivo y en una lápida creativa.

La ética de las editoriales y de los autores adquiere aquí una importancia central. Si una editorial decide utilizar IA en la elaboración de libros de texto, no debería ocultarlo ni presentarlo como si todo el contenido hubiera sido construido por autores humanos con experiencia docente. La transparencia tendría que ser una obligación básica: señalar en qué partes se usó IA, con qué finalidad, bajo qué criterios de revisión y quién asumió la responsabilidad final del contenido.

Lo mismo ocurre con los autores. No basta con entregar un texto aparentemente correcto si una parte sustantiva fue generada, organizada o sustituida por una herramienta automatizada sin revisión profunda. En los materiales escolares, la responsabilidad no puede diluirse entre la máquina, la editorial, el editor y el autor. Alguien debe responder por la pertinencia pedagógica, la calidad del contenido, el tratamiento de temas sensibles y los efectos posibles del libro en el aula. La IA puede asistir, pero no puede convertirse en una forma cómoda de evadir la responsabilidad autoral o editorial.

El uso de IA en la producción de libros de texto debería legislarse o, al menos, regularse mediante criterios claros y obligatorios. No se trata de prohibir su empleo, sino de establecer límites, condiciones y responsabilidades. Tendría que exigirse la declaración explícita del uso de IA, la revisión humana especializada, la participación de autores con experiencia docente, la evaluación pedagógica externa y la prohibición de presentar como autoría humana plena un material producido de manera automatizada. Regular su uso no sería un gesto tecnófobo, sino una forma de proteger al alumnado, al profesorado y el sentido humano del libro de texto.

El libro de texto como mediación curricular, pedagógica y docente

Para Arnaut y Giorguli (2010), el “currículum es el corazón del sistema educativo... comprende los contenidos educativos plasmados en los planes y programas de estudio, los libros de texto y otros materiales y dispositivos didácticos, la formación, la práctica docente y otros contenidos” (p. 14). Desde esta perspectiva, el libro de texto no puede entenderse como un simple contenedor de información. Forma parte del currículum porque organiza experiencias de lectura, legitima saberes, regula tonos, construye ejemplos, orienta percepciones y participa en la manera en que el alumno se aproxima al conocimiento. Por ello, su escritura implica decisiones intelectuales, didácticas y éticas que exceden por completo la simple producción verbal.

Toda reflexión sobre lo social también debe reconocer que un libro de texto representa, por sí mismo, una oportunidad para contribuir a la mejora de la sociedad, aunque jamás alcance la perfección, ni siquiera con el apoyo de la IA. Su valor no depende únicamente de que contenga información correcta, sino de la forma en que esa información se organiza, se vuelve enseñable y se relaciona con las condiciones reales de quienes aprenden y enseñan.

El libro de texto puede describirse en tres dimensiones. Primero, como artefacto pedagógico, porque decide qué se explica primero, qué se presupone, qué se repite, qué se simplifica, con qué ejemplos se acompaña un concepto y bajo qué ritmo se despliega la dificultad. Segundo, como objeto cultural, ya que no solo transmite conceptos: también nombra sujetos, organiza mundos, selecciona referentes, legitima ciertas formas de lenguaje y deja fuera otras. Tercero, como dispositivo organizacional, pues se inserta en calendarios, secuencias, evaluaciones, políticas institucionales y condiciones de producción que exceden al aula.

Estas tres dimensiones permiten ubicar mejor la discusión sobre la IA. Si el libro fuera solo información, bastaría con revisar la exactitud de los datos y la claridad de la redacción. Sin embargo, al ser mediación pedagógica, objeto cultural y pieza organizacional, su planeación exige algo más complejo que reunir contenidos y

ordenarlos con buena sintaxis. La IA puede ofrecer definiciones, ejemplos y explicaciones formalmente correctas, pero no es un actor escolar dentro de la escena en la que esos recursos serán recibidos. No sabe cómo interactúan el alumnado y el profesorado, ni comprende las condiciones concretas en que ocurre el aprendizaje. Sin duda, procesa enormes cantidades de información, pero nunca ha olfateado la Capilla Sixtina (Van Sant, 1997).

Por eso, los libros de texto no deberían ser escritos por personas ajenas a la enseñanza. Su fuerza depende, entre otras cosas, de su capacidad para relacionarse con lo que ocurre en el aula y con la forma en que trabajan quienes enseñan. Un material formalmente pulcro, pero didácticamente antipático, produce conflicto e indiferencia, no mediación. Quienes escriben libros escolares no pueden limitarse a fabricar materiales funcionales en apariencia, pero débiles en su vínculo con la práctica real. Por el contrario, necesitan comprender la vida concreta de los espacios escolares y trabajar con “contenidos significativos para el joven y para el adulto” (Schmelkes, 2010, p. 581). La finalidad es clara: hacer sentido en los lectores.

Los materiales forman parte activa de la experiencia escolar. No solo informan sobre el mundo; también modelan la relación del estudiante con el saber. Pueden hacer que el conocimiento aparezca como algo alcanzable, inteligible y abierto, o como un territorio ajeno al que solo se accede mediante fórmulas impersonales. Pueden invitar a la comprensión o favorecer la repetición; dosificar la complejidad o precipitar la confusión; cuidar el lenguaje o reproducir una frialdad que mantenga intacta la distancia entre el texto y la vida del lector. Por ello, el libro no debe verse como un producto terminado, sino como una mediación entre el conocimiento y el alumno.

Esa mediación también alcanza al profesorado. El libro de texto puede funcionar como una guía para enseñar, no en el sentido mecánico de imponer una ruta rígida, sino como apoyo para organizar la clase, abrir preguntas, explicar conceptos, introducir dificultades y cerrar aprendizajes. En muchas aulas, sobre todo cuando

se trata de docentes novatos o de profesores que apenas comienzan a familiarizarse con un campo disciplinar, el material ofrece una estructura práctica para distribuir tiempos, plantear actividades y acercar el contenido al grupo.

Cuando un libro está bien construido, transmite también la experiencia acumulada de quienes lo elaboraron. Su orden, sus ejemplos, sus actividades y su graduación de la dificultad pueden ayudar al docente a encontrar caminos posibles para trabajar con el grupo. De ahí la importancia de que sus autores conozcan realmente el aula. Un material elaborado desde una lógica fría, automatizada o ajena a la práctica educativa puede trasladar al salón formas deficientes de enseñanza: rigidez, distancia, mecanización, actividades sin sentido y poca sensibilidad frente a lo que ocurre con los estudiantes.

En la práctica, muchos libros están diseñados conforme a horas de trabajo, días de clase, bloques temáticos y momentos de apertura, desarrollo y cierre. Por eso, no solo acompañan la enseñanza: también sugieren una forma concreta de organizarla. Si el material está bien pensado, puede ayudar al docente a dosificar la dificultad, introducir actividades pertinentes y favorecer interacciones más ricas en el aula. Pero, si nace de una lógica estandarizada, impersonal y ajena a lo humano, puede conducir a una pedagogía más rígida, más cercana al cumplimiento de un guion que a la mediación viva del conocimiento.

La discusión sobre la autoría de los libros de texto adquiere, entonces, una relevancia mayor. Cuando la creatividad y la construcción de los materiales se alejan de los autores y se aproximan a la IA, no solo se compromete la calidad del contenido o su densidad pedagógica; también se altera, de manera indirecta, la práctica docente que habrá de apoyarse en ellos. Un libro escolar elaborado desde la impersonalidad puede terminar enseñando, de forma implícita, una manera igualmente automatizada de conducir la clase. La crítica, por tanto, no debe centrarse únicamente en si la IA escribe bien o mal, sino en la posibilidad de que los materiales producidos bajo su predominio extiendan su racionalidad al aula y

moldeen tanto al estudiante como al docente dentro de una experiencia formativa cada vez menos humana.

El libro de texto también participa en la relación entre profesorado y alumnado. Una actividad bien planteada puede propiciar diálogo, participación, confianza y formas de interacción que fortalecen el vínculo pedagógico. El trabajo con el libro puede producir momentos gratos dentro de la clase, espacios de interés genuino e incluso instantes de disfrute y diversión que no son ajenos al aprendizaje, sino parte de él. Cuando el material logra activar esa cercanía, el aula deja de ser únicamente un espacio de transmisión y se convierte en un espacio de relación, acompañamiento y construcción compartida del sentido del conocimiento. Por eso, el valor de los materiales escolares no se agota en su contenido temático ni en su corrección conceptual; también importa por los vínculos, ambientes y experiencias de aprendizaje que ayudan a construir.

Experiencia docente, revisión y responsabilidad frente al aula real

Debe reconocerse que algunos maestros aprovechan su experiencia docente para escribir sus propios materiales o libros, los cuales, en ocasiones, son utilizados y vendidos a sus propios alumnos. Esta práctica no debe juzgarse de manera simplista, pues puede tener una dimensión positiva. Un profesor que conoce bien su asignatura, su programa, el perfil de sus estudiantes, el contexto en el que se desenvuelven y las dificultades que aparecen cada periodo escolar puede producir un material mucho más cercano a las necesidades reales del grupo. En esos casos, el alumnado puede acceder al conocimiento de una manera más clara, situada y pertinente, porque el texto surge de la práctica cotidiana, de los errores observados, de las preguntas frecuentes, de los temas que suelen complicarse y de las estrategias que han funcionado en clase.

Un libro escrito por un docente con experiencia puede convertirse en una mediación valiosa, precisamente porque no parte de un lector imaginario, sino de estudiantes concretos. Sin embargo, esa misma cercanía también puede generar problemas. Cuando el profesor escribe, recomienda, utiliza y vende su propio material, pueden

aparecer conflictos de interés, presiones explícitas para que el alumnado lo compre o una aceptación poco crítica del texto debido a la autoridad que el docente ya tiene dentro del aula. Aunque la experiencia docente sea una fuente importante de valor pedagógico, no basta por sí sola para garantizar la calidad de un libro.

El problema se vuelve mayor cuando esos materiales no pasan por procesos serios de revisión. Muchos textos elaborados por docentes circulan sin dictamen académico, sin revisión editorial suficiente, sin evaluación pedagógica externa y sin criterios claros sobre su pertinencia. Otros reciben únicamente anotaciones de autoridades escolares o de los responsables del eje curricular correspondiente, pero eso no siempre garantiza una valoración objetiva ni profunda. A veces, la revisión se limita a verificar si el material se ajusta al programa, si cubre los contenidos mínimos o si no contradice la línea institucional, pero no necesariamente analiza con cuidado su calidad didáctica, su lenguaje, sus ejemplos, sus riesgos de interpretación, el tratamiento de temas sensibles o su impacto en la dinámica del aula.

Además, debe considerarse que quienes revisan estos materiales no siempre cuentan con la formación suficiente para valorar el contenido con profundidad. En algunos casos, los responsables de los ejes curriculares o las autoridades escolares no están debidamente capacitados, desconocen el tema específico o carecen de experiencia directa en el aula, por lo que sus observaciones pueden quedarse en un plano administrativo, superficial o meramente formal. Puede ocurrir también que la cercanía entre el autor, la escuela y las autoridades impida una revisión realmente crítica.

La elaboración de libros de texto requiere equilibrio: la experiencia docente es indispensable, pero debe acompañarse de responsabilidad autoral, revisión seria, criterios institucionales claros y una mirada externa capaz de señalar problemas que el propio autor quizá no advierte. El libro de texto no puede depender únicamente de la buena intención, la autoridad o la experiencia individual de quien lo escribe; necesita pasar por procesos de cuidado que aseguren que su uso

beneficie al alumnado y no reproduzca errores, sesgos, improvisaciones o prácticas poco convenientes dentro del aula.

La responsabilidad del autor de un libro de texto no se reduce a escribir bien ni a manejar correctamente un contenido. También consiste en pensar qué puede ocurrir en el salón de clases con aquello que se escribió. Un ejemplo mal elegido puede detonar una situación desagradable que no estaba prevista en el papel, pero que sí puede aparecer en la realidad del aula. Si en un libro se incluye un caso didáctico relacionado con la salud, el cuerpo o los hábitos de vida, y ese caso presenta a una persona con sobrepeso, el autor tendría que advertir que esa situación puede terminar convertida en apodo, burla o motivo de humillación para algún estudiante.

Lo mismo puede ocurrir cuando se utilizan ejemplos, actividades o lecturas sobre pobreza, forma de vestir, dificultades para aprender, color de piel, tipo de familia, apariencia física, higiene personal o cualquier otro rasgo que, dentro del grupo, pueda ser usado para señalar, excluir o ridiculizar a alguien. El problema no está solo en que el contenido sea correcto o útil para explicar un tema, sino en lo que ese contenido puede provocar cuando entra en contacto con un grupo real.

Esto también puede suceder con situaciones aparentemente sencillas. Si un libro presenta un caso para hablar sobre el esfuerzo escolar y utiliza como personaje a alguien que reprueba con frecuencia, el grupo puede identificar de inmediato al compañero que “siempre reprueba”. Una actividad sobre la casa, el cuarto propio o los bienes materiales, pensada quizá para trabajar la descripción del entorno, puede exhibir diferencias económicas. Una lectura sobre papá, mamá e hijos, usada para explicar la familia o la convivencia, puede incomodar a quienes viven otra estructura familiar. Un ejercicio sobre el cuerpo o la imagen personal, aunque busque promover el autocuidado, puede abrir comparaciones crueles. Una consigna sobre participación oral, diseñada para fortalecer la expresión, puede poner en evidencia a quien tiene dificultad para hablar frente al grupo. Incluso un recurso pensado para llamar la atención puede terminar desviándola por completo,

provocar risas, desorden o comentarios que se salen del tema y rompen el ritmo de la clase.

Lo anterior no se aprende solo leyendo teoría ni escribiendo con rapidez; se aprende en el aula. La experiencia docente enseña qué tipo de actividad puede funcionar, cuál puede convertirse en burla, qué palabra puede terminar en apodo, qué ejemplo puede romper el orden del grupo y qué caso puede generar incomodidad o exponer a alguien. Esta es una parte central de la responsabilidad autoral. El autor de libros escolares tiene que pensar más allá del contenido y más allá del escritorio. Debe considerar las consecuencias posibles de lo que escribe una vez que ese material llega a estudiantes reales, en grupos reales y en contextos reales.

La IA puede redactar con rapidez y construir ejemplos verosímiles, pero no tiene experiencia de aula. Por eso, no puede anticipar con la misma precisión cuándo una actividad puede derivar en burla, exhibición, incomodidad o desorden dentro del grupo. En la elaboración de libros de texto, el problema no consiste solo en escribir contenido correcto, sino en prever sus posibles efectos en la vida real del salón de clases.

Esa responsabilidad sigue dependiendo de la experiencia pedagógica y del criterio humano. La IA no siente esa responsabilidad, no carga con ella ni forma parte de sus propios criterios de funcionamiento. No le preocupa si una actividad termina lastimando a un alumno, si un ejemplo abre la puerta a la burla o si una consigna rompe el equilibrio del grupo. Su tarea es producir texto; la responsabilidad por lo que ese texto puede provocar sigue siendo humana. Precisamente por eso, en la escritura de materiales escolares no basta con la velocidad ni con la apariencia de corrección: hacen falta conciencia, experiencia y sentido de responsabilidad frente a estudiantes reales.

La responsabilidad no nace de la técnica, sino de la experiencia de aula. Quien ha estado frente a grupo conoce los ritmos reales de comprensión, los límites cognitivos del alumnado, los modos de recepción de un contenido, las reacciones

emocionales, las dificultades de vocabulario, el cansancio escolar, las formas de exclusión entre pares, los riesgos de una escritura poco cuidada y las maneras concretas de volver enseñable un contenido. Los autores con experiencia de aula no solo saben más sobre enseñanza; saben algo distinto. Saben cómo se comporta el conocimiento cuando entra en contacto con estudiantes reales.

Reconocen el momento en que el grupo se pierde, el gesto que muestra desconexión, la pregunta que revela un malentendido, el silencio que señala vergüenza, la risa que indica una humillación latente y el cansancio que vuelve inviable una explicación que, en ese momento, resultaría compleja. Ninguno de estos elementos aparece plenamente en las bases de datos ni en los patrones lingüísticos de un modelo. Forman parte de una memoria práctica del aula. De ahí que la experiencia docente no sea apenas una variable deseable del proceso autoral, sino una condición relevante para que el libro no se desligue de la realidad de la enseñanza.

En cualquier nivel educativo, el proceso formativo no se sostiene solo con contenidos correctos, sino con la capacidad de mediarlos. Esa mediación depende del profesorado. Si esto vale para el aula, también vale para el libro que pretende acompañar el trabajo que ahí se realiza. Un libro humanizado nace del cruce entre saber disciplinario y experiencia docente. Sin ese cruce, el texto corre el riesgo de convertirse en una superficie verbal sin anclaje pedagógico.

Detrás de la defensa de la experiencia docente hay una idea más amplia: el aula no puede reducirse a una guía ni a una forma abstracta de redactar. Mucho menos puede reducirse a un procesamiento algorítmico del lenguaje. En el salón de clases intervienen el contexto, la historia, la desigualdad, la cultura, el lenguaje, las trayectorias escolares, la violencia simbólica, el humor grupal, el prestigio, la vergüenza, el rechazo, la apropiación, el silencio, la resistencia y el aprendizaje real.

Todos esos elementos modifican la manera en que un contenido es recibido, entendido o rechazado por el grupo. Una explicación puede parecer clara en el

papel, pero no producir el mismo efecto cuando llega a un aula atravesada por cansancio, desigualdad, bromas internas, tensiones entre compañeros o experiencias previas de fracaso escolar. La escritura escolar no puede realizarse como si bastara con ordenar palabras o acomodar ideas de forma correcta. Hay una realidad viva en el aula que un sistema no humano simplemente no habita.

En esa realidad también influyen las condiciones materiales en las que ocurre la enseñanza. No es lo mismo trabajar una actividad en un salón sofocado por el calor, en un día de lluvia que modifica el ánimo del grupo, en un espacio frío o incómodo, o en una jornada donde la atención ya está desgastada por el ritmo escolar. Todo eso influye en la disposición para escuchar, participar, leer, escribir y procesar una explicación. Por eso, un libro de texto no puede pensarse solo desde la corrección de sus palabras, sino desde las condiciones reales en que esas palabras serán usadas.

El docente como autor del libro sin negar la tecnología

La discusión sobre la IA en la elaboración de libros de texto para la EMS no puede reducirse a una pregunta técnica sobre si escribe bien, organiza con rapidez o produce textos coherentes. El problema verdadero es otro: qué tipo de libro se está construyendo, desde qué experiencia se escribe, con qué responsabilidad se piensa al estudiante y qué lugar ocupa el ser humano en ese proceso.

Un libro de texto no es únicamente una pieza editorial ni una suma ordenada de contenidos. Es una mediación pedagógica que entra al aula, circula entre docentes y estudiantes, organiza formas de enseñar, propone modos de leer la realidad y participa en la formación de jóvenes que atraviesan una etapa decisiva de su vida. Cuando la IA se coloca en el centro de la producción educativa, el riesgo no consiste solo en generar materiales fríos, repetitivos o descontextualizados. El riesgo más profundo es normalizar la idea de que la experiencia docente puede ser reemplazada por una operación técnica.

Esa idea empobrece la autoría pedagógica, debilita el sentido humano del libro y convierte la formación en un proceso administrable, rápido y aparentemente

eficiente, pero menos sensible a la complejidad real del aula. En la EMS, donde los estudiantes viven desigualdades, dudas, tensiones familiares, presiones escolares, búsqueda de identidad y fragilidades propias de su etapa formativa, esa pérdida de sensibilidad no es menor.

La IA puede ayudar, pero no puede hacerse cargo del sentido completo de un libro escolar. Puede sugerir, ordenar, comparar, revisar y acelerar ciertos procesos, pero no puede asumir la responsabilidad de decidir qué conviene decir, cómo decirlo, qué ejemplo puede lastimar, qué actividad puede excluir, qué palabra puede abrir la puerta a la burla o qué tono puede acercar o alejar al estudiante del conocimiento. Esa responsabilidad pertenece al autor humano, especialmente a quien conoce el aula y entiende que enseñar no es solo transmitir información, sino construir condiciones para que otro ser humano pueda comprender, participar y reconocerse en su proceso formativo.

Defender el libro escrito por docentes no significa negar el avance tecnológico ni rechazar toda forma de apoyo digital. Significa establecer límites claros para que la tecnología no desplace aquello que da sentido a la educación: la experiencia, el juicio, la sensibilidad, la responsabilidad y la presencia humana. La EMS necesita libros escritos con conocimiento, pero también con cuidado; con claridad, pero también con comprensión de las condiciones reales en que los jóvenes aprenden. En ese punto se juega la diferencia entre producir materiales escolares y construir verdaderas mediaciones pedagógicas.

La incorporación de la IA en la producción editorial de libros de texto adquiere mayor complejidad cuando se observa el contexto específico de la EMS. A diferencia de la educación básica, cuyos libros gratuitos forman parte de una política pública más homogénea, en la EMS buena parte de los materiales circula dentro del mercado editorial. Esto significa que los libros se producen bajo presiones de tiempo, costo, competencia, extensión, reforma curricular y viabilidad comercial. La IA puede convertirse fácilmente en un atajo productivo: una forma de acelerar la escritura, reducir costos y responder con rapidez a las exigencias del mercado.

El problema no está en usar tecnología, sino en permitir que la lógica de la eficiencia desplace la responsabilidad pedagógica. Cuando el libro se piensa solo como producto, se corre el riesgo de olvidar que su destino no es el anaquel ni el catálogo editorial, sino el salón de clases, donde será leído, usado, interpretado y vivido por estudiantes y docentes. Por eso, resulta necesario establecer criterios éticos, pedagógicos y editoriales claros sobre el uso de IA en la elaboración de libros de texto. No se trata de prohibir su empleo, sino de impedir que se oculte, se normalice como sustituto del autor o se utilice para presentar como creación plenamente humana un contenido producido de manera automatizada.

La transparencia debe ser una condición básica. Las editoriales y los autores tendrían que señalar cuándo se utilizó IA, con qué finalidad, en qué partes del proceso, bajo qué revisión especializada y quién asumió la responsabilidad final del material. También se vuelve necesario garantizar la participación de autores con experiencia docente, la revisión pedagógica externa, dictámenes serios y criterios que valoren no solo la corrección conceptual, sino también la pertinencia didáctica, el tratamiento de temas sensibles, el lenguaje utilizado y los posibles efectos del texto en la dinámica del aula. Un libro escolar no puede quedar sostenido únicamente por la rapidez de producción ni por la apariencia de calidad.

La defensa del libro de texto humanizado en la Educación Media Superior es, al mismo tiempo, una defensa de la experiencia docente, de la responsabilidad autoral y del sentido formativo de la educación. La IA puede acompañar el trabajo humano, pero no puede reemplazar aquello que hace pedagógico a un libro: la capacidad de comprender a sus lectores, anticipar dificultades, modular el lenguaje, cuidar los ejemplos, organizar la dificultad y construir una relación sensible entre conocimiento, docente y estudiante. La EMS requiere materiales que reconozcan la humanidad de quienes aprenden, no solo su condición de usuarios o receptores de información.

Por eso, el desafío no consiste en elegir entre tecnología y tradición, sino en impedir que la tecnología vacíe de humanidad el proceso educativo. Un libro de texto

verdaderamente formativo no nace de la velocidad, sino del cuidado; no se sostiene solo en la corrección verbal, sino en la responsabilidad pedagógica; no se limita a informar, sino que acompaña, orienta y participa en la formación de personas.

Nota sobre el uso de herramientas de inteligencia artificial

Para la elaboración de este texto se utilizó ChatGPT como herramienta de apoyo en la revisión de redacción y detección de errores de escritura. Su uso no sustituyó la autoría del trabajo ni la responsabilidad académica sobre el contenido, ya que las ideas, el enfoque argumentativo, las decisiones finales y la revisión del texto fueron realizadas por el autor.

Referencias

Arellano Rodríguez, C. A. (2025). IA como nuevo impulso a la educación escolarizada y su papel en el desarrollo social. En O. Lozano Carrillo, Y. Zavala Osorio, & D. P. Couturier Bañuelos (Coords.), *El PIAESS: Una IA para la vida* (pp. 761–793). Red Mexicana de Investigadores en Estudios Organizacionales (REMINEO), A.C.

Arnaut, A., & Giorguli, S. (2010). Introducción general. En A. Arnaut & S. Giorguli (Coords.), *Educación* (Los grandes problemas de México, Vol. VII, pp. 13–32). El Colegio de México.

Feixa, C. (1999). De culturas, subculturas y estilos. En *De jóvenes, bandas y tribus: Antropología de la juventud* (Cap. 3). Ariel.

Quiñónez Salcido, A. (2025). La IA como herramienta educativa para fortalecer procesos de enseñanza-aprendizaje con enfoque social. En O. Lozano Carrillo, Y. Zavala Osorio, & D. P. Couturier Bañuelos (Coords.), *El PIAESS: Una IA para la vida* (pp. 443–464). Red Mexicana de Investigadores en Estudios Organizacionales (REMINEO), A.C.

Schmelkes, S. (2010). La educación básica de adultos. En A. Arnaut & S. Giorguli (Coords.), *Educación* (Los grandes problemas de México, Vol. VII, pp. 577–598). El Colegio de México.

Székely Pardo, M. (2010). Avances y transformaciones en la educación media superior. En A. Arnaut & S. Giorguli (Coords.), *Educación* (Los grandes problemas de México, Vol. VII, pp. 313–336). El Colegio de México.

Van Sant, G. (Director). (1997). *Good Will Hunting* [Película]. A Band Apart; Miramax; Lawrence Bender Productions.

Villa Lever, L. (2010). La educación media superior: su construcción social desde el México independiente hasta nuestros días. En A. Arnaut & S. Giorguli

(Coords.), *Educación* (Los grandes problemas de México, Vol. VII, pp. 271–312).
El Colegio de México.

Wallerstein, I. (1995). ¿El fin de qué modernidad? J. Hernández Prado,
Trad. *Sociológica*, 10(27).

CAPÍTULO XIII

REGULACIÓN JURÍDICA DE LA INTELIGENCIA ARTIFICIAL FRENTE A LOS DERECHOS HUMANOS

Legal regulation of Artificial Intelligence versus Human Rights

Dr. Jesús Manuel López Marroquín

jesus.lopezm@uadeo.mx
Universidad Autónoma de Occidente
<https://orcid.org/0009-0005-9175-9841>

Dra. Gisela Cota Yucupicio

gisela.cota@uadeo.mx
Universidad Autónoma de Occidente
<https://orcid.org/0000-0002-8030-5191>

Dra. Edna Karina Miranda Mondaca

edna.miranda@uadeo.mx
Universidad Autónoma de Occidente
<https://orcid.org/0009-0004-9976-658X>

Resumen

La Inteligencia Artificial (IA) se ha consolidado como una de las tecnologías con mayor impacto en aspectos jurídicos, políticos y sociales contemporáneos. Su implementación en ámbitos como la administración pública, la seguridad, la justicia, la salud y el mercado laboral ha transformado los procesos de toma de decisiones, al generar beneficios en eficiencia y automatización, pero también riesgos significativos para los derechos humanos.

La regulación jurídica de la IA representa una exigencia derivada del derecho internacional de los derechos humanos. Su capacidad para procesar grandes volúmenes de datos, identificar patrones complejos, automatizar procesos decisionales y predecir conductas humanas ha abierto un abanico de oportunidades inéditas para la eficiencia institucional y la innovación social; sin embargo, la ausencia de marcos regulatorios adecuados puede derivar en prácticas

discriminatorias, vulneraciones a la privacidad, afectación al debido proceso y a la dignidad humana (UNESCO, 2021; Naciones Unidas, 2024).

La incorporación progresiva de sistemas de IA en ámbitos estratégicos de la vida plantea desafíos sustantivos para los ordenamientos jurídicos, particularmente en lo que respecta a la protección y garantía de los derechos humanos. En este escenario, el derecho se ve obligado a replantear sus categorías tradicionales y a construir marcos regulatorios capaces de responder a los riesgos asociados a la automatización de decisiones, ausencia de transparencia algorítmica, los sesgos o desigualdades sistémicas y la concentración desigual del poder informacional. Se examina la regulación jurídica de la inteligencia artificial desde un enfoque basado en derechos humanos, a partir del análisis articulado de convenciones internacionales, normativa nacional en México, jurisprudencia relevante y aportaciones doctrinales especializadas, con el propósito de ofrecer una comprensión global sobre los límites, obligaciones y responsabilidades que deben orientar el desarrollo, implementación y uso de estas tecnologías en un Estado constitucional de derecho.

Palabras clave: Regulación Jurídica, Inteligencia Artificial, Derechos Humanos

Abstract

Artificial Intelligence (AI) has established itself as one of the technologies with the greatest impact on contemporary legal, political and social aspects. Its implementation in areas such as public administration, security, justice, health and the labour market has transformed decision-making processes, generating benefits in efficiency and automation, but also significant risks for human rights.

The legal regulation of AI represents a requirement derived from international human rights law. Its ability to process large volumes of data, identify complex patterns, automate decision-making processes and predict human behavior has opened up a range of unprecedented opportunities for institutional efficiency and social

innovation; however, the absence of adequate regulatory frameworks can lead to discriminatory practices, violations of privacy, and an impact on due process and human dignity (UNESCO, 2021; United Nations, 2024).

The progressive incorporation of AI systems in strategic areas of life poses substantive challenges for legal systems, particularly with regard to the protection and guarantee of human rights. In this scenario, the law is forced to rethink its traditional categories and build regulatory frameworks capable of responding to the risks associated with the automation of decisions, the absence of algorithmic transparency, systemic biases or inequalities and the unequal concentration of informational power. The legal regulation of artificial intelligence is examined from a human rights-based approach, based on the articulated analysis of international conventions, national regulations in Mexico, relevant jurisprudence and specialized doctrinal contributions, with the purpose of offering a global understanding of the limits, obligations and responsibilities that should guide the development, implementation and use of these technologies in a constitutional state of law.

Keywords: Legal Regulation, Artificial Intelligence, Human Rights

Introducción

El desarrollo acelerado de la inteligencia artificial (IA) constituye uno de los fenómenos tecnológicos más relevantes del siglo XXI, la capacidad de los sistemas de IA para procesar grandes volúmenes de datos, aprender de manera autónoma, predecir comportamientos y adoptar decisiones con efectos jurídicos directos ha transformado sustancialmente la forma en que se diseñan e implementan políticas públicas, se prestan servicios esenciales y se ejercen funciones tradicionales del poder público. En este caso, la IA ha dejado de ser un mero instrumento técnico para convertirse en un objeto regulatorio cuya operación incide de manera directa en el ejercicio de derechos fundamentales, lo que exige un análisis jurídico riguroso desde una perspectiva constitucional y de derechos humanos.

Si bien la inteligencia artificial ofrece oportunidades para mejorar la eficiencia institucional, ampliar el acceso a servicios y fortalecer capacidades estatales, también implica riesgos para los derechos humanos cuando se desarrolla o despliega sin salvaguardas jurídicas adecuadas. Entre estos riesgos se encuentran la reproducción de prácticas discriminatorias, la afectación a la privacidad y protección de datos personales, la ausencia de transparencia y explicabilidad en la toma de decisiones automatizadas, así como la debilitación de garantías fundamentales como el debido proceso y la presunción de inocencia. Frente a ello, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) ha sostenido que la protección de la dignidad y los derechos humanos debe constituir el eje rector de cualquier estrategia normativa sobre inteligencia artificial, subrayando la necesidad de establecer principios de transparencia, rendición de cuentas y supervisión humana efectiva en todo el ciclo de vida de los sistemas de IA (UNESCO, 2021). En una línea convergente, la Organización de las Naciones Unidas ha señalado que la gobernanza de la IA debe orientarse a garantizar que su desarrollo y uso no vulneren los derechos humanos, la democracia ni el Estado de derecho, enfatizando la urgencia de articular marcos regulatorios basados en estándares internacionales de protección de derechos (Naciones Unidas, 2024).

Desde la perspectiva jurídica, la Inteligencia Artificial (IA) no constituye únicamente un conjunto de herramientas tecnológicas, sino un nuevo objeto regulatorio con capacidad de incidir directamente en situaciones jurídicas concretas. Los sistemas de IA pueden clasificar, predecir y decidir, influyendo en el acceso a bienes, servicios y derechos, lo que cuestiona categorías tradicionales como la imputación de responsabilidad, la motivación de los actos y la seguridad jurídica. Dicha transformación desafía los paradigmas tradicionales del derecho, en la medida en que introduce decisiones automatizadas con escaso sustento jurídico, frecuentemente carentes de trazabilidad, explicabilidad y control humano efectivo, lo que tensiona principios fundamentales del Estado constitucional como la seguridad jurídica, el debido proceso, la rendición de cuentas y la tutela efectiva de

los derechos fundamentales (Taboada Macías, 2024; Aveiga Saldarriaga y otros., 2024).

La automatización de procesos decisorios en ámbitos sensibles como la justicia penal, la seguridad pública o la asignación de beneficios sociales, evidencia la necesidad de un replanteamiento normativo que condicione el uso de la IA a límites constitucionales claros.

El presente capítulo examina la regulación jurídica de la inteligencia artificial desde un enfoque basado en derechos humanos, mediante el análisis articulado de instrumentos internacionales relevantes, normativa nacional aplicable en México, jurisprudencia constitucional y aportaciones doctrinales especializadas. El objetivo es ofrecer una comprensión integral de los límites, obligaciones y responsabilidades que deben orientar el desarrollo, implementación y uso de la inteligencia artificial en un Estado constitucional de derecho, atendiendo tanto a su potencial transformador como a los riesgos que plantea para la dignidad humana y el ejercicio efectivo de los derechos fundamentales.

I. Marco regulatorio de la IA en México

El análisis del impacto de la Inteligencia Artificial en el orden jurídico mexicano, desde una perspectiva constitucional, debe partir necesariamente del artículo 1º de la Constitución Política de los Estados Unidos Mexicanos, como piedra angular del sistema de derechos humanos, ya que dicho precepto establece un marco normativo integral que condiciona cualquier desarrollo tecnológico con efectos jurídicos (Constitución Política de los Estados Unidos Mexicanos, 2026).

En primer lugar, el reconocimiento y vigencia de los derechos humanos implica que ninguna innovación, incluidos los sistemas de IA, pueden operar al margen de los derechos reconocidos en la Constitución y en los tratados internacionales, ni justificar su restricción indebida (OACNUDH, 2023). El hecho de que una tecnología sea innovadora o avanzada no le otorga un trato excepcional frente al orden

jurídico, es decir, los derechos humanos funcionan como un límite ineludible para cualquier desarrollo tecnológico, incluida la inteligencia artificial. Esto significa que los sistemas de IA deben diseñarse, implementarse y utilizarse respetando plenamente los derechos reconocidos tanto en la Constitución como en los tratados internacionales en materia de derechos humanos. Por lo que, la IA no puede emplearse como pretexto para restringir, debilitar o evadir garantías fundamentales como la igualdad, el debido proceso o la privacidad, ya que la innovación tecnológica está jurídicamente subordinada al respeto y protección de la dignidad humana.

En segundo término, la obligación impuesta a todas las autoridades de promover, respetar, proteger y garantizar los derechos humanos extiende la responsabilidad constitucional a las decisiones automatizadas, de modo que el uso de algoritmos no exime al Estado de su deber de diligencia, prevención y reparación frente a posibles violaciones (Constitución Política de los Estados Unidos Mexicanos, 2026; ACNUDH, 2023). A ello, se suma el principio pro-persona, que exige que la interpretación y aplicación de normas, políticas públicas y tecnologías se realice siempre en el sentido más favorable a la protección de las personas, lo cual implica privilegiar el control humano, la transparencia y la explicabilidad algorítmica cuando existan riesgos para derechos fundamentales (Suprema Corte de Justicia de la Nación, 2014).

Asimismo, la prohibición de toda forma de discriminación, expresamente reconocida en el artículo 1º constitucional, obliga a examinar críticamente los sesgos inherentes a los sistemas de IA y a prevenir que estos reproduzcan o profundicen desigualdades, particularmente en contextos de alta vulnerabilidad social (UNESCO, 2022; OECD, 2024). En consecuencia, el desarrollo e implementación de la IA debe someterse a un escrutinio constitucional estricto, orientado a la dignidad humana, la igualdad sustantiva y la preservación del Estado de derecho.

El artículo 1º de la Constitución Política de los Estados Unidos Mexicanos (2026) constituye la norma axial del sistema constitucional mexicano, al establecer el principio de universalidad, interdependencia, indivisibilidad y progresividad de los derechos humanos, así como el bloque de constitucionalidad integrado por los tratados internacionales en la materia. Su redacción impone al Estado mexicano una doble obligación: respetar y garantizar los derechos humanos de todas las personas, y adecuar su actuación normativa, administrativa y tecnológica a dichos estándares.

El principio de universalidad en el ámbito jurídico significa que los derechos humanos corresponden a todas las personas, sin distinción, y deben ser reconocidos, respetados y garantizados por las autoridades en cualquier circunstancia. Este principio parte de la premisa de que los derechos humanos son inherentes a la dignidad humana y no dependen de condiciones sociales, políticas o económicas para su reconocimiento, por lo que su vigencia no admite excepciones (Monroy Enríquez & Ramírez Ogarrio, 2025.).

La interdependencia establece que los derechos humanos no pueden entenderse ni garantizarse de manera aislada, ya que existe una relación de interconexión entre ellos que condiciona su pleno ejercicio. Es decir, la violación de un derecho incide inevitablemente en el goce de otros, lo que impone a los Estados la obligación de adoptar medidas integrales de protección, promoción y garantía de todos los derechos humanos de manera conjunta y coherente (Monroy Enríquez & Ramírez Ogarrio, 2025.).

Por su parte, el principio de indivisibilidad sostiene que los derechos humanos son inseparables y no pueden fragmentarse ni jerarquizarse, puesto que todos poseen igual valor jurídico y deben ser garantizados de forma integral. Su relevancia radica en que cada derecho potencia el ejercicio de los demás, de tal manera que la afectación de uno compromete el disfrute pleno del resto, reafirmando la necesidad de una protección indivisible del conjunto de derechos humanos (Monroy Enríquez & Ramírez Ogarrio, 2025.).

El principio de progresividad implica que el reconocimiento y la protección de los derechos humanos deben avanzar de manera constante y nunca retroceder. Este principio obliga a los Estados a ampliar gradualmente el alcance y la efectividad de los derechos, así como a abstenerse de adoptar medidas regresivas que disminuyan su nivel de protección previamente alcanzado, reforzando un compromiso permanente con el fortalecimiento del sistema de derechos humanos (Monroy Enríquez & Ramírez Ogarrio, 2025.)

Cualquier política pública, desarrollo normativo o uso estatal de tecnologías emergentes incluida la inteligencia artificial, no es neutral, sino que queda jurídicamente subordinada a la protección efectiva de los derechos humanos reconocidos constitucional y convencionalmente. Por tanto, la IA se inserta en el ámbito de aplicación directa del artículo 1º (Constitución Política de los Estados Unidos Mexicanos, 2026), en la medida en que influye o incide sobre el ejercicio de derechos fundamentales a través de los principios descritos.

Bajo este marco normativo, el análisis de la inteligencia artificial debería considerar de manera específica los principios constitucionales de los derechos humanos, comenzando por el de universalidad, consagrado en el artículo 1º (Constitución Política de los Estados Unidos Mexicanos, 2026) que implica que todas las personas gozan de los derechos humanos sin distinción alguna. En el contexto de la inteligencia artificial, este principio se ve particularmente comprometido por los riesgos de discriminación algorítmica, ya sea por sesgos en los datos, en los modelos de entrenamiento o en los criterios utilizados para la toma automatizada de decisiones.

Desde la óptica constitucional, el uso de sistemas de IA en ámbitos como programas sociales, seguridad pública, justicia, educación o salud no puede generar exclusiones indirectas, categorizaciones arbitrarias o tratos diferenciados injustificados. La universalidad exige que la regulación de la IA incorpore mecanismos preventivos de evaluación de impacto en derechos humanos, particularmente respecto del derecho a la igualdad y a la no discriminación.

Dicho artículo reconoce que los derechos humanos son interdependientes e indivisibles, y los sistemas algorítmicos suelen generar impactos simultáneos sobre múltiples derechos fundamentales, de modo que la afectación de uno puede extenderse necesariamente a otros; así, por ejemplo, la vigilancia algorítmica no solo compromete el derecho a la privacidad, sino también a la libertad de expresión, la libertad de asociación y la participación democrática.

Los algoritmos aplicados a decisiones administrativas y jurisdiccionales pueden afectar de manera concurrente el debido proceso, la certeza jurídica y el acceso efectivo a la justicia, lo que hace indispensable la adopción de un marco regulatorio integral, coherente y transversal, orientado a la protección plena de los derechos humanos. Un ejemplo claro puede observarse en el uso de un algoritmo de evaluación automática para la asignación o cancelación de un beneficio social, como una beca, pensión o apoyo económico por parte de una autoridad administrativa. Si el sistema de IA cruza bases de datos y, con base en reglas poco transparentes, determina que una persona no cumple ciertos criterios, la decisión puede adoptarse de manera automática o semiautomática, sin que el afectado conozca los criterios utilizados ni tenga oportunidad real de controvertirlos. En este supuesto, el debido proceso se ve comprometido por la falta de audiencia previa y motivación suficiente; la certeza jurídica se debilita al no existir reglas claras, previsibles y accesibles sobre el funcionamiento del algoritmo; y el acceso efectivo a la justicia resulta limitado si no se cuenta con mecanismos eficaces para impugnar la decisión o cuestionar el funcionamiento del sistema algorítmico.

El debido proceso es un derecho humano de rango constitucional que garantiza que toda persona sea juzgada conforme a procedimientos previamente establecidos, ante autoridades competentes, independientes e imparciales, y con respeto a las formalidades esenciales del procedimiento. En el orden jurídico mexicano, este derecho se fundamenta en el artículo 14 constitucional y ha sido desarrollado por la Suprema Corte de Justicia de la Nación, la cual ha sostenido que comprende, entre otros elementos, la notificación del inicio del procedimiento, la posibilidad efectiva de ofrecer pruebas y formular alegatos, así como la emisión

de una resolución debidamente fundada y motivada, constituyendo un parámetro indispensable para el acceso efectivo a la justicia (Suprema Corte de Justicia de la Nación, 2014).

Por otra parte, la certeza jurídica significa que las normas legales deben ser claras, conocidas, estables y aplicadas de manera uniforme, de modo que las personas puedan anticipar las consecuencias jurídicas de sus actos y confiar en que sus derechos estarán protegidos. Es un principio fundamental del Estado de derecho que busca evitar la arbitrariedad y garantizar seguridad en las relaciones jurídicas (Consejo de la Judicatura Federal & UNAM, 2014; Fernández, 2024).

El principio de progresividad impone al Estado la obligación de avanzar constantemente en la protección de los derechos humanos, prohibiendo cualquier medida regresiva injustificada. En materia de inteligencia artificial, este principio implica que la adopción de tecnologías automatizadas no puede autorizar retrocesos en las garantías y estándares de protección de los derechos humanos previamente consolidados, so pretexto de eficiencia, innovación o modernización administrativa.

La progresividad de los derechos humanos contempla que la incorporación de sistemas de inteligencia artificial en la actuación administrativa o jurisdiccional no puede traducirse en un retroceso de las garantías ya existentes, de modo que la automatización no debe disminuir la transparencia exigida en la toma de decisiones, restringir el acceso a recursos efectivos de impugnación ni sustituir garantías procesales bajo el argumento de eficiencia tecnológica; por el contrario, este principio obliga a que la regulación de la IA eleve los estándares de protección, integrando criterios como la explicabilidad de las decisiones automatizadas, la trazabilidad del funcionamiento del sistema, la supervisión humana y una atribución clara de responsabilidades jurídicas. Por ejemplo, si una autoridad utiliza un algoritmo para resolver solicitudes de licencias o imponer sanciones administrativas, la decisión no puede emitirse sin fundamentos claros y comprensibles para la persona afectada, ni eliminar audiencias, plazos o recursos

legales, sino que debe permitir conocer cómo operó el sistema, contar con intervención humana para revisar el resultado y garantizar medios eficaces para impugnar la resolución, reforzando así la protección de los derechos humanos.

La Constitución Política de los Estados Unidos Mexicanos (2026), no solo reconoce derechos, sino que impone al Estado mexicano la obligación de garantizarlos. Esta obligación adquiere relevancia decisiva en el ámbito de la inteligencia artificial, pues el vacío normativo o la regulación insuficiente también pueden constituir una forma de violación a los derechos humanos. Desde esta lógica, el uso regulatorio de la inteligencia artificial implica que el Estado tiene la obligación constitucional de diseñar marcos jurídicos claros que definan cómo pueden desarrollarse y utilizarse los sistemas de IA, regular tanto a autoridades públicas como actores privados cuando sus tecnologías incidan en derechos humanos, garantizar mecanismos efectivos de rendición de cuentas y reparación del daño, y asegurar la compatibilidad de dichos sistemas con el derecho internacional de los derechos humanos.

A partir de lo anterior, la ausencia o debilidad regulatoria frente a tecnologías de alto impacto puede constituir una violación indirecta a la Constitución Política y demás leyes reglamentarias, por incumplimiento del deber de protección. Por ejemplo, si una empresa privada implementa un algoritmo para evaluar perfiles crediticios sin regulación estatal adecuada y el sistema discrimina sistemáticamente a ciertos grupos, el Estado podría ser responsable por no haber establecido normas de supervisión, transparencia y mecanismos de reclamo, permitiendo que una tecnología de alto impacto opere sin salvaguardas suficientes para proteger los derechos fundamentales de las personas afectadas ni ofrecer mecanismos efectivos de control, revisión y reparación frente a decisiones automatizadas.

Por otra parte, el artículo 1º (Constitución Política de los Estados Unidos Mexicanos, 2026) consagra el principio pro persona, el cual establece que ante varias interpretaciones posibles de una norma o frente a la concurrencia de distintas disposiciones, se elija siempre aquella que otorgue la mayor protección a la

persona. Este principio adquiere una función metodológica clave en la regulación de la IA: ante la duda, debe privilegiarse la protección del derecho de la persona frente a la eficiencia tecnológica.

En la práctica, esto significa que el empleo de sistemas de inteligencia artificial por parte del Estado o de actores privados no puede anteponer la automatización a la dignidad humana, ni justificar decisiones lesivas de derechos bajo el argumento de innovación o eficiencia tecnológica; por el contrario, cualquier desarrollo o aplicación de IA debe subordinarse al respeto pleno de los derechos humanos, y, ante la ausencia o ambigüedad de normas específicas, los vacíos regulatorios deben interpretarse conforme a los estándares internacionales de derechos humanos, aplicando el principio pro persona. Por ejemplo, si una autoridad utiliza un sistema automatizado para priorizar pacientes en el acceso a servicios de salud y la normativa no regula expresamente ese uso, la decisión no puede basarse exclusivamente en un cálculo algorítmico que deshumanice a las personas atendidas, sino que debe interpretarse y ajustarse a los estándares internacionales sobre derecho a la salud, dignidad y no discriminación, incorporando criterios humanitarios, supervisión médica y mecanismos para revisar y corregir las decisiones del sistema.

La Constitución Política de los Estados Unidos Mexicanos (2026) proporciona un fundamento constitucional sólido para la regulación de la inteligencia artificial desde una perspectiva de derechos humanos. La IA, lejos de ser un ámbito ajeno al derecho constitucional, se sitúa en el centro de las obligaciones estatales de respeto, protección y garantía de los derechos fundamentales. En consecuencia, cualquier uso regulatorio de la inteligencia artificial en México debe construirse como una manifestación contemporánea del constitucionalismo humanista, en el que la tecnología esté al servicio de la persona y no a la inversa.

II. Derechos humanos en riesgo ante el uso de la inteligencia artificial

El uso creciente de la inteligencia artificial en sectores estratégicos de la vida pública y privada ha generado una preocupación sostenida en la comunidad internacional respecto de sus impactos potenciales y reales sobre los derechos humanos. Bajo esta perspectiva, la inteligencia artificial lejos de ser una herramienta neutral, puede amplificar desigualdades, reproducir prácticas discriminatorias y erosionar garantías fundamentales si no se encuentra debidamente regulada desde una perspectiva jurídica y ética centrada en la dignidad humana.

Ante esta realidad, el análisis de los derechos humanos en riesgo frente al uso de la inteligencia artificial resulta un eje central para comprender la urgencia de marcos normativos adecuados. La experiencia comparada y los informes internacionales coinciden en que los sistemas de IA pueden comprometer simultáneamente derechos como la igualdad, la privacidad, el debido proceso, la libertad de expresión o el acceso a la justicia, en consecuencia, se requiere una regulación integral, preventiva y alineada con los estándares internacionales de derechos humanos.

Derecho a la igualdad y no discriminación frente a la Inteligencia Artificial

Uno de los principales riesgos asociados al uso de la Inteligencia Artificial es su capacidad para reproducir y amplificar sesgos preexistentes. Los sistemas de IA, particularmente aquellos basados en *machine learning*, se entrenan a partir de datos históricos, los cuales no son neutrales desde el punto de vista social, sino que reflejan desigualdades acumuladas relacionadas con género, raza, origen étnico, condición socioeconómica, edad o discapacidad (Barocas & Selbst, 2016; Noble, 2018). Como resultado, los algoritmos tienden a normalizar patrones discriminatorios, otorgándoles una apariencia de objetividad técnica que dificulta su detección y cuestionamiento jurídico.

Diversos estudios han documentado cómo la aplicación de sistemas automatizados ha derivado en discriminación algorítmica, manifestada, entre otros aspectos, en procesos de selección laboral que excluyen de forma sistemática a mujeres o minorías, en modelos de crédito que penalizan a personas por su contexto social, o en sistemas de vigilancia y predicción delictiva que concentran acciones policiales en comunidades históricamente marginadas (ACNUDH, 2021; UNESCO, 2022). Estas prácticas no solo reproducen desigualdades sociales, sino que pueden profundizarlas, al operar a gran escala y con escasa posibilidad de impugnación individual.

La mayoría de estas decisiones automatizadas carecen de una explicación clara y comprensible, lo que coloca a las personas afectadas en una situación de indefensión jurídica, por lo que es necesario, llevar a cabo una regulación al respecto. La falta de transparencia algorítmica dificulta identificar si una decisión desfavorable responde a criterios legítimos o si es producto de un sesgo discriminatorio incorporado en el diseño, los datos o el funcionamiento del sistema (OECD, 2019). De este modo, la discriminación deja de manifestarse de forma explícita para adoptar modalidades indirectas y directas más complejas de acreditar, pero igualmente lesivas de derechos fundamentales.

Desde el punto de vista jurídico, estos fenómenos resultan particularmente graves, ya que vulneran el principio de igualdad ante la ley y el derecho a no ser discriminado, pilares esenciales del sistema de derechos humanos y del Estado constitucional. De esta manera, la lucha contra la discriminación algorítmica exige marcos normativos específicos, evaluaciones de impacto en derechos humanos y mecanismos de supervisión que garanticen la igualdad sustantiva en el diseño y uso de sistemas de IA (ACNUDH, 2023; OECD, 2024).

El derecho a la privacidad y a la protección de datos personales, particularmente por prácticas de vigilancia masiva y perfilamiento.

El derecho a la privacidad y a la protección de los datos personales se constituye como uno de los derechos con mayor impacto por la Inteligencia Artificial, debido a que la mayoría de los sistemas algorítmicos dependen del tratamiento masivo de información personal para su funcionamiento. La IA se alimenta de grandes volúmenes de datos, los cuales son recolectados, combinados, analizados y reutilizados a partir de múltiples fuentes, sin que las personas involucradas tengan conocimiento efectivo del alcance y las finalidades de dicho tratamiento (Zuboff, 2019; ACNUDH, 2021).

Este fenómeno plantea riesgos significativos en el derecho a la vida privada, la autodeterminación informativa y el control sobre los datos personales, ya que los sistemas de IA permiten predecir información sensible como preferencias políticas, estado de salud, hábitos de consumo o patrones de comportamiento, aun cuando dicha información no haya sido proporcionada explícitamente por los titulares de los datos (UNESCO, 2022). En este marco, la afectación a la privacidad no se limita a la recolección de datos, sino que se extiende a su procesamiento automatizado, perfilamiento y uso predictivo, que pueden generar impactos jurídicos y sociales. Por ejemplo, uno de los problemas asociados al uso de IA es la inequidad en el acceso a la información entre quienes controlan los sistemas algorítmicos y las personas cuyos datos son procesados. Esto se manifiesta en prácticas como la vigilancia masiva, el reconocimiento facial, el rastreo digital y el análisis de comportamiento a gran escala, los cuales, pueden operar sin consentimiento informado (Crawford, 2021).

En el ámbito constitucional mexicano, el derecho a la privacidad y a la protección de datos personales se encuentra reconocido de manera expresa en los artículos 6º y 16 constitucionales, en relación directa con el artículo 1º, que obliga a todas las autoridades a respetar y garantizar los derechos humanos conforme al principio pro persona. Este marco impone al Estado el deber de evitar que el uso de

tecnologías basadas en IA derive en injerencias arbitrarias o desproporcionadas en la vida privada, incluso cuando dichas tecnologías se presenten como instrumentos de eficiencia administrativa o seguridad pública (Constitución Política de los Estados Unidos Mexicanos, 2026).

La preocupación por estos riesgos ha sido reiterada por organismos internacionales, los cuales han alertado que el tratamiento automatizado de datos mediante IA puede generar efectos acumulativos de vigilancia, incompatibles con sociedades democráticas, si no se establecen límites normativos claros y mecanismos de rendición de cuentas efectivos (ACNUDH, 2023; OECD, 2019). En particular, se ha señalado que el uso de tecnologías como el reconocimiento facial en espacios públicos plantea un riesgo elevado de normalización de la vigilancia, erosionando la expectativa razonable de privacidad de las personas.

La compatibilidad entre IA y derechos humanos exige que los sistemas algorítmicos incorporen, desde su diseño, salvaguardas robustas de protección de datos, tales como evaluaciones de impacto en privacidad, consentimiento informado efectivo, anonimización cuando sea posible y control humano sobre el tratamiento de información sensible. Asimismo, resulta indispensable garantizar a las personas derechos de acceso, rectificación, cancelación y oposición, aun frente a decisiones automatizadas (UNESCO, 2022; OECD, 2024).

La Inteligencia Artificial representa un desafío para el derecho a la privacidad y la protección de datos personales, al enfrentar los límites tradicionales entre lo público y lo privado. Por lo tanto, el derecho debe reafirmar su función garantista, es decir, se limita el poder del estado y se asegura que las garantías constitucionales se cumplan en la práctica, al garantizar que la innovación tecnológica no se traduzca en una erosión progresiva de la autonomía personal, la dignidad humana y el libre desarrollo de la personalidad, que son pilares esenciales del Estado constitucional de derecho.

El derecho al debido proceso y a la tutela judicial efectiva, cuando las decisiones automatizadas carecen de explicabilidad.

El derecho al debido proceso constituye uno de los pilares fundamentales del Estado constitucional de derecho y del sistema internacional de protección de los derechos humanos, ya que garantiza que toda persona sea oída, conozca las razones de las decisiones que le afectan y cuente con medios efectivos para impugnarlas ante una autoridad competente e imparcial. La incorporación de sistemas de Inteligencia Artificial en procesos decisorios públicos y privados plantea desafíos a este derecho, especialmente cuando dichas decisiones se adoptan de forma automatizada o con una intervención humana meramente formal (Citron, 2008; Pasquale, 2015).

La tutela judicial efectiva representa el derecho de toda persona a acceder a la justicia de manera real y práctica, mediante el establecimiento de procesos imparciales y con mecanismos que garanticen no solo la emisión de una resolución fundada, sino también su ejecución efectiva. Este derecho protege el acceso a tribunales independientes e imparciales, la defensa adecuada y la obtención de una sentencia que resuelva el fondo del conflicto y se materialice plenamente, constituyendo un elemento esencial del derecho de acceso a la justicia reconocido en el artículo 17 de la Constitución Política de los Estados Unidos Mexicanos (Suprema Corte de Justicia de la Nación, 2019)

Uno de los principales problemas jurídicos asociados a la IA es la imposibilidad práctica de conocer cómo y por qué un sistema automatizado llega a determinadas decisiones mediante un proceso interno de razonamiento que no es transparente o comprensible para las personas afectadas, o en su defecto, para las autoridades que los utilizan (Burrell, 2016). Esta falta de transparencia impide a los individuos comprender las razones de una decisión adversa, obstaculizando su derecho a la defensa y al acceso efectivo a la justicia.

La explicabilidad algorítmica dentro del debido proceso implica que las decisiones automatizadas deben ser comprensibles, justificables y susceptibles de revisión, permitiendo identificar los criterios utilizados, los datos empleados y la lógica general del sistema (European Commission, 2020). Sin estos elementos, la persona afectada se encuentra en una situación de indefensión jurídica, ya que no puede cuestionar la legalidad, razonabilidad o proporcionalidad de la decisión adoptada.

El uso de IA en contextos como el derecho procesal en todas sus áreas ejemplifica con claridad estos riesgos. Diversos estudios han señalado que los sistemas de apoyo a decisiones judiciales pueden influir de manera significativa en resoluciones que afectan derechos fundamentales como la libertad personal, el derecho a la vida y aspectos patrimoniales de la persona, sin estar fundadas y motivadas jurídicamente dentro del razonamiento algorítmico (Dressel & Farid, 2018; ACNUDH, 2021).

En el ámbito mexicano, esta problemática debe analizarse a la luz del artículo 1º constitucional, que impone a todas las autoridades la obligación de respetar y garantizar los derechos humanos, así como del derecho al debido proceso reconocido en los artículos 14 y 16 constitucionales. La mediación tecnológica no exime al Estado de esta obligación ya que las decisiones automatizadas son jurídicamente imputables a la autoridad que las adopta o utiliza, aun cuando se basen en sistemas algorítmicos desarrollados por terceros. En consecuencia, el uso de IA sin mecanismos de transparencia, revisión y control humano puede constituir una violación directa al debido proceso (Constitución Política de los Estados Unidos Mexicanos, 2026).

Algunas convenciones internacionales señalan que, para ser compatibles con los derechos humanos, los sistemas de IA deben garantizar el derecho a una explicación significativa, la posibilidad de impugnar decisiones automatizadas y la existencia de un control humano efectivo, particularmente en contextos de alto impacto jurídico y social (UNESCO, 2022; OECD, 2024). La explicabilidad algorítmica, por tanto, no debe entenderse como una exigencia meramente técnica,

sino como una condición jurídica esencial para la tutela judicial efectiva y la rendición de cuentas en sociedades digitalizadas.

La expansión de la Inteligencia Artificial obliga a replantear las garantías del debido proceso, incorporando criterios de transparencia, explicabilidad y supervisión humana, sin los cuales la automatización decisonal corre el riesgo de erosionar principios básicos del constitucionalismo democrático y del Estado de derecho.

Desde una perspectiva garantista, la decisión jurisdiccional constituye un acto eminentemente humano que involucra interpretación normativa, ponderación de principios, valoración probatoria y apreciación contextual del caso concreto. La doctrina contemporánea coincide en que los sistemas de IA, aun los más avanzados, carecen de discernimiento jurídico autónomo, pues operan mediante correlaciones estadísticas y modelos probabilísticos entrenados sobre datos pasados, sin comprensión normativa ni juicio ético (De Blas, 2025).

La sustitución de la valoración judicial humana por una decisión automatizada vulnera el núcleo esencial del debido proceso, en la medida en que elimina la deliberación racional que justifica la autoridad de las resoluciones judiciales. El uso de IA en tribunales únicamente puede concebirse como una herramienta de apoyo, nunca como un sustituto del juez natural, precisamente para evitar la erosión de la independencia judicial y de la responsabilidad decisoria (UNESCO, 2025).

La falta de transparencia en los algoritmos en el uso de la IA en la administración de justicia representa un riesgo jurídico, ante la imposibilidad de conocer, comprender y explicar el razonamiento interno que conduce al resultado decisonal. Esta opacidad resulta incompatible con el derecho de defensa, pues impide a las partes conocer los criterios que influyen en la decisión y, en consecuencia, controvertirlos eficazmente (Dancy & Zalnieriute, 2025).

La jurisprudencia comparada ha evidenciado este problema en el uso de sistemas de evaluación de riesgos y decisiones administrativas automatizadas, donde la falta de acceso al código fuente, a las variables utilizadas y a los datos de entrenamiento imposibilita que una persona impugne una decisión cuyo fundamento le es

inaccesible e ininteligible (Bechara y otros., 2024). La motivación de las resoluciones judiciales constituye una garantía imprescindible del debido proceso y un presupuesto de la legitimidad democrática del poder judicial. Una decisión que no puede ser explicada de manera racional, clara y verificable rompe el vínculo entre autoridad y razón, sustituyéndolo por una lógica tecnocrática ajena al derecho.

Los sistemas algorítmicos que generan decisiones sin ofrecer fundamentos comprensibles incumplen el deber de motivación, en la medida en que impiden reconstruir el razonamiento lógico-jurídico que conduce a la resolución. La doctrina advierte que los modelos decisionales carentes de explicaciones resultan incompatibles con las exigencias del debido proceso (Giri, 2025). Desde esta perspectiva, la explicabilidad algorítmica se configura como un estándar mínimo indispensable para cualquier utilización legítima de la inteligencia artificial en el ámbito de la justicia.

La utilización de IA también plantea interrogantes sobre la posibilidad real de impugnación y la determinación de responsabilidades. Si una resolución se basa, total o parcialmente, en un sistema algorítmico, resulta imprescindible identificar quién responde por los errores, sesgos o daños causados: el programador, la entidad pública, el proveedor tecnológico o el operador judicial que la utiliza.

Desde la óptica del acceso a la justicia, la decisión final debe ser siempre imputable a una autoridad humana claramente identificable, que asuma la responsabilidad jurídica del acto y permita su revisión por instancias superiores. La clasificación de los sistemas de IA judicial como alto riesgo, adoptada recientemente en marcos regulatorios comparados, responde precisamente a la necesidad de reforzar garantías de control, supervisión humana y rendición de cuentas (OECD, 2025; UNESCO, 2025). La integración de la inteligencia artificial en procesos judiciales y administrativos solo es compatible con el debido proceso y el acceso a la justicia si se cumplen condiciones estrictas de transparencia, explicabilidad, control humano y responsabilidad jurídica. El uso de sistemas algorítmicos cuyas decisiones no

pueden comprenderse ni evaluarse críticamente no solo debilita las garantías del debido proceso, sino que también pone en riesgo la confianza y legitimidad del sistema de justicia en su conjunto, al sustituir el razonamiento jurídico por criterios puramente técnicos de eficiencia que resultan incompatibles con el Estado constitucional de derecho.

La libertad de expresión, afectada por sistemas automatizados de moderación de contenidos.

La moderación de contenidos en plataformas digitales se ha convertido en uno de los principales espacios de controversia para el ejercicio de la libertad de expresión. Frente al volumen de información que circula en entornos digitales, las plataformas recurren a sistemas automatizados basados en inteligencia artificial (IA) para detectar, priorizar, bloquear, desindexar o eliminar contenidos. Si bien estas tecnologías se presentan como soluciones para enfrentar problemas como el discurso de odio, la desinformación o el extremismo violento, su implementación plantea riesgos para los derechos humanos, en particular para la libertad de expresión y el acceso a la información (Gillespie, 2018; Llansó y otros., 2020). Jurídicamente, no solo se trata de poner límites a la libertad de expresión, sino en cómo y bajo qué garantías se imponen dichos límites cuando las decisiones son emitidas por sistemas automatizados poco transparentes, poco explicables y carentes de control efectivo (Oliva, 2020).

El derecho a la libertad de expresión está consagrado en instrumentos fundamentales como el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos (PIDCP), el artículo 13 de la Convención Americana sobre Derechos Humanos y el artículo 10 del Convenio Europeo de Derechos Humanos. Estos marcos normativos establecen que toda limitación a la expresión debe cumplir con los principios de legalidad, finalidad legítima, necesidad y proporcionalidad.

En México, la libertad de expresión esta regulada principalmente en el artículo 6 y el artículo 7 de la Constitución Política de los Estados Unidos Mexicanos. El primero garantiza el derecho a la manifestación de ideas con límites vinculados a la protección de derechos de terceros y la seguridad pública, mientras que el segundo, establece que la libertad de difundir opiniones, información e ideas a través de cualquier medio, no puede ser objeto de censura previa, aún si está sujeta a responsabilidades posteriores (Constitución Política de México, 2026).

La Relatora Especial de Naciones Unidas sobre la promoción y protección del derecho a la libertad de opinión y expresión afirma que los sistemas automatizados de moderación no cumplen adecuadamente estos estándares, especialmente cuando operan sin supervisión humana significativa y sin mecanismos efectivos de reparación (Khan, 2025). En particular, la automatización tiende a generar restricciones imprecisas y descontextualizadas, lo que deriva en la eliminación de expresiones lícitas, afectando de manera desproporcionada a periodistas, defensores de derechos humanos y minorías lingüísticas o culturales.

Los algoritmos de moderación presentan serias dificultades para interpretar elementos esenciales del discurso humano, como la ironía, la sátira, el lenguaje figurado o el contexto histórico y cultural. Esto provoca fenómenos de sobre-eliminación de contenido legítimo, así como sub-eliminación de discursos verdaderamente dañinos (Gillespie, 2018; Oliva, 2020). Las plataformas ejercen un papel de reguladores privados del discurso público, su capacidad para decidir qué expresiones permanecen visibles o son eliminadas las convierte en un guardián de acceso a quienes controlan a un espacio, recurso o información de la esfera pública digital (Gillespie, 2018).

Las plataformas digitales utilizan la regulación por contrato como supuesta libertad de elección del usuario, basada en términos y condiciones, que en ocasiones desplaza los estándares del derecho público y coloca la libertad de expresión bajo parámetros corporativos frecuentemente alejados del enfoque basado en derechos humanos. La Relatora Especial de la ONU ha señalado que esta privatización de la

gobernanza del discurso socava el principio de legalidad y debilita el Estado de derecho en el entorno digital (Khan, 2025).

La UNESCO ha señalado que el respeto a la dignidad exige que las personas conserven un control significativo sobre los sistemas que procesan su información y sobre las decisiones que se adoptan a partir de ella, lo cual implica transparencia, consentimiento informado y posibilidad de intervención humana efectiva (UNESCO, 2025).

III. Jurisprudencia internacional y estándares judiciales sobre tecnologías algorítmicas y derechos humanos

La dignidad humana también se expresa en la exigencia de valoración individualizada de cada caso concreto, principio cardinal del derecho y límite material a cualquier forma de automatización decisonal. Los algoritmos operan a partir de generalizaciones y patrones estadísticos, lo que dificulta la consideración de circunstancias personales, contextuales o excepcionales que no encajan en el modelo predefinido.

La sustitución de la valoración individual por categorías algorítmicas estandarizadas supone un fenómeno en el que las decisiones legales o judiciales pierden su carácter reflexivo, ético y contextual, al ser sustituidas o dominadas por criterios técnicos, automatizados o puramente normativos. La resolución jurídica entonces se reduce a la aplicación mecánica de reglas o algoritmos, sin considerar la complejidad de las circunstancias humanas.

Como ha señalado la doctrina comparada, la justicia no puede reducirse a una operación aritmética sin atender a la singularidad del sujeto afectado, pues ello convertiría a la persona en un objeto de gestión y no en un titular de derechos (Giri, 2025).

Actualmente la jurisprudencia internacional sobre Inteligencia Artificial (IA) frente a los derechos humanos, muestra avances normativos y doctrinales en organismos como la Unión Europea, la Organización de las Naciones Unidas y la Corte Interamericana de Derechos Humanos que establecen criterios de protección. Estos desarrollos buscan garantizar que el uso de IA respete la dignidad humana, la privacidad y la no discriminación.

Jurisprudencia interamericana

La jurisprudencia interamericana ha establecido que el uso de tecnologías por parte del Estado constituye un ámbito especialmente sensible para la protección de los derechos humanos, en particular del derecho a la vida privada, a la protección de datos personales y a la libertad de expresión.

En este marco, los casos *Escher y otros vs. Brasil* y *Tristán Donoso vs. Panamá*, resueltos por la Corte Interamericana de Derechos Humanos en 2009, representan precedentes fundamentales, ya que fijan estándares jurídicos construidos. Particularmente en el caso *Escher y otros vs. Brasil*, la Corte sostuvo que la interceptación de comunicaciones, aun cuando estuviera orientada a fines legítimos de investigación penal, constituye una injerencia grave en la vida privada y solo resulta compatible con la Convención Americana si cumple estrictamente con los principios de legalidad, necesidad y proporcionalidad. La Corte enfatizó que la protección convencional no se agota en el contenido de la comunicación, sino que abarca toda forma de intercambio informativo, independientemente del medio técnico utilizado. Este razonamiento resulta particularmente relevante frente a tecnologías contemporáneas de vigilancia digital, recolección masiva de datos, monitoreo automatizado y análisis algorítmico, cuya capacidad intrusiva es sustancialmente mayor que la evaluada en el caso concreto.

Este estándar jurisprudencial es retomado y actualizado por la Comisión Interamericana de Derechos Humanos (CIDH) en su **Informe “Inteligencia Artificial y Derechos Humanos” (2023)**, en el que se advierte que el uso de

sistemas algorítmicos por parte del Estado, especialmente en seguridad pública, justicia y políticas sociales, amplifica los riesgos de vigilancia masiva, arbitrariedad y afectaciones a la privacidad. La CIDH establece que estas tecnologías incrementan exponencialmente el poder estatal de injerencia, por lo que **la obligación de regulación clara, control judicial y supervisión humana debe ser reforzada**, retomando la lógica desarrollada por la Corte en casos como *Escher*.

Por su parte, en *Tristán Donoso vs. Panamá*, la Corte Interamericana profundizó en el vínculo entre tecnología y derechos humanos al analizar la obtención y divulgación indebida de información privada mediante medios técnicos. El Tribunal sostuvo que la recopilación, almacenamiento, tratamiento y difusión de datos personales por parte del Estado se encuentran protegidos por el derecho a la vida privada, y que la ausencia de salvaguardas normativas y controles efectivos genera responsabilidad internacional. Este criterio anticipa uno de los problemas centrales identificados por la CIDH en 2023 referente a la **falta de transparencia algorítmica**, entendida como la imposibilidad de conocer, comprender o impugnar decisiones basadas en sistemas automatizados.

El Informe CIDH 2023 reconoce que los sistemas de inteligencia artificial no son neutrales y que, cuando operan sin transparencia ni mecanismos de rendición de cuentas, afectan no solo la privacidad, sino también el debido proceso, la igualdad y la dignidad humana. En consonancia con *Tristán Donoso*, la Comisión advierte que el procesamiento automatizado de datos personales y el perfilamiento algorítmico pueden consolidar decisiones arbitrarias si no existen límites normativos claros, supervisión humana efectiva y recursos para la impugnación de tales decisiones.

Estos estándares fueron reforzados en el **Comunicado Conjunto sobre Inteligencia Artificial y Libertad de Expresión**, emitido el 7 de mayo de 2025 por los Relatores y Relatorias de libertad de expresión de la Organización de las Naciones Unidas (ONU), la Organización de los Estados Americanos (OEA), la

Organización para Seguridad y Cooperación en Europa (OSCE) y la Comisión Africana de Derechos Humanos y de los Pueblos (CADHP). En dicho comunicado, se afirma que el rápido desarrollo de la IA ha avanzado sin una adecuada integración de los derechos humanos desde las fases iniciales de diseño y despliegue, lo que ha generado impactos en la libertad de expresión, la integridad del ecosistema informativo y la deliberación democrática.

El Comunicado de 2025 retoma de manera implícita la doctrina de *Escher y Tristán Donoso* al advertir que los sistemas algorítmicos pueden constituir **formas contemporáneas de censura indirecta y vigilancia desproporcionada**, especialmente cuando se utilizan para moderar contenidos, priorizar información o perfilar personas sin base legal clara ni controles adecuados. Asimismo, enfatiza que la eficiencia técnica, la rapidez o la rentabilidad comercial **no pueden justificar restricciones amplias o automatizadas a la libertad de expresión**, reafirmando la primacía de la dignidad humana como eje rector del desarrollo tecnológico.

De esta manera, la lectura integrada de la jurisprudencia de la Corte IDH, el Informe CIDH 2023 y el Comunicado Conjunto 2025, permite afirmar que el Sistema Interamericano ha consolidado un **estándar normativo coherente y progresivo** sobre el uso estatal de tecnologías. Dicho estándar establece que no existen vacíos jurídicos frente a las tecnologías emergentes y que el avance técnico **intensifica**, y no atenúa, las obligaciones estatales de respeto, protección y garantía de los derechos humanos.

En consecuencia, desde una perspectiva de control de convencionalidad, los Estados están obligados a someter el uso de sistemas algorítmicos e inteligencia artificial a los principios de legalidad, necesidad y proporcionalidad, garantizar la transparencia y explicabilidad de las decisiones automatizadas, asegurar la supervisión humana efectiva y prevenir impactos discriminatorios o arbitrarios. Bajo este marco, la innovación tecnológica solo resulta compatible con el derecho interamericano cuando se encuentra **subordinada a la dignidad humana y al**

fortalecimiento de la democracia, y no cuando opera como un mecanismo de poder que erosiona los derechos fundamentales.

Jurisprudencia en Europa

La regulación de la inteligencia artificial en la Unión Europea representa uno de los avances normativos más significativos del constitucionalismo digital contemporáneo. El Reglamento de Inteligencia Artificial de la Unión Europea (2024), adoptado formalmente en junio de 2024, constituye la primera norma jurídica integral en el mundo destinada a regular de manera transversal el desarrollo, uso y despliegue de sistemas de inteligencia artificial. Su relevancia no radica únicamente en su carácter pionero, sino en su fundamento axiológico, centrado expresamente en la protección de los derechos humanos y las libertades fundamentales.

Desde una perspectiva doctrinal, la inteligencia artificial no puede analizarse como una tecnología neutral, sino como un dispositivo de poder normativo y fáctico, capaz de incidir de manera directa y sistemática en la dignidad humana, la privacidad, la igualdad, la autonomía personal y el acceso a la justicia (Floridi y otros., 2018). Dentro de este entorno, el enfoque de la Unión Europea parte de la premisa de que el progreso tecnológico solo es legítimo cuando se encuentra jurídicamente subordinado al Estado de Derecho y a la centralidad de la persona humana.

El Reglamento de Inteligencia Artificial (UE, 2024) es la primera norma vinculante a nivel mundial que establece un marco jurídico uniforme para el desarrollo, comercialización y uso de sistemas de IA. Este dispositivo jurídico adopta un enfoque de clasificación de sistemas de inteligencia artificial en función del nivel de riesgo que generan para las personas. Esta técnica regulatoria encuentra un claro paralelismo con la jurisprudencia consolidada del Tribunal de Justicia de la Unión Europea en materia de tratamiento automatizado de datos y vigilancia tecnológica.

Caso Digital Rights Ireland Ltd y otros

En 2006, el Parlamento Europeo y el Consejo de la Unión Europea adoptaron la Directiva 2006/24/CE, conocida como Directiva de Retención de Datos, cuyo objetivo era armonizar en los Estados miembros la obligación de los proveedores de servicios de telecomunicaciones e Internet de conservar determinados datos relativos al origen, destino y momento de las comunicaciones, con la finalidad de apoyar la prevención, investigación y persecución de delitos graves en la Unión Europea.

La legalidad de las medidas nacionales de transposición de dicha Directiva fue cuestionada en procedimientos paralelos en Irlanda y Austria. En el caso irlandés, la organización Digital Rights Ireland Ltd. alegó que la Directiva constituía la base jurídica de regímenes de vigilancia masiva, incompatibles con los derechos fundamentales a la privacidad y a la protección de datos personales. Tanto el Alto Tribunal de Irlanda como el Tribunal Constitucional de Austria consideraron que no podían evaluar la constitucionalidad de las normas internas sin antes determinar la validez de la propia Directiva, por lo que remitieron cuestiones prejudiciales al Tribunal de Justicia de la Unión Europea (TJUE), las cuales fueron acumuladas en los asuntos C-293/12 y C-594/12.

En su sentencia de 8 de abril de 2014, el TJUE analizó la compatibilidad de la Directiva con los artículos 7 y 8 de la Carta de los Derechos Fundamentales de la Unión Europea, relativos al derecho al respeto de la vida privada y al derecho a la protección de los datos personales. Si bien el Tribunal reconoció que la finalidad de la Directiva, la lucha contra la criminalidad grave, constituía un objetivo legítimo y concluyó que las medidas establecidas no superaban el test de proporcionalidad exigido por el artículo 52.1 de la Carta (Tribunal de Justicia de la Unión Europea [TJUE], 2014).

El TJUE destacó que la Directiva permitía una injerencia amplia y generalizada en los derechos fundamentales de la totalidad de la población de la Unión, al imponer la retención sistemática de datos durante períodos de entre seis y veinticuatro

meses, sin establecer criterios objetivos y claros que limitaran el alcance del acceso a los datos ni garantías suficientes respecto de su almacenamiento, uso y protección. En particular, subrayó la ausencia de reglas precisas sobre el acceso por parte de las autoridades competentes y la falta de salvaguardas efectivas frente a usos indebidos de la información.

Como resultado, el Tribunal declaró que la Directiva 2006/24/CE era incompatible con la Carta de los Derechos Fundamentales y, en consecuencia, la declaró inválida en su totalidad. Asimismo, afirmó que el legislador europeo dispone únicamente de un margen de apreciación limitado cuando las medidas adoptadas implican restricciones a derechos fundamentales, lo que justifica un control jurisdiccional estricto destinado a garantizar que el desarrollo tecnológico y las políticas de seguridad no erosionen los derechos humanos fundamentales (TJUE, 2014).

En el asunto anterior, del caso Digital Rights Ireland Ltd contra Minister for Communications, Marine and Natural Resources (2014), el Tribunal de Justicia de la Unión Europea declaró inválida la Directiva 2006/24/CE relativa a la conservación de datos, al estimar que la retención masiva e indiscriminada de datos constituía una injerencia desproporcionada en los derechos al respeto de la vida privada y a la protección de datos personales. El Tribunal subrayó que los sistemas tecnológicos que afectan de forma general a toda la población, sin diferenciación ni límites precisos, suponen una vulneración del contenido esencial de los derechos fundamentales (Tribunal de Justicia de la Unión Europea, 2014).

Este razonamiento jurisprudencial es directamente trasladable al Reglamento de Inteligencia Artificial (UE, 2024), cuyo sistema de riesgos graduales materializa un principio de proporcionalidad reforzada, conforme al cual el aumento del riesgo tecnológico exige un incremento correlativo de las salvaguardas jurídicas.

Por otra parte, El Reglamento de Inteligencia Artificial (UE, 2024) prohíbe de forma expresa determinados sistemas de inteligencia artificial considerados de riesgo inaceptable, tales como la manipulación cognitiva de personas o colectivos vulnerables, la puntuación social y la identificación biométrica remota en tiempo real

en espacios públicos. Estas prohibiciones no responden a criterios de oportunidad política, sino a la lógica propia del Derecho internacional de los derechos humanos, la cual indica que existen prácticas tecnológicas incompatibles con la dignidad humana, con independencia de su eficacia o rentabilidad.

El Tribunal Europeo de Derechos Humanos ha sostenido reiteradamente que la vigilancia tecnológica masiva genera un efecto disuasorio incompatible con una sociedad democrática. En el caso *S. y Marper contra Reino Unido* (2008), el Tribunal determinó que la conservación indiscriminada de datos biométricos vulnera el derecho al respeto de la vida privada reconocido en el artículo 8 del Convenio Europeo de Derechos Humanos. Esta doctrina fue reforzada en *Big Brother Watch y otros contra Reino Unido* (2021), donde se estableció que los sistemas de vigilancia deben estar sujetos a controles estrictos, previsibilidad legal y supervisión independiente.

El Reglamento de Inteligencia Artificial (UE, 2024) incorpora de manera normativa esta jurisprudencia, estableciendo límites absolutos que protegen el núcleo irreductible de los derechos humanos frente al poder algorítmico. Los sistemas de inteligencia artificial clasificados como de alto riesgo coinciden con sectores estratégicos desde la perspectiva de los derechos fundamentales, tales como la educación, el empleo, el acceso a servicios públicos, la aplicación de la ley, la migración y la administración de justicia. Esto impone obligaciones de evaluación previa, supervisión humana significativa, trazabilidad y mecanismos de reclamación, en consonancia con los estándares desarrollados por el Tribunal de Justicia de la Unión Europea.

En el asunto *Data Protection Commissioner contra Facebook Ireland Limited y Maximilian Schrems* (2020), el Tribunal estableció que los sistemas de tratamiento automatizado solo son compatibles con el orden jurídico europeo si garantizan un nivel de protección esencialmente equivalente al previsto en la Unión Europea, incluyendo recursos judiciales efectivos (Tribunal de Justicia de la Unión Europea, 2020). De manera complementaria, el Tribunal Europeo de Derechos Humanos ha

señalado que el uso de tecnologías automatizadas en el entorno laboral o administrativo solo es legítimo cuando respeta los principios de necesidad, idoneidad y proporcionalidad, como se estableció en *López Ribalda y otros contra España* (2019).

El Reglamento de Inteligencia Artificial (UE, 2024) introduce exigencias específicas de transparencia para los sistemas de inteligencia artificial generativa y para los contenidos sintéticos, en particular las imágenes, audios y videos generados o modificados artificialmente. Desde un enfoque de derechos humanos, estas obligaciones se vinculan directamente con el derecho a la información, la identidad personal y la tutela judicial efectiva.

El Tribunal de Justicia de la Unión Europea ha señalado que la falta de transparencia de los sistemas automatizados impide el ejercicio efectivo del derecho a impugnar decisiones que afectan a los individuos, vulnerando el derecho a un recurso efectivo (Tribunal de Justicia de la Unión Europea, 2020). De igual forma, el Tribunal Europeo de Derechos Humanos ha reconocido la necesidad de proteger la vida privada y la reputación frente a injerencias tecnológicas injustificadas, como se observa en la jurisprudencia *Von Hannover contra Alemania* (2004; 2012).

El RIA de la Unión Europea constituye una manifestación concreta del constitucionalismo digital europeo, al integrar de manera sistemática los principios derivados de la jurisprudencia del Tribunal de Justicia de la Unión Europea y del Tribunal Europeo de Derechos Humanos. A través de una regulación basada en el riesgo, la transparencia y la supervisión humana, la Unión Europea establece un modelo normativo donde la innovación tecnológica se encuentra jurídicamente condicionada por la dignidad humana, la igualdad y el control democrático del poder algorítmico.

IV. La Jurisprudencia mexicana sobre libertad de expresión y tecnologías digitales

Si bien el sistema jurídico mexicano aún no cuenta con pronunciamientos judiciales específicos sobre la moderación automatizada de contenidos mediante inteligencia artificial en plataformas digitales, la Suprema Corte de Justicia de la Nación (SCJN) ha desarrollado una línea jurisprudencial sólida, coherente y progresiva en materia de libertad de expresión en entornos digitales, cuyos principios constitucionales resultan plenamente aplicables al análisis de tecnologías automatizadas que inciden en el discurso público.

En particular, la SCJN ha reconocido que Internet y las redes sociales constituyen espacios centrales de la esfera pública contemporánea, en los que la libertad de expresión y el derecho de acceso a la información desempeñan una función claramente constitucionalista en materia de derechos humanos. A partir de esta premisa, la Corte ha establecido que toda restricción al discurso en entornos digitales debe someterse a un escrutinio reforzado de constitucionalidad, especialmente cuando se trate de expresiones vinculadas con asuntos de interés público, crítica a funcionarios o participación ciudadana.

Estos criterios permiten afirmar que los sistemas algorítmicos de moderación automatizada producen efectos jurídicamente relevantes sobre el ejercicio de la libertad de expresión, aun cuando operen a través de decisiones técnicas aparentemente neutrales. Desde una perspectiva constitucional, tales tecnologías deben ser evaluadas conforme a los principios de proporcionalidad, prohibición de censura indirecta, prevención del efecto inhibitorio, transparencia decisoria y control efectivo, así como a la exigencia de que toda limitación al discurso público sea razonable y justificable.

En relación con ello, la jurisprudencia de la SCJN sobre libertad de expresión digital y, más recientemente, sobre el uso ético y responsable de la inteligencia artificial, ofrece un marco normativo suficiente para analizar críticamente la compatibilidad

constitucional de la moderación automatizada de contenidos, aun en ausencia de pronunciamientos específicos sobre esta tecnología.

La incorporación de herramientas de inteligencia artificial (IA) en los procesos jurisdiccionales replantea los límites entre innovación tecnológica, función jurisdiccional y garantías procesales. En el sistema jurídico mexicano, este debate ha transitado de criterios aislados de autocontención judicial a la consolidación de un estándar jurisprudencial obligatorio, como lo demuestra la jurisprudencia II.2o.C. J/2 K (12a.), emitida por el Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito (Suprema Corte de Justicia de la Nación, 2026).

La jurisprudencia referida tiene como antecedente inmediato la tesis aislada II.2o.C.9 K (11a.), en la cual un Tribunal Colegiado, ante la inexistencia de una regulación normativa expresa, estableció de manera precautoria los parámetros aplicables al uso ético y responsable de la inteligencia artificial en la tramitación de recursos de queja y de revisión derivados de juicios de amparo indirecto. Dicho criterio respondió a la necesidad de salvaguardar principios fundamentales del proceso constitucional frente a la incorporación de tecnologías emergentes en el ámbito jurisdiccional.

Posteriormente, la reiteración de este criterio en diversos asuntos permitió que dichos parámetros trascendieran su carácter orientador y adquirieran fuerza obligatoria, de conformidad con el sistema de precedentes establecido en el Acuerdo General Plenario 7/2025. En este sentido, la Suprema Corte de Justicia de la Nación consolidó un marco vinculante que reconoce la validez normativa de los criterios jurisprudenciales reiterados, fortaleciendo así el control jurídico del uso de la inteligencia artificial en la función jurisdiccional (SCJN, 2026).

Desde una perspectiva de derechos humanos, la jurisprudencia establece que la utilización de herramientas de inteligencia artificial en procesos jurisdiccionales exige la observancia de cuatro elementos mínimos: (a) proporcionalidad e inocuidad; (b) protección de datos personales; (c) transparencia y explicabilidad; y (d) supervisión y decisión humanas (SCJN, 2026). Estos elementos no constituyen

simples recomendaciones técnicas, sino auténticos estándares constitucionales de actuación judicial, vinculados directamente con el derecho al debido proceso, la tutela judicial efectiva y la garantía de una judicatura independiente e imparcial.

El principio de proporcionalidad e inocuidad delimita el ámbito legítimo de uso de la inteligencia artificial, restringiéndolo a funciones estrictamente auxiliares, como la facilitación de cálculos matemáticos o razonamientos numéricos necesarios para la tramitación del proceso. La jurisprudencia enfatiza que tales herramientas no pueden extenderse al razonamiento jurídico interpretativo ni a la aplicación de normas al caso concreto, preservando así el núcleo esencial de la función jurisdiccional y evitando una delegación indebida del juicio jurídico en sistemas automatizados (SCJN, 2026).

En segundo lugar, la protección de datos personales se erige como un límite infranqueable al uso de la inteligencia artificial en sede judicial. La jurisprudencia reconoce que los expedientes judiciales contienen información sensible cuya exposición indebida puede vulnerar derechos fundamentales, por lo que el empleo de herramientas tecnológicas debe garantizar la confidencialidad, seguridad y tratamiento legítimo de los datos procesales, en consonancia con los principios constitucionales de privacidad y autodeterminación informativa (SCJN, 2025).

El tercer elemento, relativo a la transparencia y explicabilidad, representa una garantía esencial frente al riesgo de opacidad algorítmica. La jurisprudencia señala que la persona juzgadora no solo debe informar que emplea herramientas de inteligencia artificial, sino que está obligada a explicar la metodología utilizada, los datos considerados y la forma en que se obtuvo el resultado. Este estándar resulta indispensable para permitir la contradicción procesal, la revisión por órganos superiores y el control democrático de la función jurisdiccional, evitando decisiones técnicamente incomprensibles o inaccesibles para las partes (SCJN, 2026).

Finalmente, la supervisión y decisión humanas constituyen el eje rector del criterio jurisprudencial. La Corte reafirma que la inteligencia artificial debe operar exclusivamente como un instrumento auxiliar y nunca como sustituto de la

deliberación judicial. La valoración de los hechos, la interpretación normativa y la decisión final deben permanecer bajo control humano, como salvaguarda frente a sesgos algorítmicos y como garantía de que la justicia conserve su carácter racional, responsable y humanista (SCJN, 2026).

Resulta particularmente relevante que esta jurisprudencia obligatoria se fundamente en estándares internacionales en materia de ética de la inteligencia artificial, como las Directrices Éticas para una Inteligencia Artificial Fiable de la Comisión Europea y la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO.

Al respecto, se evidencia una clara apertura del Poder Judicial mexicano al diálogo normativo transnacional y propone un modelo de justicia digital compatible con los derechos humanos y el constitucionalismo democrático.

A continuación, se mencionan algunos criterios claves respecto a derechos humanos y contenidos digitales emanados del Poder Judicial de la Federación en México.

Internet como medio esencial para la libertad de expresión.

En el Amparo en Revisión 1/2017, la Segunda Sala de la SCJN reconoció expresamente que Internet constituye un medio indispensable para el ejercicio del derecho a la libertad de expresión y de acceso a la información, tanto en su dimensión individual como colectiva (SCJN, 2017). La Corte sostuvo que las medidas de bloqueo o restricción al acceso a contenidos digitales equivalen a formas de censura indirecta, prohibidas por el artículo 7º constitucional, y estableció que cualquier limitación debe superar un test estricto de legalidad, necesidad y proporcionalidad.

Este criterio resulta particularmente relevante para el análisis de la moderación automatizada de contenidos, en tanto la eliminación masiva o indiscriminada de información lícita mediante sistemas algorítmicos produce efectos equiparables a un bloqueo generalizado del discurso en línea, aun cuando no provenga directamente de una autoridad estatal.

Libertad de expresión en redes sociales y decisiones técnicas de exclusión

En la tesis aislada con Registro Digital 2022074, derivada del Amparo en Revisión 468/2019, los Tribunales Colegiados de Circuito determinaron que cuando un servidor público utiliza redes sociales como canal de comunicación gubernamental, dichas cuentas adquieren la naturaleza de foros públicos digitales, por lo que bloquear usuarios por razones críticas vulnera los derechos de libertad de expresión y acceso a la información (Tribunales Colegiados de Circuito, 2020).

Este precedente cobra especial importancia frente a los sistemas automatizados de moderación, dado que las decisiones técnicas que restringen la visibilidad, interacción o permanencia del contenido afectan directamente el núcleo del derecho a la expresión y deben ser sometidas a estándares constitucionales de control.

Estándares reforzados de protección del discurso de interés público

En 2025, la Primera Sala de la SCJN consolidó una jurisprudencia de protección reforzada de la libertad de expresión mediante las tesis 126/2025 y 128/2025, en las que estableció que las opiniones, juicios de valor y expresiones relacionadas con asuntos de interés público gozan de una presunción de protección constitucional, limitando severamente la posibilidad de imponer restricciones o sanciones (SCJN, 2025a; SCJN, 2025b).

Estos criterios son particularmente relevantes en el contexto de la moderación automatizada, ya que los sistemas algorítmicos carecen de la capacidad de diferenciar adecuadamente entre hechos verificables, opiniones y discursos protegidos, incrementando el riesgo de supresión ilegítima de expresiones constitucionalmente tuteladas.

Uso de inteligencia artificial y exigencia de control humano significativo

Respecto al uso de la inteligencia artificial y exigencia de control humano significativo, aunque fuera del ámbito directo de la moderación de contenidos, el Poder Judicial Federal ha comenzado a pronunciarse sobre el uso de inteligencia artificial en la función jurisdiccional. En 2025, un Tribunal Colegiado estableció que la IA solo puede emplearse como herramienta auxiliar, jamás como sustituto del juicio humano, y que su utilización debe respetar los principios de transparencia, explicabilidad, proporcionalidad, protección de datos personales y supervisión humana (Tribunal Colegiado en Materia Civil del Segundo Circuito, 2025).

Este criterio refuerza un principio transversal aplicable a la moderación algorítmica de que, toda decisión automatizada que incida en derechos fundamentales exige control humano significativo, especialmente cuando afecta el ejercicio de la libertad de expresión en espacios digitales.

A partir de la convergencia con estándares internacionales de derechos humanos, la jurisprudencia mexicana en materia de libertad de expresión digital resulta consistente con los estándares desarrollados por la Corte Interamericana de Derechos Humanos, pues reconoce que los avances tecnológicos no eximen al Estado ni a otros actores con poder de respetar los principios de legalidad, necesidad y proporcionalidad en la restricción de derechos fundamentales (SCJN, 2017).

V. Modelos normativos para la regulación de la inteligencia artificial

La evolución de la inteligencia artificial (IA) ha impulsado propuestas de modelos normativos diferenciados orientados a regular sus riesgos e impactos sobre los derechos humanos. Estos modelos pueden agruparse en dos categorías: la regulación vinculante de naturaleza jurídica (*hard law*) y los estándares internacionales de carácter no vinculante (*soft law*), los cuales han convergido progresivamente hacia un enfoque basado en derechos humanos como paradigma regulatorio predominante.

El término *Hard law* se refiere al conjunto de normas jurídicas formalmente vinculantes, adoptadas mediante procedimientos legislativos o constitucionales reconocidos, que imponen obligaciones jurídicas exigibles y cuya inobservancia genera consecuencias legales. En segundo término, el *soft law* se refiere a recomendaciones, declaraciones, principios, directrices, códigos de conducta y marcos éticos adoptados en el ámbito internacional o regional, que buscan establecer estándares comunes, orientar políticas públicas y anticipar respuestas regulatorias frente a tecnologías emergentes cuya naturaleza es no vinculante.

Un ejemplo de regulación vinculante es el Reglamento Europeo de Inteligencia Artificial. El Reglamento (UE) 2024/1689, conocido como Artificial Intelligence Act (AI Act), constituye el primer marco jurídico vinculante de alcance general destinado a regular la inteligencia artificial de forma integral. Adoptado por el Parlamento Europeo y el Consejo el 13 de junio de 2024, el AI Act tiene como objetivo garantizar el desarrollo y uso de sistemas de IA centrados en el ser humano y confiables, asegurando simultáneamente un alto nivel de protección de los derechos fundamentales consagrados en la Carta de Derechos Fundamentales de la Unión Europea (Unión Europea, 2024).

El modelo normativo del AI Act se estructura a partir de un enfoque basado en riesgos, que clasifica los sistemas de IA en cuatro categorías: riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo. Las prácticas consideradas de riesgo inaceptable, como determinadas formas de manipulación del comportamiento, la

puntuación social o ciertos usos de identificación biométrica remota, son expresamente prohibidas por su incompatibilidad con la dignidad humana y los derechos fundamentales. Por su parte, los sistemas de alto riesgo quedan sujetos a obligaciones reforzadas que incluyen sistemas de gestión de riesgos, gobernanza de datos, documentación técnica, transparencia, trazabilidad, supervisión humana y garantías de precisión y robustez (Unión Europea, 2024).

Particularmente para el enfoque de derechos humanos, el AI Act incorpora el principio de supervisión humana significativa, reconociendo que las decisiones automatizadas no deben desplazar el juicio humano cuando están en juego derechos fundamentales. Asimismo, su alcance proyecta estos estándares más allá del territorio europeo, consolidando al Reglamento como un referente normativo global en la gobernanza de la inteligencia artificial.

Paralelamente a los modelos vinculantes, se ha desarrollado un conjunto robusto de estándares internacionales de soft law, orientados a guiar a los Estados y a los actores privados en la regulación ética y jurídica de la inteligencia artificial. Entre ellos, destaca la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO, adoptada en 2021 por la totalidad de sus Estados miembros (UNESCO, 2021).

Esta recomendación constituye el primer instrumento normativo global sobre ética de la IA, y sitúa a la protección de los derechos humanos, la dignidad humana y las libertades fundamentales como ejes centrales del desarrollo tecnológico. Aunque no vinculante, establece principios normativos claros, tales como la proporcionalidad, la no discriminación, la transparencia, la rendición de cuentas y la supervisión humana, así como la obligación de realizar evaluaciones de impacto ético y social antes del despliegue de sistemas de IA (UNESCO, 2021).

A nivel universal, la Resolución A/RES/78/265 de la Asamblea General de las Naciones Unidas, adoptada en marzo de 2024, reafirma que los sistemas de inteligencia artificial deben ser seguros, confiables y compatibles con el derecho internacional de los derechos humanos. La resolución subraya que los beneficios de la IA solo pueden legitimarse si su desarrollo y uso respetan los derechos

fundamentales durante todo su ciclo de vida, con especial atención a los grupos en situación de vulnerabilidad (Asamblea General de las Naciones Unidas, 2024).

En conjunto, estos instrumentos de derecho blando han desempeñado un papel clave en la armonización progresiva de estándares internacionales, influyendo directamente en legislaciones nacionales y regionales, así como en iniciativas regulatorias vinculantes como el AI Act.

En este marco, la regulación de la inteligencia artificial se concibe no como una excepción al derecho, sino como una extensión del constitucionalismo digital y del derecho internacional de los derechos humanos, reafirmando que los avances tecnológicos no pueden desarrollarse al margen de los principios fundamentales de dignidad, igualdad y justicia.

Conclusiones

La regulación jurídica de la inteligencia artificial frente a los derechos humanos constituye uno de los mayores retos del derecho actual. Los modelos normativos emergentes evidencian una convergencia hacia enfoques preventivos, basados en riesgos y centrados en la dignidad humana. No obstante, su eficacia dependerá de la capacidad de los Estados para implementar estos estándares, reducir brechas y garantizar que la innovación tecnológica esté al servicio de la persona y no en detrimento de sus derechos.

La inteligencia artificial se ha consolidado como un fenómeno tecnológico que incide directamente en el ejercicio y la garantía de los derechos humanos. Lejos de tratarse de una herramienta neutra o meramente técnica, la IA configura un nuevo objeto jurídico regulatorio capaz de reordenar relaciones de poder, redefinir procesos decisionales y alterar las bases clásicas del Estado constitucional de derecho. En esta línea, se considera que la regulación jurídica de la inteligencia artificial no constituye una opción discrecional, sino una exigencia normativa derivada del derecho internacional de los derechos humanos y del constitucionalismo contemporáneo.

En primer lugar, se evidencia que la automatización de decisiones mediante sistemas algorítmicos pone en entredicho las categorías jurídicas fundamentales como la imputación de responsabilidad, la motivación de los actos, la seguridad jurídica y la tutela judicial efectiva. La falta de transparencia, explicabilidad y la ausencia de control humano en muchos sistemas de IA comprometen principios esenciales del debido proceso y debilitan los mecanismos de rendición de cuentas, particularmente cuando estas tecnologías se utilizan en ámbitos de alto impacto jurídico y social como la justicia, la seguridad pública, los programas sociales o la moderación de contenidos.

Desde la perspectiva del ordenamiento jurídico mexicano, el análisis desarrollado confirma que el artículo 1º de la Constitución Política de los Estados Unidos Mexicanos constituye el eje normativo central para la regulación de la inteligencia artificial en México. Dicho precepto no solo reconoce el bloque de constitucionalidad en materia de derechos humanos, sino que impone al Estado una obligación positiva de respeto, protección y garantía que se extiende plenamente al uso de tecnologías emergentes. En consecuencia, ningún sistema de IA, público o privado puede operar legítimamente al margen de los principios de universalidad, interdependencia, indivisibilidad, progresividad y del principio pro persona.

Los sistemas de IA, entrenados sobre datos históricos marcados por desigualdades sociales, tienden a reproducir y amplificar sesgos relacionados con género, origen étnico, condición socioeconómica o discapacidad. Esto convierte a la igualdad y la no discriminación en derechos especialmente vulnerables frente al uso no regulado de tecnologías algorítmicas, exigiendo evaluaciones de impacto en derechos humanos, mecanismos de auditoría y salvaguardas normativas específicas.

De manera complementaria, se ha demostrado que el derecho a la privacidad y a la protección de datos personales enfrenta desafíos profundos ante prácticas de vigilancia masiva, perfilamiento y tratamiento automatizado de información a gran escala. La expansión de tecnologías como el reconocimiento facial o el análisis predictivo amenaza la privacidad y la autodeterminación informativa, lo que obliga

a reafirmar la función garantista del derecho frente a dinámicas tecnológicas que tienden a normalizar la vigilancia y la captura permanente de datos.

Particular atención merece el impacto de la inteligencia artificial sobre el derecho al debido proceso y la tutela judicial efectiva. La imposibilidad de comprender y cuestionar decisiones automatizadas carentes de explicabilidad sitúa a las personas en estados de indefensión jurídica incompatibles con el Estado de derecho. El análisis doctrinal y jurisprudencial confirma que la IA solo puede ser legítimamente utilizada como herramienta de apoyo, pero nunca como sustituto del juicio humano, pues la deliberación, la motivación y la responsabilidad decisoria constituyen el núcleo irreductible del acto jurisdiccional y administrativo.

En relación con la libertad de expresión, se ha evidenciado que los sistemas automatizados de moderación de contenidos introducen riesgos significativos de censura indirecta en contenidos legítimos. La privatización de la gobernanza del discurso mediante algoritmos poco claros y reglas contractuales corporativas debilita el principio de legalidad y exige la incorporación de estándares de derechos humanos en la regulación de plataformas digitales, así como supervisión humana y mecanismos efectivos de reparación.

El análisis de la jurisprudencia internacional y nacional permite afirmar que existe una convergencia normativa progresiva en torno a los principios de legalidad, proporcionalidad y supervisión humana. Tanto la Corte Interamericana de Derechos Humanos, como el Tribunal de Justicia de la Unión Europea, el Tribunal Europeo de Derechos Humanos y la Suprema Corte de Justicia de la Nación, han establecido que los avances tecnológicos no eximen a los Estados de sus obligaciones en derechos humanos, incluso frente a tecnologías no previstas originalmente en los marcos normativos.

En este contexto, en el estudio de los modelos regulatorios se evidencia que la gobernanza de la inteligencia artificial avanza hacia un esquema híbrido, en el que instrumentos de hard law, como el Reglamento Europeo de Inteligencia Artificial, y estándares de soft law, como la Recomendación de la UNESCO y las resoluciones

de la Asamblea General de la ONU, convergen hacia un enfoque basado en derechos humanos (Human Rights-Based Approach). Este paradigma se consolida como el marco regulatorio dominante, al exigir evaluaciones de impacto, transparencia, explicabilidad, control humano efectivo, rendición de cuentas y protección reforzada de grupos en situación de vulnerabilidad.

Finalmente, la regulación jurídica de la inteligencia artificial debe concebirse como una manifestación contemporánea del constitucionalismo humanista y del derecho internacional de los derechos humanos, en el que la tecnología se subordine a la dignidad humana, la igualdad y la justicia. La IA no puede transformar el derecho sin ser, a su vez, transformada por él. Solo mediante una regulación garantista, preventiva y centrada en la persona será posible asegurar que la inteligencia artificial opere como un instrumento al servicio del desarrollo humano y no como un factor de erosión del Estado de derecho y de los derechos fundamentales.

Referencias

ACNUDH. (2021). Los riesgos de la inteligencia artificial para los derechos humanos. Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos. <https://www.idhc.org/es/noticias/los-riesgos-de-la-inteligencia-artificial-para-los-derechos-humanos/>

ACNUDH. (2023). La inteligencia artificial debe basarse en los derechos humanos. Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos. <https://www.ohchr.org/es/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner>

Asamblea General de las Naciones Unidas. (2024). Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development (Res. A/RES/78/265). Naciones Unidas. <https://documents.un.org>

Aveiga Saldarriaga, J. K., Cortes Andachi, F. A., & Zambrano Quiroz, P. H. (2024). El impacto de la inteligencia artificial en el derecho: Nuevos desafíos y oportunidades. <https://lexenlace.com/wp-content/uploads/2024/12/LEX-AC75.pdf>

Bechara, A., Vides, M., Mendoza, F., Caballero, J., & Navas, J. (2024). Inteligencia artificial y su incidencia en el debido proceso. *Jurídicas CUC*, 20(1), 436–447. <https://doi.org/10.17981/juridcuc.20.1.2024.20>

Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>

Citron, D. K. (2008). Technological due process. *Washington University Law Review*, 85(6), 1249–1313.

Comisión Interamericana de Derechos Humanos. (2023). Inteligencia artificial y derechos humanos (OEA/Ser.L/V/II). Organización de los Estados Americanos. <https://www.oas.org>

Consejo de la Judicatura Federal & UNAM. (2014). Garantía de la seguridad jurídica. Biblioteca Jurídica Virtual del Instituto de Investigaciones Jurídicas de la UNAM. <http://biblio.juridicas.unam.mx>

Constitución Política de los Estados Unidos Mexicanos. (2026). Artículos 1º, 6º, 14 y 16. Diario Oficial de la Federación.

Corte Interamericana de Derechos Humanos. (2009a). Caso Escher y otros vs. Brasil (Sentencia de 6 de julio de 2009, Serie C No. 200). <https://www.corteidh.or.cr>

Corte Interamericana de Derechos Humanos. (2009b). Caso Tristán Donoso vs. Panamá (Sentencia de 27 de enero de 2009, Serie C No. 193). <https://www.corteidh.or.cr>

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Dancy, T., & Zalnieriute, M. (2025). AI and transparency in judicial decision making. Oxford Journal of Legal Studies, 46(1), 1–34. <https://doi.org/10.1093/ojls/gqaf030>

De Blas, J. R. (2025). Artificial intelligence and judicial decision-making. European Law and Research Association. <https://www.elra.eu/wp-content/uploads/2025/11/5.-Jose-Ramon-de-Blas-Panel-AI.pdf>

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), eaao5580. <https://doi.org/10.1126/sciadv.aao5580>

European Commission. (2020). White paper on artificial intelligence: A European approach to excellence and trust. <https://commission.europa.eu/>

Fernández, A. (2024, octubre 17). La importancia de la certeza jurídica. La Razón de México. <https://www.larazon.mx>

Gillespie, T. (2018). Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press.

Giri, D. (2025). Judicial discretion in the age of algorithms: Regulating artificial intelligence in judicial decision-making. Lawful Legal. <https://lawfullegal.in/judicial-discretion-in-the-age-of-algorithms-regulating-artificial-intelligence-in-judicial-decision-making/>

Khan, I. (2025). Promotion and protection of the right to freedom of opinion and expression (A/80/341). Naciones Unidas. <https://digitallibrary.un.org/record/4088466>

Llansó, E., van Hoboken, J., Leerssen, P., & Harambam, J. (2020). Artificial intelligence, content moderation, and freedom of expression. Center for Democracy and Technology.

Monroy Enríquez, N., & Ramírez Ogarrío, J. A. (2025.). Los principios de indivisibilidad e interdependencia de los derechos humanos. Perseo. Revista Electrónica del Programa Universitario de Derechos Humanos, Universidad Nacional Autónoma de México. <https://www.pudh.unam.mx/perseo/los-principios-de-indivisibilidad-e-interdependencia-de-los-derechos-humanos>

Naciones Unidas. (2024). General Assembly resolution on safe, secure and trustworthy artificial intelligence. <https://news.un.org>

Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (OACNUDH). (2023). Artificial intelligence must be grounded in human rights. <https://www.ohchr.org/es/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner>

Oliva, T. D. (2020). Content moderation technologies: Applying human rights standards to protect freedom of expression. *Human Rights Law Review*, 20(4), 607–640. <https://doi.org/10.1093/hrlr/ngaa032>

OpenAI. (2026). ChatGPT [Modelo de lenguaje de gran tamaño]. <https://chat.openai.com/>

Organisation for Economic Co-operation and Development. (2019). Artificial intelligence in society. OECD Publishing. <https://doi.org/10.1787/eedfee77-en>

Organisation for Economic Cooperation and Development. (2024). OECD principles on artificial intelligence. <https://oecd.ai/en/ai-principles>

Organisation for Economic Cooperation and Development. (2025). AI in justice administration and access to justice. OECD Publishing. https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en/full-report/ai-in-justice-administration-and-access-to-justice_f0cbe651.html

Organización de los Estados Americanos. Comisión Interamericana de Derechos Humanos, Relatoría Especial para la Libertad de Expresión. (2025, 7 de mayo). *Comunicado conjunto sobre inteligencia artificial y libertad de expresión*. https://www.oas.org/en/iachr/jsForm/?File=/en/iachr/expression/media_center/preleases/2025/089.asp

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Perplexity AI. (2026). *Perplexity* [Modelo de lenguaje de gran tamaño]. <https://www.perplexity.ai/>

Suprema Corte de Justicia de la Nación. (2011). Expediente varios 912/2010 (Caso Radilla Pacheco). Ciudad de México: SCJN.

Suprema Corte de Justicia de la Nación. (2014). Derecho humano al debido proceso. Elementos que lo integran (Registro digital 2005401). Gaceta del Semanario Judicial de la Federación, Décima Época, Primera Sala. <https://juristeca.com/mx/scjn/tesis-aisladas/2014/1/registro-2005401-derecho-humano-al-debido-proceso>

Suprema Corte de Justicia de la Nación. (2017). Amparo en revisión 1/2017. Segunda Sala, Semanario Judicial de la Federación, Décima Época.

Suprema Corte de Justicia de la Nación. (2019). Amparo en revisión 1005/2018. Segunda Sala. Semanario Judicial de la Federación. <https://www.scjn.gob.mx>

Suprema Corte de Justicia de la Nación. (2019). Derecho a la tutela judicial efectiva. Alcance del principio de justicia completa respecto al cumplimiento de las sentencias (Registro digital 2019663). Gaceta del Semanario Judicial de la Federación, Décima Época, Segunda Sala. <https://juristeca.com/mx/scjn/tesis-aisladas/2019/4/registro-2019663-derecho-a-la-tutela-judicial-efectiva>

Suprema Corte de Justicia de la Nación. (2021). Cuadernos de jurisprudencia núm. 13: Libertad de expresión y medios de comunicación. Centro de Estudios Constitucionales. <https://www.scjn.gob.mx>

Suprema Corte de Justicia de la Nación. (2025). Inteligencia artificial aplicada en procesos jurisdiccionales. Elementos mínimos que deben observarse para su uso ético y responsable con perspectiva de derechos humanos (Tesis aislada II.2o.C.9 K [11a.], Registro digital 2031010). Gaceta del Semanario Judicial de la Federación, Undécima Época, Libro 52, Tomo IV, Vol. 2, 1706–1709.

Suprema Corte de Justicia de la Nación. (2025a). Jurisprudencia 126/2025: Libertad de expresión. Opiniones y juicios de valor sobre asuntos de interés público. Primera Sala, Semanario Judicial de la Federación.

Suprema Corte de Justicia de la Nación. (2025b). Jurisprudencia 128/2025: Daño moral y libertad de expresión. Estándar de real malicia. Primera Sala, Semanario Judicial de la Federación.

Suprema Corte de Justicia de la Nación. (2026). Acuerdo General Plenario 7/2025, por el que se establece el sistema de precedentes obligatorios. SCJN.

Suprema Corte de Justicia de la Nación. (2026). Inteligencia artificial aplicada en procesos jurisdiccionales. Elementos mínimos que deben observarse para su uso ético y responsable con perspectiva de derechos humanos (Jurisprudencia II.2o.C. J/2 K [12a.], Registro digital 2031640). Gaceta del Semanario Judicial de la Federación, Duodécima Época, Libro 5, Tomo II, Vol. 1, 599–604.

Taboada Macías, I. (2024). Los riesgos a los derechos humanos por la inteligencia artificial: Su intento de mitigación en la EU AI Act. Cuestiones Constitucionales. Revista Mexicana de Derecho Constitucional, 26(52), e19226. <https://doi.org/10.22201/ij.24484881e.2025.52.19226>

Tribunal Colegiado en Materia Civil del Segundo Circuito. (2025). Criterios para el uso ético y responsable de la inteligencia artificial en procesos judiciales. Semanario Judicial de la Federación.

Tribunal de Justicia de la Unión Europea. (2014). Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources y Kärntner Landesregierung y otros (Asuntos acumulados C-293/12 y C-594/12). <https://dpcuria.eu/case?reference=C-293/12>

Tribunal de Justicia de la Unión Europea. (2020). Data Protection Commissioner contra Facebook Ireland Limited y Maximilian Schrems (Asunto C-311/18). <https://eur-lex.europa.eu>

Tribunal Europeo de Derechos Humanos. (2004). Von Hannover contra Alemania. <https://hudoc.echr.coe.int>

Tribunal Europeo de Derechos Humanos. (2008). S. y Marper contra Reino Unido. <https://hudoc.echr.coe.int>

Tribunal Europeo de Derechos Humanos. (2012). Von Hannover contra Alemania (n.º 2). <https://hudoc.echr.coe.int>

Tribunal Europeo de Derechos Humanos. (2019). López Ribalda y otros contra España. <https://hudoc.echr.coe.int>

Tribunal Europeo de Derechos Humanos. (2021). Big Brother Watch y otros contra Reino Unido. <https://hudoc.echr.coe.int>

Tribunales Colegiados de Circuito. (2020). Libertad de expresión y derecho de acceso a la información en redes sociales de Internet (Registro digital 2022074). Semanario Judicial de la Federación, Décima Época.

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. <https://www.unesco.org>

UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>

UNESCO. (2025). Guidelines for the use of AI systems in courts and tribunals. United Nations Educational, Scientific and Cultural Organization. <https://jusmundi.com/en/document/rule/en-unesco-united-nations-educational-scientific-and-cultural-organization-guidelines-for-the-use-of-ai-systems-in-courts-and-tribunals-2025>

Unión Europea. (2024). *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial)*. Diario Oficial de la Unión Europea, L, 2024/1689. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs.

CAPÍTULO XIV

REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL A NIVEL GLOBAL: ENFOQUES Y MODELOS CONTEMPORÁNEOS

Global Regulation of Artificial Intelligence: Contemporary Approaches and
Models

Marlene Atenea Grimaldo Ramírez

ramireznea51@gmail.com

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0009-0009-7784-941X>

Resumen

La inteligencia artificial (IA) se ha consolidado como una de las tecnologías más influyentes del siglo XXI, generando profundas transformaciones en distintos ámbitos sociales y económicos. Sin embargo, su expansión ha planteado desafíos relacionados con la protección de los derechos humanos, la privacidad y la transparencia. En este contexto, diversos países han desarrollado marcos regulatorios para gobernar su uso.

El presente capítulo analiza las principales regulaciones existentes a nivel global en materia de inteligencia artificial, identificando sus enfoques y diferencias. A través de un análisis comparativo, se examinan los modelos regulatorios de la Unión Europea, Estados Unidos, China, Canadá, Reino Unido, Brasil y México, así como los principios éticos internacionales promovidos por la UNESCO. Se concluye que, pese a la diversidad normativa, existe una convergencia global en torno a principios fundamentales como la transparencia, la supervisión humana y la protección de derechos.

Palabras clave: Inteligencia artificial, regulación, gobernanza digital, derechos humanos, ética tecnológica.

Abstract

Artificial intelligence (AI) has established itself as one of the most influential technologies of the 21st century, generating profound transformations in different social and economic areas. However, its expansion has posed challenges related to the protection of human rights, privacy, and transparency. In this context, several countries have developed regulatory frameworks to govern their use.

This chapter analyzes the main regulations existing globally in the field of artificial intelligence, identifying their approaches and differences. Through a comparative analysis, the regulatory models of the European Union, the United States, China, Canada, the United Kingdom, Brazil and Mexico are examined, as well as the international ethical principles promoted by UNESCO. It concludes that, despite the regulatory diversity, there is a global convergence around fundamental principles such as transparency, human oversight and the protection of rights.

Keywords: Artificial intelligence, regulation, digital governance, human rights, technological ethics.

Introducción

La inteligencia artificial (IA) ha emergido como una de las tecnologías más disruptivas en los últimos años, transformando profundamente la manera en que las sociedades producen, interactúan y toman decisiones en su entorno. Recalcando que su implementación se ha expandido rápidamente en muchos sectores estratégicos, impulsando procesos de automatización, análisis de datos y generación de conocimiento sin precedentes.

Por ejemplo, en el sector salud, la IA se utiliza para el diagnóstico asistido mediante imágenes médicas, la predicción de enfermedades y el desarrollo de tratamientos personalizados; en la educación, permite la creación de sistemas de aprendizaje adaptativo que ajustan contenidos según el ritmo y desempeño del estudiante; en materia de seguridad, se emplea en sistemas de videovigilancia inteligente, reconocimiento facial y análisis predictivo del delito; en la industria, optimiza procesos productivos a través de la automatización, el mantenimiento predictivo y la logística inteligente; y, finalmente, en la administración pública, facilita la toma de decisiones basada en datos, la digitalización de servicios gubernamentales y la mejora en la gestión de políticas públicas.

No obstante, este desarrollo acelerado ha traído consigo una serie de desafíos éticos, jurídicos, sociales y tecnológicos que han encendido el debate a nivel internacional. Estos desafíos no solo se relacionan con los riesgos inherentes al uso de la inteligencia artificial, sino también con las profundas desigualdades en su desarrollo e implementación entre distintos países y regiones. Mientras que algunas naciones cuentan con infraestructura tecnológica avanzada, acceso a grandes volúmenes de datos y capacidades de innovación, otras enfrentan limitaciones en conectividad, recursos técnicos y formación especializada, lo que amplía la brecha digital y condiciona el impacto de la IA en sus sociedades.

A ello se suma la insuficiente alfabetización digital en diversos sectores de la población, lo que dificulta la comprensión, adopción y uso crítico de estas tecnologías. Esta situación genera barreras adicionales, particularmente en grupos vulnerables o históricamente marginados, que pueden quedar excluidos de los beneficios de la inteligencia artificial o incluso verse afectados de manera desproporcionada por sus riesgos. Asimismo, existen resistencias sociales derivadas de la desconfianza hacia la tecnología, el desconocimiento de su funcionamiento o el temor a la sustitución laboral, lo que evidencia que el desarrollo de la IA no es únicamente un desafío técnico, sino también social y cultural.

Asimismo, existen retos asociados a la gobernanza de estas tecnologías, como la

falta de marcos regulatorios claros, la dificultad para supervisar sistemas complejos y la ausencia de estándares internacionales uniformes. A ello se suman problemáticas específicas como la discriminación algorítmica, la opacidad en los procesos de toma de decisiones, la vulneración de la privacidad, la manipulación de información y la concentración del poder tecnológico en grandes corporaciones. Estas condiciones han puesto en evidencia la necesidad de establecer mecanismos regulatorios que permitan garantizar un uso responsable, equitativo y alineado con los derechos humanos, considerando que “las tecnologías de la IA pueden ser de gran utilidad para la humanidad y que todos los países pueden beneficiarse de ellas, pero que también suscitan preocupaciones éticas fundamentales, por ejemplo, en relación con los sesgos que pueden incorporar y exacerbar, lo que puede llegar a provocar discriminación, desigualdad, brechas digitales y exclusión y suponer una amenaza para la diversidad cultural, social y biológica, así como generar divisiones sociales o económicas” (UNESCO, 2021, p. 5).

En respuesta a estos desafíos, distintos países y organismos internacionales han comenzado a desarrollar estrategias regulatorias para gobernar la inteligencia artificial. Sin embargo, estas regulaciones no son homogéneas, sino que reflejan distintas visiones sobre el papel de la tecnología en la sociedad. En este contexto, es posible identificar al menos tres grandes modelos regulatorios a nivel global: el modelo europeo, basado en la gestión del riesgo y la protección de los derechos fundamentales; el modelo estadounidense, orientado a la innovación y la autorregulación; y el modelo chino, centrado en el control estatal de las tecnologías digitales y la seguridad nacional. A su vez, han surgido enfoques intermedios o híbridos en países como Canadá, Reino Unido y Brasil, que buscan equilibrar el desarrollo tecnológico con la mitigación de riesgos. En América Latina, y particularmente en México, la regulación de la inteligencia artificial se encuentra en una etapa de construcción, con un enfoque inicial de carácter ético y orientador — como lo reflejan los Principios de Chapultepec—, los cuales no constituyen un marco normativo vinculante, pero sí establecen bases para el desarrollo de futuras regulaciones en la materia.

Bajo este panorama, surge la siguiente pregunta de investigación: **¿qué regulaciones existen en el mundo para el uso de la inteligencia artificial?** Para responder a esta interrogante, el presente capítulo tiene como objetivo analizar comparativamente los principales marcos regulatorios internacionales, identificando sus características, diferencias y puntos de convergencia.

La relevancia de este análisis radica en que la regulación de la IA no solo responde a cuestiones técnicas o jurídicas, sino que también refleja decisiones políticas, económicas y éticas sobre el futuro de las sociedades digitales. En este sentido, comprender las distintas aproximaciones regulatorias permite identificar tendencias globales, así como los desafíos y oportunidades que enfrentan países en desarrollo, como México, en la construcción de sus propios marcos normativos.

Finalmente, este capítulo se estructura en tres apartados principales: en primer lugar, se abordan los fundamentos éticos internacionales que han orientado la regulación de la IA; en segundo lugar, se analizan los principales modelos regulatorios a nivel global; y, en tercer lugar, se presenta un análisis comparativo que permite identificar convergencias y divergencias entre los distintos enfoques. A partir de ello, se plantean reflexiones finales sobre los retos actuales de la gobernanza de la inteligencia artificial.

Desarrollo

1. Fundamentos éticos internacionales de la IA

El desarrollo de la inteligencia artificial a nivel global no ha ocurrido de manera aislada, sino que ha estado acompañado por la construcción de marcos éticos internacionales que buscan orientar su diseño, implementación y uso hacia el respeto de los derechos humanos, el bienestar social y su aplicación responsable. Esto responde, en gran medida, a la preocupación creciente sobre los efectos que estas tecnologías pueden tener en la vida cotidiana de las personas, especialmente cuando su avance supera la capacidad de regulación de muchos países.

En este contexto, antes de que existieran leyes formales o regulaciones vinculantes, diversos organismos internacionales comenzaron a establecer principios que sirvieran como una guía inicial. Instituciones como la UNESCO y la OCDE han desempeñado un papel fundamental en este proceso, promoviendo recomendaciones y lineamientos orientados a garantizar que la inteligencia artificial se desarrolle de manera responsable, segura y alineada con la protección de los usuarios.

Esto implica no solo prevenir riesgos como la vulneración de la privacidad o el uso indebido de datos personales, sino también asegurar que los sistemas de IA no generen decisiones discriminatorias, opacas o difíciles de cuestionar. Asimismo, se busca que los usuarios tengan claridad sobre cuándo están interactuando con sistemas automatizados y que existan mecanismos de supervisión y rendición de cuentas que permitan identificar responsabilidades en caso de afectaciones.

Más que imponer reglas obligatorias, estos marcos éticos han funcionado como una especie de brújula que orienta a gobiernos, empresas y sociedad en general sobre cómo utilizar la inteligencia artificial de forma adecuada. Su importancia radica en que han permitido generar acuerdos básicos a nivel internacional sobre temas como la transparencia, la equidad y la supervisión humana, sentando así las bases para el desarrollo de regulaciones más formales en distintos países.

Uno de los instrumentos más relevantes en este ámbito es la Recomendación sobre la Ética de la Inteligencia Artificial, adoptada en 2021, la cual constituye el primer marco normativo de carácter global en materia de ética de la IA. Este documento establece que el desarrollo y uso de la inteligencia artificial debe estar centrado en el ser humano, promoviendo valores como la dignidad, la equidad, la inclusión y la sostenibilidad (UNESCO, 2021).

En este sentido, la recomendación plantea que los sistemas de IA deben diseñarse e implementarse de manera transparente, responsable y supervisada, garantizando que sus decisiones puedan ser explicables y auditables. Asimismo, destaca la importancia de prevenir riesgos asociados como la discriminación algorítmica, la

exclusión social y la vulneración de la privacidad, enfatizando la necesidad de evaluar el impacto de estos sistemas a lo largo de todo su ciclo de vida.

En concordancia con lo anterior, la UNESCO subraya la centralidad de los derechos humanos en el desarrollo de la inteligencia artificial, al señalar que “los derechos humanos y las libertades fundamentales han de ser respetados, protegidos y promovidos a lo largo del ciclo de vida de los sistemas de IA” (UNESCO, 2021, p. 19). Esto refuerza la idea de que la ética no debe considerarse un elemento aislado, sino un principio transversal que guía todas las etapas del desarrollo y uso de la IA.

A partir de estos principios, diversos países han comenzado a desarrollar sus propios marcos regulatorios, adaptando estas directrices a sus contextos políticos, económicos y sociales. De este modo, los fundamentos éticos internacionales no solo han servido como guía, sino también como punto de partida para la construcción de modelos regulatorios diferenciados a nivel global.

2. El modelo europeo: regulación basada en riesgos

El modelo europeo de regulación de la inteligencia artificial se ha consolidado como uno de los más avanzados y estructurados a nivel global, al proponer un enfoque centrado en la protección de los derechos fundamentales y la gestión de riesgos. A diferencia de otros modelos que priorizan la innovación o el control estatal, la Unión Europea ha optado por establecer un marco normativo que busca equilibrar el desarrollo tecnológico con la seguridad, la confianza y el bienestar de las personas.

En este sentido, el eje central de la regulación europea radica en la clasificación de los sistemas de inteligencia artificial según su nivel de riesgo. Este enfoque permite diferenciar entre aplicaciones que representan un impacto mínimo y aquellas que pueden afectar de manera significativa la vida de las personas, particularmente en ámbitos sensibles como la salud, la educación, el empleo o la seguridad. A partir de esta clasificación, se establecen distintas obligaciones para los desarrolladores y proveedores de sistemas de IA, siendo más estrictas en los casos de mayor riesgo.

En este sentido, el propio Reglamento establece que los sistemas de IA deben clasificarse “en función de los riesgos que plantean” (European Commission, 2021, p. 12, traducción propia).

De manera particular, el Reglamento de Inteligencia Artificial (*AI Act*) introduce categorías como los sistemas de riesgo inaceptable, que son directamente prohibidos; los sistemas de alto riesgo, que deben cumplir con requisitos estrictos de transparencia, supervisión humana y evaluación de impacto; y aquellos de riesgo limitado o mínimo, que están sujetos a obligaciones más flexibles. Este esquema refleja un enfoque preventivo, orientado a anticipar posibles daños antes de que estos ocurran, ya que el Reglamento señala que los sistemas de alto riesgo están sujetos a “requisitos específicos para mitigar los riesgos para la salud, la seguridad y los derechos fundamentales” (European Commission, 2021, p. 14, traducción propia).

Para comprender mejor esta clasificación, es posible sistematizar los distintos niveles de riesgo y sus principales implicaciones regulatorias en la siguiente tabla. Esta diferenciación permite identificar no solo el grado de impacto de los sistemas de inteligencia artificial, sino también las obligaciones específicas que deben cumplir en cada caso, evidenciando el carácter proporcional del modelo europeo.

Nivel de riesgo	Descripción	Ejemplos	Requisitos principales
Riesgo inaceptable	Sistemas que representan una amenaza para los derechos fundamentales o la seguridad	Puntuación social, manipulación subliminal	✗ Prohibidos
Alto riesgo	Sistemas que afectan directamente derechos, seguridad o decisiones relevantes	IA en salud, empleo, educación, justicia	✓ Evaluación de riesgos ✓ Supervisión humana ✓ Alta calidad de datos ✓ Documentación y trazabilidad ✓ Seguridad y ciberseguridad
Riesgo limitado	Sistemas con interacción directa con usuarios, pero menor impacto	Chatbots, IA generativa	✓ Obligación de transparencia (informar que es IA)
Riesgo mínimo	Sistemas sin impacto significativo en derechos o seguridad	Filtros de spam, videojuegos	✓ Sin obligaciones específicas

Esta clasificación permite comprender cómo el modelo europeo establece una regulación diferenciada en función del impacto potencial de los sistemas de inteligencia artificial, evitando imponer cargas innecesarias a aplicaciones de bajo

riesgo, al tiempo que refuerza los controles en aquellas que pueden afectar directamente a las personas.

Asimismo, este modelo pone un énfasis especial en la transparencia y la rendición de cuentas, estableciendo que los sistemas de inteligencia artificial deben ser comprensibles y explicables para los usuarios. Esto implica no solo informar sobre su uso, sino también garantizar que las decisiones automatizadas puedan ser interpretadas, justificadas y, en su caso, impugnadas. En este sentido, la transparencia se convierte en un elemento fundamental para generar confianza en estas tecnologías, mientras que la rendición de cuentas permite delimitar responsabilidades cuando un sistema produce resultados erróneos o afecta derechos fundamentales.

Este enfoque adquiere especial relevancia en contextos donde la inteligencia artificial incide directamente en la vida de las personas, como en el acceso al empleo, los servicios públicos o la seguridad. Por ello, el Reglamento establece la necesidad de incorporar mecanismos de supervisión humana en los sistemas de mayor riesgo, evitando que las decisiones queden completamente automatizadas y asegurando un control efectivo sobre su funcionamiento.

En conjunto, este enfoque basado en riesgos no solo permite una regulación diferenciada y proporcional, sino que también posiciona a la Unión Europea como un referente internacional en la construcción de una inteligencia artificial confiable, centrada en el ser humano y orientada a la protección de los derechos fundamentales.

En este marco, la implementación del Reglamento se ve reforzada por la creación de organismos especializados como la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), la cual constituye una de las primeras instituciones a nivel europeo dedicadas específicamente a la supervisión de sistemas de IA. Este organismo tiene como objetivo principal garantizar el cumplimiento de la normativa, así como vigilar que el desarrollo y uso de la inteligencia artificial se realice

conforme a los principios de transparencia, seguridad y respeto a los derechos fundamentales.

En este sentido, la AESIA tiene como finalidad “liderar y coordinar la supervisión del desarrollo de la inteligencia artificial en España, asegurando que se aplique de forma ética, segura e inclusiva” (Gobierno de España, s. f.).

Entre sus funciones destacan la evaluación de riesgos, la supervisión de sistemas de alto riesgo, la emisión de recomendaciones y el impulso de buenas prácticas en el uso de la inteligencia artificial. Asimismo, la AESIA desempeña un papel relevante en la coordinación entre autoridades nacionales y europeas, contribuyendo a la construcción de un ecosistema regulatorio más coherente y armonizado.

De esta manera, la creación de la AESIA refleja la transición de un modelo meramente normativo hacia uno de gobernanza activa, en el que no solo se establecen reglas, sino que también se desarrollan mecanismos institucionales para su implementación, seguimiento y mejora continua. En conjunto, tanto el Reglamento de Inteligencia Artificial como las estructuras institucionales que lo acompañan consolidan al modelo europeo como un referente en la regulación de la IA, al articular de manera integral la dimensión normativa, ética y operativa en la gobernanza de estas tecnologías.

3. El modelo estadounidense: innovación y regulación flexible

A diferencia del modelo europeo, que se caracteriza por un enfoque normativo estructurado y basado en la gestión de riesgos, el modelo estadounidense de regulación de la inteligencia artificial se distingue por su carácter flexible y orientado a la innovación. En lugar de establecer una legislación integral y uniforme, Estados Unidos ha optado por un enfoque descentralizado, en el que distintas agencias y sectores desarrollan lineamientos específicos según sus necesidades. Este modelo responde, en gran medida, a lo que se ha definido como un enfoque basado en principios, en el que “las personas deben estar protegidas frente a sistemas automatizados inseguros o ineficaces” (The White House, 2022, traducción propia).

Este modelo se centra en la intención de no limitar el desarrollo tecnológico ni afectar la competitividad de las empresas, especialmente en un contexto donde la inteligencia artificial se ha convertido en un elemento estratégico para el crecimiento económico y el liderazgo global. En este sentido, la regulación en Estados Unidos busca equilibrar la promoción de la innovación con la mitigación de riesgos, aunque con un menor grado de intervención estatal en comparación con el modelo europeo. En línea con ello, la política estadounidense ha enfatizado la necesidad de “promover una innovación responsable que proteja a las personas y fortalezca el liderazgo global en inteligencia artificial” (The White House, 2023, traducción propia).

Esta aproximación se caracteriza por su naturaleza descentralizada, en la que distintas agencias y sectores —como el ámbito financiero, de salud o tecnológico— establecen lineamientos específicos según sus necesidades, en lugar de contar con una legislación única y general. Asimismo, se basa principalmente en principios, guías y políticas públicas que orientan el desarrollo de la IA, más que en normas obligatorias de carácter integral.

Uno de los instrumentos más relevantes en este enfoque es el *Blueprint for an AI Bill of Rights*, el cual establece una serie de principios orientados a proteger a las personas frente a los posibles riesgos derivados del uso de sistemas automatizados. Más que constituir una regulación vinculante, este documento funciona como una guía que orienta el desarrollo y uso de la inteligencia artificial desde una perspectiva centrada en el usuario, al señalar que “las personas deben estar protegidas frente a sistemas automatizados inseguros o ineficaces” (The White House, 2022, traducción propia).

Entre sus principales planteamientos se encuentran la protección frente a sistemas inseguros o ineficaces, la prevención de la discriminación algorítmica, la garantía de la privacidad de los datos y el derecho de las personas a ser informadas cuando interactúan con sistemas de inteligencia artificial. Estos principios buscan asegurar que la IA sea utilizada de manera justa, transparente y segura, promoviendo

condiciones en las que los usuarios puedan comprender, cuestionar y confiar en las decisiones automatizadas.

La política estadounidense en materia de inteligencia artificial se ha visto reforzada por la emisión de órdenes ejecutivas que buscan establecer lineamientos generales para el desarrollo seguro, confiable y competitivo de estas tecnologías. Estas disposiciones reflejan un enfoque basado en principios y guías, más que en regulaciones estrictamente vinculantes, lo que permite una mayor adaptabilidad frente a los cambios tecnológicos. En este sentido, se ha destacado la necesidad de “garantizar un desarrollo seguro, confiable y responsable de la inteligencia artificial” (The White House, 2023, traducción propia).

Sin embargo, este modelo también ha estado influido por decisiones políticas y estrategias gubernamentales que priorizan el liderazgo tecnológico del país, en algunos casos por encima de la consolidación de un marco regulatorio integral. En este contexto, la inteligencia artificial se ha convertido en un elemento clave dentro de la competencia geopolítica global, lo que ha llevado a que su desarrollo sea impulsado como una ventaja estratégica.

Este enfoque ha generado debates en torno a si la flexibilidad regulatoria favorece realmente la innovación sin comprometer la protección de los derechos de los usuarios, o si, por el contrario, la ausencia de normas más estrictas puede derivar en riesgos no controlados. De este modo, el caso estadounidense refleja una tensión constante entre innovación, competitividad y regulación.

Por lo que, este modelo también ha sido objeto de críticas, principalmente por la ausencia de un marco regulatorio integral que establezca obligaciones claras y homogéneas. A diferencia de la Unión Europea, donde existe una regulación formal y obligatoria, en Estados Unidos la protección de los usuarios depende en gran medida de la interpretación y aplicación de principios por parte de distintas instituciones. Esta situación ha sido señalada en diversos análisis, los cuales advierten que la fragmentación regulatoria puede dificultar la aplicación coherente de estándares en materia de inteligencia artificial (OECD, 2023).

Esta situación puede generar inconsistencias en la forma en que se regulan los sistemas de inteligencia artificial, ya que cada sector o agencia puede adoptar criterios distintos en función de sus prioridades y capacidades. Como consecuencia, algunos ámbitos pueden contar con mayores niveles de protección, mientras que otros permanecen con vacíos regulatorios, lo que dificulta garantizar estándares uniformes de seguridad, transparencia y rendición de cuentas.

Asimismo, la falta de un marco centralizado plantea desafíos en la supervisión efectiva de los sistemas de IA, especialmente en aquellos casos donde su impacto trasciende sectores o jurisdicciones. Esto ha dado lugar a un debate creciente sobre la necesidad de avanzar hacia esquemas regulatorios más integrales que permitan equilibrar la flexibilidad del modelo con una mayor protección de los derechos de los usuarios.

En el caso de Estados Unidos, esta discusión adquiere una dimensión adicional, ya que la inteligencia artificial no solo es entendida como una herramienta tecnológica, sino también como un recurso estratégico dentro de la competencia global. En este sentido, su desarrollo se encuentra estrechamente vinculado con objetivos de liderazgo económico, innovación y posicionamiento internacional. Como se ha señalado en la política estadounidense reciente, la inteligencia artificial constituye un elemento clave para “mantener el liderazgo global y la seguridad nacional” (The White House, 2023, traducción propia).

Bajo esta lógica, la ausencia de una regulación integral puede interpretarse como parte de una estrategia que prioriza la rapidez en el desarrollo tecnológico frente a la imposición de restricciones normativas. No obstante, este enfoque también plantea desafíos importantes, ya que podría ampliar las brechas tecnológicas entre países y generar riesgos asociados al uso de la IA en ámbitos sensibles como la seguridad o la defensa.

4. El modelo chino: control estatal y seguridad

Comparado con los modelos europeo y estadounidense, el enfoque de China en materia de inteligencia artificial se distingue por una fuerte participación del Estado en su desarrollo, regulación y supervisión. Mientras que en la Unión Europea se prioriza la protección de los derechos fundamentales y en Estados Unidos se favorece la innovación mediante esquemas flexibles, el modelo chino se orienta hacia una gobernanza centralizada en la que el Estado desempeña un papel determinante en la definición de los límites y usos de estas tecnologías.

En este contexto, la inteligencia artificial no solo es concebida como una herramienta de innovación tecnológica, sino también como un instrumento estratégico vinculado al desarrollo económico, la estabilidad social y el fortalecimiento del poder estatal. De este modo, la regulación de la IA en China se integra dentro de una visión más amplia de gestión y control de la información, en la que la tecnología se utiliza como un medio para consolidar objetivos políticos y sociales a largo plazo.

Este enfoque combina el impulso a la innovación con un control regulatorio robusto, en el que el gobierno establece directrices claras sobre el uso de la inteligencia artificial, especialmente en ámbitos considerados sensibles. A través de diversas normativas, se han implementado mecanismos para supervisar el contenido generado por IA, regular el uso de algoritmos y controlar la difusión de información en plataformas digitales.

Una de las características más relevantes de este modelo es la regulación específica de sistemas algorítmicos y tecnologías generativas, como los sistemas de recomendación y los contenidos sintéticos (*deepfakes*). Estos sistemas ocupan un papel central en el ecosistema digital contemporáneo, ya que influyen directamente en la forma en que las personas acceden a la información, consumen contenidos y participan en espacios digitales.

En el caso de los sistemas de recomendación, ampliamente utilizados en plataformas digitales, su regulación responde a la necesidad de controlar los

criterios mediante los cuales se prioriza, filtra y distribuye la información. En este sentido, la normativa establece que los proveedores deben “fortalecer la gestión de la información y garantizar la orientación correcta de la opinión pública” (Cyberspace Administration of China, 2022, traducción propia), lo que evidencia el papel estratégico de estos sistemas dentro del modelo chino.

Por otro lado, la regulación de tecnologías generativas responde al potencial que estas herramientas tienen para producir información manipulada o engañosa con un alto grado de realismo. En este marco, se establece que los proveedores deben “etiquetar claramente el contenido generado o sintetizado para evitar confusión o engaño” (Cyberspace Administration of China, 2023, traducción propia), lo que refleja una preocupación directa por la veracidad de la información y la seguridad informativa.

Este esquema se complementa con un conjunto más amplio de obligaciones dirigidas a los sistemas algorítmicos, que incluyen el registro de algoritmos ante las autoridades, la transparencia en su funcionamiento y la implementación de mecanismos para prevenir la difusión de contenidos que puedan considerarse perjudiciales para el orden público o la seguridad nacional. En este sentido, se establece que los proveedores deben “cumplir con las normas de gestión de algoritmos y garantizar la seguridad de la información” (Cyberspace Administration of China, 2022, traducción propia), lo que evidencia el carácter normativo y operativo de este modelo.

Desde esta perspectiva, la regulación en China no se limita a aspectos técnicos, sino que responde a una lógica más amplia de control sobre el flujo de información y el comportamiento en entornos digitales. En la práctica, esto implica que las empresas no solo deben cumplir con estándares técnicos, sino también alinearse con directrices estatales respecto al tipo de contenido que puede ser promovido o restringido, reforzando así el papel del Estado como actor central en la gobernanza tecnológica.

Asimismo, este enfoque se articula con una estructura de gobernanza altamente centralizada, en la que el Estado desempeña un papel predominante en la supervisión y control de la inteligencia artificial. Esta centralización permite una aplicación más directa, uniforme y rápida de las regulaciones, facilitando la implementación de políticas a gran escala y el monitoreo constante de los sistemas de IA, con el objetivo de garantizar su alineación con las prioridades políticas y sociales del gobierno.

No obstante, este modelo también ha generado diversas críticas, particularmente en relación con su impacto en las libertades individuales. La amplia capacidad de supervisión estatal, apoyada en el uso de tecnologías de inteligencia artificial, ha suscitado preocupaciones sobre la privacidad de los datos, la libertad de expresión y el uso de sistemas de vigilancia masiva. En este contexto, la IA no solo se limita a fines tecnológicos o económicos, sino que también se integra en mecanismos de control social, lo que puede restringir la autonomía de los individuos y la diversidad de opiniones en entornos digitales.

Desde una perspectiva más amplia, este modelo plantea un dilema entre eficiencia regulatoria y protección de derechos: si bien permite un control más efectivo de los riesgos asociados a la inteligencia artificial, también puede derivar en un uso intensivo de estas tecnologías con fines de supervisión social. De este modo, el caso de China evidencia cómo la gobernanza de la IA puede estar estrechamente vinculada con estructuras políticas y modelos de organización del Estado, influyendo directamente en el equilibrio entre innovación, control y libertades individuales, lo que pone de manifiesto que la regulación de la inteligencia artificial no solo responde a criterios técnicos, sino también a dinámicas de poder y organización estatal.

5. Modelos híbridos: Canadá, Reino Unido y Brasil

Frente a los enfoques más definidos de la Unión Europea, Estados Unidos y China, han surgido modelos regulatorios intermedios que buscan equilibrar la promoción de la innovación con la mitigación de los riesgos asociados al uso de la inteligencia artificial. En este contexto, países como Canada, United Kingdom y Brazil han adoptado enfoques que combinan elementos de regulación formal con marcos flexibles basados en principios, configurando así modelos híbridos de gobernanza.

Estos modelos reflejan una tendencia hacia la búsqueda de equilibrio entre el control regulatorio y la adaptabilidad tecnológica, evitando tanto la rigidez normativa como la ausencia de supervisión. Más que representar soluciones intermedias, constituyen aproximaciones que intentan integrar las ventajas de los distintos enfoques globales, ajustándose a las particularidades institucionales y contextuales de cada país.

En el caso de Canada, se ha desarrollado una propuesta normativa a través del *Artificial Intelligence and Data Act (AIDA)*, la cual establece obligaciones específicas para los sistemas de alto impacto, incluyendo evaluaciones de riesgo, medidas de mitigación y mecanismos de supervisión. Este enfoque se centra en regular de manera más estricta aquellos sistemas que pueden generar efectos significativos en los derechos de las personas, particularmente en ámbitos sensibles, sin imponer restricciones generalizadas al desarrollo tecnológico. En este sentido, se reconoce que los sistemas de IA de alto impacto deben ser “identificados, evaluados y gestionados para mitigar posibles daños” (Government of Canada, 2022, traducción propia).

Asimismo, el modelo canadiense se inscribe dentro de una estrategia más amplia de gobernanza responsable de la inteligencia artificial, en la que se promueve el desarrollo ético y seguro de estas tecnologías. Un ejemplo de ello es el *Algorithmic Impact Assessment (AIA)*, instrumento utilizado en el sector público para evaluar los riesgos de los sistemas automatizados antes de su implementación, lo que evidencia un enfoque preventivo orientado a la protección de los usuarios (Treasury

Board of Canada Secretariat, 2019). De este modo, Canadá adopta una lógica intermedia que equilibra la protección de derechos con la promoción de la innovación, configurándose como un ejemplo de regulación proporcional basada en el nivel de impacto de los sistemas de IA.

Por su parte, el United Kingdom ha optado por un modelo flexible basado en principios, en el que la regulación de la inteligencia artificial se delega a autoridades sectoriales ya existentes, como los reguladores en materia de competencia, protección de datos o servicios financieros. En lugar de establecer una legislación única y centralizada, el gobierno ha definido un conjunto de principios rectores — como la seguridad, la transparencia, la equidad, la rendición de cuentas y la gobernanza— que orientan la actuación de los distintos reguladores, permitiendo una adaptación más dinámica frente a los rápidos avances tecnológicos (UK Government, 2023).

Este enfoque busca evitar la sobre regulación y fomentar la innovación, promoviendo un entorno propicio para el desarrollo tecnológico sin imponer cargas normativas excesivas. En este sentido, se plantea que la regulación de la inteligencia artificial debe ser “pro-innovación y adaptable a contextos específicos” (UK Government, 2023, traducción propia). No obstante, la dependencia de múltiples autoridades puede generar desafíos en términos de coherencia y uniformidad regulatoria, ya que cada sector puede interpretar y aplicar los principios de manera distinta, lo que puede derivar en un marco fragmentado.

En América Latina, Brazil ha avanzado en la construcción de un marco regulatorio a través del Proyecto de Ley 2.338/23, el cual se inspira en el modelo europeo al incorporar un enfoque basado en riesgos y la protección de los derechos fundamentales. Este planteamiento propone clasificar los sistemas de inteligencia artificial según su nivel de impacto, estableciendo mayores obligaciones para aquellos considerados de alto riesgo, particularmente en ámbitos sensibles como la seguridad, el acceso a servicios o la toma de decisiones automatizadas. En este sentido, la iniciativa busca garantizar que los sistemas de IA sean desarrollados y

utilizados de manera “segura, transparente y respetuosa de los derechos fundamentales” (Congreso Nacional de Brasil, 2023, traducción propia).

No obstante, la propuesta también incorpora elementos de flexibilidad orientados a adaptarse a las particularidades del contexto regional, como las capacidades institucionales, la infraestructura tecnológica y las condiciones socioeconómicas. En este sentido, el modelo brasileño no replica de manera estricta el esquema europeo, sino que lo adapta, buscando equilibrar la adopción de estándares internacionales con las necesidades locales y evitando la imposición de un marco excesivamente rígido.

En conjunto, estos modelos híbridos reflejan una tendencia hacia la búsqueda de equilibrio entre innovación y regulación, evitando tanto la rigidez normativa como la ausencia de control. En este sentido, se posicionan como puntos de convergencia entre los principales modelos regulatorios a nivel global —basados en derechos, innovación y control estatal— al articular la protección de derechos, el desarrollo tecnológico y la viabilidad institucional en contextos diversos.

6. México: un modelo en construcción

En el caso de México, la regulación de la inteligencia artificial se encuentra en una etapa incipiente, caracterizada por la ausencia de un marco normativo integral y la presencia de iniciativas de carácter ético y orientador. A diferencia de otros modelos internacionales, el país aún no cuenta con una legislación específica que regule de manera sistemática el desarrollo, implementación y uso de la inteligencia artificial, lo que ha dado lugar a un enfoque fragmentado en su gobernanza.

En este contexto, han surgido instrumentos como los denominados *Principios de Chapultepec sobre Inteligencia Artificial*, los cuales constituyen uno de los primeros esfuerzos en el país por establecer lineamientos que orienten el desarrollo y uso responsable de estas tecnologías. Estos principios buscan sentar las bases de una

gobernanza ética de la inteligencia artificial, colocando en el centro la protección de los derechos humanos y el bienestar social.

De manera específica, los *Principios de Chapultepec* se estructuran en torno a ejes fundamentales como la centralidad de la persona, la equidad, la inclusión, la transparencia, la rendición de cuentas y la responsabilidad en el uso de la tecnología. A partir de estos principios, se plantea que los sistemas de inteligencia artificial deben diseñarse e implementarse de forma que no reproduzcan ni amplifiquen desigualdades existentes, especialmente en contextos marcados por brechas sociales y digitales como el mexicano. Asimismo, se enfatiza que el uso de la inteligencia artificial debe ser comprensible para los usuarios, promoviendo la explicabilidad de los sistemas y el acceso a información clara sobre su funcionamiento (Gobierno de México, 2024).

Otro aspecto relevante de estos principios es la incorporación de la dimensión social en el desarrollo tecnológico, al reconocer que la inteligencia artificial no es neutral, sino que puede tener impactos diferenciados en distintos grupos de la población. En este sentido, se promueve la necesidad de evaluar de manera anticipada los posibles riesgos asociados al uso de estos sistemas, así como de establecer mecanismos que permitan prevenir la discriminación algorítmica, proteger la privacidad de los datos y garantizar el respeto a los derechos fundamentales.

De igual forma, los *Principios de Chapultepec* subrayan la importancia de la colaboración entre actores públicos, privados y académicos, reconociendo que la gobernanza de la inteligencia artificial requiere de un enfoque multisectorial. Esta perspectiva busca fomentar la construcción de capacidades institucionales y técnicas, así como el desarrollo de marcos de cooperación que permitan enfrentar los desafíos asociados a estas tecnologías de manera coordinada.

Desde una perspectiva jurídica, estos principios se inscriben dentro del ámbito del *soft law*, lo que implica que funcionan como lineamientos orientadores sin establecer obligaciones vinculantes ni mecanismos formales de sanción. Esta característica limita su capacidad de implementación efectiva, ya que su

cumplimiento depende en gran medida de la voluntad de los actores involucrados y de su adopción voluntaria en distintos sectores.

Sin embargo, su relevancia radica en que constituyen un punto de partida para la construcción de un marco regulatorio más amplio, al introducir en el debate público elementos clave como la ética, la responsabilidad y la protección de derechos en el uso de la inteligencia artificial. En este sentido, los *Principios de Chapultepec* permiten visibilizar la necesidad de avanzar hacia esquemas normativos más robustos que traduzcan estos lineamientos en obligaciones concretas, acompañadas de mecanismos de supervisión, evaluación de riesgos y rendición de cuentas.

En este contexto, el caso mexicano enfrenta el reto de articular estos principios con políticas públicas y marcos legales que respondan a sus condiciones específicas, como la brecha digital, la desigualdad social, la limitada infraestructura tecnológica y las capacidades institucionales aún en desarrollo. De este modo, la construcción de una regulación en materia de inteligencia artificial en México no solo implica adoptar estándares internacionales, sino también diseñar un modelo propio que sea sensible a su realidad social.

7. Conclusión

El análisis de las distintas regulaciones de la inteligencia artificial a nivel global permite identificar la coexistencia de enfoques diferenciados que responden a contextos políticos, económicos y sociales específicos. Lejos de configurarse como un campo homogéneo, la gobernanza de la IA se ha desarrollado a partir de estrategias diversas que reflejan prioridades distintas en torno a la innovación tecnológica, la protección de los derechos fundamentales y el control del entorno digital.

En este marco, el modelo de la Unión Europea se distingue por su enfoque normativo estructurado, basado en la gestión de riesgos y en una fuerte orientación hacia la protección de derechos. En contraste, Estados Unidos privilegia la

innovación y la competitividad mediante esquemas flexibles y descentralizados, lo que favorece la adaptabilidad tecnológica, aunque con retos en términos de coherencia regulatoria. Por su parte, China adopta un modelo de gobernanza centralizada en el que el Estado asume un papel predominante en la supervisión y control de la inteligencia artificial, integrando su desarrollo dentro de objetivos más amplios de estabilidad social y control informativo.

A su vez, los modelos híbridos, como los observados en Canadá, Reino Unido y Brasil, evidencian una tendencia hacia la búsqueda de equilibrio entre regulación e innovación, combinando elementos de distintos enfoques y adaptándolos a sus contextos nacionales. Estos modelos ponen de manifiesto que la regulación de la inteligencia artificial no puede desvincularse de las capacidades institucionales, las condiciones socioeconómicas y las prioridades estratégicas de cada país.

En este contexto, el caso de México refleja un modelo en construcción, caracterizado por la ausencia de una regulación integral y la presencia de instrumentos de carácter ético y orientador. Si bien esta situación plantea desafíos relevantes en términos de protección de derechos y supervisión de los sistemas de IA, también abre la posibilidad de diseñar un marco normativo que incorpore las mejores prácticas internacionales y responda de manera más adecuada a las particularidades del entorno nacional.

En términos generales, el estudio comparado de estos modelos permite concluir que la regulación de la inteligencia artificial trasciende el ámbito técnico o jurídico, configurándose como un fenómeno profundamente político, económico y geoestratégico. En este sentido, las decisiones regulatorias no solo determinan el desarrollo tecnológico, sino también la distribución del poder, el acceso a oportunidades y la protección de los derechos en las sociedades digitales.

Por ello, avanzar hacia esquemas de gobernanza más integrales, flexibles y coordinados a nivel internacional se presenta como un desafío fundamental para el futuro. Dichos esquemas deberán no solo fomentar la innovación, sino también garantizar que el desarrollo de la inteligencia artificial se oriente hacia el bienestar

social, la reducción de desigualdades y la protección efectiva de los derechos fundamentales, en un contexto global marcado por profundas asimetrías tecnológicas.

Referencias

Cyberspace Administration of China. (2022). *Internet Information Service Algorithmic Recommendation Management Provisions*. <http://www.cac.gov.cn>

Cyberspace Administration of China. (2023). *Interim Measures for the Management of Generative Artificial Intelligence Services*. <http://www.cac.gov.cn>

Congresso Nacional do Brasil. (2023). *Projeto de Lei No. 2.338/2023*. <https://www.camara.leg.br>

European Commission. (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>

European Parliament. (2024). *Artificial Intelligence Act*. <https://www.europarl.europa.eu>

Gobierno de España. (s. f.). *Agencia Española de Supervisión de la Inteligencia Artificial (AESIA)*. <https://aesia.digital.gob.es/es>

Gobierno de México. (2024). *Principios de Chapultepec sobre Inteligencia Artificial*. <https://www.gob.mx>

Government of Canada. (2022). *Artificial Intelligence and Data Act (AIDA)*. <https://ised-isde.canada.ca>

Ley Federal de Protección de Datos Personales en Posesión de los Particulares. (2010). *Diario Oficial de la Federación*. <https://www.diputados.gob.mx>

Ley General de Transparencia y Acceso a la Información Pública. (2015). *Diario Oficial de la Federación*. <https://www.diputados.gob.mx>

OECD. (2022). *OECD Digital Economy Outlook 2022*. <https://www.oecd.org/digital/oecd-digital-economy-outlook-2022.htm>

OECD. (2023). *Framework for the classification of AI systems*.
<https://www.oecd.org>

The White House. (2022). *Blueprint for an AI Bill of Rights*.
<https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/>

The White House. (2023). *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*.
<https://www.whitehouse.gov>

Treasury Board of Canada Secretariat. (2019). *Directive on Automated Decision-Making and Algorithmic Impact Assessment*. <https://www.canada.ca>

UK Government. (2023). *A pro-innovation approach to AI regulation*.
<https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach>

UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

CAPÍTULO XV

ÉTICA, INVESTIGACIÓN E INTELIGENCIA ARTIFICIAL GENERATIVA EN LA EDUCACIÓN SUPERIOR LATINOAMERICANA: FUNDAMENTOS FILOSÓFICOS PARA UNA ORIENTACIÓN DE POLÍTICA UNIVERSITARIA.

Ethics, research and Generative Artificial Intelligence in Latin American
Higher Education: philosophical foundations for a university policy
orientation.

Jorge Alberto Rosas Castro

jrosasc@azc.uam.mx

Departamento de Administración, Unidad Azcapotzalco,
Universidad Autónoma Metropolitana
<https://orcid.org/0000-0002-9924-5731>

Gabriela Camacho Deonicio

gabriela.cd92@gmail.com

Coordinadora Editorial de la Red Mexicana de Investigadores en
Estudios Organizacionales (REMINEO), A. C.
0000-0000-0000-0000

Resumen

El capítulo analiza críticamente el impacto de la Inteligencia Artificial Generativa en los procesos de investigación dentro de la Educación Superior latinoamericana, sosteniendo que su incorporación no puede evaluarse únicamente desde criterios de eficiencia técnica, sino desde fundamentos éticos, filosóficos y académicos. El texto plantea que la investigación universitaria implica responsabilidad intelectual, juicio crítico, integridad científica y compromiso público con el conocimiento, elementos que pueden verse debilitados cuando las herramientas generativas desplazan operaciones propias del pensamiento humano.

Desde distintas tradiciones filosóficas y enfoques decoloniales, el capítulo sostiene que la Inteligencia Artificial Generativa debe entenderse como una mediación

sociotécnica que reorganiza prácticas investigativas, modifica relaciones con el conocimiento y exige nuevas formas de supervisión ética e institucional. Asimismo, identifica dilemas relacionados con la autoría intelectual, la transparencia metodológica, los sesgos algorítmicos, la homogeneización del discurso académico y la posible reproducción de desigualdades epistémicas en América Latina.

Finalmente, el texto propone fortalecer una formación ética, crítica y digital del investigador, así como desarrollar marcos de gobernanza universitaria orientados a garantizar transparencia, supervisión humana, rendición de cuentas e integridad científica. Se concluye que el uso de la Inteligencia Artificial Generativa solo es éticamente legítimo cuando permanece subordinado al juicio humano y a la responsabilidad académica, evitando que la automatización sustituya el pensamiento crítico y la deliberación intelectual.

Palabras clave: Inteligencia Artificial Generativa, ética de la investigación, integridad científica, Educación Superior, pensamiento crítico, gobernanza universitaria, transparencia metodológica, autoría intelectual, colonialidad del saber, justicia epistémica, formación digital, responsabilidad ética, mediación sociotécnica, universidad latinoamericana, inteligencia artificial en investigación

Abstract

The chapter critically analyzes the impact of Generative Artificial Intelligence on research processes within Latin American Higher Education, arguing that its incorporation cannot be evaluated only from criteria of technical efficiency, but from ethical, philosophical and academic foundations. The text states that university research implies intellectual responsibility, critical judgment, scientific integrity, and public commitment to knowledge, elements that can be weakened when generative tools displace operations of human thought.

From different philosophical traditions and decolonial approaches, the chapter argues that Generative Artificial Intelligence should be understood as a sociotechnical mediation that reorganizes research practices, modifies relationships

with knowledge, and requires new forms of ethical and institutional supervision. It also identifies dilemmas related to intellectual authorship, methodological transparency, algorithmic biases, the homogenization of academic discourse, and the possible reproduction of epistemic inequalities in Latin America.

Finally, the text proposes to strengthen ethical, critical and digital training for researchers, as well as to develop university governance frameworks aimed at guaranteeing transparency, human supervision, accountability and scientific integrity. It is concluded that the use of Generative Artificial Intelligence is only ethically legitimate when it remains subordinate to human judgment and academic responsibility, preventing automation from replacing critical thinking and intellectual deliberation.

Keywords: Generative Artificial Intelligence, research ethics, scientific integrity, Higher Education, critical thinking, university governance, methodological transparency, intellectual authorship, coloniality of knowledge, epistemic justice, digital training, ethical responsibility, sociotechnical mediation, Latin American university, artificial intelligence in research

Introducción

La expansión reciente de la Inteligencia Artificial Generativa (IAG) ha modificado de manera significativa los modos en que se busca, organiza, procesa y presenta la información en la Educación Superior. En el ámbito de la investigación, estas herramientas intervienen cada vez con mayor frecuencia en tareas como la exploración bibliográfica, la delimitación temática, la síntesis documental, la redacción preliminar de borradores, la revisión de estilo e incluso en apoyos iniciales para el análisis de datos. Esta presencia creciente ha sido celebrada por su capacidad para acelerar procesos y asistir en el tratamiento de grandes volúmenes de información; sin embargo, también ha puesto en tensión principios fundamentales de la vida académica, entre ellos la autoría intelectual, la

transparencia metodológica, la verificabilidad del conocimiento, la integridad científica y la responsabilidad del investigador (Chan & Colloton, 2024; UNESCO, 2023).

En este contexto, el debate sobre la IAG no puede reducirse a una dicotomía entre aceptación entusiasta y rechazo precautorio. El problema de fondo no es únicamente si estas tecnologías son útiles, sino qué transformaciones introducen en la producción del conocimiento y bajo qué condiciones su incorporación puede considerarse éticamente legítima. La investigación universitaria no es una actividad meramente instrumental: supone juicio crítico, deliberación, contrastación de fuentes, construcción argumentativa y responsabilidad pública sobre lo afirmado. Cuando una herramienta generativa participa en fases sustantivas de ese proceso, se vuelve necesario preguntar si funciona como apoyo legítimo o si comienza a desplazar indebidamente operaciones intelectuales que deberían permanecer bajo control humano. En otras palabras, el problema no radica solo en la herramienta, sino en la relación que investigadores e instituciones establecen con ella.

Esta discusión adquiere particular relevancia en la Educación Superior latinoamericana. La adopción de IAG ocurre en contextos atravesados por desigualdades de infraestructura, acceso desigual a tecnologías, dependencia de plataformas desarrolladas fuera de la región y persistencia de jerarquías epistémicas que suelen privilegiar marcos teóricos, corpus y lenguajes dominantes del Norte global. Por ello, una reflexión ética sobre investigación e IAG en América Latina no puede limitarse a trasplantar lineamientos generales, sino que debe examinar críticamente sus condiciones concretas de uso, sus posibles efectos en la formación de investigadores y su relación con problemas de colonialidad del saber, justicia epistémica y responsabilidad social del conocimiento (Dussel, 1998; Quijano, 2000; Walsh, 2007).

Desde esta perspectiva, el presente capítulo parte de una distinción fundamental entre moral y ética. La moral remite al conjunto de normas, costumbres, valores y

creencias que orientan de hecho la conducta de sujetos y comunidades; la ética, en cambio, constituye la reflexión crítica que examina la validez, el alcance y los límites de esos marcos normativos. Esta distinción es decisiva porque la discusión sobre IAG en la investigación universitaria no puede agotarse en reglas de uso o en códigos de conducta. Requiere una fundamentación filosófica que permita preguntarse por la justicia, la responsabilidad, la dignidad, la verdad y el sentido público del conocimiento en una época en que la automatización del lenguaje amenaza con confundirse con la actividad misma del pensar.

El capítulo sostiene que la IAG debe ser entendida no solo como una herramienta técnica, sino como una mediación sociotécnica que reordena tareas intelectuales, altera prácticas investigativas y exige una reconsideración de la responsabilidad científica. Desde este enfoque, la cuestión central ya no es si la IAG “piensa” como un ser humano, sino bajo qué criterios debe ser diseñada, regulada y utilizada para no debilitar la autonomía intelectual del investigador ni erosionar la integridad académica. Asimismo, se sostiene que la legitimidad ética de su uso depende de una articulación entre cuatro dimensiones: una fundamentación filosófica de la ética, una comprensión de la universidad como espacio formativo, una defensa de la integridad científica en la investigación y una orientación institucional capaz de traducirse en criterios de gobernanza universitaria.

Sobre esta base, el capítulo se organiza en siete apartados. El primero recupera la ética como fundamento filosófico de la acción humana, dialogando con tradiciones como la ética de las virtudes, la ética del deber, la ética de la responsabilidad y la ética de la liberación. El segundo examina la relación entre ética, educación y responsabilidad formativa de la universidad, particularmente en la Educación Superior latinoamericana. El tercero aborda la ética de la investigación e integridad científica como marco inmediato para comprender los desafíos contemporáneos del trabajo académico. El cuarto analiza la IAG como mediación sociotécnica y desarrolla los principales dilemas éticos asociados a su uso en investigación. El quinto se concentra en la formación ética, crítica y digital del investigador, entendida

como condición indispensable para un uso responsable de estas herramientas. El sexto plantea la necesidad de marcos de gobernanza universitaria que orienten institucionalmente la incorporación de la IAG. Finalmente, las conclusiones articulan estos niveles para proponer una orientación de política universitaria que subordine la innovación tecnológica a la integridad científica, al juicio crítico y a la responsabilidad social del conocimiento.

En este sentido, el problema que orienta el capítulo puede formularse así: la incorporación de la IAG en los procesos de investigación en la Educación Superior tiende a acelerar operaciones de búsqueda, síntesis y redacción, pero al mismo tiempo puede debilitar el pensamiento crítico, desdibujar la autoría, introducir opacidad metodológica, amplificar sesgos y favorecer una racionalidad académica centrada en la productividad antes que en la integridad científica. De ahí surge la pregunta central: ¿en qué medida el uso de la IAG en los procesos de investigación en la Educación Superior latinoamericana desplaza, debilita o reconfigura el pensamiento crítico, el juicio intelectual y la responsabilidad ética de quienes investigan, y bajo qué condiciones puede integrarse sin sacrificar la integridad científica en favor de la velocidad, la automatización y la eficiencia?

La tesis que se sostiene es la siguiente: la IAG no sustituye legítimamente el pensamiento crítico ni la responsabilidad ética del investigador; su uso solo es defendible cuando permanece subordinado al juicio humano, a la integridad científica y a marcos de transparencia, atribución, supervisión y rendición de cuentas. Por ello, la ética no debe añadirse al final del proceso, sino incorporarse desde el diseño, desarrollo, regulación y uso de la IAG en la investigación universitaria. Desde esta proposición, el capítulo busca contribuir a una discusión que no se limite a describir riesgos, sino que ofrezca fundamentos filosóficos y criterios institucionales para orientar políticas universitarias en materia de investigación e IAG en la Educación Superior latinoamericana.

La ética como fundamento filosófico de la acción humana

Como se explicó en la introducción, aunque en el lenguaje cotidiano ética y moral suelen emplearse como sinónimos, en el plano filosófico conviene distinguirlas, porque el problema no se reduce a verificar si existen reglas institucionales de uso, sino a preguntar por qué ciertas prácticas deben considerarse justas, responsables o inadmisibles en la producción de conocimiento. La ética, en este sentido, no describe simplemente lo que una comunidad académica hace o tolera, sino que interroga las razones por las cuales determinadas acciones deberían aceptarse, limitarse o rechazarse (Alexander & Moore, 2021; Kraut, 2022).

Desde sus orígenes, la filosofía moral ha intentado responder a la pregunta por la vida buena y por la acción justa. En la tradición aristotélica, la ética se vincula con la formación del carácter y con el cultivo de virtudes que permiten deliberar prudentemente acerca de lo que conviene hacer en situaciones concretas. Aristóteles sitúa en la *phronesis* o prudencia una capacidad decisiva: juzgar rectamente en el ámbito de la acción, evitando tanto el exceso como el defecto. Esta perspectiva resulta especialmente fecunda para el presente capítulo, porque la investigación universitaria no es una actividad puramente mecánica ni una mera aplicación de procedimientos. Investigar implica seleccionar problemas, evaluar fuentes, interpretar datos, establecer jerarquías de relevancia y responder por los efectos de lo que se afirma. En todos esos momentos, el investigador no solo opera técnicamente: delibera, valora y ejerce juicio práctico. Por ello, la ética de la investigación no puede descansar únicamente en reglas externas; requiere también virtudes intelectuales y morales que sostengan la práctica científica desde dentro (Aristóteles, 2009; Hursthouse & Pettigrove, 2023; Kraut, 2022).

Junto a esta tradición, la ética deontológica ofrece otra clave decisiva. En ella, el centro no está puesto principalmente en la formación del carácter, sino en la validez racional de principios que orientan la acción y establecen deberes, prohibiciones y responsabilidades. Desde este horizonte, no todo lo técnicamente posible es

moralmente aceptable. La utilidad o la eficiencia no bastan para justificar una práctica si esta vulnera la dignidad de las personas, oculta información relevante, falsea procedimientos o instrumentaliza a otros como simples medios. Esta línea de pensamiento resulta especialmente importante para el análisis de la IAG, porque permite cuestionar una racionalidad según la cual la velocidad de producción, la economía de tiempo o la optimización del trabajo bastarían para legitimar el uso de sistemas generativos en investigación. La ética del deber recuerda que la verdad, la transparencia, la honestidad y el respeto por el otro no pueden sacrificarse en nombre del rendimiento técnico (Alexander & Moore, 2021).

La reflexión ética contemporánea ha debido, además, enfrentarse al problema específico de la técnica y de su creciente poder sobre la vida y el futuro. En este punto, Hans Jonas ocupa un lugar fundamental. Su ética de la responsabilidad parte del hecho de que la acción humana mediada por tecnología puede tener hoy efectos de gran escala e incluso irreversibles. En consecuencia, la ética ya no puede limitarse a regular relaciones inmediatas entre sujetos, sino que debe atender también las consecuencias amplificadas de la acción técnica sobre seres humanos, otros seres vivos, comunidades y ecosistemas. Esta perspectiva resulta especialmente pertinente para pensar la IAG en el contexto universitario, pues obliga a desplazar la discusión desde la mera utilidad instrumental hacia la pregunta por sus efectos sobre la autonomía intelectual, la justicia epistémica, la integridad científica y la vida colectiva. Lo decisivo ya no es solo si una tecnología “funciona”, sino qué tipo de mundo contribuye a producir y bajo qué responsabilidades puede ser legítimamente utilizada (Jonas, 1995).

Para una reflexión situada en América Latina, esta base filosófica necesita ampliarse con una perspectiva crítica de la dominación y de la exclusión. Enrique Dussel aporta en este punto una contribución decisiva. En su ética de la liberación, el criterio último de validez ética se encuentra en la afirmación y reproducción de la vida humana, especialmente de la vida de quienes han sido colocados en condiciones de exterioridad por sistemas de poder. La ética no puede quedar

restringida a principios abstractos formulados desde posiciones dominantes; debe partir también del sufrimiento, la exclusión y la negación de quienes resultan afectados por estructuras históricas de explotación y subordinación. Aplicada al tema de este capítulo, esta perspectiva obliga a preguntar si el uso de la IAG en investigación fortalece condiciones de vida, conocimiento y justicia, o si reproduce nuevas formas de exclusión, desresponsabilización académica y colonialidad tecnológica. Desde esta mirada, la ética no es un adorno normativo añadido al final del proceso, sino una crítica de las mediaciones que amenazan la dignidad, la autonomía y la vida de sujetos y comunidades (Dussel, 1998, 2013).

Esta apertura dialoga, a su vez, con los enfoques decoloniales. Si bien la ética filosófica ha buscado con frecuencia principios de validez amplia, en contextos latinoamericanos resulta indispensable reconocer que la producción de conocimiento ha estado atravesada por relaciones históricas de colonialidad del poder y del saber. De ahí que una reflexión ética sobre la IAG no pueda limitarse a discutir normas generales de uso; debe también examinar qué epistemologías privilegian los sistemas generativos, qué tradiciones intelectuales desplazan y qué asimetrías globales reactivan en la circulación del conocimiento académico. La ética, entendida desde este horizonte, no solo pregunta qué debe hacerse, sino desde dónde se produce el juicio, quiénes quedan autorizados a definir lo válido y qué voces son sistemáticamente relegadas. Este elemento es especialmente relevante en una investigación orientada a la Educación Superior latinoamericana, donde las cuestiones de justicia epistémica y soberanía intelectual no pueden separarse del debate sobre tecnología.

En el ámbito iberoamericano, Adela Cortina ha insistido en que la ética no debe entenderse como una cuestión puramente subjetiva o privada, sino como una tarea de fundamentación racional y de construcción de mínimos compartidos para la vida pública. Su aporte resulta valioso porque permite articular la dimensión personal de la ética con su dimensión institucional y social. La investigación universitaria, vista desde esta perspectiva, no es solo una práctica individual de especialistas; es

también una actividad pública cuyos resultados inciden en políticas, instituciones, comunidades y formas de vida. Por ello, la ética aplicada a la investigación requiere tanto responsabilidad personal como marcos compartidos de justicia, integridad y deliberación. La introducción de IAG en este terreno vuelve todavía más urgente esa articulación, ya que las decisiones sobre diseño, uso y regulación de estas herramientas no afectan solo a usuarios aislados, sino a culturas académicas enteras y, en última instancia, al sentido social del conocimiento (Cortina, 2013).

A partir de estas tradiciones, puede sostenerse que la ética constituye el fundamento filosófico de la acción humana porque permite evaluar racionalmente las prácticas a la luz de categorías como bien, justicia, deber, virtud, responsabilidad y dignidad. Cuanto mayor es el poder técnico disponible, más necesaria se vuelve esta reflexión. Tal afirmación adquiere una importancia particular en el caso de la IAG, cuya capacidad para producir texto, sintetizar información, sugerir categorías y modelar decisiones puede generar la falsa impresión de que el pensamiento crítico humano resulta prescindible. Sin embargo, precisamente porque estas tecnologías amplifican la velocidad y el alcance del procesamiento de información, se vuelve más urgente discernir qué fines persiguen, qué sesgos incorporan, qué intereses refuerzan y qué relación establecen entre conocimiento, verdad y poder. La ética, por tanto, no queda desplazada por la automatización: se vuelve más necesaria que antes.

Desde esta óptica, la pregunta central no es si la IAG “piensa” como un ser humano, sino bajo qué criterios debe ser diseñada, regulada y utilizada para no erosionar la responsabilidad moral e intelectual de quienes investigan. La IAG no es un sujeto moral, ni puede asumir deberes, culpa o autoría; pero sí puede ser configurada, entrenada y empleada de maneras que reproduzcan o amplifiquen valores, sesgos e intenciones humanas. Por ello, la reflexión ética debe acompañar todo el ciclo tecnológico: diseño, entrenamiento, implementación, uso académico, presentación de resultados y efectos sociales. Esta base filosófica es la que permite pasar, en el siguiente apartado, de una ética general de la acción humana a una reflexión más

específica sobre la universidad como institución formativa y sobre la responsabilidad educativa de la Educación Superior frente a la innovación tecnológica.

Ética, educación y responsabilidad formativa de la universidad

La universidad no es únicamente una organización destinada a transmitir información especializada ni una estructura encargada de certificar competencias profesionales. Constituye, en un sentido más profundo, una institución social orientada a la formación integral de sujetos capaces de comprender críticamente su tiempo, producir conocimiento relevante y actuar responsablemente frente a problemas públicos. Por ello, la relación entre ética y educación no debe concebirse como un complemento marginal del currículo, sino como una dimensión constitutiva del quehacer universitario. Formar implica siempre orientar criterios de juicio, hábitos intelectuales, disposiciones morales y modos de relación con los otros, con la verdad y con el mundo. De ahí que toda reflexión sobre el uso de la IAG en investigación universitaria deba reconocer primero que la universidad no solo incorpora tecnologías: también forma a quienes habrán de utilizarlas, legitimarlas y eventualmente gobernarlas.

En América Latina, esta afirmación adquiere un sentido específico. Las instituciones de Educación Superior operan en contextos marcados por desigualdades estructurales, brechas de acceso, asimetrías tecnológicas y tensiones entre cobertura, calidad, financiamiento y pertinencia social. En este escenario, la adopción de tecnologías digitales y de sistemas de IAG no puede evaluarse solo por su capacidad para aumentar eficiencia o productividad. También debe preguntarse si fortalece o debilita la misión formativa de la universidad, si amplía capacidades críticas o si genera nuevas dependencias, y si contribuye a democratizar el conocimiento o reproduce desigualdades ya existentes. La cuestión ética, por tanto, no es externa a la innovación tecnológica: se encuentra en el centro mismo de la decisión universitaria sobre qué tipo de formación desea promover.

Desde la pedagogía crítica, Paulo Freire ofrece una clave decisiva para comprender este problema. Educar, en su horizonte, no equivale a depositar contenidos en sujetos pasivos, sino a promover procesos de conciencia, diálogo, problematización y transformación. Llevada al terreno contemporáneo, esta idea obliga a desconfiar de cualquier incorporación tecnológica que reduzca la actividad intelectual a la recepción acrítica de respuestas rápidas o a la administración instrumental de información. Si la universidad acepta sin mediación reflexiva el uso de la IAG, corre el riesgo de favorecer prácticas de heteronomía cognitiva: estudiantes e investigadores podrían delegar en la herramienta operaciones que son constitutivas del pensamiento mismo, como delimitar un problema, juzgar una fuente, construir una interpretación o sostener una conclusión. En consecuencia, el valor formativo de la IAG no depende solo de lo que puede hacer, sino del modo en que su uso se inserta en prácticas educativas orientadas a la autonomía intelectual (Freire, 2005).

Esta preocupación puede leerse también a la luz de Martha Nussbaum. Su defensa de la educación humanista insiste en que una sociedad democrática necesita sujetos capaces de razonar críticamente, imaginar la situación del otro y resistir la reducción de la educación a fines meramente utilitarios. Este planteamiento resulta particularmente pertinente en el debate sobre IAG, porque la expansión de herramientas generativas puede reforzar una lógica de rendimiento según la cual vale más producir más rápido que comprender mejor, y vale más responder con inmediatez que sostener procesos rigurosos de deliberación. Frente a ello, la universidad debe preservar espacios de lectura atenta, escritura reflexiva, argumentación compleja y discusión ética. De otro modo, la innovación tecnológica podría terminar debilitando precisamente aquello que la Educación Superior debería proteger: el juicio prudente, la imaginación crítica y la responsabilidad pública del conocimiento (Nussbaum, 2010).

La responsabilidad formativa de la universidad se expresa con particular fuerza en la formación de investigadores. Investigar no consiste solo en dominar métodos y técnicas; implica también desarrollar disposiciones éticas como honestidad

intelectual, rigor, apertura a la crítica, respeto por las fuentes, cuidado de los datos y conciencia de los efectos del conocimiento producido. En este sentido, la ética de la investigación no puede improvisarse al final del proceso ni limitarse a códigos de conducta. Debe incorporarse desde la formación misma de quienes investigan, como parte de una cultura académica que valore la integridad científica por encima de la apariencia de productividad. Esto es aún más urgente cuando se trabaja con IAG, pues la facilidad para generar borradores, resúmenes o esquemas bien estructurados puede inducir a una falsa sensación de suficiencia intelectual si no se cuenta con hábitos de verificación, contraste y juicio crítico.

La literatura reciente sobre IA en Educación Superior converge en este punto. La guía de la UNESCO sobre IA generativa en educación e investigación insiste en que estas tecnologías no deben desplazar la agencia humana ni sustituir la deliberación intelectual, sino fortalecer capacidades bajo principios de supervisión, transparencia y responsabilidad. De forma convergente, Cecilia K. Y. Chan propone que toda política institucional sobre IA en la universidad articule dimensiones pedagógicas, operativas y de gobernanza, incluyendo alfabetización en IA, criterios claros de uso y acompañamiento para docentes, estudiantes e investigadores. Estas orientaciones son particularmente valiosas porque muestran que la incorporación ética de la IAG no depende exclusivamente del comportamiento individual, sino de un ecosistema formativo que haga posible su uso crítico, contextualizado y responsable.

En el contexto mexicano y latinoamericano, el diagnóstico de la UNAM sobre uso y desarrollo ético de la inteligencia artificial en la universidad refuerza este argumento al subrayar que principios como honestidad, integridad, autoría intelectual, transparencia y responsabilidad siguen siendo ejes vertebrales para pensar la docencia y la investigación mediadas por IA. Ello permite sostener que la universidad tiene una doble responsabilidad: por un lado, formar sujetos capaces de utilizar la IAG sin abdicar de su juicio; por otro, generar condiciones

institucionales para que ese uso no erosione la legitimidad del conocimiento producido.

Desde una perspectiva decolonial, esta responsabilidad formativa también exige examinar críticamente qué saberes, voces y tradiciones quedan privilegiados o desplazados por la mediación tecnológica. La universidad latinoamericana no debería limitarse a enseñar a usar sistemas generativos producidos fuera de la región; debe también formar capacidad crítica para detectar sesgos epistémicos, ausencias sistemáticas y formas de colonialidad del saber reproducidas por herramientas entrenadas sobre corpus desiguales. En este punto, Catherine Walsh y Aníbal Quijano resultan especialmente sugerentes: la educación crítica en América Latina no puede separarse de la interrogación por las jerarquías de conocimiento que organizan lo visible y lo invisible, lo central y lo periférico. Formar para el uso ético de IAG implica, entonces, enseñar también a resistir la naturalización de marcos hegemónicos de producción de verdad (Quijano, 2000; Walsh, 2007).

Por ello, la relación entre ética y educación universitaria puede formularse con precisión: la universidad cumple una función ética no solo cuando enseña normas o códigos, sino cuando configura ambientes intelectuales en los que pensar críticamente, argumentar con rigor, reconocer al otro y responder por lo que se afirma se vuelven prácticas habituales. La IAG, en este marco, puede ser un recurso valioso para apoyar procesos de búsqueda, organización y tratamiento preliminar de información, pero solo en la medida en que quede subordinada a fines formativos y científicos más amplios. Cuando su uso desplaza el esfuerzo de comprensión, homogeneiza la escritura, debilita la relación crítica con las fuentes o promueve una cultura de la respuesta inmediata, la universidad deja de formar y comienza a adaptarse pasivamente a una lógica tecnológica ajena a su misión.

En síntesis, la responsabilidad formativa de la universidad frente a la IAG no consiste en prohibirla ni en celebrarla sin reservas, sino en construir criterios para

su integración crítica. Esto implica enseñar a distinguir entre asistencia y sustitución, entre apoyo instrumental y delegación indebida del pensamiento, entre uso transparente y opacidad metodológica, entre eficiencia técnica e integridad académica. En la Educación Superior latinoamericana, donde la adopción tecnológica se cruza con desigualdades persistentes y dependencias estructurales, esta responsabilidad adquiere un peso aún mayor. Sobre esta base, la discusión puede avanzar hacia una cuestión más específica: cómo se articulan ética de la investigación e integridad científica cuando la producción de conocimiento comienza a estar mediada por sistemas capaces de redactar, sintetizar, sugerir y modelar decisiones.

Ética de la investigación e integridad científica en Educación Superior

La ética de la investigación constituye un campo específico de la ética aplicada que examina los principios, valores y responsabilidades que deben orientar la producción, validación, comunicación y uso del conocimiento. En el ámbito de la Educación Superior, esta cuestión adquiere una importancia decisiva, porque las universidades no solo generan saberes, sino que también forman a quienes habrán de producirlos, evaluarlos, difundirlos y aplicarlos socialmente. Por ello, la ética de la investigación no puede reducirse a una serie de requisitos administrativos o a un conjunto de prohibiciones formales; remite, más profundamente, a las condiciones de legitimidad del conocimiento científico y académico. Investigar éticamente significa responder no solo por la corrección técnica de un procedimiento, sino también por la honestidad intelectual del proceso, por la relación responsable con las fuentes, por el tratamiento justo de los datos, por la transparencia metodológica y por las consecuencias sociales del conocimiento producido (Resnik, 2020; Steneck, 2006).

En este sentido, la noción de integridad científica constituye uno de los ejes centrales de la ética de la investigación. La integridad no se limita a evitar conductas abiertamente fraudulentas, como la fabricación, falsificación o plagio, aunque estos

comportamientos sean expresiones paradigmáticas de su ruptura. También comprende hábitos positivos: rigor, trazabilidad, honestidad en la presentación de hallazgos, reconocimiento adecuado de autorías y colaboraciones, apertura a la revisión crítica, y disposición a corregir errores cuando estos se detectan. Así entendida, la integridad científica es una práctica de coherencia entre métodos, resultados, interpretación y comunicación académica. Su importancia es aún mayor en la Educación Superior, porque la credibilidad de la investigación universitaria depende de la confianza pública en que los conocimientos que allí se producen no están guiados por engaño, simulación o manipulación, sino por un compromiso efectivo con la búsqueda responsable de la verdad (Steneck, 2006; ALLEA, 2023).

En contextos universitarios, la ética de la investigación se articula además con la formación de nuevas generaciones de investigadores. Esto significa que la integridad científica no debe entenderse solo como un estándar externo de evaluación, sino como una dimensión formativa. Los estudiantes de licenciatura y posgrado, así como los investigadores en formación, aprenden a investigar no únicamente por medio de manuales metodológicos, sino también a través de culturas institucionales, estilos de tutoría, prácticas de escritura, normas de atribución y formas de relación con la evidencia. Desde esta perspectiva, la universidad tiene la responsabilidad de generar un ecosistema académico donde la investigación rigurosa no sea vista como obstáculo para la productividad, sino como condición de legitimidad del trabajo científico. Tal responsabilidad resulta especialmente importante en un momento en el que herramientas de IAG ofrecen posibilidades inéditas de acelerar procesos de búsqueda, síntesis, redacción y análisis, pero también abren nuevos riesgos de simulación intelectual y debilitamiento del juicio crítico (UNESCO, 2023; Chan & Colloton, 2024).

La irrupción de la IAG en los procesos de investigación obliga, por ello, a repensar la relación entre ética e integridad científica. Tradicionalmente, la ética de la investigación se ha ocupado de problemas como el respeto a sujetos participantes, el consentimiento informado, la confidencialidad, el cuidado de animales no

humanos, la honestidad en el manejo de datos, los conflictos de interés o la autoría académica. Estas preocupaciones no desaparecen con la IAG; por el contrario, se reconfiguran y en algunos casos se intensifican. Cuando un sistema generativo interviene en la delimitación del tema, en la revisión de literatura, en la redacción de borradores, en la organización de referencias o en la interpretación preliminar de datos, surge una pregunta decisiva: ¿en qué punto el uso de una herramienta de apoyo deja de ser asistencia legítima y se convierte en desplazamiento indebido del trabajo intelectual que debía realizar el investigador? La dificultad de responder esta pregunta muestra que la ética de la investigación ya no puede pensarse al margen de la mediación sociotécnica que hoy condiciona la producción del conocimiento (UNESCO, 2023; Castañeda de León y otros., 2025).

Uno de los núcleos más sensibles de esta discusión es la autoría intelectual. La integridad científica exige que quien firma un texto responda efectivamente por su contenido, por sus fuentes, por la validez de sus afirmaciones y por la corrección de sus procedimientos. En este punto, distintas instancias editoriales han convergido en una idea fundamental: la IAG no puede ser considerada autora, porque no es sujeto moral ni puede asumir responsabilidad pública por lo que produce. Tanto COPE como ICMJE han insistido en que la responsabilidad sigue recayendo en personas humanas, incluso cuando se utilicen herramientas de IA en alguna fase del trabajo académico. Esta postura no es menor: protege la noción de autoría como forma de responsabilidad, no solo como atribución nominal. Si la firma de un texto deja de expresar responsabilidad humana, entonces la integridad del proceso investigativo se erosiona en su base misma (Committee on Publication Ethics [COPE], 2023; International Committee of Medical Journal Editors [ICMJE], 2024).

A ello se suma el problema de la transparencia metodológica. Un rasgo esencial de la investigación íntegra consiste en hacer visibles, hasta donde sea posible, las decisiones que sostienen el proceso de indagación: cómo se definió el corpus, qué criterios de selección se utilizaron, qué técnicas de análisis se aplicaron, qué

limitaciones se reconocen y cómo se construyeron las interpretaciones. La IAG introduce aquí una tensión particular, porque muchas de sus respuestas aparecen con un alto grado de fluidez discursiva, pero sin que el usuario conozca con claridad la ruta exacta por la que se produjo una síntesis, una clasificación o una formulación determinada. Por ello, el uso ético de estas herramientas en investigación exige una práctica reforzada de explicitación: cuándo se usaron, para qué tareas, con qué alcances, bajo qué criterios de verificación y con qué límites. Sin esa transparencia, la incorporación de IAG corre el riesgo de volverse una zona opaca dentro del proceso metodológico, afectando la verificabilidad del conocimiento generado (UNESCO, 2023; OECD, 2024).

Otro punto crucial se relaciona con la confiabilidad de la información. Es sabido que los sistemas de IAG pueden generar referencias inexistentes, atribuciones incorrectas, resúmenes engañosamente plausibles o formulaciones que simulan precisión sin sostenerse en evidencia verificable. En investigación, este problema es especialmente grave, porque el lenguaje bien estructurado puede encubrir errores sustantivos. La ética de la investigación exige entonces una actitud activa de verificación. No basta con que la herramienta “parezca convincente”; el investigador debe contrastar fuentes, revisar datos, comprobar citas y ejercer juicio crítico sobre cada uso. En este terreno, la IAG no elimina la carga de la interpretación, sino que la desplaza y la hace más exigente: ahora no solo hay que analizar literatura y datos, sino también evaluar críticamente la calidad y la legitimidad de las respuestas generadas por sistemas automatizados (Chan & Colloton, 2024; UNESCO, 2023).

La cuestión de los sesgos también ocupa un lugar central. Ningún proceso de investigación está completamente exento de sesgos humanos; sin embargo, la IAG puede amplificar, estabilizar o naturalizar sesgos presentes en los datos con los que fue entrenada, en los patrones dominantes de publicación o en marcos culturales hegemónicos. En el contexto de la Educación Superior latinoamericana, este problema tiene una resonancia particular. Si gran parte de los sistemas generativos

se nutre de corpus producidos en contextos anglófonos o del Norte global, sus respuestas pueden invisibilizar tradiciones intelectuales, categorías críticas y experiencias históricas del Sur global. Aquí el diálogo con enfoques decoloniales resulta indispensable: la ética de la investigación no solo debe preguntar si un contenido es técnicamente correcto, sino también qué voces omite, qué jerarquías reproduce y qué formas de colonialidad del saber pueden reactivarse a través de la automatización del conocimiento (Dussel, 1998; Quijano, 2000; Walsh, 2007). Desde esta mirada, la integridad científica no se reduce al cumplimiento formal de reglas, sino que exige atención a las asimetrías epistémicas que atraviesan la producción académica.

En esta línea, Enrique Dussel ofrece una contribución particularmente valiosa. Su ética de la liberación sitúa el criterio ético fundamental en la afirmación y reproducción de la vida humana y en la responsabilidad frente a quienes quedan colocados en posición de exterioridad o exclusión por sistemas de dominación. Trasladado al campo de la investigación universitaria, ello obliga a examinar si el uso de la IAG fortalece procesos de conocimiento al servicio de la vida y la justicia, o si, por el contrario, contribuye a nuevas formas de exclusión, subordinación epistémica y desresponsabilización académica. Esta perspectiva enriquece la noción de integridad científica, porque recuerda que la honestidad metodológica es necesaria, pero no suficiente: también importa preguntarse para quién se produce conocimiento, desde qué marcos de poder y con qué efectos sobre sujetos, comunidades y ecosistemas (Dussel, 1998).

Por ello, en el contexto actual, la ética de la investigación en Educación Superior debe ampliarse en una doble dirección. Por un lado, debe preservar principios clásicos de integridad: honestidad, rigor, autoría responsable, transparencia, verificabilidad y cuidado de los datos. Por otro, debe incorporar problemas emergentes vinculados con la mediación algorítmica del conocimiento: opacidad, automatización de tareas intelectuales, dependencia cognitiva, sesgos estructurales, uso indebido de información sensible, homogeneización discursiva y

debilitamiento del juicio crítico. Esta ampliación no implica abandonar la tradición de la ética de la investigación, sino actualizarla frente a un escenario en el que la producción académica está siendo transformada por tecnologías capaces de intervenir en múltiples fases del trabajo científico (ALLEA, 2023; UNESCO, 2021, 2023).

En suma, la integridad científica en la era de la IAG exige defender una idea fuerte de investigación universitaria: una práctica intelectual, pública y responsable en la que la tecnología puede asistir, pero no sustituir la deliberación, la verificación, la autoría ni la responsabilidad de quienes investigan. La universidad no debe responder a este desafío con prohibiciones abstractas ni con aceptación acrítica, sino con formación, criterios, lineamientos y culturas académicas capaces de distinguir entre apoyo legítimo y delegación indebida del pensamiento. Sobre esta base, el siguiente paso del capítulo consiste en examinar con mayor precisión a la IAG no solo como herramienta, sino como mediación sociotécnica que reordena tiempos, tareas, decisiones y formas de autoridad en los procesos de investigación.

La IAG como mediación sociotécnica y sus dilemas éticos en los procesos de investigación

Comprender la Inteligencia Artificial Generativa (IAG) únicamente como una herramienta de apoyo técnico resulta insuficiente para analizar su impacto en los procesos de investigación en la Educación Superior. Más que un simple instrumento neutral, la IAG debe entenderse como una mediación sociotécnica, es decir, como un dispositivo tecnológico que interviene en prácticas humanas complejas, reorganiza tiempos, redistribuye tareas intelectuales, condiciona formas de interacción con la información y modifica los criterios desde los cuales se produce, valida y comunica el conocimiento. Bajo esta perspectiva, la IAG no actúa al margen del investigador ni constituye un elemento externo al proceso investigativo; se inserta dentro de él, influye en sus secuencias y altera, en distintos grados, la

relación entre sujeto, técnica, lenguaje, evidencia y juicio crítico (Chan & Colloton, 2024; UNESCO, 2023).

La idea de mediación sociotécnica permite superar dos simplificaciones frecuentes. La primera consiste en considerar que la tecnología es un ente puramente instrumental cuya bondad o maldad depende solo del uso que le den los individuos. La segunda, en sentido contrario, atribuye a la tecnología una suerte de agencia autónoma capaz de sustituir por completo la capacidad humana de pensar, decidir y responsabilizarse. Ninguna de estas posiciones resulta adecuada. La IAG no es moralmente autónoma, pero tampoco es neutral en un sentido estricto, ya que su funcionamiento depende de arquitecturas, datos de entrenamiento, criterios de diseño, filtros, lógicas de optimización y contextos de uso que expresan decisiones humanas previas. De ahí que su presencia en la investigación no sea inocua: orienta ciertas operaciones, privilegia determinados tipos de respuesta, sugiere formas de organización discursiva y puede reforzar marcos de inteligibilidad dominantes, invisibilizando otros (UNESCO, 2021; OECD, 2024).

En los procesos de investigación universitaria, esta mediación se observa desde etapas tempranas. La IAG puede intervenir en la delimitación del tema, en la formulación de preguntas iniciales, en la exploración de antecedentes, en la síntesis de literatura, en la redacción de esquemas, en la sugerencia de categorías analíticas, en la revisión de estilo e incluso en la preparación de versiones preliminares de resultados. Tales apoyos pueden ser valiosos cuando se utilizan como recursos instrumentales sometidos a verificación y juicio humano. Sin embargo, justamente porque participan en fases sustantivas del trabajo investigativo, estas tecnologías también pueden producir efectos epistemológicos y éticos de fondo. Una de las cuestiones centrales es que la rapidez con que generan respuestas puede inducir a confundir celeridad operativa con solidez intelectual, o bien plausibilidad discursiva con verdad científicamente sustentada (Chan & Colloton, 2024; Castañeda de León y otros., 2025).

Desde esta perspectiva, el primer dilema ético importante es el del desplazamiento del pensamiento crítico. La investigación académica exige problematizar, comparar, discriminar, interpretar, dudar y argumentar. No se trata solo de reunir información, sino de construir una posición razonada frente a un problema. La IAG, por su capacidad para ofrecer respuestas fluidas y aparentemente coherentes, puede facilitar tareas preliminares de exploración; pero también puede propiciar una delegación excesiva de operaciones que forman parte del trabajo intelectual mismo. Cuando el investigador acepta sin contraste suficiente una síntesis generada, adopta marcos interpretativos sugeridos sin examinarlos o traslada a la herramienta la responsabilidad de organizar el problema, se corre el riesgo de sustituir la deliberación por automatismos discursivos. En ese punto, la IAG deja de ser asistencia para convertirse en una mediación que debilita la autonomía cognitiva del investigador (Freire, 1970/2005; UNESCO, 2023).

Un segundo dilema remite a la autoridad epistémica. Tradicionalmente, en la investigación universitaria la legitimidad de una afirmación se vincula con la posibilidad de rastrear su fundamento: fuentes, datos, método, discusión teórica, criterios de análisis. La IAG altera parcialmente esta lógica, porque puede producir respuestas convincentes sin hacer visible de manera transparente la ruta específica por la cual llegó a ellas. Así, la fluidez del texto generado puede conferirle una apariencia de autoridad que no necesariamente corresponde a una fundamentación robusta. Esta situación es particularmente delicada en la investigación, ya que la forma lingüística puede enmascarar vacíos argumentativos, errores de interpretación o referencias inexistentes. El riesgo no radica solo en que el sistema “se equivoque”, sino en que el investigador deje de interrogar críticamente la base de esa aparente certeza. La ética de la investigación exige, por ello, mantener una relación de sospecha metódica frente a las salidas generadas y reafirmar que ninguna respuesta automatizada exime de la obligación de verificar, contextualizar y juzgar (Resnik, 2020; UNESCO, 2023).

Un tercer problema es la opacidad metodológica. En investigación, la integridad científica se sostiene en buena medida sobre la posibilidad de explicitar procedimientos, decisiones y límites. Cuando la IAG se incorpora al proceso sin declaración clara de uso, sin registro de qué tareas realizó o sin delimitación de sus alcances, se introduce una zona opaca que afecta la trazabilidad del trabajo académico. Esto es especialmente importante cuando la herramienta participa en actividades como el resumen de literatura, la codificación preliminar de información cualitativa, la reorganización de datos o la redacción de apartados enteros. No se trata de prohibir tales usos de manera indiscriminada, sino de reconocer que, en ausencia de transparencia, se compromete la posibilidad de evaluar con rigor la autoría, la metodología y la confiabilidad de los resultados. De ahí que distintos marcos de orientación insistan en la necesidad de declarar de forma explícita cuándo y cómo se utilizaron herramientas generativas en el proceso de investigación (COPE, 2023; ICMJE, 2024).

Un cuarto dilema ético aparece en relación con los sesgos algorítmicos y las asimetrías epistémicas. La IAG aprende a partir de grandes volúmenes de datos y reproduce patrones lingüísticos, conceptuales y culturales presentes en esos corpus. Esto implica que sus respuestas no solo pueden contener errores, sino también reflejar jerarquías previas en la circulación global del conocimiento. En el contexto latinoamericano, esta cuestión es especialmente relevante, pues muchos sistemas de IAG se alimentan mayoritariamente de materiales producidos en inglés o desde marcos teóricos dominantes del Norte global. El resultado puede ser una reproducción automática de perspectivas hegemónicas que desplazan autores, conceptos y experiencias históricas situadas en América Latina o en otros espacios subalternizados. Aquí la crítica decolonial resulta indispensable: no basta con preguntar si una respuesta es formalmente correcta, sino también qué epistemologías privilegia, qué voces deja fuera y qué relaciones de poder reinscribe en la producción académica (Quijano, 2000; Walsh, 2007; Dussel, 1998).

A esta problemática se añade el riesgo de la homogeneización del discurso académico. La IAG tiende a producir textos fluidos, ordenados y estilísticamente aceptables, pero con frecuencia cercanos a formas de escritura promedio, previsibles o excesivamente estandarizadas. Si este rasgo se normaliza en la investigación universitaria, puede debilitar la diversidad argumentativa, aplanar estilos de pensamiento y favorecer una cultura de la corrección formal desvinculada de la originalidad teórica. El problema no es solo estético. La escritura académica es parte del trabajo intelectual mismo: en ella se prueba la capacidad de articular conceptos, justificar inferencias, sostener una voz propia y hacer visible el recorrido del pensamiento. Cuando esa escritura se delega de manera excesiva a sistemas generativos, el investigador corre el riesgo de presentar un texto formalmente pulido, pero débil en apropiación conceptual y en elaboración crítica. En este punto, la mediación sociotécnica incide no solo en la forma de escribir, sino en la forma de pensar y de presentarse como sujeto de conocimiento (Nussbaum, 2010; Chan & Colloton, 2024).

Un quinto dilema concierne al uso de datos sensibles y a la confidencialidad. En investigación, especialmente en ciencias sociales, educativas, administrativas o de salud, no es extraño trabajar con entrevistas, observaciones, expedientes, documentos internos, materiales institucionales o bases de datos que contienen información sensible. Introducir estos materiales en plataformas generativas sin consentimiento, anonimización adecuada o resguardo institucional puede vulnerar derechos, comprometer la privacidad de personas y colectivos, y exponer información que no debería circular fuera del marco de investigación autorizado. Aquí la ética del uso de IAG no puede limitarse a la calidad del texto producido; debe incluir una valoración estricta sobre protección de datos, confidencialidad, consentimiento y responsabilidad institucional en el tratamiento de la información (UNESCO, 2021; European Union, 2024).

Existe además un dilema más profundo, vinculado con la racionalidad que orienta la investigación. La IAG encaja con especial facilidad en una cultura académica

atravesada por métricas, urgencias productivistas y presiones de visibilidad. En contextos donde se espera publicar más, escribir más rápido y responder con inmediatez, la herramienta puede ser celebrada como solución eficiente. Sin embargo, este mismo encuadre puede desfigurar el sentido de la investigación universitaria, reduciéndola a generación acelerada de productos y debilitando su dimensión reflexiva, pública y formativa. Desde una perspectiva ética, la pregunta relevante no es solo qué puede hacer la IAG por la investigación, sino qué modelo de universidad e investigación fortalece su uso predominante: uno centrado en la prisa, la apariencia de competencia y la eficiencia cuantificable, o uno comprometido con la comprensión rigurosa, la crítica y la responsabilidad social del conocimiento (Freire, 1970/2005; Nussbaum, 2010).

Todo lo anterior permite afirmar que la IAG, entendida como mediación sociotécnica, no debe ser evaluada únicamente por su funcionalidad, sino por los efectos que produce en la estructura misma del trabajo investigativo. Su incorporación puede ser éticamente defendible cuando se utiliza para apoyar tareas delimitadas, bajo supervisión humana, con transparencia, con verificación rigurosa y sin desplazar la responsabilidad autoral e intelectual del investigador. Por el contrario, su uso se vuelve éticamente problemático cuando opera como sustituto del juicio, cuando introduce opacidad en la metodología, cuando amplifica sesgos no examinados, cuando compromete datos sensibles o cuando refuerza una lógica de productividad que vacía de contenido crítico a la investigación. En otras palabras, el problema ético no reside en la mera existencia de la herramienta, sino en el tipo de relación que la universidad y los investigadores establecen con ella.

Desde esta base, la discusión debe avanzar hacia una cuestión decisiva: si la IAG transforma los modos de investigar, entonces la universidad necesita formar investigadores capaces de dialogar críticamente con estas tecnologías sin quedar subordinados a ellas. Esto lleva al siguiente apartado: la necesidad de una formación ética, crítica y digital del investigador como condición para el uso responsable de la IAG en la Educación Superior latinoamericana.

Formación ética, crítica y digital del investigador

La incorporación de la Inteligencia Artificial Generativa (IAG) en los procesos de investigación en la Educación Superior ha puesto en evidencia que el problema no se resuelve exclusivamente mediante regulaciones externas o lineamientos institucionales. Aun cuando estos sean necesarios, su eficacia depende en gran medida de la formación de quienes investigan. Por ello, una de las condiciones centrales para el uso responsable de la IAG consiste en fortalecer una formación ética, crítica y digital del investigador, entendida no como un complemento instrumental, sino como una dimensión constitutiva de la práctica académica contemporánea. Investigar en un entorno tecnológicamente mediado exige hoy no solo dominar métodos y marcos teóricos, sino también desarrollar capacidades para evaluar herramientas digitales, reconocer sus límites, identificar sus sesgos, transparentar su uso y sostener el juicio intelectual frente a la automatización de tareas complejas (UNESCO, 2023; Chan & Colloton, 2024).

Esta exigencia formativa puede comprenderse a partir de una idea básica: la tecnología no sustituye la responsabilidad del sujeto que investiga. Por el contrario, cuanto mayor es la capacidad técnica de una herramienta para asistir en la búsqueda, síntesis, redacción o análisis de información, mayor debe ser la competencia ética y crítica de quien la utiliza. De lo contrario, la promesa de apoyo se convierte en riesgo de dependencia cognitiva. La IAG puede facilitar operaciones preliminares o acelerar ciertos procesos, pero no elimina la necesidad de discernir entre lo relevante y lo accesorio, entre una fuente sólida y una dudosa, entre una interpretación plausible y una conclusión justificada. En este sentido, la formación del investigador debe preservar la centralidad del juicio humano como instancia última de validación y responsabilidad, evitando que la fluidez de las respuestas generadas se confunda con rigor científico o con comprensión profunda del problema investigado (Resnik, 2020; UNESCO, 2023).

Desde una perspectiva pedagógica, esta formación remite a la necesidad de sostener la autonomía intelectual como finalidad universitaria. Paulo Freire insistió en que educar implica formar sujetos capaces de leer críticamente el mundo y no meros receptores de contenidos. Trasladado al presente debate, ello significa que la interacción con herramientas generativas no puede reducirse a aprender comandos eficaces o a perfeccionar la formulación de instrucciones, sino que debe orientarse a fortalecer capacidades de interrogación, problematización y crítica. Un investigador éticamente formado no es aquel que solo sabe obtener respuestas rápidas, sino aquel que puede examinar por qué una respuesta debe ponerse en duda, qué supuestos contiene, qué ausencias revela y qué efectos puede tener su aceptación acrítica. La alfabetización digital, en este punto, solo es suficiente cuando está articulada con una pedagogía de la pregunta y no con una simple cultura de la respuesta inmediata (Freire, 1970/2005).

De manera convergente, la formación ética del investigador exige recuperar la dimensión del carácter intelectual. Desde la tradición aristotélica, la práctica del conocimiento no se agota en la posesión de técnicas, sino que requiere virtudes intelectuales y morales como prudencia, honestidad, justicia y templanza. En la era de la IAG, esta perspectiva conserva plena vigencia. La prudencia resulta indispensable para distinguir cuándo una herramienta puede ser usada como apoyo legítimo y cuándo su intervención desborda el ámbito aceptable de asistencia; la honestidad intelectual es necesaria para reconocer con transparencia los alcances de su uso; la justicia obliga a interrogar sesgos y exclusiones; y la templanza previene el deslumbramiento tecnocrático que lleva a sacrificar rigor por rapidez. Así, la formación del investigador no debe limitarse a competencias técnicas, sino incluir una disposición ética estable que le permita decidir responsablemente en escenarios de creciente complejidad tecnológica (Aristóteles, 2009; Cortina, 2013).

Esta exigencia se vuelve especialmente importante en el ámbito latinoamericano. Las instituciones de Educación Superior de la región se encuentran atravesadas por desigualdades en infraestructura, conectividad, acceso a bases de datos,

formación especializada y disponibilidad de herramientas tecnológicas. En ese contexto, hablar de formación digital no puede significar únicamente enseñar a usar plataformas de IAG, sino también problematizar las condiciones materiales y epistémicas de su adopción. La alfabetización en IAG requiere comprender que estas herramientas no operan en el vacío: responden a modelos entrenados con corpus desiguales, expresan sesgos de circulación global del conocimiento y pueden favorecer la subordinación de tradiciones teóricas periféricas si no se cuenta con capacidad crítica para identificarlas. Por ello, la formación del investigador en América Latina debe incorporar una sensibilidad decolonial que permita examinar qué voces quedan desplazadas, qué categorías dominan en las respuestas automatizadas y cómo evitar que la automatización del conocimiento refuerce jerarquías epistémicas históricas (Quijano, 2000; Walsh, 2007; Dussel, 1998).

A partir de ello, conviene distinguir tres dimensiones complementarias de esta formación. La primera es la formación ética, que implica desarrollar criterio para reconocer deberes de honestidad, autoría, confidencialidad, responsabilidad y cuidado en la investigación. La segunda es la formación crítica, que supone la capacidad de analizar supuestos, contrastar información, resistir respuestas aparentemente concluyentes y sostener la pregunta como núcleo del trabajo científico. La tercera es la formación digital, entendida como el conjunto de conocimientos y habilidades necesarias para comprender cómo funcionan las herramientas tecnológicas, qué pueden y qué no pueden hacer, qué riesgos implican y cómo incorporarlas de manera reflexiva a prácticas académicas concretas. Separadas, estas dimensiones resultan insuficientes; articuladas, permiten formar investigadores capaces de usar la IAG sin abdicar de su autonomía intelectual ni de su responsabilidad académica (Chan, 2023; UNESCO, 2023).

La noción de alfabetización en IA propuesta por Chan y desarrollada en el campo de la Educación Superior resulta particularmente útil aquí. No se trata solo de familiarizar a estudiantes y docentes con aplicaciones tecnológicas, sino de

comprender principios básicos de funcionamiento, identificar alcances y limitaciones, y desarrollar conciencia sobre implicaciones éticas, sociales y educativas del uso de la IA. Esta perspectiva es especialmente pertinente para la investigación, donde los errores no suelen aparecer como fallas visibles inmediatas, sino como distorsiones acumuladas: referencias inexistentes, interpretaciones superficiales, dependencia excesiva de formulaciones estándar, sesgos no advertidos o pérdida de apropiación conceptual del texto. La alfabetización en IAG, entonces, debe incluir una dimensión metodológica y epistemológica: no basta con saber usar la herramienta; es necesario saber interrogarla y, sobre todo, saber cuándo no utilizarla (Chan & Colloton, 2024; Chan, 2023).

En el mismo sentido, la guía de la UNESCO sobre IA generativa en educación e investigación insiste en que el desarrollo de capacidades humanas debe preceder y acompañar la adopción tecnológica. Entre esas capacidades se encuentran la verificación de información, la transparencia en el uso, la protección de datos, la comprensión de límites técnicos y la capacidad de asumir supervisión humana significativa. Lo relevante de este planteamiento es que desplaza el foco desde la fascinación por la herramienta hacia la formación del sujeto que interactúa con ella. En otras palabras, la cuestión no es solamente qué tan avanzada sea una IAG, sino qué tan preparado está el investigador para usarla sin convertirla en sustituto de su pensamiento ni en coartada para prácticas académicamente débiles o éticamente opacas (UNESCO, 2023).

Esta formación también debe incluir una reflexión explícita sobre los usos éticos y no éticos de la IAG en investigación. Su uso puede considerarse éticamente defendible cuando sirve, por ejemplo, para explorar palabras clave, organizar notas de lectura, mejorar claridad expositiva, sugerir formas preliminares de síntesis o ayudar a revisar estilo, siempre bajo supervisión humana, con contraste de fuentes y con transparencia sobre sus alcances. En cambio, se vuelve éticamente problemático cuando se usa para sustituir la revisión de literatura, inventar referencias, redactar resultados no verificados, procesar información confidencial

sin resguardo adecuado o presentar como elaboración propia un contenido cuya apropiación intelectual es mínima. La formación del investigador debe trabajar precisamente sobre estas distinciones, no a partir de prohibiciones abstractas, sino mediante ejercicios concretos de análisis, evaluación de casos y toma razonada de decisiones (COPE, 2023; ICMJE, 2024; Castañeda de León y otros., 2025).

Desde el punto de vista institucional, lo anterior supone que la formación ética, crítica y digital no puede quedar librada al esfuerzo individual. La universidad debe crear condiciones para que esta formación sea posible: seminarios, lineamientos, talleres de alfabetización en IA, discusión metodológica, acompañamiento tutorial, revisión colegiada y construcción de criterios compartidos para el uso académico de la IAG. Esto es particularmente importante en posgrados y cuerpos académicos, donde la presión por publicar, cumplir metas y producir resultados visibles puede favorecer usos apresurados de estas herramientas. Sin una cultura institucional que valore el juicio crítico y la integridad científica por encima de la mera productividad, la IAG tenderá a insertarse en la investigación como acelerador de prácticas ya frágiles, en lugar de convertirse en apoyo para una actividad académica más reflexiva y rigurosa (Chan, 2023; ALLEA, 2023).

En este punto resulta útil recuperar a Martha Nussbaum, quien advierte sobre el peligro de subordinar la educación superior a criterios exclusivamente utilitarios. Si la formación del investigador se redefine solo en función de eficiencia, competencia tecnológica y rapidez en la producción de resultados, se corre el riesgo de empobrecer justamente aquellas capacidades que sostienen la vida académica: argumentar con profundidad, leer de forma atenta, imaginar perspectivas diversas, reconocer la vulnerabilidad del otro y deliberar con responsabilidad. La IAG no debe volverse criterio de reducción del esfuerzo reflexivo, sino ocasión para reafirmar el valor del pensamiento complejo frente a la tentación de la automatización total (Nussbaum, 2010).

Por otra parte, una perspectiva decolonial aporta aquí un elemento fundamental: la formación crítica del investigador debe incluir la capacidad de reconocer cómo los sistemas generativos pueden reproducir colonialidades del saber. En muchos casos, las respuestas de la IAG privilegian autores consagrados en circuitos dominantes, categorías hegemónicas de análisis y formas estandarizadas de escritura académica, invisibilizando tradiciones intelectuales latinoamericanas, indígenas, populares o críticas. Formar investigadores críticamente alfabetizados en IAG implica, por tanto, enseñarles a detectar estas ausencias y a intervenir deliberadamente para recuperar marcos situados de comprensión. La ética de la investigación en América Latina no puede quedar reducida al uso técnicamente correcto de la herramienta; debe incluir una vigilancia sobre las jerarquías epistémicas que el propio uso de la IAG puede reactivar o consolidar (Quijano, 2000; Walsh, 2007; Dussel, 1998).

En suma, la formación ética, crítica y digital del investigador constituye una condición indispensable para que la IAG pueda ser incorporada de manera legítima en la investigación en Educación Superior. Sin ella, la herramienta corre el riesgo de desplazar el juicio, homogeneizar la escritura, debilitar la autoría intelectual y profundizar sesgos epistémicos. Con ella, en cambio, puede convertirse en un recurso de apoyo sometido a criterios humanos de responsabilidad, verificación, justicia y sentido público del conocimiento. La universidad, en consecuencia, no debe limitarse a regular el uso de la IAG, sino asumir la tarea más profunda de formar investigadores capaces de interactuar con estas tecnologías sin quedar subordinados a ellas. Sobre esta base, el siguiente paso del capítulo consiste en examinar cómo esa formación debe traducirse en marcos de gobernanza universitaria y en criterios institucionales para el uso responsable de la IAG.

Gobernanza universitaria y criterios institucionales para el uso responsable de la IAG

La incorporación de la Inteligencia Artificial Generativa (IAG) en la investigación en Educación Superior no puede quedar reducida a decisiones individuales, improvisaciones metodológicas o respuestas aisladas de docentes e investigadores. Si, como se ha sostenido en las secciones previas, la IAG constituye una mediación sociotécnica que reconfigura tareas, tiempos, modos de acceso a la información, prácticas de escritura y criterios de validación del conocimiento, entonces su uso exige también una respuesta institucional. En este punto adquiere centralidad la noción de gobernanza universitaria, entendida como el conjunto de principios, normas, estructuras, procedimientos y acuerdos mediante los cuales una institución orienta, regula, acompaña y evalúa prácticas académicas vinculadas con tecnologías emergentes. Desde esta perspectiva, la ética del uso de la IAG no depende únicamente de la buena voluntad de quien investiga, sino de la capacidad de la universidad para crear marcos de orientación que articulen responsabilidad, transparencia, protección de derechos, formación y rendición de cuentas (UNESCO, 2021, 2023; Chan, 2023).

Hablar de gobernanza universitaria en materia de IAG implica reconocer, en primer término, que la tecnología no se inserta en un vacío normativo. Toda universidad posee ya culturas institucionales, reglas explícitas e implícitas, jerarquías, prácticas editoriales, rutinas de evaluación y modos de entender la autoría, la evidencia y la legitimidad del conocimiento. La llegada de la IAG tensiona ese entramado y revela, con particular fuerza, la insuficiencia de marcos pensados para un escenario anterior a la expansión de sistemas capaces de generar texto, sintetizar información, sugerir argumentos o asistir en el tratamiento preliminar de datos. Por ello, la respuesta universitaria no puede limitarse a la vigilancia del fraude ni a la emisión de prohibiciones genéricas. Requiere una reflexión más amplia sobre qué modelo de investigación desea sostener la institución y bajo qué condiciones la IAG puede contribuir a ese horizonte sin erosionar la integridad científica ni la misión

formativa de la universidad (Chan & Colloton, 2024; Castañeda de León y otros., 2025).

En el plano internacional, distintos referentes convergen en algunos principios básicos que resultan pertinentes para la construcción de políticas universitarias. La Recomendación de la UNESCO sobre la ética de la inteligencia artificial subraya la necesidad de transparencia, supervisión humana, justicia, no discriminación, seguridad, privacidad, responsabilidad y protección de la dignidad humana y del entorno (UNESCO, 2021). De modo convergente, la guía de la UNESCO para IA generativa en educación e investigación insiste en que la incorporación de estas tecnologías debe priorizar el fortalecimiento de capacidades humanas, la protección de datos, la claridad sobre usos permitidos y la supervisión significativa por parte de personas e instituciones (UNESCO, 2023). Por su parte, la OCDE ha reafirmado principios orientados a una IA confiable, centrada en valores democráticos, derechos humanos, robustez técnica, transparencia y accountability (OECD, 2024). En el ámbito regulatorio, el AI Act de la Unión Europea ha consolidado una perspectiva en la que la evaluación de riesgos, la documentación, la trazabilidad y la responsabilidad institucional ocupan un lugar determinante (European Union, 2024). Aunque estos marcos no pueden trasladarse mecánicamente a toda universidad latinoamericana, sí ofrecen criterios orientadores para pensar políticas situadas y responsables.

Desde el ámbito específico de la Educación Superior, Cecilia Ka Yuk Chan ha propuesto un AI Ecological Education Policy Framework que resulta especialmente útil para este capítulo, porque articula tres dimensiones: gobernanza, operación y pedagogía. La dimensión de gobernanza se refiere a la formulación de políticas, estructuras de decisión, criterios éticos y mecanismos de regulación; la dimensión operativa alude a infraestructura, soporte técnico, seguridad, uso de datos y viabilidad institucional; y la dimensión pedagógica vincula la IA con formación, evaluación y transformación de prácticas de enseñanza y aprendizaje (Chan, 2023). Aunque su propuesta no se dirige exclusivamente a la investigación, ofrece una

base muy valiosa para pensar la universidad como ecosistema y no solo como suma de usuarios individuales. En el caso de la investigación, este enfoque permite sostener que el uso responsable de IAG requiere una articulación entre lineamientos, recursos, formación y cultura académica, y no una simple serie de advertencias generales.

En América Latina, esta necesidad de gobernanza adquiere características propias. Las universidades de la región enfrentan asimetrías materiales y epistémicas significativas: desigual acceso a plataformas, restricciones presupuestarias, brechas en conectividad, escasa infraestructura para protección de datos, debilidades en formación digital avanzada y, en muchos casos, ausencia de políticas explícitas sobre el uso académico de la IAG. A ello se suma una dependencia tecnológica de sistemas diseñados y entrenados principalmente fuera de la región, lo cual plantea preguntas de soberanía tecnológica, subordinación epistémica y adecuación contextual. Por esta razón, las políticas universitarias latinoamericanas no pueden limitarse a importar modelos normativos del Norte global; necesitan traducir principios generales a condiciones institucionales específicas, atendiendo tanto a los desafíos éticos comunes como a las particularidades de la región en materia de desigualdad, colonialidad del saber y vulnerabilidad institucional (Quijano, 2000; Walsh, 2007; Castañeda de León y otros., 2025).

A partir de ello, una orientación de política universitaria para el uso responsable de la IAG en investigación debería considerar, al menos, cinco núcleos de acción institucional. El primero es la claridad normativa. La universidad debe definir de manera explícita qué usos de la IAG son aceptables, cuáles requieren declaración expresa, cuáles están restringidos y cuáles resultan incompatibles con la integridad científica. Esta claridad no debe adoptar la forma de una lista exhaustiva e inflexible, pues la tecnología cambia con rapidez, pero sí debe ofrecer criterios suficientemente precisos para orientar decisiones en casos concretos. Entre ellos resultan centrales la transparencia en el uso, la responsabilidad autoral humana, la

verificación de contenidos, la protección de datos sensibles y la prohibición de prácticas engañosas o de simulación intelectual (COPE, 2023; ICMJE, 2024).

El segundo núcleo es la declaración de uso y la trazabilidad metodológica. En investigación universitaria, el uso de la IAG debe poder ser identificado y evaluado. Esto no significa burocratizar innecesariamente el trabajo académico, pero sí reconocer que la opacidad en la intervención de estas herramientas afecta la posibilidad de valorar la autoría, la metodología y la confiabilidad del texto final. Las instituciones pueden establecer lineamientos para que los investigadores especifiquen si la IAG fue empleada en búsqueda exploratoria, síntesis preliminar, revisión de estilo, apoyo analítico o cualquier otra fase del proceso. Esta explicitación no pretende penalizar el uso, sino volverlo evaluable y éticamente visible.

El tercer núcleo es la protección de datos, privacidad y seguridad de la información. Muchas investigaciones universitarias trabajan con materiales que no pueden ser cargados indiscriminadamente en plataformas generativas: entrevistas, expedientes, bases institucionales, documentos internos, información médica, administrativa o personal. Una política responsable debe ofrecer lineamientos claros sobre anonimización, consentimiento, resguardo, almacenamiento y uso de herramientas autorizadas o seguras. En este punto, la gobernanza universitaria se vincula directamente con deberes jurídicos y éticos, ya que la exposición indebida de datos no solo compromete la integridad del proceso investigativo, sino también derechos de personas y colectivos (UNESCO, 2021; European Union, 2024).

El cuarto núcleo es la formación y el acompañamiento institucional. Como se ha sostenido en la sección anterior, ninguna política será efectiva si quienes investigan no cuentan con capacidades éticas, críticas y digitales para comprender y evaluar la IAG. Por ello, la gobernanza universitaria debe incluir dispositivos de formación continua: talleres, seminarios, guías metodológicas, espacios de deliberación, apoyo tutorial, criterios editoriales y actualización periódica de lineamientos. Esta

formación no debería dirigirse solo a estudiantes, sino también a docentes, tutores, comités de ética, editores de revistas institucionales y cuerpos colegiados, ya que todos ellos participan, en distintos momentos, en la producción y validación del conocimiento académico (Chan, 2023; UNESCO, 2023).

El quinto núcleo es la evaluación y revisión constante de la política institucional. Dada la velocidad de cambio de los sistemas de IAG, ninguna normativa puede considerarse cerrada o definitiva. La universidad debe asumir que sus lineamientos requerirán monitoreo, revisión y ajuste continuo. Esto implica observar cómo se están usando efectivamente estas herramientas, qué problemas emergen, qué vacíos normativos persisten y qué efectos no previstos aparecen en la práctica. La gobernanza, en este sentido, no es un acto único de regulación, sino un proceso reflexivo y adaptativo. Esta idea es especialmente importante para evitar respuestas reactivas, moralizantes o meramente punitivas, que suelen llegar tarde o resultar ineficaces frente a la complejidad del fenómeno.

Ahora bien, una política universitaria éticamente consistente no debe limitarse a la gestión del riesgo. También necesita afirmar un horizonte normativo positivo sobre el sentido de la investigación. Esto significa recordar que la universidad no incorpora IAG para reemplazar la vida intelectual, sino para apoyar, bajo ciertas condiciones, tareas que contribuyan a una investigación más rigurosa, más responsable y socialmente más relevante. Desde esta óptica, la gobernanza universitaria debe defender, al menos, cuatro criterios sustantivos: la centralidad del juicio humano, la integridad científica, la justicia epistémica y la responsabilidad social del conocimiento. La centralidad del juicio humano impide tratar la IAG como sustituto del pensar; la integridad científica preserva autoría, verificabilidad y honestidad metodológica; la justicia epistémica obliga a vigilar sesgos y exclusiones; y la responsabilidad social recuerda que la investigación universitaria tiene efectos públicos y no puede desvincularse de la vida colectiva.

En el contexto latinoamericano, este último punto tiene un peso particular. Una orientación de política universitaria sobre IAG no debería contentarse con regular procedimientos internos; también tendría que interrogar la función social de la universidad frente a tecnologías cuyo desarrollo y control suelen estar concentrados en grandes corporaciones y contextos geopolíticos asimétricos. La pregunta no es solo cómo usar bien la IAG dentro de la institución, sino también qué tipo de conocimiento quiere producir la universidad, a quién beneficia, qué dependencias reproduce y qué formas de autonomía intelectual y tecnológica está dispuesta a defender. En este punto, una perspectiva inspirada en Dussel permite recordar que toda política ética debe evaluarse por su capacidad de afirmar la vida, reducir la exclusión y resistir formas de dominación que convierten a sujetos y comunidades en simples medios para fines ajenos (Dussel, 1998).

En suma, la gobernanza universitaria constituye una condición indispensable para que el uso de la IAG en investigación pueda considerarse legítimo desde el punto de vista ético. Sin ella, las decisiones quedan dispersas, la responsabilidad se diluye y los riesgos se trasladan al ámbito individual. Con ella, en cambio, la universidad puede construir criterios compartidos, resguardar derechos, formar capacidades y orientar institucionalmente la incorporación de la tecnología. No se trata de clausurar la innovación, sino de subordinarla a un proyecto universitario capaz de preservar la integridad científica, la autonomía intelectual y el sentido público del conocimiento. Sobre esta base, el capítulo puede cerrarse con una reflexión final que articule los distintos niveles trabajados: fundamento filosófico, responsabilidad formativa, ética de la investigación, mediación sociotécnica, formación del investigador y gobernanza institucional.

Conclusiones

La expansión de la Inteligencia Artificial Generativa (IAG) en la Educación Superior ha abierto una discusión que no puede resolverse en términos binarios de aceptación o rechazo tecnológico. A lo largo del capítulo se ha sostenido que la

cuestión de fondo no radica en la mera presencia de herramientas generativas en la investigación, sino en la transformación que estas introducen en las condiciones mismas de producción del conocimiento. La IAG interviene en la búsqueda, organización, síntesis, redacción y tratamiento preliminar de información; por ello, su incorporación modifica ritmos de trabajo, redistribuye tareas intelectuales y tensiona nociones fundamentales como autoría, transparencia metodológica, verificabilidad, integridad científica y responsabilidad académica. En consecuencia, la discusión sobre su uso no es meramente técnica, sino filosófica, educativa e institucional.

Desde una perspectiva filosófica, el análisis permitió recuperar la ética como reflexión crítica sobre la acción humana y no como simple repertorio de reglas externas. Esta base resultó necesaria para mostrar que la investigación universitaria no puede juzgarse solo por su eficacia operativa, sino por la calidad moral e intelectual de las decisiones que la sostienen. La ética de las virtudes permitió destacar la centralidad de la prudencia, la honestidad intelectual y el carácter del investigador; la ética del deber subrayó la imposibilidad de justificar cualquier práctica en nombre de la utilidad o la eficiencia; y la ética de la responsabilidad recordó que la acción técnica debe evaluarse también por sus consecuencias sobre la vida, la justicia y el mundo compartido. Incorporar a Dussel y a perspectivas decoloniales amplió esta comprensión al insistir en que toda reflexión ética sobre conocimiento y tecnología debe preguntarse por sus efectos sobre quienes quedan subordinados, invisibilizados o excluidos por estructuras históricas de poder (Aristóteles, 2009; Cortina, 2013; Dussel, 1998; Jonas, 1995).

En el plano educativo, el capítulo defendió que la universidad no es solo un espacio de transmisión de saberes especializados, sino una institución social de formación ética, intelectual y política. Bajo esta mirada, la IAG no puede evaluarse únicamente por su capacidad para acelerar procesos o aumentar productividad, sino por su impacto en la autonomía cognitiva, el juicio crítico y la misión formativa de la Educación Superior. La universidad, especialmente en América Latina, enfrenta el

desafío de incorporar tecnologías emergentes sin renunciar a su responsabilidad de formar sujetos capaces de pensar, argumentar, contrastar y responder por lo que producen. Por ello, la ética de la IAG en investigación se encuentra inseparablemente unida a la pedagogía universitaria, a la defensa del pensamiento crítico y a la preservación de una cultura académica que no confunda rapidez con comprensión ni automatización con inteligencia (Freire, 2005; Nussbaum, 2010; UNESCO, 2023).

En relación con la ética de la investigación, se argumentó que la integridad científica constituye hoy uno de los criterios decisivos para valorar el uso de la IAG. La investigación universitaria exige honestidad, transparencia, trazabilidad, cuidado de los datos, reconocimiento de autorías y disposición a la verificación. La IAG no elimina estas exigencias; por el contrario, las vuelve más estrictas, porque puede generar textos convincentes, pero también referencias inexistentes, sesgos no advertidos, interpretaciones superficiales o formulaciones plausibles sin sustento verificable. De ahí que el uso ético de estas herramientas dependa de que permanezcan subordinadas al juicio humano y a prácticas explícitas de revisión, contraste y responsabilidad. En este punto, marcos recientes convergen en una idea central: la IA no puede asumir autoría ni responsabilidad moral; estas siguen recayendo en personas humanas e instituciones.

Asimismo, el capítulo sostuvo que la IAG debe entenderse como una mediación sociotécnica y no como una herramienta neutral. Esto significa que sus efectos dependen de diseños, datos de entrenamiento, filtros, marcos de uso y relaciones institucionales que expresan decisiones humanas previas. En América Latina, esta constatación obliga a considerar no solo riesgos de fraude o mala práctica individual, sino también problemas de dependencia tecnológica, colonialidad del saber, homogeneización discursiva y subordinación epistémica frente a corpus y categorías dominantes del Norte global. Desde esta perspectiva, la ética del uso de IAG en investigación no puede limitarse a la corrección formal de procedimientos,

sino que debe examinar qué conocimientos privilegia, qué voces desplaza y qué relaciones de poder reproduce.

Sobre esa base, se afirmó que la formación ética, crítica y digital del investigador es una condición indispensable para el uso responsable de la IAG. Ninguna política institucional será suficiente si quienes investigan no cuentan con capacidades para comprender cómo funciona la herramienta, reconocer sus límites, detectar sesgos, proteger información sensible, verificar resultados y decidir cuándo su uso resulta pertinente y cuándo constituye una delegación indebida del pensamiento. La alfabetización en IA, por tanto, no debe reducirse a habilidades operativas; ha de incluir formación metodológica, ética y epistemológica. La universidad tiene aquí una responsabilidad insustituible: no solo regular el uso de la IAG, sino formar investigadores capaces de interactuar con ella sin abdicar de su autonomía intelectual.

Finalmente, el capítulo mostró que la respuesta ética al uso de la IAG en investigación debe traducirse en gobernanza universitaria. La incorporación de estas tecnologías no puede quedar librada a criterios individuales dispersos ni a improvisaciones coyunturales. Se requieren marcos institucionales que articulen claridad normativa, declaración de uso, protección de datos, formación continua, evaluación periódica y criterios compartidos sobre integridad científica. No se trata de clausurar la innovación, sino de subordinarla a un proyecto universitario centrado en la responsabilidad, la justicia epistémica y el sentido público del conocimiento. Para la Educación Superior latinoamericana, esto implica además construir orientaciones propias, situadas, sensibles a las desigualdades materiales y epistémicas de la región, y capaces de evitar que la adopción de IAG refuerce dependencias tecnológicas o nuevas formas de exclusión.

En síntesis, la IAG no sustituye legítimamente el pensamiento crítico, la autoría intelectual ni la responsabilidad ética del investigador. Puede convertirse en un aliado valioso para el tratamiento de grandes volúmenes de información y el apoyo

a problemas complejos, pero solo cuando su uso permanece sometido a la deliberación humana, a la transparencia metodológica, a la integridad científica y a marcos institucionales de responsabilidad. Por ello, la orientación de política universitaria que aquí se propone no parte de una desconfianza abstracta frente a la tecnología, sino de una convicción más profunda: la investigación en Educación Superior solo conserva su legitimidad si la innovación permanece subordinada a la ética, al juicio crítico y a la defensa de la vida humana y no humana como horizonte último del conocimiento.

Referencias

- ALLEA. (2023). *The European code of conduct for research integrity* (Rev. ed.). ALLEA.
- Aristóteles. (2009). *Ética nicomáquea* (J. Pallí Bonet, Trad.). Gredos.
- Castañeda de León, L. M., Ramírez Molina, A. Y., Castillejos Reyes, J. M., & Ventura Miranda, M. T. (2025). *Uso y desarrollo ético de la inteligencia artificial en la universidad: Docencia e investigación*. Universidad Nacional Autónoma de México.
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20, Article 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Chan, C. K. Y., & Colloton, T. (2024). *Generative AI in higher education: The ChatGPT effect*. Routledge. <https://doi.org/10.4324/9781003459026>
- Committee on Publication Ethics. (2023, February 13). *Authorship and AI tools*. <https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools>
- Cortina, A. (2013). *¿Para qué sirve realmente la ética?* Paidós.
- Dussel, E. (1998). *Ética de la liberación en la edad de la globalización y de la exclusión*. Trotta.
- European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- Freire, P. (2005). *Pedagogía del oprimido* (2.^a ed.). Siglo XXI Editores.

- International Committee of Medical Journal Editors. (2026, January). *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals*. <https://www.icmje.org/recommendations/>
- Jonas, H. (1995). *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica* (J. M. Fernández Retenaga, Trad.). Herder.
- Nussbaum, M. C. (2010). *Sin fines de lucro: Por qué la democracia necesita de las humanidades*. Katz.
- Organisation for Economic Co-operation and Development. (2024). *Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449)*.
<https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>
- Quijano, A. (2000). Colonialidad del poder, eurocentrismo y América Latina. En E. Lander (Comp.), *La colonialidad del saber: Eurocentrismo y ciencias sociales. Perspectivas latinoamericanas* (pp. 201–246). CLACSO.
- Resnik, D. B. (2020, December 23). *What is ethics in research & why is it important?* National Institute of Environmental Health Sciences.
<https://www.niehs.nih.gov/research/resources/bioethics/whatis>
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics*, 12(1), 53–74.
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.
<https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.
<https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Walsh, C. (2007). Interculturalidad y colonialidad del poder. Un pensamiento y posicionamiento “otro” desde la diferencia colonial. En S. Castro-Gómez &

R. Grosfoguel (Comps.), *El giro decolonial: Reflexiones para una diversidad epistémica más allá del capitalismo global* (pp. 47–62). Siglo del Hombre Editores; Universidad Central, Instituto de Estudios Sociales Contemporáneos; Pontificia Universidad Javeriana, Instituto Pensar.

CAPÍTULO XVI

LA ILUSIÓN DE LA NEUTRALIDAD: INTELIGENCIA ARTIFICIAL, PODER Y SESGO EN LA CONFIGURACIÓN DE LA REALIDAD DIGITAL

The illusion of neutrality: artificial intelligence, power and bias in the configuration of digital reality

Victor Daniel Santos Hortelano

coduamer@gmail.com

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0009-0001-8042-5943>

RESUMEN

La inteligencia artificial (IA) ha sido presentada frecuentemente como una tecnología objetiva, eficiente e imparcial, capaz de optimizar procesos y mejorar la toma de decisiones. Sin embargo, esta percepción oculta su carácter profundamente condicionado por los intereses de quienes la diseñan, implementan y controlan. El presente trabajo analiza la IA como un dispositivo técnico inscrito en estructuras de poder, cuya operación no es neutral, sino orientada por lógicas corporativas, estatales e ideológicas. A partir de un enfoque teórico sustentado en la psicopolítica de Byung-Chul Han, y mediante el análisis de casos contemporáneos como el uso de IA en conflictos geopolíticos, campañas políticas y procesos de desinformación, se argumenta que estos sistemas no solo procesan información, sino que participan activamente en la construcción de la realidad social. En este sentido, se plantea que la regulación de la IA enfrenta límites estructurales, en tanto intenta normar tecnologías cuya función está intrínsecamente vinculada al ejercicio del poder.

Palabras clave: inteligencia artificial, poder, psicopolítica, sesgo algorítmico, regulación

Abstract

Artificial intelligence (AI) has often been presented as an objective, efficient and unbiased technology, capable of optimizing processes and improving decision-making. However, this perception hides its nature deeply conditioned by the interests of those who design, implement and control it. This paper analyzes AI as a technical device inscribed in power structures, whose operation is not neutral, but guided by corporate, state, and ideological logics. Based on a theoretical approach based on Byung-Chul Han's psychopolitics, and through the analysis of contemporary cases such as the use of AI in geopolitical conflicts, political campaigns and disinformation processes, it is argued that these systems not only process information, but also actively participate in the construction of social reality. In this sense, it is argued that the regulation of AI faces structural limits, as it tries to regulate technologies whose function is intrinsically linked to the exercise of power.

Keywords: artificial intelligence, power, psychopolitics, algorithmic bias, regulation

1. INTRODUCCIÓN

En los últimos años, la inteligencia artificial (IA) ha adquirido un papel central en la reorganización de múltiples esferas sociales, económicas y políticas. Su implementación en ámbitos como la seguridad, la comunicación, el mercado laboral y la toma de decisiones estratégicas ha sido acompañada por un discurso que la presenta como una tecnología objetiva, eficiente y, sobre todo, imparcial. Esta narrativa ha contribuido a consolidar la idea de que los sistemas algorítmicos operan bajo criterios estrictamente técnicos, alejados de intereses humanos o estructuras de poder.

No obstante, diversos estudios y acontecimientos recientes han comenzado a cuestionar esta supuesta neutralidad. Desde el uso de datos personales para influir en procesos electorales, como en el caso de Cambridge Analytica (Hu, 2020), hasta la implementación de sistemas de inteligencia artificial en contextos de conflicto

geopolítico (Carbajal, 2026), se evidencia que la IA no actúa en un vacío técnico, sino enmarcada en intereses específicos que orientan su funcionamiento.

En este contexto, el presente artículo parte de la premisa de que la inteligencia artificial no puede entenderse únicamente como una herramienta tecnológica, sino como un dispositivo inserto en dinámicas de poder. En lugar de operar de manera autónoma o neutral, estos sistemas reproducen, amplifican y, en algunos casos, sofistican los intereses de los actores que los desarrollan y controlan, ya sean corporaciones tecnológicas, Estados o grupos con capacidad de influencia.

Para abordar esta problemática, se recurre al concepto de psicopolítica propuesto por Byung-Chul Han, quien señala que las formas contemporáneas de poder no operan principalmente a través de la coerción, sino mediante la manipulación de la información, la percepción y la subjetividad (Han, 2021) . Desde esta perspectiva, la inteligencia artificial no solo organiza datos, sino que interviene en la configuración de la realidad social, condicionando lo que se percibe como verdadero, relevante o legítimo.

A partir de este marco, el objetivo del artículo es analizar la inteligencia artificial como un sistema no neutral, cuya operación está atravesada por intereses estructurales, así como examinar las implicaciones de esta condición para los procesos de regulación. Se plantea que la dificultad para regular la IA no radica únicamente en su complejidad técnica, sino en su naturaleza como herramienta de poder.

2. LA ILUSIÓN DE NEUTRALIDAD EN LA INTELIGENCIA ARTIFICIAL

Uno de los supuestos más extendidos en torno a la inteligencia artificial es su carácter neutral. Esta idea se sustenta en la percepción de que los algoritmos procesan información de manera objetiva, basándose en datos y criterios lógicos que eliminan la subjetividad humana. Sin embargo, esta concepción omite que tanto los datos como los modelos algorítmicos son producto de decisiones humanas, lo que introduce inevitablemente sesgos, intenciones y marcos interpretativos

específicos. Para comprender mejor esta falta de neutralidad, conviene recordar que los datos históricos con los que se entrenan los sistemas de IA suelen reflejar desigualdades y prejuicios preexistentes en la sociedad. Por ejemplo, si un algoritmo de selección de personal se entrena con currículos de empresas donde históricamente han predominado ciertos perfiles demográficos, el sistema tenderá a replicar esos patrones de exclusión. De igual forma, la elección de qué variables incluir —o excluir— en un modelo responde a criterios subjetivos de los diseñadores, como la disponibilidad de datos o los objetivos comerciales. Así, lejos de ser un mero espejo de la realidad, la IA actúa como un filtro que prioriza ciertas dimensiones de los fenómenos sobre otras.

En este sentido, la neutralidad de la inteligencia artificial puede entenderse más como una construcción discursiva que como una realidad técnica. Los sistemas de IA no solo dependen de los datos con los que son entrenados, sino también de los objetivos definidos por sus desarrolladores, los cuales responden a intereses económicos, políticos o institucionales. Como señala Han (2021), en el contexto del capitalismo digital, el procesamiento de información no es un ejercicio inocente, sino una forma de intervención en la psique y el comportamiento de los individuos. Esta intervención se manifiesta, por ejemplo, en los sistemas de recomendación de plataformas como YouTube, Netflix o TikTok, que no buscan simplemente reflejar los gustos del usuario, sino modelar y modificar sus preferencias para maximizar el tiempo de consumo. Del mismo modo, los algoritmos de búsqueda en internet pueden priorizar ciertos resultados sobre otros en función de acuerdos comerciales o criterios de relevancia definidos de manera arbitraria. Por tanto, el discurso de la neutralidad cumple una función legitimadora: al presentar a la IA como un mero instrumento técnico, se ocultan las relaciones de poder y los intereses particulares que guían su operación. Esta ocultación no es accidental, pues permite que las decisiones algorítmicas se perciban como inevitables o naturales, dificultando su cuestionamiento.

Casos recientes evidencian esta problemática. Por ejemplo, el uso de inteligencia artificial en contextos de conflicto ha mostrado que estos sistemas pueden participar en la producción y validación de narrativas específicas. En algunos escenarios, herramientas de IA han sido señaladas por reproducir o priorizar ciertas versiones de los hechos, contribuyendo a la construcción de marcos interpretativos que favorecen determinadas posiciones políticas o estratégicas. De manera similar, en el ámbito de la información digital, la proliferación de contenidos generados o filtrados por sistemas automatizados ha intensificado la circulación de desinformación, dificultando la distinción entre hechos verificables y narrativas construidas. Un ejemplo concreto se observa en el uso de bots y cuentas automatizadas en redes sociales durante procesos electorales o conflictos armados. Estos sistemas pueden amplificar artificialmente ciertos mensajes, creando la impresión de que una opinión es mayoritaria cuando en realidad ha sido orquestada. Asimismo, los algoritmos de moderación de contenido, al carecer de contexto cultural o político suficiente, han eliminado publicaciones legítimas mientras permiten la circulación de discursos de odio disfrazados. En el ámbito de los conflictos internacionales, se ha documentado cómo sistemas de IA utilizados para analizar imágenes satelitales o inteligencia de fuentes abiertas (OSINT) pueden llegar a conclusiones diferentes según los parámetros de entrenamiento elegidos, lo que demuestra que incluso herramientas aparentemente técnicas incorporan sesgos interpretativos.

El caso de Cambridge Analytica representa un antecedente significativo en este proceso. A través del uso de datos obtenidos de redes sociales, esta empresa desarrolló perfiles psicológicos de millones de usuarios, con el objetivo de influir en sus decisiones políticas mediante campañas altamente segmentadas (Hu, 2020). Este episodio evidenció que el uso de datos no se limita a la descripción de comportamientos, sino que puede orientarse activamente a su modificación. Más allá de Cambridge Analytica, este modelo de microsegmentación conductual se ha generalizado en la publicidad política y comercial actual. Partidos políticos, consultoras y gobiernos utilizan herramientas de IA para identificar votantes

indecisos, activar emociones específicas mediante mensajes personalizados y suprimir la participación de ciertos grupos a través de campañas de desánimo. Lo relevante es que estos sistemas no operan como observadores neutrales del electorado: al clasificar a los ciudadanos en perfiles como "temerosos", "esperanzados" o "desconfiados", ya están construyendo una interpretación particular de la realidad social, y esa interpretación orienta intervenciones concretas. Así, la supuesta neutralidad algorítmica se disuelve frente a la evidencia de que la IA es una herramienta de persuasión y control.

De manera paralela, el desarrollo de modelos de inteligencia artificial en distintos contextos nacionales ha puesto de manifiesto que estos sistemas también pueden reflejar agendas estatales. La omisión o reinterpretación de ciertos eventos históricos en plataformas digitales, así como la moderación selectiva de contenidos, sugieren que la IA puede funcionar como un mecanismo de control narrativo, contribuyendo a la configuración de versiones específicas de la realidad. Por ejemplo, en ciertos países se han diseñado sistemas de IA para monitorear el discurso público en línea y sugerir la eliminación de contenidos que contradigan la versión oficial de hechos históricos. En otros casos, los asistentes virtuales y motores de búsqueda nacionales ofrecen respuestas divergentes sobre un mismo acontecimiento dependiendo de la localización del usuario, lo que introduce una fragmentación geopolítica del conocimiento. Incluso en los grandes modelos de lenguaje (como ChatGPT o Gemini), se ha observado que las respuestas pueden variar según la región o incorporar sesgos culturales provenientes de los datos de entrenamiento mayoritariamente occidentales. Esto no es un simple defecto técnico, sino un reflejo de que la IA no existe en un vacío político: su desarrollo y despliegue responden a intereses estratégicos, económicos e ideológicos.

En conjunto, estos elementos permiten sostener que la inteligencia artificial no opera como un sistema neutral, sino como una tecnología situada, cuyo funcionamiento está condicionado por los intereses y estructuras de poder en las que se inserta. Esta condición no solo cuestiona la narrativa de objetividad que la

rodea, sino que también plantea interrogantes fundamentales sobre su impacto en la construcción de la verdad y en los procesos democráticos. Entre esos interrogantes destacan: ¿cómo podemos garantizar la rendición de cuentas cuando un sistema de IA toma decisiones que afectan derechos fundamentales? ¿Qué mecanismos de auditoría independiente son necesarios para evidenciar los sesgos ocultos en los algoritmos? ¿Y cómo formar a ciudadanos y funcionarios en una comprensión crítica de la tecnología, de modo que no se dejen engañar por la ilusión de neutralidad? Responder a estas preguntas exige, en primer lugar, abandonar la idea ingenua de que la IA es un mero espejo de la realidad. Por el contrario, hay que reconocer que cada dato recopilado, cada variable seleccionada y cada criterio de optimización incorpora una decisión cargada de valor. La neutralidad, entonces, no es una propiedad técnica de los sistemas, sino un discurso que oculta las relaciones de poder. Desmontar esa ilusión es el primer paso para construir una gobernanza de la inteligencia artificial más democrática, participativa y transparente.

3. INTELIGENCIA ARTIFICIAL Y PSICOPOLÍTICA: EL PODER QUE OPERA DESDE LA LIBERTAD

El análisis de la inteligencia artificial como un sistema no neutral requiere un marco teórico que permita comprender las formas contemporáneas de poder más allá de los modelos tradicionales de coerción. En este sentido, el concepto de psicopolítica desarrollado por Byung-Chul Han resulta particularmente pertinente, al describir un tipo de dominación que no se impone desde la prohibición o la violencia, sino desde la gestión de la información, la subjetividad y la percepción (Han, 2021). Para entender mejor esta distinción, conviene recordar que el poder disciplinario propio de los siglos XIX y XX, analizado por Michel Foucault, se ejercía a través de instituciones visibles como la prisión, el cuartel o la fábrica, donde el castigo y la vigilancia directa moldeaban los cuerpos y las conductas. En cambio, la psicopolítica descrita por Han opera en un registro diferente: no busca encerrar ni prohibir, sino seducir, persuadir y guiar desde el interior de la propia subjetividad.

Este cambio resulta fundamental para comprender por qué las herramientas de inteligencia artificial pueden ser tan efectivas sin necesidad de ejercer una coerción explícita.

A diferencia de las formas disciplinarias del poder, que operaban mediante la vigilancia y el castigo, la psicopolítica se caracteriza por su capacidad de actuar de manera invisible, integrándose en los propios procesos de decisión de los individuos. Según Han (2021), el poder contemporáneo no busca restringir la libertad, sino explotarla, orientando las elecciones de los sujetos sin que estos perciban dicha intervención como una imposición externa. Esta forma de poder resulta más eficiente precisamente porque se presenta como libertad. Un ejemplo cotidiano de este mecanismo se encuentra en las interfaces de las plataformas digitales: el usuario tiene la aparente libertad de elegir qué contenido ver, pero el algoritmo ya ha preseleccionado y jerarquizado las opciones a partir de sus datos de navegación. Así, la libertad de elección no desaparece, pero queda enmarcada dentro de un universo de posibilidades predefinido por intereses ajenos al individuo. La eficacia del poder psicopolítico radica precisamente en esta invisibilidad: cuanto menos se percibe como una restricción, más difícil resulta resistirse a él.

En este contexto, la inteligencia artificial se configura como una herramienta central de la psicopolítica. Los sistemas basados en Big Data permiten no solo recolectar grandes volúmenes de información, sino también analizarlos para identificar patrones de comportamiento, preferencias y tendencias. Este conocimiento no es neutral, sino que se convierte en un instrumento de intervención sobre la conducta individual y colectiva. Como señala Han (2021), el Big Data posibilita una forma de dominación que actúa a nivel prerreflexivo, anticipándose a las decisiones de los sujetos y condicionando su margen de acción. Esto significa que los sistemas de IA no esperan a que el individuo manifieste una preferencia para luego satisfacerla; por el contrario, modelan probabilísticamente lo que el individuo probablemente deseará en el futuro, y actúan en consecuencia. De esta manera, la intervención psicopolítica se produce antes incluso de que el sujeto tome conciencia de su propia

intención. El margen de acción real, entonces, se reduce a la confirmación o al ligero ajuste de las opciones ya filtradas por el algoritmo.

Desde esta perspectiva, la inteligencia artificial no se limita a procesar información, sino que participa activamente en la configuración de la realidad social. Al determinar qué contenidos se muestran, cuáles se priorizan y cómo se interpretan, estos sistemas influyen en la percepción de lo verdadero, lo relevante y lo legítimo. De este modo, la IA opera como un dispositivo de mediación que no solo organiza la información, sino que estructura las condiciones mismas bajo las cuales esta es comprendida. Por ejemplo, cuando un motor de búsqueda devuelve resultados ordenados según criterios algorítmicos de "relevancia", lo que está haciendo es definir tácitamente qué fuentes, autores o perspectivas merecen ser vistas primero. Esta definición, aunque se presenta como un cálculo objetivo de popularidad o autoridad, incorpora siempre decisiones de diseño y parámetros de optimización que reflejan intereses comerciales o ideológicos. En consecuencia, la realidad que percibe un usuario de estos sistemas no es la realidad en su totalidad, sino una versión filtrada y jerarquizada conforme a lógicas que escapan a su control consciente.

Un elemento clave de la psicopolítica es su capacidad para funcionar sin generar resistencia. A diferencia de las formas de poder visibles, que pueden ser cuestionadas o confrontadas, el poder psicopolítico se disuelve en la cotidianidad de las prácticas digitales. La interacción constante con plataformas, algoritmos y sistemas automatizados produce una forma de participación voluntaria en los procesos de control, donde los individuos contribuyen activamente a la generación de los datos que posteriormente serán utilizados para influir en su comportamiento (Han, 2021). Esta paradoja resulta central: cuanto más interactúa una persona con las redes sociales, los motores de búsqueda o los asistentes virtuales, más información proporciona sobre sus gustos, miedos, deseos y rutinas. Esa información, procesada por sistemas de IA, retroalimenta los modelos predictivos que determinan qué estímulos recibirá a continuación. El resultado es un bucle en

el que la libertad individual —expresada como clics, likes, búsquedas o pausas— se convierte en la materia prima para una forma de dominación suave, que no necesita la coerción porque ha logrado que los propios individuos ofrezcan voluntariamente los datos con los que serán gobernados. Como advierte Han, el poder psicopolítico no encierra, sino que seduce; no prohíbe, sino que recomienda; no castiga, sino que premia adhiriéndose a las preferencias detectadas.

Esta lógica resulta fundamental para comprender por qué la inteligencia artificial no puede considerarse un sistema imparcial. Su funcionamiento no solo está condicionado por los intereses de quienes la diseñan, sino que también se integra en dinámicas más amplias de producción de subjetividad. En lugar de imponer una visión de la realidad, la IA contribuye a configurarla de manera progresiva, a través de mecanismos que operan en el nivel de la información, la atención y la percepción. Así, no se trata de un aparato externo que distorsiona una realidad preexistente, sino de un dispositivo que participa en la construcción misma de aquello que los individuos experimentan como real. Esta dimensión constructiva de la IA es la que la convierte en una herramienta psicopolítica por excelencia: no necesita falsificar los hechos cuando puede alterar las condiciones de su aparición y reconocimiento.

En este sentido, la inteligencia artificial se inscribe en una transformación más amplia del poder en la era digital, donde la capacidad de influir en la forma en que los individuos piensan, sienten y deciden adquiere una centralidad sin precedentes. La aparente neutralidad de estos sistemas encubre, en realidad, su función como herramientas de intervención psicopolítica, lo que plantea desafíos significativos para su comprensión y regulación. Entre estos desafíos destaca la necesidad de desarrollar una alfabetización algorítmica que permita a los ciudadanos identificar cuándo están siendo objeto de intervenciones psicopolíticas, así como la urgencia de establecer mecanismos de auditoría independiente que evalúen no solo los resultados visibles de los sistemas de IA, sino también los presupuestos y objetivos que guían su diseño. Solo reconociendo que el poder puede operar desde la

libertad —aprovechando nuestras elecciones cotidianas en lugar de anularlas— será posible diseñar respuestas políticas, éticas y jurídicas a la altura del fenómeno. De lo contrario, la ilusión de neutralidad seguirá cumpliendo su función legitimadora, mientras la psicopolítica algorítmica moldea silenciosamente el mundo social.

4. INTELIGENCIA ARTIFICIAL EN CONTEXTOS DE CONFLICTO: ENTRE LA AUTOMATIZACIÓN Y LA DECISIÓN POLÍTICA

El uso de inteligencia artificial en contextos de conflicto representa uno de los escenarios más reveladores de su carácter no neutral. En estos entornos, donde las decisiones tienen implicaciones directas sobre la vida humana y la estabilidad geopolítica, la incorporación de sistemas automatizados pone en evidencia que la IA no opera únicamente como una herramienta técnica, sino como un elemento integrado en estrategias políticas y militares. Para comprender la magnitud de esta transformación, conviene recordar que los conflictos armados han sido históricamente campos de prueba y aceleración de innovaciones tecnológicas, desde la aviación en la Primera Guerra Mundial hasta la energía nuclear en la Segunda Guerra Mundial. Sin embargo, la inteligencia artificial introduce una novedad cualitativa: no solo potencia la capacidad destructiva o de inteligencia, sino que interviene en los propios procesos de decisión, recomendando o incluso ejecutando acciones con una autonomía creciente. Esta capacidad de agencia algorítmica es lo que distingue a la IA de tecnologías militares anteriores y lo que la convierte en un objeto de especial atención ética y política.

Un caso ilustrativo es el uso de inteligencia artificial por parte de instituciones militares para el análisis de información, la identificación de objetivos y la toma de decisiones operativas. De acuerdo con Carbajal (2026), el Pentágono habría utilizado modelos de IA desarrollados por empresas privadas en el contexto de acciones militares, lo que sugiere una estrecha relación entre el desarrollo tecnológico y los intereses estratégicos del Estado. Este tipo de colaboración plantea interrogantes sobre la autonomía de los sistemas de IA, así como sobre la responsabilidad en su uso. La participación de empresas privadas en el diseño y

entrenamiento de estos modelos introduce una capa adicional de complejidad, pues dichas entidades no están sujetas a los mismos marcos de rendición de cuentas que las instituciones estatales. Además, los algoritmos entrenados con datos procedentes de entornos de combate pueden incorporar sesgos operativos que luego resulten difíciles de detectar y corregir. Por ejemplo, un sistema de reconocimiento de objetivos entrenado con imágenes de una geografía o tipo de infraestructura específica podría tener un rendimiento deficiente —o peligrosamente erróneo— al ser desplegado en un contexto diferente. Esta dependencia de los datos de entrenamiento, lejos de ser un mero detalle técnico, se convierte en una cuestión de vida o muerte en el campo de batalla.

La incorporación de inteligencia artificial en escenarios de conflicto también introduce una dimensión temporal que complejiza aún más su regulación. Como señalan diversos análisis contemporáneos, la velocidad con la que estos sistemas pueden procesar información y generar recomendaciones supera ampliamente la capacidad humana de deliberación. Esto implica que las decisiones pueden ser tomadas en tiempos cada vez más reducidos, limitando los espacios de reflexión ética y política. Esta aceleración decisoria tiene consecuencias profundas: lo que antes requería horas o días de análisis por parte de equipos de inteligencia humana, ahora puede ser procesado en segundos por un algoritmo. El problema es que esa velocidad no viene acompañada necesariamente de una mayor precisión o de un juicio contextual adecuado. Un sistema de IA puede recomendar un ataque basándose en correlaciones estadísticas que un analista humano, con tiempo para examinar los matices, descartaría como insuficientes o engañosas. Sin embargo, la presión operativa y la confianza mal entendida en la objetividad algorítmica pueden llevar a que estas recomendaciones se ejecuten sin el debido escrutinio. Así, la promesa de eficiencia tecnológica choca con la exigencia de deliberación democrática y responsabilidad ética, generando una tensión que aún no ha sido resuelta.

En este contexto, la inteligencia artificial no sustituye completamente la toma de decisiones humanas, pero sí reconfigura sus condiciones. Al proporcionar análisis, predicciones o sugerencias, estos sistemas influyen en las decisiones finales, orientando las acciones de los actores involucrados. Esta mediación tecnológica introduce un nuevo nivel de complejidad, en el que la responsabilidad se distribuye entre humanos y sistemas automatizados. Surgen entonces preguntas difíciles de responder: si un comandante sigue la recomendación de un sistema de IA y esa decisión causa bajas civiles desproporcionadas, ¿quién es responsable? ¿El desarrollador del algoritmo, el operador humano o la institución militar que desplegó el sistema? Los marcos legales actuales, incluyendo el derecho internacional humanitario, fueron diseñados para un mundo donde las decisiones eran tomadas exclusivamente por agentes humanos. La irrupción de sistemas autónomos e híbridos ha creado un vacío normativo que los tratados existentes no logran cubrir. Por ello, diversas organizaciones internacionales han comenzado a debatir la necesidad de un nuevo instrumento legal que regule, o incluso prohíba, ciertos usos de la IA en conflictos armados, especialmente aquellos que implican sistemas de armas autónomas letales (LAWS, por sus siglas en inglés).

Por otro lado, el uso de inteligencia artificial en conflictos no se limita al ámbito militar. También se extiende a la esfera de la información, donde los sistemas automatizados pueden contribuir a la difusión de narrativas específicas, la amplificación de ciertos discursos o la deslegitimación de otros. En este sentido, la IA participa en la construcción de marcos interpretativos que influyen en la percepción pública de los conflictos, lo que refuerza su papel como herramienta de poder. Durante un conflicto armado, la opinión pública nacional e internacional es un campo de batalla tanto como el territorio físico. Los sistemas de IA pueden ser utilizados para generar y distribuir contenido propagandístico a escala, microsegmentar audiencias con mensajes adaptados a sus sesgos cognitivos, o detectar y neutralizar narrativas consideradas adversas. Herramientas de análisis de sentimiento basadas en IA permiten monitorear en tiempo real la reacción de la población a ciertos eventos, ajustando las estrategias comunicacionales de manera

ágil. Esta dimensión informacional de la guerra asistida por IA ha sido denominada por algunos analistas como "ciberguerra cognitiva", y sus efectos pueden ser tan devastadores como los de un ataque convencional, erosionando la confianza en las instituciones, polarizando a la sociedad y desestabilizando democracias sin necesidad de disparar un solo tiro.

Estos elementos permiten afirmar que la inteligencia artificial, lejos de ser un instrumento neutral en contextos de conflicto, se integra en dinámicas estratégicas que responden a intereses específicos. Su uso no solo transforma la manera en que se toman decisiones, sino también la forma en que se construyen y comunican los acontecimientos, lo que plantea desafíos significativos para su regulación y control. Entre estos desafíos, cabe mencionar la necesidad de establecer mecanismos de supervisión humana significativa (meaningful human control) sobre cualquier sistema de IA utilizado en decisiones que afecten vidas humanas, así como la exigencia de que los algoritmos militares sean auditables y transparentes. Asimismo, resulta urgente avanzar hacia acuerdos internacionales que limiten la autonomía decisoria de la IA en el campo de batalla, prohibiendo aquellos usos que vulneren principios fundamentales del derecho humanitario, como la distinción entre combatientes y civiles o la proporcionalidad de los ataques. Sin una regulación robusta y con mecanismos de verificación efectivos, el despliegue de inteligencia artificial en conflictos seguirá expandiéndose, acelerando la toma de decisiones y erosionando la responsabilidad humana, en una senda cuyas consecuencias últimas aún no podemos anticipar completamente.

5. INTELIGENCIA ARTIFICIAL, MANIPULACIÓN Y CONTROL DE LA INFORMACIÓN

La relación entre inteligencia artificial y control de la información constituye uno de los ámbitos donde su carácter no neutral se manifiesta con mayor claridad. Lejos de limitarse a organizar o facilitar el acceso a datos, los sistemas de IA participan activamente en la selección, jerarquización e interpretación de la información, influyendo en la manera en que los individuos comprenden la realidad. Para

dimensionar adecuadamente esta transformación, conviene recordar que los medios tradicionales (prensa, radio, televisión) también ejercían una función de filtro editorial, pero sus criterios de selección eran, al menos en principio, identificables y atribuibles a actores concretos (directores de periódicos, consejos de redacción, reguladores estatales). En cambio, los sistemas de IA introducen una opacidad estructural: las decisiones sobre qué información mostrar, en qué orden y con qué énfasis responden a modelos matemáticos complejos que ni siquiera sus propios desarrolladores pueden explicar completamente. Esta opacidad, lejos de ser un defecto secundario, constituye un rasgo central que potencia la capacidad de influencia de estos sistemas, pues los hace más difíciles de cuestionar y regular.

Uno de los antecedentes más significativos de esta problemática es el caso de Cambridge Analytica. A partir de la recopilación masiva de datos personales obtenidos a través de redes sociales, esta empresa desarrolló perfiles psicográficos de millones de usuarios con el objetivo de influir en su comportamiento político. Como señala Hu (2020), el valor de estos datos no radicaba únicamente en su volumen, sino en su capacidad para identificar patrones emocionales, preferencias y vulnerabilidades, lo que permitió diseñar estrategias de comunicación altamente segmentadas. Este caso evidenció que el uso de datos no solo permite describir a los individuos, sino intervenir directamente en sus decisiones. Lo ocurrido con Cambridge Analytica no fue un episodio aislado o una anomalía; más bien, representó el modelo de negocio y de intervención política que se generalizaría en los años siguientes. La combinación de Big Data, psicometría y mensajería personalizada demostró que era posible influir en el voto, en las actitudes y en los comportamientos sin necesidad de coerción o propaganda masiva, simplemente enviando a cada persona el mensaje que sus propios datos indicaban que sería más efectivo. Esta lección no se perdió: desde entonces, partidos políticos, consultoras y gobiernos de todo el mundo han desarrollado capacidades similares, perfeccionando las técnicas de microsegmentación conductual.

En este sentido, la inteligencia artificial amplifica una lógica preexistente: el uso estratégico de la información con fines de influencia. Sin embargo, lo hace a una escala y con una precisión sin precedentes. Los sistemas contemporáneos no solo analizan comportamientos pasados, sino que también generan contenidos, recomiendan información y priorizan narrativas, lo que les otorga un papel activo en la construcción del entorno informativo. Mientras que Cambridge Analytica se basaba fundamentalmente en la explotación de datos existentes, los modelos actuales de IA generativa (como los grandes modelos de lenguaje) pueden producir contenido original —textos, imágenes, vídeos, incluso conversaciones enteras— que imita perfectamente la apariencia de autenticidad humana. Esto significa que no solo se pueden seleccionar los mensajes adecuados para cada persona, sino que se pueden crear mensajes personalizados en masa, a un costo cercano a cero. La combinación de microsegmentación y generación automatizada de contenido constituye una de las herramientas de manipulación informativa más poderosas jamás desarrolladas, y sus efectos apenas comienzan a ser comprendidos.

Esta capacidad se vuelve particularmente relevante en contextos de desinformación. La proliferación de contenidos generados o filtrados por sistemas automatizados ha contribuido a difuminar las fronteras entre información verificada y narrativas construidas. Diversos análisis recientes han señalado que algunas plataformas basadas en inteligencia artificial han reproducido o amplificado versiones parciales o inexactas de acontecimientos, lo que influye en la percepción pública de los mismos (RTVE, 2026; BBC, s.f.). El fenómeno de los "deepfakes" (vídeos falsos hiperrealistas generados por IA) ha recibido una atención considerable, pero quizás más preocupante es la capacidad de la IA para producir desinformación de bajo costo a gran escala: noticias falsas, reseñas fraudulentas, testimonios inventados o documentos apócrifos que pueden circular viralmente antes de que cualquier verificador de hechos pueda desmentirlos. Peor aún, los propios sistemas de verificación de hechos basados en IA pueden ser engañados por otros sistemas de IA, generando una carrera armamentística informativa en la que la capacidad de generar contenido falso avanza al menos tan rápido como la

capacidad de detectarlo. Este escenario de incertidumbre generalizada erosiona la confianza en todas las fuentes de información, creando un caldo de cultivo para el cinismo y la desmovilización política.

Asimismo, el desarrollo de modelos de inteligencia artificial en distintos contextos nacionales ha puesto de manifiesto la relación entre estos sistemas y agendas políticas específicas. En algunos casos, se ha observado la omisión o reinterpretación de hechos históricos y acontecimientos contemporáneos en función de marcos narrativos alineados con intereses estatales. Este tipo de prácticas no solo afecta el acceso a la información, sino que también incide en la construcción de la memoria colectiva y en la legitimidad de las instituciones. Cuando un buscador o un asistente virtual financiado por un gobierno ofrece sistemáticamente una versión de los hechos que excluye críticas o matices, no está simplemente "priorizando resultados": está ejerciendo una forma de control narrativo comparable a la censura estatal de regímenes autoritarios, pero con la apariencia de neutralidad algorítmica. La diferencia crucial es que, mientras la censura tradicional es visible y puede ser denunciada como tal, la mediación algorítmica se presenta como un mero servicio técnico, lo que dificulta su identificación como mecanismo de control. Esta invisibilidad es lo que hace de la IA una herramienta particularmente eficaz para la gestión de la información en contextos donde la legitimidad democrática sigue siendo un requisito formal, aunque su contenido se vacíe progresivamente.

Un elemento particularmente relevante en este proceso es la capacidad de la inteligencia artificial para operar de manera invisible. A diferencia de los medios tradicionales, donde la intervención editorial puede ser identificada, los sistemas algorítmicos presentan sus resultados como productos de procesos automáticos y objetivos. Esta apariencia de neutralidad refuerza su capacidad de influencia, ya que reduce la percepción de que la información ha sido filtrada o dirigida. Un lector puede sospechar del sesgo de un periódico conocido, pero tiende a confiar en que el resultado de una búsqueda en Google o la respuesta de ChatGPT es

simplemente "lo que arroja el algoritmo", como si ese algoritmo fuera un espejo de la realidad en lugar de un artefacto construido por decisiones humanas con intereses concretos. Esta asimetría cognitiva entre la suspicacia hacia los medios humanos y la credulidad hacia los medios automatizados es explotada activamente por quienes diseñan y despliegan sistemas de IA con fines de influencia. Por ello, una de las tareas más urgentes de la educación en el siglo XXI es precisamente desmontar esta ilusión de neutralidad algorítmica, enseñando a los ciudadanos a preguntarse siempre: ¿quién diseñó este sistema, con qué datos fue entrenado, y qué intereses sirve su funcionamiento?

Desde la perspectiva de la psicopolítica, esta dinámica puede entenderse como una forma de intervención en la subjetividad. Como plantea Han (2021), el control contemporáneo no se ejerce principalmente a través de la imposición, sino mediante la gestión de la información y la orientación de la atención. En este marco, la inteligencia artificial actúa como un dispositivo que no solo transmite información, sino que configura los marcos a partir de los cuales esta es interpretada. No se trata simplemente de que la IA pueda mostrar u ocultar ciertos hechos; se trata de que puede moldear las categorías mismas con las que los individuos distinguen un hecho relevante de uno irrelevante, una fuente confiable de una sospechosa, un argumento válido de una falacia. Cuando los sistemas de IA determinan qué tipo de contenido mantiene la atención de un usuario, están entrenando a ese usuario en ciertos hábitos perceptivos y cognitivos: la brevedad sobre la profundidad, la novedad sobre la consistencia, la emocionalidad sobre la argumentación. Estos hábitos, una vez internalizados, permanecen incluso cuando el individuo se desconecta de la plataforma. En este sentido, la IA no solo influye en lo que la gente cree, sino en cómo la gente piensa.

En conjunto, estos elementos permiten sostener que la inteligencia artificial no solo participa en la circulación de información, sino en su producción y estructuración. Su capacidad para influir en la percepción de la realidad la convierte en una herramienta central en las dinámicas contemporáneas de poder, donde el control

ya no se limita a los recursos materiales, sino que se extiende al ámbito de la información y la subjetividad. Frente a este panorama, resulta insuficiente confiar en soluciones puramente técnicas, como algoritmos de detección de desinformación o sistemas de verificación automática. Estas herramientas, aunque útiles, son parte del mismo ecosistema que se busca regular y pueden generar una falsa sensación de seguridad. La respuesta adecuada requiere, más bien, una combinación de medidas: regulación estatal que exija transparencia algorítmica y auditorías independientes; desarrollo de alternativas tecnológicas de código abierto y gobernanza democrática; promoción de una ciudadanía alfabetizada críticamente en competencias digitales; y, sobre todo, un debate público amplio sobre qué tipo de ecosistema informativo queremos construir. La tecnología no es neutral, pero la política puede decidir qué usos de la tecnología se fomentan, cuáles se limitan y cuáles se prohíben. Recuperar la soberanía sobre la información en la era de la inteligencia artificial es uno de los desafíos políticos más apremiantes de nuestro tiempo.

6. INTELIGENCIA ARTIFICIAL Y GEOPOLÍTICA: DISPUTA POR EL CONTROL TECNOLÓGICO

El desarrollo y despliegue de la inteligencia artificial no ocurre en un vacío técnico, sino en un escenario de competencia geopolítica donde Estados y corporaciones buscan consolidar ventajas estratégicas. En este contexto, la IA se configura como un recurso clave de poder, comparable a otras tecnologías críticas en momentos históricos previos, pero con la particularidad de incidir directamente en la producción de información, la economía digital y la seguridad nacional. Para comprender la magnitud de esta disputa, conviene recordar que el control de tecnologías estratégicas ha sido históricamente un factor determinante en la configuración del orden internacional: la navegación oceánica en los siglos XV y XVI, la máquina de vapor durante la Revolución Industrial, la energía nuclear durante la Guerra Fría o las tecnologías de la información en las décadas recientes. Sin embargo, la inteligencia artificial presenta una diferencia crucial respecto de estas tecnologías

anteriores: no se limita a potenciar la capacidad productiva o destructiva de una nación, sino que incide directamente en la cognición, la percepción y la toma de decisiones de poblaciones enteras. Quien domine la IA no solo producirá más riqueza o mejores armas, sino que podrá moldear la subjetividad de millones de personas a escala planetaria, lo que eleva la competencia tecnológica a un nivel de intensidad y relevancia sin precedentes.

La creciente tensión en torno al control de tecnologías de IA se manifiesta, por ejemplo, en la regulación de inversiones, restricciones a adquisiciones y políticas industriales orientadas a proteger capacidades nacionales. Decisiones como el bloqueo de adquisiciones de empresas de IA por motivos de seguridad evidencian que el desarrollo tecnológico no se entiende únicamente como un proceso económico, sino como un asunto estratégico vinculado a la soberanía y la influencia internacional (El Economista, 2026). Estados Unidos, China y la Unión Europea han implementado en los últimos años medidas dirigidas a limitar la transferencia de tecnología sensible, restringir inversiones extranjeras en sectores estratégicos y promover el desarrollo de capacidades propias. China, por ejemplo, ha establecido planes nacionales para convertirse en líder mundial en inteligencia artificial para 2030, invirtiendo masivamente en investigación, infraestructura de datos y formación de talento. Estados Unidos, por su parte, ha respondido con restricciones a la exportación de semiconductores avanzados y software de IA, buscando mantener su ventaja competitiva. La Unión Europea, aunque con un enfoque más centrado en la regulación y los derechos fundamentales, también ha lanzado iniciativas para reducir su dependencia tecnológica y fomentar el desarrollo de IA de confianza. Esta fragmentación del espacio tecnológico global sugiere que estamos asistiendo al surgimiento de regímenes tecnológicos diferenciados, con estándares, infraestructuras y modelos de gobernanza cada vez más divergentes, lo que algunos analistas han denominado la "fractura digital" o el "splinternet".

En paralelo, las grandes corporaciones tecnológicas han adquirido un papel protagónico en esta disputa. Empresas dedicadas al desarrollo de modelos

avanzados de inteligencia artificial no solo compiten entre sí, sino que también establecen relaciones complejas con los Estados, que pueden oscilar entre la cooperación y la tensión. La colaboración entre instituciones gubernamentales y compañías privadas en el desarrollo o implementación de sistemas de IA — particularmente en ámbitos como la defensa— pone de manifiesto una convergencia de intereses que difumina las fronteras entre lo público y lo privado (Nexos, 2026). Esta convergencia no es nueva: durante la Guerra Fría, el complejo militar-industrial ya articulaba una estrecha relación entre el Pentágono, las universidades y las corporaciones de defensa. Sin embargo, lo característico del presente es que las corporaciones tecnológicas poseen no solo recursos económicos comparables a los de muchos Estados, sino también capacidades técnicas, acceso a datos masivos y una presencia global que las convierte en actores geopolíticos por derecho propio. Empresas como Google, Microsoft, Amazon, Meta, OpenAI y sus equivalentes chinos como Baidu, Alibaba o Tencent acumulan poder suficiente para influir en políticas públicas, condicionar regulaciones y definir estándares técnicos que afectan a miles de millones de usuarios. En este contexto, la distinción tradicional entre poder estatal y poder corporativo se vuelve cada vez más difícil de sostener, dando paso a una ecología de poder híbrida donde ambos tipos de actores cooperan y compiten simultáneamente.

Esta convergencia tiene implicaciones significativas. Por un lado, refuerza la capacidad de los Estados para acceder a tecnologías avanzadas sin necesidad de desarrollarlas completamente de manera autónoma. Por otro, otorga a las corporaciones un grado de influencia sobre decisiones estratégicas que tradicionalmente correspondían exclusivamente al ámbito estatal. En este sentido, la inteligencia artificial se convierte en un espacio de negociación y disputa donde distintos actores buscan imponer sus intereses. Surgen así nuevas formas de poder que operan a través de la propiedad de los datos, el control de la infraestructura computacional (por ejemplo, los centros de datos que albergan los servidores necesarios para entrenar grandes modelos) y la definición de protocolos y

estándares técnicos. Como ha ocurrido en otras áreas tecnológicas, quien establece los estándares en IA define las reglas del juego para todos los demás, lo que convierte la estandarización en un campo de batalla geopolítico adicional. Por ello, no es casual que las grandes potencias estén invirtiendo en sus propios modelos fundacionales (foundation models) y en infraestructuras de cómputo dedicadas, buscando reducir su dependencia de proveedores extranjeros y asegurar su autonomía estratégica en un dominio considerado crítico para la seguridad nacional y la competitividad económica.

Asimismo, el desarrollo de modelos de IA en contextos nacionales específicos ha evidenciado la existencia de enfoques diferenciados en torno a la gestión de la información. Algunos sistemas han sido diseñados o ajustados para alinearse con marcos normativos y narrativas particulares, lo que refleja la influencia de agendas políticas en su funcionamiento. Este fenómeno sugiere que la inteligencia artificial no solo es un producto tecnológico, sino también un instrumento que puede ser configurado para reproducir visiones del mundo específicas. Por ejemplo, los grandes modelos de lenguaje entrenados predominantemente con datos en inglés procedentes de fuentes occidentales incorporan sesgos culturales, éticos y epistemológicos propios de ese contexto, lo que puede generar fricciones cuando son desplegados en entornos socioculturales diferentes. A su vez, algunos países han desarrollado modelos de IA explícitamente diseñados para reflejar sus propias narrativas históricas, valores normativos y prioridades políticas, lo que introduce una dimensión de competencia cultural y civilizatoria en la disputa por la IA. Esta divergencia plantea interrogantes sobre la posibilidad de mantener un espacio informativo global compartido: si los ciudadanos de diferentes países acceden a asistentes de IA que ofrecen versiones incompatibles de los hechos, ¿sobre qué base será posible el debate internacional, la cooperación científica o incluso la resolución pacífica de conflictos? La geopolítica de la inteligencia artificial no solo decide quién lidera tecnológicamente, sino qué versión de la realidad se globaliza.

Desde una perspectiva más amplia, la competencia por el liderazgo en inteligencia artificial también responde a su potencial para redefinir las dinámicas económicas globales. La capacidad de desarrollar y controlar sistemas avanzados de IA implica ventajas en sectores clave como la automatización, la innovación tecnológica y el análisis de datos. En consecuencia, los países y empresas que logren posicionarse en este ámbito adquieren una ventaja significativa en la economía digital. Se estima que la inteligencia artificial podría contribuir con hasta 15,7 billones de dólares a la economía global para 2030, lo que equivale aproximadamente al PIB actual de China e India combinados. Quien capture la mayor parte de ese valor económico no solo incrementará su riqueza, sino que podrá reinvertirla en investigación y desarrollo, consolidando así su liderazgo en un ciclo de retroalimentación positiva. Por el contrario, los países que queden rezagados en esta carrera podrían enfrentar no solo pérdidas económicas, sino también una creciente dependencia tecnológica, erosión de su soberanía digital y vulnerabilidad ante decisiones adoptadas por actores externos. Esta dimensión económica de la geopolítica de la IA explica la intensidad de las políticas industriales desplegadas en los últimos años, con subsidios masivos, incentivos fiscales y medidas proteccionistas que recuerdan a las estrategias de desarrollo de industrias nacientes de los siglos XIX y XX.

En este escenario, la inteligencia artificial deja de ser únicamente una herramienta funcional y se convierte en un elemento estructural del poder contemporáneo. Su desarrollo y control están atravesados por intereses estratégicos que condicionan tanto su diseño como su implementación, lo que refuerza la idea de que estos sistemas no operan de manera neutral, sino en función de agendas específicas. Frente a este panorama, resulta ingenuo pensar que la gobernanza de la IA puede resolverse exclusivamente mediante códigos de conducta voluntarios o autorregulación empresarial. La disputa geopolítica por el control tecnológico exige respuestas a nivel interestatal: acuerdos multilaterales que establezcan límites a ciertos usos de la IA (particularmente en el ámbito militar), mecanismos de verificación y cumplimiento, así como espacios de cooperación científica y técnica que permitan compartir beneficios sin ceder soberanía. Organismos como Naciones

Unidas, la OCDE o el G7 han comenzado a abordar estas cuestiones, pero los acuerdos alcanzados hasta el momento son mayoritariamente declarativos y carecen de mecanismos de enforcement efectivos. La historia sugiere que las tecnologías transformadoras tienden a generar nuevas configuraciones de poder antes de ser efectivamente reguladas, y no hay razones para pensar que la inteligencia artificial será una excepción. Sin embargo, comprender su dimensión geopolítica es el primer paso para diseñar estrategias que permitan navegar esta disputa sin sacrificar los valores democráticos, los derechos humanos y la cooperación internacional que constituyen los pilares del orden global construido tras 1945.

7. LOS LÍMITES ESTRUCTURALES DE LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL

El reconocimiento de la inteligencia artificial como una tecnología atravesada por intereses políticos, económicos e ideológicos plantea desafíos significativos para su regulación. Si bien en los últimos años se han impulsado diversas iniciativas orientadas a establecer marcos normativos para el desarrollo y uso de la IA, estas enfrentan limitaciones estructurales que dificultan su efectividad. Para comprender cabalmente estas limitaciones, es necesario distinguir entre las dificultades habituales que enfrenta cualquier intento de regular una tecnología emergente y los obstáculos específicos que derivan de la propia naturaleza de la inteligencia artificial como dispositivo de poder. Las primeras pueden ser abordadas con ajustes técnicos e institucionales; los segundos, en cambio, cuestionan el supuesto mismo de que sea posible regular eficazmente un fenómeno que opera en gran medida en los intersticios invisibles de la vida cotidiana. Esta distinción es crucial, pues de ella depende el realismo de las expectativas depositadas en las iniciativas regulatorias actuales.

En primer lugar, la velocidad del desarrollo tecnológico supera con frecuencia la capacidad de los marcos regulatorios para adaptarse. La innovación en inteligencia artificial se caracteriza por su dinamismo, lo que implica que las normativas suelen

formularse de manera reactiva, una vez que los sistemas ya han sido implementados y sus efectos comienzan a manifestarse. Esta asimetría temporal limita el alcance de la regulación, que difícilmente puede anticipar los usos y consecuencias de tecnologías en constante evolución. El ciclo típico de elaboración normativa —que incluye identificación del problema, consultas públicas, redacción de propuestas, trámites legislativos y plazos de entrada en vigor— puede extenderse fácilmente por varios años. En ese mismo periodo, la tecnología puede haber experimentado dos o tres saltos generacionales, volviendo obsoletas muchas de las disposiciones recién aprobadas. Por ejemplo, cuando el Reglamento General de Protección de Datos (GDPR) de la Unión Europea entró en vigor en 2018, los modelos de lenguaje a gran escala como los que hoy conocemos apenas comenzaban a desarrollarse. Cuatro años después, la irrupción de sistemas como ChatGPT planteó preguntas sobre el tratamiento de datos personales en el entrenamiento de modelos que el GDPR no había previsto explícitamente. Esta brecha entre el tiempo tecnológico y el tiempo institucional no es accidental: mientras que la innovación responde a lógicas de mercado que premian la rapidez y la ventaja competitiva, la regulación responde a lógicas democráticas que requieren deliberación, transparencia y rendición de cuentas. Ambas temporalidades son legítimas, pero su desajuste estructural genera un déficit de gobernanza que ningún diseño normativo parece capaz de resolver completamente.

En segundo lugar, la naturaleza transnacional de la inteligencia artificial complica la aplicación de regulaciones locales. Los sistemas de IA son desarrollados, desplegados y utilizados en múltiples jurisdicciones, lo que genera tensiones entre marcos normativos nacionales y dinámicas globales. Esta situación permite que actores con capacidad tecnológica operen en espacios donde la regulación es más laxa o inexistente, lo que reduce la efectividad de los esfuerzos regulatorios. Una empresa con sede en un país con legislación poco estricta puede desarrollar un sistema de IA, entrenarlo con datos obtenidos globalmente, ofrecer sus servicios a través de Internet a usuarios ubicados en jurisdicciones con regulaciones más

exigentes y, sin embargo, no estar sujeta a estas últimas más que de manera marginal. Los intentos de aplicar extraterritorialmente ciertas normativas (como el GDPR europeo) han tenido un éxito relativo, pero dependen de la cooperación internacional y de la capacidad de los Estados para hacer cumplir sus decisiones más allá de sus fronteras. En un mundo donde los flujos de datos atraviesan continuamente los límites territoriales, la regulación estatal de la IA se asemeja a intentar pescar con redes de agujeros cada vez más grandes: siempre habrá fugas que permitan eludir el control. Esta situación favorece una carrera hacia el mínimo común denominador regulatorio, donde los países compiten por ofrecer las condiciones más laxas para atraer inversión tecnológica, generando una presión a la baja sobre los estándares de protección de derechos fundamentales.

Sin embargo, más allá de estas dificultades técnicas y jurídicas, existe un problema más profundo: la inteligencia artificial forma parte de estructuras de poder que condicionan su desarrollo y uso. En este sentido, la regulación no se enfrenta únicamente a una tecnología compleja, sino a intereses consolidados que pueden influir en los procesos normativos. La participación de corporaciones tecnológicas en la definición de estándares y políticas públicas introduce un elemento de conflicto de intereses que puede limitar el alcance de la regulación. No se trata de atribuir intencionalidad maliciosa a los actores corporativos, sino de reconocer que las empresas tienen un interés legítimo en maximizar sus beneficios y que, por tanto, tenderán a promover marcos regulatorios que no obstaculicen sus modelos de negocio. Este fenómeno, conocido en la literatura como "captura regulatoria", es particularmente agudo en el campo de la IA, donde el conocimiento técnico necesario para diseñar regulaciones efectivas está concentrado en las propias empresas que serán reguladas. Los legisladores y funcionarios públicos, con frecuencia carentes de formación técnica especializada, dependen de la información y el asesoramiento que les proporcionan los mismos actores que deberían supervisar. Esta asimetría cognitiva no invalida la regulación, pero ciertamente sesga su dirección: es mucho más probable que se regulen aspectos periféricos o que no afectan el núcleo del negocio tecnológico (como los requisitos

de transparencia sobre el origen de los datos) y mucho menos probable que se cuestionen aspectos estructurales (como la propiedad de los datos o la concentración del poder computacional). En este sentido, la regulación de la IA corre el riesgo de convertirse en un ejercicio de legitimación cosmética, que crea la impresión de control sin alterar significativamente las relaciones de poder subyacentes.

Desde la perspectiva de la psicopolítica, este problema adquiere una dimensión adicional. Como señala Han (2021), las formas contemporáneas de poder operan de manera difusa, integrándose en las prácticas cotidianas y en los procesos de subjetivación. En este contexto, la regulación de la inteligencia artificial no solo implica establecer límites técnicos o legales, sino intervenir en dinámicas de producción de sentido que son menos visibles y más difíciles de controlar. El derecho y la regulación están diseñados para abordar acciones y consecuencias observables, no para intervenir en la formación de la subjetividad. Cuando un sistema de IA recomienda contenidos que refuerzan la polarización política de un usuario, el daño no es inmediatamente cuantificable ni atribuible de manera clara. Sin embargo, la acumulación de estas microintervenciones a lo largo del tiempo puede transformar radicalmente el espectro de opiniones, actitudes y disposiciones emocionales de una población. La regulación tradicional carece de instrumentos para abordar este tipo de efectos emergentes, que resultan de la interacción compleja entre sistemas técnicos, hábitos individuales y dinámicas sociales. Por ello, el poder psicopolítico de la IA tiende a permanecer fuera del alcance de los mecanismos normativos existentes, no por negligencia de los reguladores, sino por la inadecuación estructural de las herramientas regulatorias disponibles para el tipo de fenómeno al que se enfrentan.

Esto plantea una pregunta fundamental: ¿es posible regular una tecnología cuya eficacia radica precisamente en su capacidad para operar de manera invisible? La inteligencia artificial no solo ejecuta funciones, sino que participa en la configuración de las condiciones bajo las cuales se toman decisiones, se construyen opiniones y

se percibe la realidad. Esta característica limita la capacidad de la regulación para abordar sus efectos más profundos. Lo que hace que la IA sea tan poderosa como herramienta de influencia es también lo que la hace tan difícil de regular: su intervención no es percibida como tal por quienes la sufren, y sus efectos se manifiestan no como eventos discretos sino como transformaciones graduales del entorno informativo y emocional. Un ciudadano puede notar que un anuncio político lo ha manipulado; difícilmente notará que el algoritmo de su feed ha estado moldeando silenciosamente sus preferencias durante meses. La regulación que exige etiquetar contenidos publicitarios o revelar la identidad del patrocinador de un mensaje político presupone que la manipulación es reconocible como tal. Pero cuando la manipulación se produce a nivel prerreflexivo, a través de la arquitectura misma de las plataformas y algoritmos, la exigencia de transparencia pierde gran parte de su eficacia. No porque los datos no sean técnicamente accesibles, sino porque el mecanismo de influencia es estructural y ambiental, no intencional y discreto. Desde esta perspectiva, la regulación de la IA se enfrenta a un problema análogo al de regular la contaminación atmosférica: los efectos son fácilmente perceptibles agregadamente, pero atribuirlos a fuentes específicas y diseñar intervenciones efectivas es técnicamente muy complejo, y los afectados tienen pocas posibilidades de identificar al responsable de su malestar.

En consecuencia, la regulación de la inteligencia artificial tiende a centrarse en aspectos visibles —como la protección de datos, la transparencia o la responsabilidad—, mientras que sus dimensiones más estructurales permanecen en gran medida fuera de su alcance. Esto no implica que la regulación sea irrelevante, sino que sus efectos son necesariamente parciales. Es importante no caer ni en el extremo de la impotencia regulatoria (nada se puede hacer) ni en el extremo del optimismo tecnocrático (basta con la ley adecuada). La evidencia disponible sugiere que ciertos tipos de regulación pueden ser eficaces para abordar problemas específicos y acotados: por ejemplo, la exigencia de consentimiento informado para la recopilación de datos personales, la obligación de realizar auditorías de sesgos en sistemas de contratación automatizada o el

establecimiento de mecanismos de supervisión humana en decisiones algorítmicas de alto riesgo. Sin embargo, sería un error confundir estos avances necesarios con una solución integral al problema del poder psicopolítico de la IA. La regulación puede poner límites, establecer sanciones y crear incentivos, pero no puede, por sí misma, transformar la estructura de relaciones de poder que hace que la IA sea una herramienta tan eficaz para la producción de subjetividad. Para abordar esa dimensión estructural se requieren intervenciones adicionales, que van más allá del derecho y la política regulatoria: educación crítica, fomento de alternativas tecnológicas descentralizadas, fortalecimiento del espacio público no mediado algorítmicamente y, quizás lo más importante, una ciudadanía activa y movilizadora que exija cuentas tanto a los gobiernos como a las corporaciones.

En última instancia, los límites de la regulación de la inteligencia artificial no pueden entenderse únicamente como fallas en el diseño normativo, sino como el reflejo de una tensión más amplia entre el poder tecnológico y la capacidad de control institucional. La inteligencia artificial, en tanto dispositivo de poder, no solo plantea la necesidad de regulación, sino que también cuestiona las posibilidades mismas de ejercerla de manera efectiva. Esta paradoja no debe conducir al quietismo o a la resignación, sino a una redefinición del problema: no se trata de preguntar "¿cómo podemos regular la IA?", como si esta fuera un objeto externo que los legisladores pueden moldear a voluntad, sino de preguntar "¿qué formas de organización política, económica y social pueden contener el poder de la IA sin reproducir sus lógicas?". Responder a esta pregunta exige ir más allá de los marcos regulatorios tradicionales y aventurarse en el terreno de la democracia económica, la soberanía tecnológica y el diseño institucional experimental. Quizás el límite más profundo de la regulación de la inteligencia artificial sea que esta, tal como está concebida actualmente, presupone un Estado fuerte, autónomo y capacitado para imponer sus decisiones a actores poderosos. Pero en la era de la globalización digital, este presupuesto es cada vez menos realista. Por ello, repensar la regulación de la IA implica, en última instancia, repensar el lugar del Estado y la

democracia en un mundo donde el poder ya no reside exclusiva ni principalmente en instituciones territoriales.

8. DISCUSIÓN: ENTRE LA TECNOLOGÍA Y EL PODER

El análisis desarrollado a lo largo de este trabajo permite replantear la manera en que se concibe la inteligencia artificial en el debate contemporáneo. Más que una herramienta técnica orientada a la optimización de procesos, la IA emerge como un dispositivo profundamente imbricado en estructuras de poder que condicionan tanto su funcionamiento como sus efectos sociales.

La idea de una inteligencia artificial neutral se sostiene, en gran medida, sobre una concepción tecnocrática que separa la tecnología de los contextos en los que se produce. Sin embargo, como se ha argumentado, los sistemas de IA no solo procesan datos, sino que los seleccionan, interpretan y priorizan en función de criterios definidos por actores específicos. En este sentido, la tecnología no elimina la subjetividad, sino que la reconfigura bajo nuevas formas.

Los casos analizados muestran que la inteligencia artificial puede operar como un mecanismo de influencia tanto en contextos políticos como en escenarios de conflicto. Desde la segmentación de audiencias en campañas electorales hasta la construcción de narrativas en entornos geopolíticos, la IA participa en la producción de marcos interpretativos que inciden en la percepción de la realidad. Este proceso no se presenta como una imposición explícita, sino como una mediación aparentemente neutral, lo que refuerza su eficacia.

Desde la perspectiva de la psicopolítica, este fenómeno puede entenderse como una transformación en las formas de dominación. Como plantea Han (2021), el poder contemporáneo no se ejerce principalmente a través de la coerción, sino mediante la orientación de la conducta y la gestión de la información. En este contexto, la inteligencia artificial no solo actúa sobre los individuos, sino que interviene en las condiciones mismas bajo las cuales estos perciben y comprenden el mundo.

Esto conduce a una tensión central: si la inteligencia artificial está estructuralmente vinculada a intereses de poder, ¿es posible concebirla como un sistema imparcial? La evidencia sugiere que dicha imparcialidad no solo es difícil de alcanzar, sino que podría constituir una expectativa problemática, en tanto invisibiliza los procesos de diseño, selección y control que intervienen en el desarrollo de estos sistemas.

Asimismo, la discusión sobre la regulación adquiere una nueva dimensión. No se trata únicamente de establecer límites normativos, sino de reconocer que la tecnología que se busca regular forma parte de las mismas estructuras que influyen en la construcción de dichos marcos regulatorios. Esto no implica que la regulación sea inviable, pero sí que su alcance está condicionado por factores que exceden lo técnico y lo jurídico.

En este sentido, más que plantear la regulación como una solución definitiva, resulta pertinente entenderla como un proceso parcial, que debe complementarse con una reflexión crítica sobre las relaciones entre tecnología, poder y sociedad. La inteligencia artificial no puede analizarse únicamente desde su funcionalidad, sino desde las dinámicas de poder que la atraviesan y que determinan sus usos y efectos.

9. CONCLUSIONES

La inteligencia artificial se ha consolidado como una de las tecnologías más influyentes del siglo XXI, transformando múltiples dimensiones de la vida social. No obstante, su creciente presencia ha estado acompañada por una narrativa que la presenta como un sistema neutral, objetivo y autónomo. El análisis desarrollado en este trabajo permite cuestionar esta percepción, al mostrar que la IA opera dentro de marcos definidos por intereses corporativos, estatales e ideológicos.

A partir del enfoque de la psicopolítica, se ha argumentado que la inteligencia artificial no solo procesa información, sino que participa en la configuración de la realidad social, influyendo en la percepción, la toma de decisiones y la construcción de sentido. Este papel la posiciona como un dispositivo de poder que actúa de

manera sutil, integrándose en las prácticas cotidianas y operando bajo la apariencia de neutralidad.

Los casos analizados, que abarcan desde el uso de datos en campañas políticas hasta la implementación de sistemas de IA en contextos de conflicto y la gestión de la información en entornos digitales, evidencian que estos sistemas no funcionan de manera independiente de los intereses que los configuran. Por el contrario, reproducen y amplifican dichas agendas, lo que refuerza su carácter no neutral.

En este contexto, la regulación de la inteligencia artificial enfrenta límites estructurales. La velocidad del desarrollo tecnológico, su carácter transnacional y su inserción en dinámicas de poder dificultan la construcción de marcos normativos plenamente efectivos. Si bien la regulación sigue siendo necesaria, su alcance es necesariamente parcial, en tanto no puede abordar completamente las dimensiones más profundas del fenómeno.

En última instancia, comprender la inteligencia artificial implica reconocerla no solo como una innovación tecnológica, sino como un elemento central en la reconfiguración del poder en la era digital. Este reconocimiento no busca desestimar sus beneficios potenciales, sino situarlos en un contexto más amplio, donde sus implicaciones sociales, políticas y éticas puedan ser analizadas de manera crítica.

Referencias

British Broadcasting Corporation. (s.f.). [Artículo sobre inteligencia artificial y desinformación]. BBC

Mundo. <https://www.bbc.com/mundo/articles/cddnjd34p7no>

Carbajal, J. (2026). El Pentágono usó la IA de Anthropic para atacar a Irán, a pesar de ser vetada por Trump horas antes. *Wired en Español*. <https://es.wired.com/articulos/el-pentagono-uso-la-ia-de-anthropic-para-atacar-a-iran-a-pegar-de-ser-vetada-por-trump-horas-antes>

El Economista. (2026). China bloquea adquisición de firma de IA Manus por parte de Meta. *El Economista*. <https://www.eleconomista.com.mx/tecnologia/china-bloquea-adquisicion-firma-ia-manus-parte-meta-20260427-810666.html>

Han, B.-C. (2021). *Psicopolítica: Neoliberalismo y nuevas técnicas de poder*. Herder.

Hu, M. (2020). Cambridge Analytica's black box. *Big Data & Society*. <https://journals.sagepub.com/doi/pdf/10.1177/2053951720938091>

Nexos. (2026). Anthropic vs. el Pentágono y la lucha sobre quién controla la IA. *Nexos*. <https://eljuegodelacorte.nexos.com.mx/anthropic-vs-el-pentagono-y-la-lucha-sobre-quien-controla-la-ia/>

RTVE. (2026). Por qué haces caso a Grok: así desinforma la IA sobre la guerra en Irán. *RTVE Noticias*. <https://www.rtve.es/noticias/20260310/por-haces-caso-a-grok-asi-desinforma-ia-sobre-guerra-iran-no-solo/16973187.shtml>

CAPÍTULO XVII

GOBERNANZA DE LA IA CON PROPÓSITO: UN MARCO PARA EL IMPACTO SOCIAL UNIVERSITARIO

AI Governance with Purpose: A Framework for University Social Impact

Dra. Janette Brito Laredo

jbrito@uabc.edu.mx

Universidad Autónoma de Baja California, México

<https://orcid.org/0000-0001-8715-0770>

Resumen

La integración de la Inteligencia Artificial (IA) en la educación superior ha transitado mayoritariamente entre la experimentación técnica y la optimización administrativa, dejando un vacío crítico en la gestión de proyectos orientados al bienestar social. El presente trabajo propone la "Gobernanza de la IA con Propósito", un marco habilitador y transversal diseñado para guiar la implementación de tecnologías emergentes en proyectos con impacto social positivo, específicamente dirigidos a sectores vulnerables.

A diferencia de los modelos regulatorios tradicionales de corte restrictivo, esta propuesta se define como una arquitectura de acompañamiento que permite a todas las disciplinas universitarias —desde las humanidades hasta las ciencias de la salud— orquestar soluciones tecnológicas con un enfoque humanista. El marco se fundamenta en tres ejes rectores: la intencionalidad social como requisito de viabilidad, la transparencia explicable para los beneficiarios y la soberanía tecnológica comunitaria.

A través de protocolos de co-diseño y métricas basadas en el bienestar humano (en lugar de la eficiencia algorítmica), esta gobernanza busca institucionalizar una

"IA para la Vida". El resultado es un modelo de gestión que no solo protege a las poblaciones marginadas de los riesgos del extractivismo de datos, sino que posiciona a la universidad como un actor clave en la creación de un futuro tecnológico digno, equitativo y profundamente humano.

Palabras clave: Gobernanza de la IA, Impacto Social, Sectores Vulnerables, Gestión Universitaria, Ética Transversal, IA con Propósito

Abstract

The integration of Artificial Intelligence (AI) in higher education has mostly moved between technical experimentation and administrative optimization, leaving a critical gap in the management of projects aimed at social welfare. This paper proposes the "Governance of AI with Purpose", an enabling and transversal framework designed to guide the implementation of emerging technologies in projects with positive social impact, specifically aimed at vulnerable sectors.

Unlike traditional restrictive regulatory models, this proposal is defined as an accompanying architecture that allows all university disciplines – from the humanities to the health sciences – to orchestrate technological solutions with a humanistic approach. The framework is based on three guiding axes: social intentionality as a requirement for viability, transparency that can be explained to the beneficiaries and community technological sovereignty.

Through co-design protocols and metrics based on human well-being (rather than algorithmic efficiency), this governance seeks to institutionalize an "AI for Life." The result is a management model that not only protects marginalized populations from the risks of data extractivism, but positions the university as a key player in the creation of a dignified, equitable, and deeply humane technological future.

Keywords: AI Governance, Social Impact, Vulnerable Sectors, University Management, Cross-Cutting Ethics, AI with Purpose

I. INTRODUCCIÓN

La Inteligencia Artificial (IA) ha dejado de ser una promesa de la ciencia ficción para convertirse en la infraestructura invisible que moldea la realidad contemporánea. Sin embargo, su despliegue en el ámbito universitario mexicano y latinoamericano enfrenta una paradoja crítica: mientras la capacidad técnica de procesamiento crece exponencialmente, la claridad sobre su propósito social parece diluirse. En la actualidad, la mayoría de las instituciones de educación superior gestionan la IA bajo una lógica puramente instrumental o administrativa, ignorando su potencial —y su riesgo— como herramienta de transformación en los sectores más vulnerables de la sociedad.

El despliegue global de la Inteligencia Artificial ha seguido una trayectoria dictada por las leyes del mercado y la eficiencia operativa. En menos de una década, hemos sido testigos de una explosión tecnológica donde la IA se ha utilizado con éxito para propósitos de vigilancia, marketing predictivo, automatización industrial y entretenimiento masivo. Sin embargo, esta omnipresencia técnica contrasta con una ausencia de propósito humano en los sectores que más necesitan de la innovación: las poblaciones en situación de vulnerabilidad.

La problemática reside en que la IA, al ser desarrollada mayoritariamente bajo lógicas de rentabilidad, ha ignorado los problemas que no ofrecen un retorno económico inmediato. Esto ha generado una tecnología de élite, diseñada por y para contextos de abundancia, que resulta ciega a las necesidades de los sectores marginados.

- IA para la eficiencia, no para la equidad: Mientras existen algoritmos capaces de predecir el comportamiento de compra de un usuario con precisión quirúrgica, hay un vacío de herramientas que ayuden a un pequeño productor agrícola a mitigar los efectos del cambio climático o que permitan a comunidades indígenas acceder a servicios de salud en su propia lengua.

- La IA como herramienta de exclusión: En muchos casos, cuando la IA ha llegado a los sectores vulnerables, no lo ha hecho para "ayudar", sino para vigilar o clasificar. Desde algoritmos de evaluación crediticia que niegan préstamos por códigos postales hasta sistemas de reconocimiento facial con sesgos raciales, la tecnología ha funcionado más como una barrera que como un puente.

Dentro de la universidad, esta problemática se refleja en el predominio de proyectos de "IA por la IA". Con frecuencia, los esfuerzos académicos se agotan en la optimización de parámetros técnicos (mejorar la precisión de un modelo en un 1%) sin preguntarse si ese modelo tiene una aplicación real en la resolución de una injusticia social.

La falta de un marco de gobernanza ha permitido que:

1. Se ignoren los contextos locales: Se importan modelos y bases de datos que no reflejan la realidad de los sectores vulnerables mexicanos.
2. Se priorice el "*paper*" sobre la "causa": El éxito de una investigación se mide por la publicación en revistas internacionales y no por el impacto positivo generado en la comunidad donde se recolectaron los datos.
3. Inexistencia de una metodología de vinculación: Incluso cuando existe la voluntad de ayudar, los investigadores carecen de un protocolo de gobernanza que les indique cómo integrar la IA en procesos de economía solidaria o derechos humanos sin causar daños colaterales.

La problemática, por tanto, no es la IA en sí misma, sino el aislamiento ético de su uso. Si la universidad no interviene con un marco regulatorio que exija y facilite la orientación de estos proyectos hacia sectores vulnerables, la IA seguirá siendo una herramienta poderosa pero indiferente al sufrimiento y a las carencias humanas. Es necesario pasar de una IA que "puede hacerlo todo" a una IA que "decida hacer lo correcto".

Objetivos de la Investigación

Para dar respuesta a la problemática planteada y cumplir con el propósito de este estudio, se han definido los siguientes objetivos:

Objetivo General

Diseñar un Marco de Gobernanza Transversal para la IA con Propósito, que sirva como modelo habilitador para la gestión de proyectos universitarios orientados a generar un impacto social positivo y tangible en sectores vulnerables.

Objetivos Específicos

1. Sustentar el paradigma de la "IA para la Vida" mediante una revisión crítica de la literatura sobre justicia algorítmica y soberanía tecnológica, contrastándolo con los modelos extractivos vigentes.
2. Desarrollar un protocolo de gestión transversal y co-diseño que estandarice los pasos administrativos y éticos para que cualquier disciplina universitaria pueda implementar proyectos de IA con pertinencia social.
3. Establecer métricas de impacto basadas en el bienestar humano que permitan evaluar el éxito de los proyectos según la reducción de brechas de desigualdad y el empoderamiento de las comunidades beneficiarias.

Justificación

La pertinencia de esta investigación emana de la necesidad crítica de dotar a las instituciones de educación superior de una brújula ética y operativa en un momento de expansión tecnológica sin precedentes. Mientras la Inteligencia Artificial (IA) se ha consolidado como un motor de eficiencia para sectores financieros, comerciales y de vigilancia, su aplicación para el fortalecimiento de los sectores vulnerables permanece como una tarea pendiente y desarticulada.

La justificación de este marco de gobernanza reside en la actualización de la Tercera Función Universitaria: la extensión y vinculación social. En el siglo XXI, la vinculación no puede limitarse a la presencia física; debe manifestarse como una presencia digital responsable. Si la universidad no lidera la creación de proyectos de IA con propósito, abandona su responsabilidad de actuar como contrapeso ante el uso extractivo de la tecnología. Este trabajo se justifica al proponer que la IA sea el nuevo lenguaje de la justicia social, permitiendo que el conocimiento experto se democratice y llegue a quienes históricamente han sido excluidos del progreso técnico.

Así mismo es urgente transitar de una "IA de consumo" a una "IA para la Vida". Esta investigación se justifica al ofrecer una alternativa al colonialismo de datos, donde los sectores vulnerables suelen ser sujetos de experimentación o proveedores de información para modelos ajenos a su realidad. Al establecer una gobernanza con propósito, la universidad garantiza la Soberanía Tecnológica: asegura que los datos y soluciones generados permanezcan bajo la custodia y el beneficio directo de las comunidades, transformando la tecnología en un bien común y no en una herramienta de desplazamiento o control.

Finalmente, esta propuesta se justifica por la necesidad de redefinir la excelencia académica. En un entorno donde el éxito se mide a menudo por indicadores de productividad editorial, es imperativo introducir métricas de Impacto Social Positivo (ISP). El marco de gobernanza propuesto otorga validez institucional a la innovación solidaria, incentivando que investigadores y estudiantes de todas las disciplinas utilicen la IA no para "hacer más", sino para "hacer el bien", consolidando a la universidad como un ecosistema de innovación al servicio de la dignidad humana.

La arquitectura de este proyecto se despliega a través de una ruta crítica que transita de la abstracción teórica a la operatividad administrativa, estructurándose en cuatro ejes fundamentales que permiten desglosar la Gobernanza de la IA con

Propósito de manera integral. En primer lugar, se establece una Fundamentación Teórica que construye un diálogo crítico entre la justicia algorítmica y el paradigma de la "IA para la Vida" frente a los modelos extractivos. Posteriormente, se detalla una Metodología de corte cualitativo y propositivo, diseñada para identificar las brechas de gestión institucional y trazar el diseño de una arquitectura habilitadora. El núcleo de la investigación reside en los Resultados, donde se presenta formalmente el modelo de gobernanza, describiendo los protocolos de co-diseño multidisciplinario y el ciclo de vida de los proyectos con pertinencia social. Finalmente, el capítulo culmina con las Conclusiones y Recomendaciones, donde se ofrecen estrategias para institucionalizar este marco en la educación superior, proyectando un futuro donde la excelencia académica sea indisoluble del impacto social positivo y la justicia digital.

II. REVISIÓN DE LA LITERATURA

La construcción de un marco de gobernanza para la Inteligencia Artificial (IA) con propósito social no puede partir de un vacío conceptual. Requiere, por el contrario, una sólida cimentación que articule la responsabilidad histórica de la universidad, los modelos económicos alternativos y las nuevas filosofías del derecho tecnológico. La presente revisión analiza cómo la educación superior debe evolucionar para convertir la tecnología en un bien común, alejándose de los modelos puramente comerciales o restrictivos.

La Función Social de la Universidad en la Era Digital: Hacia una Responsabilidad Social Tecnológica

La universidad, particularmente en el contexto latinoamericano, ha sido definida por su "Tercera Función": la extensión y vinculación con la sociedad. Según Tünnermann (2008), esta función no es un apéndice de la docencia o la investigación, sino el eje que otorga pertinencia a ambas. En la era digital, esta misión enfrenta un nuevo desafío: la brecha de datos y el analfabetismo algorítmico.

La Función Social de la Universidad hoy debe entenderse como la garantía de que el avance científico no profundice las desigualdades existentes. Ordorika (2006) sostiene que la universidad pública debe actuar como un espacio de resistencia frente a las fuerzas del mercado que buscan privatizar el conocimiento. Aplicado a la IA, esto implica que las instituciones no deben limitarse a formar técnicos para la industria, sino a desarrollar "intelectuales orgánicos" capaces de diseñar sistemas que resuelvan carencias en salud, educación y justicia para los sectores más vulnerables.

La responsabilidad social universitaria (RSU) se transforma así en Responsabilidad Social Tecnológica. Esta perspectiva sugiere que la universidad tiene la obligación ética de intervenir en el ciclo de vida de la IA, asegurando que su implementación en territorio sea transparente y beneficie directamente a la comunidad, evitando que los sectores marginados sean vistos meramente como fuentes de extracción de datos (Couldry & Mejias, 2019).

Economía Social y Solidaria (ESS) y Tecnología: El Referente para la IA con Propósito

Para que la IA tenga un impacto social positivo, su gestión debe alinearse con principios económicos que prioricen la reproducción de la vida sobre la acumulación de capital. La Economía Social y Solidaria (ESS) ofrece el marco ideal para esta transición. Coraggio (2011) define la economía social como un sistema que busca la satisfacción de las necesidades de todos los miembros de la comunidad mediante el trabajo asociado y la solidaridad.

Al vincular la tecnología con la ESS, surge el concepto de Tecnología Social. Según Dagnino (2014), la tecnología social es aquella desarrollada de manera participativa, diseñada para ser apropiada por los usuarios y orientada a la resolución de problemas locales. Una gobernanza de la IA con propósito debe adoptar estos principios de la ESS para:

1. Fomentar la Propiedad Colectiva: Evitar la dependencia de infraestructuras privadas y fomentar el uso de software libre y datos abiertos.
2. Soberanía Tecnológica: Garantizar que las comunidades tengan el control sobre los algoritmos que afectan sus vidas (Janssen y otros., 2020).
3. Gestión Democrática: Incluir a los beneficiarios en el co-diseño de las soluciones, tal como propone el modelo de innovación solidaria.

La IA, bajo la óptica de la ESS, deja de ser una herramienta de optimización de ganancias para convertirse en un instrumento de "ayuda mutua digital", permitiendo que organizaciones comunitarias y sectores vulnerables accedan a niveles de eficiencia y predicción que antes eran exclusivos de las grandes corporaciones.

Gobernanza Habilitadora vs. Gobernanza Restrictiva: El Cambio de Paradigma

La literatura jurídica actual sobre la IA suele dividirse en dos grandes corrientes. Por un lado, la Gobernanza Restrictiva, ejemplificada por el *AI Act* de la Unión Europea, que se centra en la gestión de riesgos, la prohibición de usos peligrosos y la protección de la privacidad (European Commission, 2021). Si bien este enfoque es necesario para prevenir abusos, a menudo resulta insuficiente para fomentar el impacto social en países en desarrollo, ya que su lógica es defensiva.

Por otro lado, la Gobernanza Habilitadora (o proactiva) propone que la regulación no debe ser un freno, sino una infraestructura que facilite la innovación con ética. Janssen y Helbig (2019) argumentan que una gobernanza efectiva en la era digital debe ser ágil y orientada a la creación de valor público. En el entorno universitario, una gobernanza habilitadora se traduce en:

- Protocolos de Acompañamiento: En lugar de listas de prohibiciones, se ofrecen guías para que investigadores de todas las disciplinas puedan integrar IA de forma segura y ética.

- Transversalidad Operativa: La gobernanza rompe los silos técnicos y permite que expertos en humanidades o salud lideren proyectos tecnológicos.
- Flexibilidad Institucional: La capacidad de adaptar la normativa a las necesidades cambiantes de los sectores vulnerables sin perder el rigor ético.

Mientras la gobernanza restrictiva pregunta "¿qué no podemos hacer?", la Gobernanza con Propósito pregunta "¿cómo podemos usar esto para ayudar?". Este cambio de paradigma es lo que permite que la IA pase de ser una amenaza latente a una herramienta de empoderamiento comunitario.

La revisión de la literatura permite concluir que la IA no es un fenómeno aislado de la estructura social. Su éxito en el ámbito universitario no se medirá por la potencia de sus algoritmos, sino por su capacidad de integrarse en la misión histórica de la universidad y en los principios de la economía solidaria. La transición de una gobernanza restrictiva a una habilitadora es el paso lógico para una institución que busca ser un agente de cambio. Los autores analizados coinciden en que la tecnología debe ser gestionada como un bien público, donde la justicia algorítmica y la soberanía de datos sean la base sobre la cual se construya el bienestar de los sectores vulnerables. Este sustento teórico es el que da validez al modelo de gobernanza que se propondrá en los capítulos posteriores, asegurando que la innovación sea, ante todo, un acto de responsabilidad humana.

III. METODOLOGÍA

La construcción de un marco de gobernanza para la Inteligencia Artificial (IA) con propósito social requiere un abordaje metodológico que sea capaz de integrar la complejidad técnica de los algoritmos con la sensibilidad ética del impacto humano. El presente estudio no se limita a observar la realidad, sino que busca transformarla mediante la creación de un objeto de gestión institucional. Por ello, se adopta un enfoque cualitativo y propositivo, orientado al diseño de políticas y protocolos que faciliten la vinculación universitaria con sectores vulnerables.

Como señala Hevner y otros. (2004) en su teoría de la *Ciencia del Diseño*, la investigación en sistemas y gestión debe buscar no solo la comprensión del entorno, sino la creación de "artefactos" (en este caso, un marco de gobernanza) que resuelvan problemas organizacionales y sociales detectados. Esta metodología permite que el resultado final sea una herramienta operativa y no solo un documento teórico, asegurando que la tecnología se gestione como un bien común.

Paradigma y Diseño de la Investigación

El estudio se inscribe en el paradigma crítico-social, el cual sostiene que la investigación debe ser un medio para la emancipación y la justicia social (Guba & Lincoln, 1994). Bajo esta lógica, la metodología se aleja de la supuesta neutralidad tecnológica para reconocer que la IA, si no es gobernada bajo principios éticos, tiende a replicar estructuras de poder y exclusión.

El diseño es no experimental, descriptivo y transversal. Se describe la situación actual de la gobernanza de la IA en la universidad —caracterizada por un vacío de propósito— y se propone una solución estructurada. La investigación es de carácter cualitativo, utilizando el análisis documental y la síntesis conceptual como herramientas principales. Siguiendo a Hernández-Sampieri y otros. (2018), este enfoque permite explorar los significados que las instituciones otorgan a la innovación y cómo estos pueden alinearse con el bienestar comunitario.

Ruta Crítica: Fases de la Investigación

Para garantizar la solidez del modelo, la metodología se despliega en cuatro fases consecutivas:

Fase 1: Diagnóstico y Análisis de Brechas (Exploratoria)

Se realizó un análisis documental de los marcos regulatorios internacionales y locales, así como de los reglamentos internos de vinculación universitaria. El

objetivo fue identificar la "brecha de propósito": la distancia entre las capacidades técnicas y las exigencias normativas de retorno social. Se utilizó el Análisis Crítico del Discurso (Fairclough, 2013) para entender cómo los conceptos de "eficiencia" han desplazado a la "justicia social" en la gestión tecnológica.

Fase 2: Fundamentación y Marco de Referencia (Conceptual)

Se seleccionaron los pilares teóricos: Justicia Algorítmica, Economía Solidaria y Responsabilidad Social Universitaria. Esta fase definió los criterios de validación que el marco de gobernanza exigirá a los proyectos, integrando el modelo de Diseño Centrado en la Justicia (Costanza-Chock, 2020).

Fase 3: Diseño de la Arquitectura de Gobernanza (Propositiva).

Esta fase constituye el núcleo de la investigación, en la cual se modela el flujo de gestión institucional orientado a la implementación de proyectos con propósito social. La arquitectura propuesta se conceptualiza como una Gobernanza Habilitadora, estructurada en cinco etapas críticas que articulan de manera secuencial y coherente la transición desde la ideación hasta la generación de impacto social.

Como se ilustra en la Figura 1, este modelo permite integrar procesos de validación, articulación interdisciplinaria, co-diseño, desarrollo e impacto, asegurando la pertinencia, viabilidad y sostenibilidad de las iniciativas.



Figura 1:

Flujo de gobernanza universitaria

- **1. Validación e Intencionalidad Social:** Esta etapa constituye un filtro de entrada orientado a garantizar que los proyectos surjan de necesidades reales, particularmente en sectores en situación de vulnerabilidad. En este sentido, no se permite el avance de iniciativas que no demuestren un propósito de bienestar humano tangible y verificable. Como se muestra en la Figura 2, este proceso se operacionaliza mediante una evaluación secuencial que considera criterios de justicia social, vinculación comunitaria e impacto tecnológico, asegurando que cada propuesta no solo sea viable, sino también socialmente pertinente y éticamente fundamentada antes de avanzar a las siguientes fases del modelo.



Figura 2:

Diagrama de Flujo de la Etapa de Validación de Intencionalidad

- **2. Triage Multidisciplinario:** La gobernanza actúa como un puente que integra dimensiones técnicas, éticas y sociales. Aquí se asegura que investigadores de humanidades y ciencias sociales tengan el mismo peso en la toma de decisiones que los desarrolladores técnicos.

- **3. Co-Diseño en Territorio:** En esta fase se establecen protocolos participativos que garantizan la incorporación activa de la comunidad beneficiaria en el diseño de la solución de inteligencia artificial. Este enfoque busca evitar dinámicas extractivas y promover procesos colaborativos basados en el diálogo horizontal, asegurando que la tecnología desarrollada sea culturalmente pertinente, contextualizada y socialmente legítima. Como se muestra en la Figura 3, el co-diseño se articula como un ciclo iterativo que integra escucha, traducción,

prototipado y transferencia, fortaleciendo la apropiación comunitaria y la sostenibilidad de la solución.



Figura 3:

Esquema del Modelo de Co-diseño con Sectores Vulnerables

- **4. Desarrollo Ético y Monitoreo:** En esta fase se implementan mecanismos sistemáticos de supervisión orientados a garantizar el funcionamiento responsable de la solución de inteligencia artificial. Se incorporan auditorías periódicas para identificar y mitigar posibles sesgos algorítmicos, así como estrategias de transparencia y explicabilidad que permitan comprender el comportamiento del sistema. De este modo, se asegura que el desarrollo tecnológico se mantenga alineado con principios éticos, normativos y de rendición de cuentas.

- **5. Evaluación de Impacto y Soberanía:** En esta etapa final se valora el éxito de la intervención a partir de indicadores centrados en la autonomía, el bienestar y la mejora tangible en las condiciones de vida de la comunidad. A diferencia de enfoques tradicionales, la evaluación no se limita al desempeño técnico, sino que incorpora métricas de apropiación social y empoderamiento.

Asimismo, se establece un protocolo de transferencia orientado a garantizar la soberanía tecnológica, mediante el cual la comunidad asume la custodia de sus datos y de la solución desarrollada. Como se muestra en la **Figura 4**, este proceso consolida el tránsito desde la implementación hacia la sostenibilidad, asegurando la continuidad, legitimidad y control comunitario de la tecnología.



Figura 4:

Matriz de Indicadores de Bienestar Humano vs. Métricas Técnicas

Fase 4: Validación Teórica y Refinamiento (Evaluativa)

El modelo se sometió a una **triangulación teórica**, contrastándolo con los principios del Programa Institucional de Economía Social y Solidaria, asegurando que la propuesta sea compatible con la misión histórica de la universidad pública y su compromiso con la sociedad.

Técnicas e Instrumentos de Recolección

La técnica principal fue el Análisis Documental Sistemático. El instrumento utilizado fue una Matriz de Análisis de Contenido, donde se categorizaron los documentos (planes de desarrollo, leyes de IA, marcos éticos) según tres variables fundamentales:

1. Nivel de Control Ético: Capacidad de supervisar impactos negativos.
2. Nivel de Habilidad: Facilidades para la colaboración interdisciplinaria.
3. Presencia de Sectores Vulnerables: Inclusión explícita de grupos marginados.

Alcance y Limitaciones

El alcance es institucional y estratégico, diseñado para ser replicable en cualquier institución de educación superior que busque dotar de sentido social a sus áreas de innovación. Una limitación es la velocidad del cambio tecnológico; no obstante, al centrar la gobernanza en la ética humana y no en herramientas de software específicas, el marco mantiene su vigencia frente a la evolución de la IA.

IV. RESULTADOS

El resultado central de esta investigación es el diseño de un Marco de Gobernanza Habilitadora, un modelo que transforma la administración de la tecnología en un ejercicio de ética aplicada. A diferencia de los marcos regulatorios tradicionales que operan bajo una lógica de control o restricción, este modelo propone una estructura que acompaña la creación tecnológica desde su concepción hasta su despliegue en territorio. El objetivo no es fiscalizar la herramienta, sino garantizar que la Inteligencia Artificial (IA) funcione como un acelerador de derechos y bienestar para los sectores históricamente marginados. A continuación, se detallan los componentes que integran el modelo, explicando cómo cada uno de ellos emana directamente del proceso metodológico.

4.1. El Comité Interdisciplinario (CEIIA) como Respuesta al Diálogo de Saberes

Como consecuencia directa de la etapa metodológica orientada al triaje multidisciplinario, surge la propuesta del Comité Interdisciplinario de IA con Propósito (CEIIA). El hallazgo fundamental en esta etapa fue que la gobernanza técnica suele operar en un vacío ético cuando no existe un contrapeso de las humanidades. Por ello, el CEIIA se materializa no como un órgano burocrático, sino como un eje de convergencia el cual no pertenece a una facultad técnica, sino que se sitúa como un eje transversal que conecta la investigación, la administración y la vinculación social de la universidad.

El CEIIA rompe los silos académicos mediante una composición híbrida:

- Defensoría de Justicia Algorítmica: Perfiles de Derecho y Ética encargados de supervisar el cumplimiento de los derechos humanos digitales.
- Arquitectura de Innovación Solidaria: Expertos en ingeniería que aseguran la factibilidad, seguridad y el uso de estándares abiertos.
- Mediación Sociocomunitaria: Especialistas en ciencias sociales que validan la pertinencia cultural y el contacto con los sectores vulnerables.
- Voz Ciudadana: Representantes de las comunidades beneficiarias con capacidad de veto sobre proyectos que consideren invasivos o inútiles.

Este último componente es quizás el hallazgo más disruptivo: la metodología demostró que, sin poder de decisión real por parte de los beneficiarios, la gobernanza es meramente simbólica.

Instrumentos Operativos: De la Intencionalidad al Co-Diseño

El proceso de diseño de la arquitectura permitió identificar que la ética universitaria requiere de instrumentos tangibles para ser aplicable. Así, derivados

de la fase metodológica de **Validación de Intencionalidad**, se presentan tres instrumentos que guían al investigador durante todo el ciclo de vida del proyecto.

1. **Matriz de Intencionalidad Social.** Este documento administrativo es la materialización del primer filtro ético: obliga al investigador a declarar la brecha de desigualdad específica que busca mitigar antes de iniciar cualquier desarrollo. Este instrumento transforma la intención en un compromiso auditable, asegurando que el proyecto no sea una simple experimentación técnica, sino una respuesta a un problema de justicia social.

2. **Protocolo de Co-Diseño.** Este manual operativo es la respuesta metodológica al riesgo de extractivismo de datos. Establece una ruta de trabajo horizontal donde la comunidad beneficiaria participa en la definición del problema, la validación de los requerimientos y las pruebas de usabilidad en territorio. El hallazgo clave aquí es que la tecnología solo es inclusiva cuando se adapta a los niveles de alfabetización digital y contextos lingüísticos de quienes la utilizan.

3. **Manual de Transparencia Explicable** es el resultado de la etapa de monitoreo ético. Este instrumento garantiza que los algoritmos universitarios funcionen como "cajas blancas", permitiendo que cualquier ciudadano comprenda la lógica detrás de una decisión automatizada, fortaleciendo así la confianza institucional y la democracia digital.

Como culminación operativa de este marco, se ha desarrollado una propuesta normativa detallada (ver Anexo 1: Borrador de Reglamento de IA con Propósito), la cual traduce los principios de justicia algorítmica en artículos legales aplicables.

El Nuevo Paradigma de Evaluación: El Impacto en el Bienestar

Uno de los resultados más potentes de esta investigación es el desplazamiento de las métricas tradicionales de éxito. La metodología, en su fase de evaluación de impacto, reveló que indicadores como el *accuracy* o la precisión estadística son

insuficientes para medir el retorno social de la ciencia. Como consecuencia, el modelo propone el Índice de Impacto en el Bienestar (IB). Este sistema de evaluación se sustenta en tres indicadores emergentes:

1. Índice de Autonomía, que mide en qué medida la IA ha reducido la dependencia de los sectores vulnerables hacia intermediarios o sistemas burocráticos opresivos.

2. Soberanía de Datos, que garantiza que la información generada por la comunidad permanezca bajo su custodia y la de la universidad, impidiendo su comercialización por entes externos.

3. Equidad Algorítmica funciona como una auditoría de salida que certifica la ausencia de sesgos discriminatorios. Estos indicadores transforman la noción de "excelencia académica", alineándola con la dignidad humana y no con la productividad comercial.

Escenarios de Validación: Casos Hipotéticos de Aplicación

Para evidenciar el potencial transformador del modelo de gobernanza propuesto, se destacan tres casos hipotéticos con alto impacto estructural, en tanto abordan dimensiones críticas de exclusión: acceso a la justicia, inclusión financiera y redistribución de recursos públicos. Estos casos ilustran cómo la inteligencia artificial, bajo un enfoque ético y de co-diseño, puede generar cambios sostenibles en contextos de vulnerabilidad.

Caso 1: Justicia Lingüística y Jurídica ("Voces de la Tierra")

Este proyecto representa un avance significativo en el acceso equitativo a la justicia para comunidades indígenas. Mediante el uso de Procesamiento de Lenguaje Natural (PLN), el sistema no solo traduce términos legales, sino que incorpora elementos semánticos y culturales propios de cada comunidad, permitiendo una interpretación contextualizada del lenguaje jurídico. Esto reduce

las barreras estructurales que históricamente han limitado la defensa legal de estos grupos. El valor diferencial radica en su enfoque de co-diseño, donde líderes comunitarios participan en la validación del sistema, asegurando que la herramienta refleje adecuadamente los significados culturales. Su impacto trasciende lo individual, al fortalecer la confianza institucional y promover una justicia más inclusiva. La mejora en la tasa de resoluciones favorables y la reducción de errores de interpretación legal constituyen indicadores clave de éxito.

Caso 2: Micro-crédito Soberano y Justicia Crediticia

Este caso aborda uno de los problemas más persistentes en economías emergentes: la exclusión del sistema financiero formal. A través de un modelo de inteligencia artificial que integra datos alternativos —como historial comunitario, redes de cooperación y comportamiento social—, se construye un sistema de evaluación crediticia más justo y representativo. La innovación principal radica en desplazar la lógica tradicional basada exclusivamente en historial bancario, incorporando métricas de confianza social que reflejan mejor la realidad de los sectores informales. La universidad, como actor institucional, garantiza la legitimidad del sistema y reduce el riesgo percibido por entidades financieras. Esto permite ofrecer créditos en condiciones accesibles, disminuyendo la dependencia de esquemas de financiamiento abusivos. El impacto se manifiesta en el fortalecimiento de microemprendimientos, la estabilidad económica de los hogares y la generación de circuitos financieros más inclusivos, contribuyendo a una redistribución más equitativa de oportunidades.

Caso 3: Captura de Surplus Social y simplificación burocrática

Este proyecto incide directamente en la eficiencia del Estado y en la equidad en la distribución de recursos públicos. A través de un asistente inteligente de acompañamiento ciudadano, se automatiza la identificación de derechos no ejercidos, facilitando el acceso a programas sociales, subsidios y beneficios fiscales. El sistema reduce la fricción administrativa —uno de los principales

obstáculos para las poblaciones vulnerables— al eliminar barreras como la complejidad de los trámites, la falta de información y los costos de intermediación. Además, introduce una lógica proactiva, donde el Estado, mediado por la tecnología, “encuentra” al ciudadano en lugar de esperar que este navegue sistemas burocráticos complejos. El impacto es altamente significativo, ya que permite recuperar recursos públicos no distribuidos (surplus social) y canalizarlos directamente hacia quienes más los necesitan. Esto no solo mejora el ingreso disponible de los hogares, sino que fortalece la legitimidad institucional y la eficiencia del gasto público.

En resumen, estos casos no solo ilustran aplicaciones aisladas de la inteligencia artificial, sino que evidencian su potencial como infraestructura de transformación social cuando es orientada por un modelo de gobernanza ética, participativa y centrada en el bienestar humano. Lejos de reproducir lógicas extractivas o de maximización puramente económica, la IA se reconfigura aquí como un instrumento de redistribución, capaz de ampliar derechos, reducir asimetrías estructurales y fortalecer las capacidades de agencia de las comunidades. Estos ejemplos demuestran que el verdadero valor de la inteligencia artificial no radica únicamente en su sofisticación técnica, sino en su capacidad de incidir en las condiciones materiales de vida, particularmente en contextos de exclusión. En este sentido, el modelo propuesto permite trascender la dicotomía entre innovación tecnológica y justicia social, integrándolas en una misma lógica operativa.

Los hallazgos de este capítulo permiten concluir que la Gobernanza de la IA con Propósito dota a la universidad de una metodología replicable para convertir la sofisticación técnica en un motor de soberanía y justicia. Al alinear cada componente orgánico e instrumental con la ruta crítica metodológica, se confirma que la tecnología puede y debe ser gestionada como un bien común.

El éxito de esta arquitectura no reside en la complejidad del algoritmo, sino en su capacidad para actuar como un puente entre el conocimiento universitario y las

necesidades más apremiantes de los sectores vulnerables. En última instancia, estos resultados posicionan a la institución educativa como un actor mediador indispensable y garante de legitimidad en el diseño de un futuro digital donde la excelencia tecnológica se defina, ante todo, por su contribución a la dignidad, la autonomía y la justicia social de todas las personas, independientemente de su campo de estudio.

V. CONCLUSIONES Y RECOMENDACIONES

La presente investigación permite concluir que la Inteligencia Artificial no es un fenómeno técnico aislado, sino un vector de poder que, sin una gobernanza adecuada, tiende a profundizar las asimetrías sociales preexistentes. En este sentido, se reafirma que la universidad no puede permanecer como un actor pasivo, sino que tiene la responsabilidad histórica de intervenir en el ciclo de vida de los algoritmos, asegurando que su diseño, implementación y evaluación respondan a una lógica de justicia y no meramente de mercado.

A lo largo del estudio se ha demostrado que la Gobernanza con Propósito constituye el mecanismo necesario para transitar de una “IA de vigilancia y extracción” hacia una “IA para la Vida”. Esta transición no es opcional, sino una respuesta ineludible de las instituciones de educación superior frente a la crisis contemporánea de privacidad, la explotación de datos y la reproducción de desigualdades mediante sistemas algorítmicos. En coherencia con este planteamiento, el desarrollo del Marco de Gobernanza Transversal propuesto evidencia que es posible articular una infraestructura institucional que no solo regule, sino que habilite la innovación con sentido social. Asimismo, la fundamentación teórica, el diseño del protocolo de gestión y la incorporación de métricas centradas en el bienestar humano permiten constatar que los objetivos de la investigación han sido alcanzados, al ofrecer un modelo operativo que integra justicia algorítmica, soberanía tecnológica y participación comunitaria.

De este modo, el modelo propuesto no solo dota a las universidades de una arquitectura ética, sino que redefine el sentido mismo de la innovación tecnológica, posibilitando que esta sea, simultáneamente, de excelencia técnica y de alta sensibilidad social. En última instancia, se plantea que la verdadera innovación no radica únicamente en la sofisticación de los algoritmos, sino en su capacidad de contribuir de manera tangible a la dignidad, el bienestar y el empoderamiento de los sectores más vulnerables.

A partir de los hallazgos y del modelo de Gobernanza con Propósito desarrollado, se vuelve necesario traducir sus principios en orientaciones concretas que permitan su implementación efectiva tanto en el ámbito universitario como en el espacio público. En este sentido, las siguientes recomendaciones no deben entenderse como lineamientos accesorios, sino como condiciones operativas indispensables para garantizar que la IA mantenga su orientación hacia el bienestar social, la justicia algorítmica y la soberanía tecnológica. Asimismo, estas directrices buscan facilitar la apropiación del modelo por parte de investigadores, gestores e instituciones, asegurando su viabilidad práctica, su sostenibilidad en el tiempo y su potencial de incidencia más allá del entorno académico.

Recomendaciones de Uso y Operatividad

Para los investigadores, académicos y gestores universitarios que busquen implementar este marco, se recomiendan las siguientes directrices de uso:

- **Adopción de la Transversalidad como Método:** No se debe iniciar un proyecto de IA sin la integración de una mirada humanística desde la primera línea de código. La recomendación es que todo equipo de desarrollo cuente con al menos un "enlace ético" encargado de vigilar los sesgos de entrenamiento.
- **Priorización de la Soberanía de Datos:** Se recomienda el uso de infraestructuras propias y protocolos de datos abiertos. El investigador debe asegurar que el conocimiento generado no quede "atrapado" bajo licencias

privativas que impidan que las comunidades beneficiarias sigan utilizando la herramienta de forma autónoma.

- Implementación de Auditorías de Sesgo Continuas: El uso de la IA no es un proceso estático. Se recomienda que, una vez desplegada la solución en territorio, se realicen evaluaciones trimestrales de impacto para detectar si el algoritmo ha comenzado a generar exclusiones imprevistas.

Para la implementación institucional de estas políticas, se sugiere tomar como base el modelo normativo propuesto en el presente trabajo (ver Anexo 1), el cual garantiza la soberanía de datos y la transparencia explicable.

Recomendaciones de Política Pública

El alcance de este modelo trasciende las paredes de la universidad y busca incidir en la agenda nacional. Se proponen las siguientes recomendaciones de política pública:

1. Creación de un Observatorio Nacional de IA con Propósito: Se sugiere que los organismos de ciencia y tecnología adopten esta gobernanza para supervisar todos los proyectos financiados con recursos públicos, asegurando que el Retorno Social de la Inversión (SROI) sea el principal indicador de éxito.

2. Incentivos Fiscales para la Innovación Solidaria: Proponer marcos legales que otorguen beneficios a las empresas y universidades que desarrollen IA orientada a sectores vulnerables bajo los estándares de transparencia y co-diseño aquí presentados.

3. Regulación de la Custodia de Datos Comunitarios: Impulsar leyes que reconozcan a los sectores marginados como dueños colectivos de sus datos, impidiendo el extractivismo por parte de plataformas transnacionales.

Es fundamental destacar que esta línea de investigación se alinea plenamente con la visión contemporánea que orienta el desarrollo científico en México. El Sistema Nacional de Investigadores e Investigadoras (SNII) y los marcos actuales

de política científica nacional han colocado en el centro la necesidad de que el conocimiento trascienda su dimensión teórica para convertirse en un instrumento de Soberanía Nacional y Bienestar Social. En este contexto, se impulsa una transformación profunda en el quehacer académico: pasar de una ciencia acumulativa, centrada en la producción de publicaciones, a una ciencia con incidencia real, capaz de integrarse de manera activa en los Programas Nacionales Estratégicos (Pronaces).

El modelo de Gobernanza de la IA con Propósito propuesto en esta investigación responde de manera directa a este llamado, al ofrecer una ruta clara para que la innovación tecnológica en el país no solo sea competitiva, sino socialmente pertinente. Al articular principios de justicia algorítmica, soberanía tecnológica y participación comunitaria, el modelo posiciona a la universidad como un actor clave en la construcción de soluciones orientadas a la equidad, la reducción de la pobreza y el fortalecimiento de una democracia digital inclusiva. En este sentido, gobernar la IA con propósito trasciende el ámbito técnico para convertirse en un acto de resistencia creativa frente a los modelos dominantes de extracción y mercantilización del conocimiento. Implica apostar por una universidad que no solo forma especialistas, sino ciudadanos críticos capaces de orientar el desarrollo tecnológico hacia el bien común. Así, la verdadera medida de una potencia tecnológica deja de estar definida por la capacidad de procesamiento o la sofisticación de sus sistemas, y se redefine en función de su capacidad para generar bienestar, dignidad y oportunidades, asegurando que el avance tecnológico no deje a nadie atrás.

ANEXO: Borrador del Reglamento de IA con Propósito

ARTÍCULO 1. OBJETO. El presente reglamento tiene por objeto establecer las bases éticas, técnicas y administrativas para la gestión de proyectos de Inteligencia Artificial (IA) que utilicen recursos o infraestructura universitaria, priorizando el impacto social en sectores vulnerables.

ARTÍCULO 2. DEL COMITÉ (CEIIA). Se crea el Comité Interdisciplinario de IA con Propósito, el cual será el órgano máximo de validación ética. Ningún proyecto de IA con financiamiento institucional podrá ejecutarse sin el dictamen favorable de este Comité.

ARTÍCULO 3. INTENCIONALIDAD SOCIAL. Todo proyecto deberá presentar una matriz de impacto social que identifique: a) El sector vulnerable beneficiario, b) El derecho que se busca fortalecer, c) El plan de mitigación de riesgos de exclusión.

ARTÍCULO 4. SOBERANÍA DE DATOS. Los datos recolectados en comunidades marginadas se consideran **Bien Común**. La Universidad garantiza su custodia y prohíbe su venta o cesión a terceros con fines de lucro. Al finalizar el proyecto, la comunidad tendrá acceso total a los modelos resultantes.

ARTÍCULO 5. TRANSPARENCIA Y EXPLICABILIDAD. Se prohíbe el uso de sistemas de "caja negra" en decisiones que afecten derechos fundamentales. Todo algoritmo debe ser explicable en lenguaje no técnico para sus usuarios y beneficiarios.

ARTÍCULO 6. MÉTRICAS DE EVALUACIÓN. El éxito del proyecto se medirá mediante el Índice de Bienestar Humano (IBH), evaluando el incremento en la autonomía, el ahorro de costos para el usuario y la mejora en la calidad de vida de la población objetivo.

ARTÍCULO 7. CO-DISEÑO OBLIGATORIO. No se validará ningún proyecto que no demuestre la participación de los beneficiarios en la etapa de diagnóstico y diseño de la solución técnica.

Referencias Bibliográficas

- Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we need. MIT Press.
- Couldry, N., & Mejias, U. A. (2019). The costs of connection: How data is colonizing human life and appropriating it for capitalism. Stanford University Press.
- Coraggio, J. L. (2011). Economía social y solidaria: El trabajo antes que el capital. Abya-Yala.
- Dagnino, R. (2014). Tecnología social: Herramienta para construir otra sociedad. Universidad Nacional de Quilmes.
- European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). <https://eur-lex.europa.eu>
- Fairclough, N. (2013). Critical discourse analysis: The critical study of language (2nd ed.). Routledge.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. En N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 105–117). Sage.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2018). Metodología de la investigación (6.^a ed.). McGraw-Hill.
- Janssen, M., & Helbig, N. (2019). Innovating and changing the policy-cycle: Policy-makers be prepared! Government Information Quarterly, 36(4), 101–115.

- Janssen, M., Brous, P., Estevez, E., Barbosa, L., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly*, 37(3), 101–493.
- Ordorika, I. (2006). *La disputa por el campus: Poder, política y autonomía en la UNAM*. UNAM.
- Tünnermann, C. (2008). *La universidad necesaria para el siglo XXI*. Editorial Hispamer.

CAPÍTULO XVIII

REGULACIÓN ACTIVA DEL INDIVIDUO FRENTE A LA INTELIGENCIA ARTIFICIAL: EL SUJETO COMO PRIMER AGENTE REGULADOR EN LA ERA DE LA IA GENERATIVA

Active regulation of the individual in the face of artificial intelligence: the subject as the first regulatory agent in the era of generative AI

Fernando Javier Vega Riveros

fer@flebo.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0001-8715-0770>

Resumen

El capítulo analiza la inteligencia artificial generativa desde una perspectiva crítica centrada en la relación entre tecnología, juicio humano y dependencia organizacional. Se sostiene que la regulación de la IA no puede reducirse exclusivamente a normas jurídicas, políticas institucionales o estándares corporativos, sino que comienza en la capacidad del individuo para ejercer pensamiento crítico, validar contextualmente los resultados y mantener autonomía frente a sistemas generativos. A partir de fundamentos teóricos provenientes de la filosofía, la teoría crítica y los estudios sobre tecnología, se distingue entre uso y consumo de herramientas digitales, entre producción automática y formación de criterio, así como entre información procesada y conocimiento situado. El análisis muestra cómo la adopción acrítica de IA puede generar erosión del criterio experto, debilitamiento del capital social y subordinación operativa a plataformas tecnológicas, aun cuando incremente la velocidad de producción. Asimismo, se examinan sectores como salud, educación y servicios financieros, donde la delegación de juicio tiene consecuencias materiales relevantes. Frente a ello, se propone la regulación activa del individuo como una práctica basada en validación

contextual, suspensión del juicio, responsabilidad sobre decisiones y subordinación instrumental de la tecnología a fines humanos. Finalmente, se argumenta que la regulación externa requiere sujetos capaces de interpretar, aplicar y cuestionar los sistemas de IA en contextos concretos, por lo que la formación crítica y la gobernanza organizacional se vuelven condiciones indispensables para una integración madura de la inteligencia artificial generativa.

Palabras clave: Inteligencia artificial generativa, regulación activa, pensamiento crítico, dependencia algorítmica, autonomía del juicio, validación contextual, gobernanza tecnológica, criterio experto, ética digital, organizaciones

Abstract

The chapter analyzes generative artificial intelligence from a critical perspective focused on the relationship between technology, human judgment, and organizational dependence. It is argued that the regulation of AI cannot be reduced exclusively to legal norms, institutional policies or corporate standards, but begins with the individual's ability to exercise critical thinking, contextually validate results and maintain autonomy in the face of generative systems. Based on theoretical foundations from philosophy, critical theory and technology studies, a distinction is made between the use and consumption of digital tools, between automatic production and the formation of criteria, as well as between processed information and situated knowledge. The analysis shows how the uncritical adoption of AI can lead to erosion of expert judgment, weakening of social capital, and operational subordination to technology platforms, even as it increases production speed. Likewise, sectors such as health, education and financial services are examined, where the delegation of judgment has relevant material consequences. In view of this, the active regulation of the individual is proposed as a practice based on contextual validation, suspension of judgment, responsibility for decisions and instrumental subordination of technology to human ends. Finally, it is argued that external regulation requires subjects capable of interpreting, applying and questioning AI systems in specific contexts, so critical training and organizational

governance become indispensable conditions for a mature integration of generative artificial intelligence.

Keywords: Generative artificial intelligence, active regulation, critical thinking, algorithmic dependence, autonomy of judgment, contextual validation, technological governance, expert judgment, digital ethics, organizations

INTRODUCCIÓN

La inteligencia artificial generativa ha sido presentada como una de las grandes tecnologías de transformación del siglo XXI. En el discurso público, empresarial y académico, suele aparecer asociada con eficiencia, innovación, automatización, democratización del conocimiento y aumento de productividad. La promesa parece clara: producir más, decidir más rápido, reducir tiempos operativos, ampliar capacidades humanas y acelerar procesos que antes dependían de estructuras lentas, costosas o excesivamente burocráticas. Sin embargo, detrás de esa narrativa se encuentra una tensión menos visible: la misma herramienta que promete liberar tiempo puede erosionar criterio; la misma tecnología que promete autonomía puede producir dependencia; la misma infraestructura que promete acceso al conocimiento puede concentrar poder en plataformas privadas que median, organizan y condicionan la forma en que se produce información.

Este capítulo parte de una preocupación concreta: la regulación de la inteligencia artificial no puede pensarse únicamente como una cuestión jurídica, institucional o técnica. Naturalmente, las leyes, los marcos normativos, las políticas públicas y los estándares corporativos son necesarios. Sería ingenuo negar su importancia. Pero también sería insuficiente creer que la regulación empieza y termina ahí. Antes de que una ley limite el uso de un sistema, antes de que una empresa redacte una política interna, antes de que un comité establezca criterios de implementación, existe un momento más elemental: el instante en que una persona decide cómo relacionarse con la herramienta. Ese momento no es menor. Ahí se juega la primera forma de regulación: la regulación activa del individuo.

Por regulación activa del individuo entiendo la capacidad del sujeto para mantener despierto su juicio, reconocer los límites de la herramienta, validar sus resultados con el contexto real, subordinar la tecnología a fines humanos y asumir responsabilidad sobre las decisiones que toma con apoyo de sistemas generativos. No se trata de rechazar la inteligencia artificial ni de romantizar un pasado sin automatización. Tampoco se trata de adoptar una postura ingenuamente tecnofóbica. El problema no es la existencia de la IA, sino la forma en que se incorpora a la vida organizacional y profesional cuando falta pensamiento crítico. La IA puede ampliar capacidades, pero también puede sustituir procesos reflexivos si se usa como atajo permanente. Puede apoyar el trabajo experto, pero también puede simular experiencia donde no existe. Puede acelerar la producción de documentos, imágenes, reportes o campañas, pero no garantiza que esos productos tengan sentido, valor o pertinencia.

La tesis central de este capítulo sostiene que la dependencia algorítmica no es inevitable. No surge de la tecnología por sí misma, sino de una adopción acrítica que confunde velocidad con productividad, plausibilidad con verdad, información con conocimiento y automatización con mejora organizacional. Frente a esa confusión, la primera respuesta regulatoria no proviene exclusivamente del Estado ni de la empresa, sino del individuo capaz de ejercer juicio. Esta afirmación no desplaza la necesidad de regulación externa; la antecede. Una organización puede tener políticas formales de uso responsable de IA y, aun así, operar de forma subordinada si sus miembros no han desarrollado criterio. Del mismo modo, una persona puede usar herramientas generativas de manera productiva sin caer en dependencia, siempre que conserve la facultad de contrastar, decidir y responder por aquello que produce.

El análisis se organiza en tres partes. La primera desarrolla los fundamentos teóricos que permiten distinguir entre uso y consumo, entre acción y ejecución, entre conocimiento tácito e información procesada. La segunda analiza la paradoja operativa que aparece en las organizaciones cuando la IA reduce tiempos visibles, pero puede erosionar capital social, transmisión de criterio y aprendizaje colectivo.

La tercera propone la regulación activa del individuo como una práctica concreta de responsabilidad, validación contextual y subordinación instrumental de la tecnología. El objetivo no es ofrecer una defensa abstracta de la autonomía humana, sino mostrar por qué esa autonomía se vuelve condición práctica para cualquier adopción madura de inteligencia artificial generativa.

I. FUNDAMENTOS TEÓRICOS PARA UNA RELACIÓN CRÍTICA CON LA IA

1.1. Más allá de las dicotomías baratas

Buena parte de la discusión contemporánea sobre inteligencia artificial cae en reducciones pobres. Se pregunta si la IA es buena o mala, si alguien está a favor o en contra, si representa progreso o amenaza. Estas oposiciones son atractivas porque simplifican una realidad incómoda; permiten tomar posición rápidamente, pero cancelan el análisis. La IA generativa no puede comprenderse desde un esquema binario. Es una tecnología con capacidades reales, intereses económicos concretos, limitaciones epistemológicas, impactos laborales, efectos culturales y consecuencias organizacionales. Reducirla a entusiasmo o rechazo empobrece la discusión.

El pensamiento crítico comienza cuando se resiste esa simplificación. No consiste en negar la utilidad de la herramienta ni en sospechar de todo avance técnico. Consiste en formular preguntas que el discurso dominante evita: eficiencia para quién? ¿a costa de qué proceso? ¿qué se gana cuando se automatiza? ¿qué se pierde al automatizar? ¿deja de discutir? ¿proceso? ¿capacidades se amplían? y ¿cuáles se debilitan? ¿qué decisiones siguen siendo humanas y cuáles se desplazan silenciosamente hacia la plataforma? Estas preguntas no buscan paralizar la adopción tecnológica. Buscan impedir que la adopción sea gobernada únicamente por ansiedad competitiva, presión de mercado o fascinación por la novedad.

Byung-Chul Han ha descrito la contemporaneidad como una infocracia: un régimen en el que el poder no opera principalmente mediante prohibición, sino mediante circulación, producción y saturación de información (Han, 2022). En ese contexto,

la IA generativa adquiere una función decisiva. No solo responde preguntas; produce discursos, organiza narrativas, propone significados y moldea expectativas. Si una organización adopta IA sin preguntarse por las condiciones de producción de esos discursos, corre el riesgo de asumir como natural lo que en realidad está mediado por modelos, datos, incentivos comerciales y decisiones de diseño.

La crítica, entonces, no es una actitud negativa. Es una forma de higiene intelectual. Permite distinguir entre valor real y simulación de valor. Una herramienta puede producir textos formalmente correctos, imágenes atractivas o reportes convincentes sin que ello implique comprensión, pertinencia o verdad. La forma puede mejorar mientras el juicio se debilita. Esta es una de las tensiones centrales de la IA generativa: la superficie del trabajo puede volverse más profesional al mismo tiempo que el proceso que formaba criterio se vuelve más pobre.

1.2. Dialéctica de la promesa tecnológica

La dialéctica permite observar la inteligencia artificial no como un objeto aislado, sino como una fuerza que entra en relación con prácticas previas, tensiones organizacionales e intereses económicos. En términos generales, el método dialéctico reconoce que la realidad se mueve por contradicciones, tensiones históricas y relaciones materiales que no pueden comprenderse de manera aislada (Hegel, 1807/1977; Marx, 1867/1982). Una tecnología no se suma simplemente a una organización; modifica relaciones, redistribuye poder, altera tiempos, redefine expectativas y produce nuevas dependencias. La introducción de IA no es una adición neutral. Es una intervención en una estructura social.

En el discurso empresarial, la IA aparece como antítesis de la lentitud: frente a procesos manuales, ofrece automatización; frente a documentos tardados, ofrece producción inmediata; frente a equipos saturados, ofrece asistencia permanente. Pero la síntesis no ocurre automáticamente. Para que exista una integración madura, la organización debe reconocer qué procesos pueden acelerarse sin pérdida sustantiva y cuáles requieren conversación, deliberación o experiencia

acumulada. Cuando esa distinción no se hace, la antítesis tecnológica no produce una síntesis superior; simplemente reemplaza procesos humanos por outputs rápidos.

Esta diferencia es esencial. No toda aceleración representa progreso. Hay procesos cuyo valor se encuentra precisamente en el tiempo que exigen. La discusión de un reporte, la construcción de una estrategia, la revisión de una campaña o la elaboración de un criterio clínico o financiero no son únicamente pasos para llegar a un documento. Son momentos de formación de juicio. Si la IA elimina el tiempo muerto sin eliminar el pensamiento, puede ser útil. Si elimina el pensamiento porque el output parece suficiente, produce empobrecimiento. La dialéctica ayuda a ver esa contradicción: una herramienta puede hacer más eficiente un proceso y, simultáneamente, debilitar las capacidades que hacían valioso ese proceso.

La pregunta relevante no es si la IA acelera. Acelera. La pregunta es qué tipo de aceleración produce. Una aceleración subordinada al juicio humano puede liberar tiempo para tareas de mayor valor. Una aceleración que sustituye el juicio puede volver a la organización más frágil, aunque sus indicadores superficiales parezcan mejorar. La regulación activa del individuo surge justamente como práctica capaz de distinguir una cosa de la otra.

1.3. Uso y consumo: una distinción necesaria

Una de las confusiones más importantes en la adopción de tecnologías digitales es la confusión entre uso y consumo. Consumir una plataforma significa dejarse conducir por su lógica: desplazarse por contenidos sugeridos, aceptar recomendaciones, seguir flujos diseñados por terceros, producir datos, pagar suscripciones y repetir patrones de interacción que responden más al interés de la plataforma que al propósito del usuario. Usar una herramienta, en cambio, implica subordinarla a un fin propio. El sujeto conserva dirección. Decide cuándo, cómo, para qué y con qué límites la utiliza.

Esta distinción es aplicable a la IA generativa. Una persona consume IA cuando le pide respuestas para evitar pensar, cuando acepta resultados sin validarlos, cuando

sustituye el diálogo con colegas por producción automática, cuando confunde la facilidad de obtener una respuesta con la calidad de haber elaborado una decisión. Una persona usa IA cuando la incorpora como instrumento de exploración, contraste, borrador, simulación o apoyo, sin entregar la responsabilidad del juicio. La diferencia no está en la herramienta, sino en la posición del sujeto frente a ella.

Ahora bien, afirmar que el sujeto decide no significa que la herramienta sea neutral. Esta advertencia coincide con la tesis de Winner, según la cual los artefactos tecnológicos no deben entenderse como instrumentos políticamente neutros, pues incorporan formas de organización, poder y orientación social (Winner, 2020). Las tecnologías tienen diseño, finalidad, interfaces, incentivos y modos preferentes de uso. Una plataforma de IA generativa no existe fuera de un modelo económico. Está construida para retener usuarios, producir valor comercial, generar dependencia de servicios, recopilar interacciones y posicionarse como infraestructura indispensable. Por eso no basta con decir que “todo depende de cómo se use”. También es necesario reconocer que la herramienta empuja ciertos usos, facilita ciertas prácticas e invisibiliza ciertos costos.

La regulación activa del individuo exige esta doble conciencia: reconocer la agencia personal y reconocer la orientación material de la tecnología. Si se exagera la agencia, se ignoran las estructuras de poder. Si se exagera la estructura, se niega la responsabilidad individual. El punto crítico está en sostener ambas dimensiones. La IA puede estar diseñada para el consumo, pero el usuario puede resistir esa lógica mediante prácticas de uso deliberado. Esa resistencia no es meramente moral; es operativa. Define si la herramienta será auxiliar de un proceso humano o si el proceso humano será reorganizado alrededor de la herramienta.

1.4. Autonomía del juicio y capacidad de decir no

Hannah Arendt distingue entre labor, trabajo y acción. La acción, en su sentido más alto, implica iniciar algo nuevo, intervenir en el mundo y asumir responsabilidad por las consecuencias de lo que se hace (Arendt, 1958). Esta idea resulta útil para pensar la IA porque permite distinguir entre ejecutar instrucciones y actuar con

juicio. Una persona que acepta sin examen el resultado de una máquina no actúa plenamente; ejecuta. Una persona que evalúa, corrige, rechaza o reorienta ese resultado ejerce acción.

La autonomía del juicio se expresa, ante todo, en la capacidad de decir no. No usar IA para cierto proceso. No aceptar un texto plausible. No automatizar una conversación que debe ocurrir entre personas. No publicar una campaña solo porque la imagen se ve profesional. No reducir una decisión estratégica a una síntesis generada por un sistema. Decir no no es rechazo tecnológico; es delimitación responsable. Toda herramienta madura requiere límites. La falta de límites convierte la utilidad en dependencia.

Esta autonomía es anterior a la regulación externa porque ninguna política puede prever todos los casos concretos. Una norma puede indicar que los resultados deben verificarse, pero solo el sujeto situado sabe qué significa verificar en su contexto. Una empresa puede establecer que la IA no debe sustituir decisiones críticas, pero es el profesional quien debe reconocer cuándo un proceso aparentemente operativo contiene juicio experto. Un marco legal puede exigir transparencia, pero la transparencia formal no garantiza comprensión real. El individuo sigue siendo el punto donde la norma se encarna o se vacía.

Amartya Sen permite ampliar esta idea desde la teoría de las capacidades. El desarrollo no consiste únicamente en tener más recursos, sino en ampliar libertades reales para actuar de acuerdo con fines valiosos (Sen, 1999). Bajo esa mirada, la IA no debería evaluarse solo por la cantidad de tareas que permite hacer, sino por las capacidades que fortalece o debilita. Si una herramienta aumenta la producción de documentos pero disminuye la capacidad de argumentar, discernir, aprender o responsabilizarse, su contribución al desarrollo humano es ambigua. Puede haber más recursos técnicos y menos agencia real.

1.5. Contexto de ventana y contexto de vida

Los sistemas generativos trabajan con ventanas de contexto: fragmentos de texto, instrucciones, documentos, ejemplos, datos o conversaciones que pueden

procesar para generar una respuesta. Esa capacidad ha aumentado de forma notable, pero no debe confundirse con el contexto de vida. El contexto humano es más amplio, dinámico y encarnado. Incluye historia organizacional, relaciones de poder, emociones, expectativas, conflictos, lealtades, tabúes, reglas no escritas, condiciones económicas, coyuntura política, cultura laboral y experiencia acumulada. Nada de eso entra de manera completa en una ventana de contexto.

La validación contextual consiste en contrastar el output de la IA con ese contexto vivo. La pregunta no es únicamente si el texto es coherente, sino si funciona aquí, ahora, para estas personas, con esta historia, bajo estas restricciones, frente a estos riesgos. Un correo puede estar bien redactado y ser políticamente inoportuno. Una campaña puede ser visualmente atractiva y contradecir la memoria de una marca. Un reporte puede ordenar datos correctamente y omitir una tensión interna que todos en la organización conocen. La IA puede producir forma; el contexto decide pertinencia.

Michael Polanyi formuló una idea decisiva: sabemos más de lo que podemos decir (Polanyi, 1966/2009). El conocimiento tácito no se agota en manuales, instrucciones ni bases de datos. Está incorporado en la experiencia. Se manifiesta en intuiciones, hábitos, sensibilidad para detectar riesgos, comprensión de personas y capacidad de anticipar consecuencias. Una organización que pretende “subir todo su contexto” a una IA confunde documentación con cultura. Los documentos ayudan, pero no sustituyen la experiencia vivida. La empresa no es solo lo que está escrito; es también lo que se practica, se calla, se negocia y se intuye.

La regulación activa exige proteger ese conocimiento tácito. No para conservarlo como privilegio opaco, sino para integrarlo conscientemente a los procesos de validación. La IA puede asistir en la sistematización de información, pero no debe reemplazar la conversación con quienes poseen experiencia contextual. Una decisión responsable combina procesamiento técnico con juicio situado. Cuando

esa combinación se rompe, la organización puede parecer más moderna y al mismo tiempo volverse menos inteligente.

1.6. Criterio experto y erosión silenciosa

El criterio experto no es solo acumulación de datos. Es la capacidad de ponderar variables visibles e invisibles para actuar adecuadamente en una situación concreta. Un experto sabe qué información importa, qué excepción debe considerarse, qué regla no escrita condiciona una decisión, qué antecedente puede alterar la interpretación de un resultado. En una organización, el criterio experto suele estar distribuido en personas que conocen la historia de los procesos, las relaciones internas, los errores previos y los límites reales de ejecución.

La IA generativa puede producir outputs que aparentan criterio. Redacta con seguridad, organiza ideas, ofrece recomendaciones, jerarquiza prioridades y simula una voz experta. Ahí reside parte de su poder y de su riesgo. Para un usuario sin experiencia, el resultado puede parecer suficiente. Para un profesional con criterio, el mismo resultado puede ser apenas un punto de partida. La diferencia no está en el texto generado, sino en la capacidad de evaluarlo.

La erosión del criterio experto ocurre cuando las organizaciones sustituyen procesos de aprendizaje por outputs inmediatos. Un colaborador junior puede producir en minutos una pieza que antes requería revisión, conversación y guía. Eso parece ganancia. Pero si el junior no entiende por qué se eligió una estrategia, por qué se descartó una alternativa, por qué cierto tono es inadecuado o por qué un cliente puede reaccionar negativamente, la organización no está formando talento. Está produciendo materiales. A corto plazo, la diferencia puede ser poco visible. A mediano plazo, aparece una dependencia: sin la IA o sin quien valide sus resultados, el equipo no sabe decidir.

Foucault recuerda que donde hay poder hay resistencia, y que la crítica es una práctica de libertad (Foucault, 1975, 1987). En el contexto de la IA, resistir no significa obstaculizar la tecnología. Significa impedir que la lógica de automatización vacíe las prácticas de formación, deliberación y responsabilidad. La

resistencia puede consistir en mantener espacios de discusión, exigir justificación de decisiones, documentar no solo resultados sino criterios, formar a los equipos en preguntas y no únicamente en prompts, o reservar ciertos procesos para conversación humana. La regulación activa es, en este sentido, una práctica de resistencia cotidiana.

1.7. Plausibilidad, conocimiento y suspensión del juicio

La IA generativa produce plausibilidad. Sus respuestas suelen sonar razonables, ordenadas y convincentes. Pero plausibilidad no equivale a verdad. Esta advertencia ha sido desarrollada en la crítica a los grandes modelos de lenguaje, donde se señala que estos sistemas pueden producir secuencias lingüísticas coherentes sin que ello implique comprensión, responsabilidad o conocimiento del mundo (Bender y otros., 2021). Un texto puede estar bien estructurado y contener errores; puede citar conceptos de manera superficial; puede formular una respuesta aceptable para un lector general e insuficiente para una decisión especializada. El riesgo aumenta porque la forma lingüística de la IA tiende a producir confianza. La seguridad del tono puede encubrir la fragilidad del contenido.

Por eso la suspensión del juicio es una práctica central. Suspender el juicio no significa carecer de postura. Significa impedir que la primera impresión gobierne la decisión. Ante un output generado por IA, el sujeto debe convertir la respuesta en objeto de análisis: ¿qué afirma?, ¿qué presupone?, ¿qué omite?, ¿qué evidencia requiere?, ¿qué sesgo puede contener?, ¿qué consecuencias tendría aplicarla?, ¿qué parte depende de información que la máquina no tiene? Esta pausa es regulatoria. Introduce una distancia entre producción automática y acción humana.

Gadamer permite comprender esta práctica desde la hermenéutica. La comprensión ocurre en una fusión de horizontes, donde el intérprete reconoce sus propios prejuicios y se deja interpelar por el texto sin desaparecer en él (Gadamer, 1975). Frente a la IA, esta idea adquiere una dimensión práctica: el usuario no debe proyectar inmediatamente sus deseos sobre el output ni aceptarlo como autoridad

externa. Debe dialogar con él. La respuesta generada es un interlocutor limitado, no un juez final.

También es necesario distinguir datos, información y conocimiento. La IA procesa datos y produce información organizada. El conocimiento, en cambio, implica una relación viva entre sujeto, objeto, experiencia, interpretación y responsabilidad. Una máquina puede generar una explicación sobre una canción, una enfermedad, una estrategia o un conflicto laboral. Pero entender esas realidades supone historia, sensibilidad, experiencia y consecuencias. El conocimiento no se reduce a la disponibilidad de enunciados correctos. Incluye saber cuándo aplicar, cuándo no aplicar, cuándo preguntar más y cuándo detenerse.

II. ANÁLISIS EMPÍRICO: LA PARADOJA EN LAS ORGANIZACIONES

2.1. Velocidad sin productividad real

En la práctica organizacional, la promesa más visible de la IA generativa es la velocidad. Un documento que antes tomaba tres horas puede producirse en una. Una campaña que requería varios borradores puede generarse en minutos. Un correo, una presentación, un reporte o una síntesis pueden aparecer casi de inmediato. Esta capacidad es real y no debe negarse. La IA efectivamente reduce tiempos de producción en muchas tareas. El problema aparece cuando esa reducción se interpreta automáticamente como productividad.

Productividad no significa únicamente producir más en menos tiempo. Significa generar valor de manera sostenible. En una organización, el valor no reside solo en el output final, sino en el proceso que permite hacerlo pertinente: conversación, aprendizaje, discusión, transmisión de criterio, corrección colectiva, comprensión de objetivos, desarrollo de capacidades y apropiación del sentido del trabajo. Cuando un proceso se acelera, es necesario preguntar qué elementos se conservaron y cuáles se perdieron.

Un ejemplo permite observar la paradoja. En una pequeña o mediana empresa, elaborar un documento estratégico podía requerir varias horas de intercambio.

Durante ese tiempo, el responsable explicaba al equipo por qué se tomaba cierta decisión, qué antecedentes importaban, qué debía evitarse, qué tono era adecuado y qué implicaciones tendría cada alternativa. El documento final era importante, pero también lo era la conversación que lo producía. Ahí se formaba criterio. Ahí se transmitía cultura organizacional. Ahí se aprendía a pensar dentro de una práctica.

Con IA, el proceso puede comprimirse. El responsable escribe un prompt, obtiene un borrador, corrige, ajusta y envía. El documento sale más rápido. Sin embargo, el equipo puede quedar fuera del razonamiento. No escucha la explicación, no pregunta, no observa las dudas, no comprende los criterios de selección. La organización gana un documento y pierde una oportunidad de aprendizaje. Si esto ocurre ocasionalmente, puede ser una optimización razonable. Si se vuelve norma, produce una erosión silenciosa de capacidades internas.

Esta distinción es central para evitar una lectura ingenua de la eficiencia. La IA puede hacer más liviana una tarea sin mejorar la organización. Incluso puede permitir producir más piezas de bajo valor, más reportes que nadie discute, más campañas que no nacen de una estrategia, más correos que no resuelven tensiones. La métrica visible aumenta; el valor real no necesariamente. En ese punto, la velocidad se vuelve una forma de simulación productiva.

2.2. Lo que no se puede codificar

Las organizaciones suelen imaginarse como estructuras formales: organigramas, procesos, manuales, misiones, visiones, políticas, indicadores. Pero su funcionamiento real depende de elementos menos visibles: confianza, lealtades, historia compartida, conflictos, reputaciones, aprendizajes informales, acuerdos tácitos y sensibilidades culturales. La IA puede procesar documentos sobre una empresa, pero no vive su cultura. Puede resumir políticas internas, pero no capta por completo cómo se aplican, cuándo se incumplen, quién tiene autoridad informal o qué temas generan resistencia.

Este punto es especialmente relevante en pequeñas y medianas empresas, donde los procesos suelen depender de personas concretas y relaciones directas. En esos

entornos, el conocimiento no está completamente institucionalizado. Se transmite en conversaciones, reuniones, correcciones, gestos y experiencias. Un colaborador aprende no solo por leer un manual, sino por observar cómo decide alguien con experiencia. Aprende qué se prioriza realmente, qué cliente requiere trato especial, qué problema se repite, qué tipo de solución ha fallado antes.

Cuando se intenta sustituir ese contexto por documentos de instrucciones para IA, se incurre en una reducción. Sistematizar es útil, pero no equivale a comprender. Un archivo puede decir cómo debe hablar la marca; una persona con experiencia sabe cuándo conviene romper ese tono. Un manual puede señalar un procedimiento; un experto sabe cuándo el procedimiento es insuficiente por una condición excepcional. Una base de datos puede registrar antecedentes; la memoria organizacional interpreta su significado.

La IA puede ayudar a codificar parte de ese conocimiento. Puede ordenar información dispersa, detectar patrones, generar manuales, crear matrices de decisión o apoyar procesos de documentación. Pero debe hacerlo al servicio de la organización, no como sustituto de su vida interna. Si la empresa cree que todo su contexto puede convertirse en instrucciones, corre el riesgo de empobrecer su propia comprensión. La cultura no es un prompt largo. La cultura es una práctica social.

2.3. Subordinación sistemática a plataformas

La subordinación a la IA no ocurre solo cuando una máquina toma decisiones autónomas. Ocurre de manera más cotidiana cuando la organización reorganiza su trabajo alrededor de herramientas que empiezan a dictar ritmos, formatos, expectativas y criterios. La pregunta deja de ser “¿qué necesita nuestro cliente?” y se convierte en “¿qué podemos producir con esta herramienta?”. El medio ocupa el lugar del fin.

La subordinación también aparece en la proliferación de suscripciones y plataformas. Cada problema parece exigir una nueva aplicación: una para redactar, otra para diseñar, otra para resumir, otra para automatizar, otra para analizar datos,

otra para atender clientes. Esta acumulación puede crear una dependencia financiera y operativa. Si la organización deja de pagar, ciertos procesos se detienen. Si la plataforma cambia sus condiciones, el flujo interno se altera. Si la herramienta modifica su modelo, la empresa debe adaptarse. En apariencia, la IA aumenta capacidad; en realidad, puede aumentar vulnerabilidad.

El problema no es pagar herramientas. Toda organización utiliza infraestructura externa. El problema es no distinguir entre infraestructura que amplía capacidades e infraestructura que captura procesos esenciales sin transferencia de conocimiento. Una herramienta bien incorporada debe dejar a la organización más capaz, no más incapaz sin ella. Debe ayudar a desarrollar criterio, no reemplazarlo. Debe apoyar procesos de valor, no convertir la producción de outputs en fin en sí mismo.

La subordinación se evidencia cuando los miembros de la organización ya no pueden explicar por qué se tomó una decisión, solo mostrar qué generó la IA. También cuando se publican materiales sin discusión porque “se ven bien”. O cuando la estrategia se reduce a pedir ideas a un modelo sin contrastarlas con mercado, experiencia, historia de marca o realidad operativa. En esos casos, la plataforma no es herramienta; es autoridad práctica.

2.4. Salud, educación y servicios financieros: sectores donde el juicio importa

La preocupación por la regulación activa se vuelve más grave en sectores donde el juicio humano tiene consecuencias materiales directas. En salud, una recomendación mal validada puede afectar decisiones clínicas, comunicación con pacientes o interpretación de información médica. Aunque la IA pueda apoyar educación, redacción, orientación o administración, no debe sustituir la responsabilidad profesional. La plausibilidad de una explicación médica puede ser peligrosa si el usuario no distingue entre información general y decisión clínica situada.

En educación, la IA puede facilitar planeación, retroalimentación, generación de materiales y acompañamiento del aprendizaje. Pero también puede erosionar

procesos formativos si estudiantes y docentes sustituyen reflexión por respuesta automática. El objetivo educativo no es producir textos correctos, sino formar capacidades de comprensión, argumentación, lectura crítica y juicio. Si la IA se usa para evitar esas capacidades, la institución puede parecer más eficiente mientras forma menos.

En servicios financieros, la tensión es igualmente relevante. Decisiones sobre crédito, inversión, patrimonio o riesgo requieren interpretación contextual, prudencia y responsabilidad. Un análisis generado por IA puede ordenar información, pero no debe reemplazar la comprensión de objetivos del cliente, horizonte temporal, tolerancia al riesgo, situación familiar, condiciones legales o coyuntura económica. En estos campos, la regulación activa no es un ideal abstracto; es una condición de diligencia profesional.

Estos sectores muestran que el problema no está en usar IA, sino en delegar juicio. La herramienta puede apoyar tareas administrativas, generar borradores, traducir información, simular escenarios o mejorar comunicación. Pero la decisión responsable exige validación humana. Cuanto mayor es el impacto de una decisión sobre la vida de otros, mayor debe ser la exigencia de criterio.

2.5. La erosión del capital social

Además del criterio individual, la adopción acrítica de IA puede erosionar capital social dentro de las organizaciones. El capital social está compuesto por confianza, cooperación, comunicación informal, reconocimiento mutuo y sentido de pertenencia. No aparece fácilmente en indicadores, pero sostiene buena parte del funcionamiento real de una empresa. Cuando los procesos se automatizan sin cuidado, disminuyen espacios de interacción donde se construyen esos vínculos.

La conversación laboral no es pérdida de tiempo por definición. En muchas ocasiones es el lugar donde se detectan problemas, se corrigen malentendidos, se transmiten aprendizajes y se construye identidad colectiva. Si todo se reduce a mensajes generados, documentos automáticos y flujos impersonales, la

organización puede volverse más rápida y menos cohesionada. La eficiencia técnica no compensa necesariamente la pérdida de confianza.

Esto no significa defender reuniones innecesarias ni burocracia. Significa distinguir entre tiempo improductivo y tiempo formativo. Hay conversaciones que deben eliminarse porque no agregan valor. Pero hay conversaciones que son el valor. Un equipo que entiende por qué trabaja de cierta manera puede adaptarse mejor que un equipo que solo recibe outputs. La regulación activa también implica proteger deliberadamente espacios de pensamiento compartido.

III. REGULACIÓN ACTIVA DEL INDIVIDUO

3.1. Definición y alcance

La regulación activa del individuo puede definirse como el conjunto de prácticas mediante las cuales una persona mantiene su autonomía de juicio frente a sistemas de inteligencia artificial generativa. Incluye pensamiento crítico, validación contextual, suspensión del juicio, conciencia de límites, responsabilidad sobre decisiones, protección del proceso formativo y subordinación de la herramienta a fines humanos. No es una política escrita, aunque puede inspirar políticas. Es una disposición práctica.

Esta regulación se llama activa porque no consiste en obedecer pasivamente una norma externa. Requiere intervención constante. Cada vez que se usa IA, el sujeto decide si acepta, corrige, rechaza, pregunta más, contrasta fuentes, consulta a otros, contextualiza o detiene el proceso. La herramienta genera; el sujeto regula. La máquina propone; el sujeto responde. Sin esa intervención, el uso se convierte en consumo y la asistencia en subordinación.

El alcance de esta propuesta es deliberadamente individual, pero no individualista. El individuo no aparece como sujeto aislado, sino como primer punto de responsabilidad dentro de redes organizacionales, institucionales y sociales. Una empresa puede fomentar o impedir esta regulación. Una cultura laboral puede premiar la validación o castigar la pausa. Un mercado puede empujar hacia

velocidad acrítica. Aun así, la decisión concreta de no delegar el juicio ocurre en la persona que usa la herramienta.

La regulación activa tampoco sustituye la regulación jurídica. La complementa desde abajo. Una ley puede establecer obligaciones de transparencia, privacidad, seguridad o no discriminación. Pero esas obligaciones requieren sujetos capaces de reconocer riesgos en situaciones concretas. Sin criterio individual, la norma se vuelve cumplimiento formal. Con criterio individual, la norma puede convertirse en práctica viva.

3.2. Pensamiento crítico como práctica cotidiana

El pensamiento crítico debe dejar de entenderse como adorno académico. En la era de la IA generativa, es una competencia operativa. Una persona que no sabe preguntar, contrastar, argumentar o detectar supuestos queda expuesta a la plausibilidad del modelo. Puede recibir una respuesta incorrecta y aceptarla porque suena profesional. Puede publicar una idea superficial porque está bien redactada. Puede tomar una decisión pobre porque la herramienta organizó la información de manera convincente.

Por eso la formación en IA no debería limitarse a enseñar prompts. Enseñar a escribir instrucciones útiles es necesario, pero insuficiente. También deben enseñarse criterios de validación, análisis de sesgos, distinción entre fuentes, pensamiento lógico, estructura argumentativa, lectura crítica y responsabilidad profesional. Una persona que sabe usar prompts pero no sabe pensar queda técnicamente habilitada y epistemológicamente vulnerable.

La práctica cotidiana del pensamiento crítico puede traducirse en preguntas simples: ¿qué problema intento resolver?, ¿por qué necesito IA para esto?, ¿qué parte del proceso no debo automatizar?, ¿qué información le falta al modelo?, ¿quién debe validar el resultado?, ¿qué consecuencias tendría equivocarme?, ¿qué estoy dejando de aprender si delego esta tarea?, ¿el output mejora realmente el valor para el cliente, paciente, estudiante o usuario? Estas preguntas funcionan como microregulaciones. Interrumpen la inercia.

La regulación activa se vuelve más sólida cuando estas preguntas se institucionalizan como hábitos. No se trata de convertir cada uso de IA en un ritual pesado, sino de crear una cultura de validación proporcional al riesgo. Para tareas menores, bastará una revisión básica. Para decisiones sensibles, se requerirá revisión experta, contraste documental y deliberación humana. El nivel de regulación debe corresponder al impacto de la decisión.

3.3. Validación contextual como método

La validación contextual puede organizarse como método en tres momentos. Primero, identificación del tipo de output: texto, recomendación, análisis, síntesis, imagen, decisión sugerida, clasificación o explicación. No todos los outputs tienen el mismo riesgo. Un borrador creativo no exige la misma validación que una recomendación clínica o financiera. Segundo, contraste con el contexto: condiciones reales, personas involucradas, antecedentes, restricciones, cultura organizacional, datos disponibles y consecuencias previsibles. Tercero, decisión responsable: aceptar, modificar, rechazar o someter a revisión.

Este método devuelve al sujeto el control del proceso. La IA deja de ser una fuente de verdad y se convierte en material de trabajo. Un texto generado no se publica; se edita. Una recomendación no se obedece; se evalúa. Una síntesis no sustituye la lectura; la orienta. Una estrategia no se adopta por sonar inteligente; se confronta con mercado, recursos, capacidades y fines.

La validación contextual también exige reconocer cuándo no se sabe. Una de las formas más peligrosas de dependencia algorítmica ocurre cuando el usuario usa IA para cubrir su falta de conocimiento sin buscar validación externa. La herramienta puede generar seguridad artificial. Frente a temas técnicos, legales, médicos, financieros o académicos especializados, la regulación activa implica humildad epistemológica: saber que un output plausible no convierte al usuario en experto.

La validación no debe confundirse con desconfianza absoluta. Si cada resultado se rechaza por principio, la herramienta pierde utilidad. La regulación activa busca una confianza calibrada. Se puede confiar en la IA para tareas de apoyo, exploración o

estructuración, pero no para sustituir juicio en áreas donde la experiencia, la evidencia y la responsabilidad son indispensables. La confianza madura no es ciega ni paranoica; es proporcional.

3.4. Proceso sobre resultado

Uno de los criterios más importantes para regular el uso de IA es preguntar si el valor de una tarea está en el resultado o en el proceso. Hay tareas donde el resultado importa más que el camino: limpiar formato, ordenar una lista, resumir información de baja complejidad, generar variaciones de estilo o convertir datos en una tabla. En esos casos, automatizar puede ser razonable. Hay otras tareas donde el proceso es parte del valor: discutir una estrategia, formar criterio, enseñar a escribir, interpretar un caso, construir postura ética, evaluar riesgos o decidir una línea de acción. Automatizar ahí puede empobrecer.

Esta distinción permite evitar un error común: creer que todo aquello que puede acelerarse debe acelerarse. La posibilidad técnica no equivale a conveniencia práctica. Una clase puede resumirse automáticamente, pero el aprendizaje exige esfuerzo de comprensión. Una reflexión puede redactarse con IA, pero la formación ética exige confrontación interna. Un reporte puede generarse en segundos, pero la capacidad directiva se forma al decidir qué medir, cómo interpretar y qué hacer con los hallazgos.

El proceso importa porque produce sujetos capaces. Una organización no solo necesita documentos; necesita personas que entiendan por qué esos documentos existen. No solo necesita campañas; necesita equipos que comprendan al cliente. No solo necesita reportes; necesita directivos capaces de interpretar. Cuando la IA sustituye el proceso, el resultado puede permanecer mientras la capacidad humana se degrada. La empresa se vuelve dependiente de outputs externos porque dejó de formar juicio interno.

La regulación activa, por tanto, no siempre consiste en usar IA mejor. A veces consiste en no usarla. O en usarla después de una primera reflexión humana. O en pedir al equipo que argumente antes de generar. O en utilizarla como contraparte

crítica después de haber construido una postura propia. El orden importa. Si la IA piensa primero, el sujeto puede limitarse a reaccionar. Si el sujeto piensa primero, la IA puede ampliar, contrastar o desafiar.

3.5. Subordinar la herramienta a fines humanos

La pregunta central no es qué puede hacer la IA, sino qué queremos hacer nosotros con ella. Esta inversión es decisiva. La lógica de mercado tiende a partir de la capacidad de la herramienta: como puede redactar, redactemos más; como puede generar imágenes, generemos más; como puede automatizar, automaticemos más. La regulación activa parte de los fines: ¿qué valor buscamos producir?, ¿qué problema real queremos resolver?, ¿qué experiencia humana queremos mejorar?, ¿qué capacidad organizacional queremos fortalecer?

Subordinar la herramienta significa que la IA debe servir al propósito, no definirlo. En una empresa, el propósito puede ser mejorar la experiencia del cliente, reducir errores administrativos, liberar tiempo para atención humana, fortalecer análisis directivo o facilitar aprendizaje interno. Si la herramienta contribuye a eso, tiene sentido. Si solo aumenta la cantidad de outputs sin mejorar valor, puede ser ruido productivo. La abundancia de contenido no equivale a estrategia.

Este principio tiene consecuencias prácticas. Antes de implementar IA, una organización debería identificar procesos donde existe dolor real, costo claro, oportunidad de mejora y criterio humano disponible para validar. No basta con adoptar herramientas porque están de moda. Tampoco basta con capacitar al personal en funciones generales. La implementación madura requiere mapa de procesos, evaluación de riesgos, selección de casos de uso, responsables de validación y criterios de éxito. La IA debe entrar donde aporta valor verificable.

A nivel individual, subordinar la herramienta exige intención. El usuario debe poder explicar por qué la usa. Si la respuesta es “porque es más rápido”, falta análisis. La velocidad puede ser una razón, pero no la única. Debe preguntarse qué se hará con el tiempo liberado, qué capacidades se conservarán, qué riesgos se introducen

y qué criterios se usarán para validar. Sin esa reflexión, la velocidad se convierte en consumo.

3.6. Responsabilidad y trazabilidad del juicio

Una práctica regulatoria clave consiste en conservar trazabilidad del juicio humano. No basta con guardar el output. Debe registrarse, al menos en procesos relevantes, qué se pidió, qué se aceptó, qué se modificó, quién validó y bajo qué criterio. Esta trazabilidad no tiene que ser burocrática en todos los casos, pero sí proporcional al riesgo. En decisiones sensibles, permite reconstruir responsabilidad.

La trazabilidad evita una forma frecuente de evasión: atribuir a la IA errores que en realidad son humanos. La máquina no responde moralmente por publicar información incorrecta, tomar una decisión imprudente o excluir una variable crítica. Quien usa, aprueba o implementa el resultado conserva responsabilidad. La regulación activa implica asumir esa responsabilidad sin esconderse detrás de la herramienta.

También permite aprender. Si una recomendación generada falla, la organización puede revisar si el problema estuvo en el prompt, en los datos, en la falta de validación, en un sesgo del modelo o en una interpretación humana deficiente. Sin trazabilidad, el error se vuelve opaco. Con trazabilidad, se convierte en aprendizaje institucional.

3.7. Formación humana para la era algorítmica

Si la IA generativa se vuelve parte de la vida profesional, la formación no debe centrarse únicamente en habilidades técnicas. Las organizaciones necesitan formar juicio. Eso incluye lectura crítica, escritura argumentativa, lógica básica, ética profesional, comprensión de sesgos, análisis de contexto, deliberación, interpretación de datos y capacidad de tomar decisiones bajo incertidumbre. Paradójicamente, cuanto más potente se vuelve la herramienta, más importante se vuelve la formación humana.

Capacitar en IA sin capacitar en pensamiento crítico puede producir usuarios veloces pero dependientes. Sabrán pedir, pero no evaluar. Sabrán generar, pero

no decidir. Sabrán automatizar, pero no comprender. La alfabetización algorítmica debe integrar tres dimensiones: técnica, epistemológica y ética. La técnica enseña cómo funciona y cómo se opera. La epistemológica enseña qué tipo de conocimiento produce y cuáles son sus límites. La ética enseña qué responsabilidades surgen al usarla.

En organizaciones pequeñas y medianas, esta formación puede ser sencilla pero constante. Reuniones de revisión de casos, análisis de errores, discusión de outputs, protocolos de validación y espacios de reflexión pueden ser suficientes para empezar. No se requiere convertir a todos en filósofos de la tecnología, pero sí evitar que la IA se convierta en una autoridad incuestionada. El objetivo no es miedo; es madurez.

3.8. Hacia una cultura de regulación distribuida

La regulación activa del individuo puede escalar hacia una cultura organizacional. Cuando varias personas practican validación, suspensión del juicio y responsabilidad, la organización desarrolla regulación distribuida. No depende solo de un comité o de una política, sino de hábitos compartidos. Cada miembro se vuelve un punto de control, no para bloquear, sino para mejorar.

Una cultura de regulación distribuida debe evitar dos extremos. El primero es el entusiasmo ingenuo: adoptar IA en todo, sin distinguir riesgos ni procesos. El segundo es la prohibición defensiva: impedir usos útiles por miedo o desconocimiento. Entre ambos extremos existe una posición más rigurosa: experimentar con criterios, medir valor real, proteger procesos formativos y ajustar continuamente. La regulación activa favorece esa posición intermedia.

Esta cultura también debe reconocer jerarquías de riesgo. No todos los usos requieren la misma supervisión. Generar ideas para una publicación interna no equivale a producir una recomendación médica, financiera o jurídica. La organización debe clasificar usos, definir responsables y establecer niveles de validación. El individuo regula en cada interacción, pero la empresa debe crear

condiciones para que esa regulación sea posible y no sea castigada por “ir más lento”.

IV. IMPLICACIONES PARA UNA REGULACIÓN MÁS AMPLIA

4.1. La regulación externa necesita sujetos reguladores

La propuesta de regulación activa del individuo no debe leerse como sustitución de la regulación pública. Sería un error desplazar toda la responsabilidad al usuario cuando existen empresas, Estados, mercados e infraestructuras que condicionan materialmente el uso de IA. La concentración de poder tecnológico, la opacidad de ciertos modelos, el uso de datos personales, la posible discriminación algorítmica y la dependencia de plataformas globales requieren marcos normativos robustos, como el Reglamento Europeo de Inteligencia Artificial y las recomendaciones internacionales sobre ética, derechos humanos, transparencia y supervisión humana (European Parliament and Council of the European Union, 2024; UNESCO, 2022). No obstante, dichos marcos serán insuficientes si se diseñan como estructuras puramente externas, desconectadas de las prácticas cotidianas de quienes usan la tecnología.

Toda regulación necesita sujetos capaces de interpretarla y aplicarla. Una política de uso responsable puede establecer que los resultados de la IA deben ser revisados por una persona competente. Pero esa regla no produce, por sí misma, competencia. Una ley puede exigir transparencia en sistemas automatizados. Pero la transparencia formal no garantiza que los usuarios comprendan las implicaciones de un modelo o que sepan detectar cuándo un resultado debe ser cuestionado. Una empresa puede prohibir el uso de IA en ciertas tareas críticas. Pero si su cultura premia únicamente rapidez, los empleados encontrarán formas informales de usarla para cumplir expectativas imposibles.

Por eso la regulación activa tiene una función mediadora. Conecta la norma externa con la práctica real. Traduce principios generales en decisiones situadas. Permite que conceptos como responsabilidad, transparencia, seguridad, explicabilidad o

supervisión humana no se queden en lenguaje corporativo, sino que orienten acciones concretas. La supervisión humana, por ejemplo, no significa que una persona vea rápidamente un resultado antes de aprobarlo. Significa que esa persona tiene conocimiento suficiente, tiempo suficiente y autoridad suficiente para rechazarlo o modificarlo. Si no cuenta con esas condiciones, la supervisión humana se convierte en simulacro.

La regulación externa debe reconocer esta dimensión. No basta con exigir que haya “human in the loop” si el humano en el circuito carece de criterio, está presionado por tiempos imposibles o no entiende el sistema que supuestamente supervisa. La presencia humana puede ser meramente decorativa. La regulación activa exige que esa presencia sea sustantiva: capacidad real de intervención, comprensión proporcional del riesgo y responsabilidad efectiva. En este sentido, la supervisión humana no debe entenderse como una presencia meramente formal, sino como una capacidad efectiva de gestión del riesgo. Esta exigencia es coherente con los enfoques contemporáneos de gobernanza de IA, que insisten en que la confianza, la seguridad, la explicabilidad y la rendición de cuentas requieren prácticas organizacionales verificables, no solo declaraciones de principio (National Institute of Standards and Technology, 2023).

4.2. Gobernanza organizacional de la IA

Las organizaciones que buscan implementar IA de manera madura deberían construir una gobernanza interna que no se reduzca a permisos y prohibiciones. Esta gobernanza puede dialogar con marcos técnicos contemporáneos como el AI Risk Management Framework del NIST, que propone gestionar riesgos, responsabilidades, trazabilidad y confiabilidad en el uso organizacional de sistemas de inteligencia artificial (National Institute of Standards and Technology, 2023). La gobernanza debe responder cuatro preguntas: para qué se usará la IA, quién valida sus resultados, qué riesgos se aceptan y qué capacidades humanas se protegerán. Sin estas preguntas, la implementación suele seguir una lógica desordenada: cada área adopta herramientas por separado, se multiplican suscripciones, se generan outputs sin trazabilidad y nadie evalúa si la organización realmente mejora.

Una gobernanza mínima puede incluir clasificación de usos. Los usos de bajo riesgo incluyen tareas como corrección de estilo, generación de ideas iniciales, organización de notas o apoyo administrativo. Los usos de riesgo medio incluyen análisis de información interna, redacción de comunicaciones sensibles, apoyo en decisiones comerciales o elaboración de materiales educativos. Los usos de alto riesgo incluyen decisiones clínicas, financieras, jurídicas, laborales, académicas evaluativas o cualquier proceso que afecte derechos, salud, patrimonio o reputación de personas. Cada nivel requiere validación distinta.

También se requiere definir responsables. Una organización no puede delegar la responsabilidad en la herramienta ni en “el área de tecnología” cuando las decisiones pertenecen a áreas sustantivas. Si se usa IA para comunicación institucional, el área responsable de comunicación debe validar tono, pertinencia y riesgo reputacional. Si se usa en educación, docentes o responsables académicos deben validar calidad formativa. Si se usa en salud, profesionales competentes deben delimitar estrictamente usos permitidos y prohibidos. La IA no elimina responsabilidad funcional; la vuelve más exigente.

La gobernanza también debe proteger el aprendizaje interno. Un criterio recomendable es identificar procesos que no deben automatizarse completamente porque tienen valor formativo. Por ejemplo, la elaboración de diagnósticos estratégicos, la revisión de errores, la discusión de casos difíciles, la construcción de posturas institucionales o la formación de nuevos colaboradores. La IA puede apoyar esos procesos, pero no sustituirlos. Una organización que automatiza todo aquello que enseña termina debilitando su propia capacidad futura.

Finalmente, la gobernanza debe incluir evaluación de valor real. No basta con medir cuántos documentos se producen o cuánto tiempo se ahorra. Deben observarse indicadores cualitativos y estratégicos: calidad de decisiones, satisfacción de usuarios, reducción de errores, aprendizaje del equipo, claridad de procesos, disminución de retrabajos, fortalecimiento de criterio y mejora en experiencia del

cliente o paciente. Si la IA aumenta volumen pero no mejora valor, la implementación debe corregirse.

4.3. Criterios prácticos para decidir cuándo usar IA

Una forma concreta de aplicar la regulación activa consiste en establecer criterios previos de decisión. Antes de usar IA, el individuo puede preguntarse: ¿esta tarea requiere juicio experto?, ¿el resultado afectará a otras personas?, ¿hay riesgo de daño si el output es incorrecto?, ¿cuento con conocimiento suficiente para validarlo?, ¿la IA me ayudará a pensar mejor o solo evitará que piense?, ¿el proceso tiene valor formativo?, ¿es necesario documentar cómo se llegó al resultado?

Si la tarea no requiere juicio especializado, tiene bajo impacto y el usuario puede revisar fácilmente el resultado, el uso de IA puede ser directo. Si la tarea requiere interpretación contextual, afecta a terceros o implica riesgo reputacional, financiero, académico, legal o clínico, el uso debe ser asistido y validado. Si la tarea exige deliberación ética, formación de criterio o responsabilidad profesional alta, la IA debe ocupar un lugar secundario: apoyo, contraste o simulación, nunca decisión final.

También conviene distinguir entre IA como generador, IA como crítico e IA como espejo. Como generador, produce borradores, ideas o estructuras iniciales. Como crítico, ayuda a detectar inconsistencias, contraargumentos, riesgos u omisiones. Como espejo, permite al usuario observar su propia postura al confrontarla con otra formulación. Estos tres usos no tienen el mismo valor. El uso generador puede facilitar, pero también fomentar dependencia. El uso crítico suele ser más formativo, porque obliga al sujeto a defender, corregir o profundizar su pensamiento. El uso como espejo puede ser especialmente útil en escritura, investigación y estrategia, siempre que el sujeto mantenga autoría intelectual.

Un principio operativo sería: no pedir a la IA que reemplace aquello que se necesita aprender. Si un estudiante debe aprender a argumentar, usar IA para entregar un ensayo completo evita el aprendizaje. Si un directivo necesita comprender

indicadores, pedir un resumen automático puede ser insuficiente. Si un médico necesita explicar un procedimiento a un paciente, usar IA para mejorar claridad puede ser útil, pero delegar la comprensión clínica sería irresponsable. La herramienta debe apoyar el desarrollo de capacidades, no encubrir su ausencia.

4.4. Riesgos de una falsa democratización

Uno de los argumentos más frecuentes a favor de la IA generativa es que democratiza el acceso a capacidades antes reservadas a expertos: escribir mejor, diseñar materiales, analizar datos, programar, traducir o aprender temas complejos. Esta afirmación contiene una parte de verdad. Muchas personas pueden producir resultados iniciales que antes eran inaccesibles. Sin embargo, la democratización de la producción no equivale a democratización del criterio. Poder generar un documento jurídico no convierte al usuario en abogado. Poder producir una explicación médica no convierte al usuario en médico. Poder elaborar un análisis financiero no implica comprender riesgo.

La falsa democratización aparece cuando el acceso a outputs se confunde con adquisición de competencia. La IA puede reducir barreras de entrada, pero también puede crear ilusiones de dominio. El usuario ve un resultado bien redactado y cree poseer una capacidad que en realidad pertenece al sistema. Esta ilusión es peligrosa porque dificulta reconocer límites. El verdadero acceso al conocimiento exige formación, práctica, corrección, error y experiencia. La IA puede acompañar ese proceso, pero no sustituirlo.

En términos sociales, esta falsa democratización puede profundizar desigualdades. Quienes ya poseen criterio experto usarán IA para ampliar su capacidad. Quienes carecen de criterio pueden depender más de outputs que no saben evaluar. La misma herramienta puede empoderar a unos y subordinar a otros. Por eso la regulación activa tiene una dimensión educativa y política. No basta con dar acceso a tecnología; hay que formar capacidad de juicio para que el acceso no se convierta en dependencia.

4.5. Autoría, responsabilidad y voz propia

La IA generativa también transforma la relación entre autoría y producción. Un texto puede estar formalmente escrito por una persona y materialmente generado en gran parte por una máquina. Esta situación obliga a distinguir entre producción lingüística y autoría intelectual. La autoría no se reduce a teclear palabras. Implica postura, selección, responsabilidad, criterio y capacidad de responder por lo dicho. Una persona puede apoyarse en IA para redactar y seguir siendo autora si conserva dirección conceptual, valida argumentos, corrige sentido y asume responsabilidad. Pero si acepta el texto sin comprenderlo, su autoría se debilita.

En trabajos académicos, profesionales o institucionales, este punto es decisivo. La IA puede mejorar claridad, ordenar ideas o ayudar a superar bloqueos de escritura. Pero la voz propia no debe desaparecer. La escritura no es solo vehículo de ideas ya formadas; también es proceso de pensamiento. Al escribir, el autor descubre matices, contradicciones, límites y conexiones. Si delega completamente la escritura, puede perder parte del pensamiento que la escritura genera.

La regulación activa protege la voz propia. Exige que el usuario se pregunte si el texto final dice realmente lo que piensa, si puede defender cada afirmación, si entiende las referencias utilizadas, si reconoce los límites de su argumento y si el estilo no ha sido esterilizado por una prosa genérica. En la era de la IA, la autenticidad intelectual no consistirá en escribir sin herramientas, sino en no abdicar de la propia responsabilidad interpretativa.

4.6. Una ética de la pausa

La cultura digital premia inmediatez. Responder rápido, publicar rápido, producir rápido, automatizar rápido. La IA intensifica esa presión porque hace posible lo que antes tomaba tiempo. Frente a esto, la regulación activa propone una ética de la pausa. Pausar no significa detener el progreso. Significa introducir un intervalo de juicio entre la generación y la acción. Ese intervalo puede ser breve o amplio según el riesgo, pero debe existir.

La pausa permite revisar, contextualizar y preguntar. Permite reconocer si el output seduce por su forma pero carece de fondo. Permite consultar a alguien con más experiencia. Permite detectar que una decisión requiere conversación y no solo eficiencia. La pausa es una forma de soberanía del sujeto frente a la velocidad de la máquina.

Esta ética de la pausa puede parecer contraria a la productividad, pero en realidad protege la productividad real. Muchos errores organizacionales surgen por actuar con rapidez sobre información insuficiente. Un correo mal enviado, una campaña inoportuna, una recomendación mal validada o una decisión automatizada sin contexto pueden generar costos mayores que el tiempo ahorrado. Pausar es una inversión en calidad de juicio.

En síntesis, la regulación activa del individuo no es una postura romántica ni una resistencia nostálgica frente a la tecnología. Es una propuesta práctica para que la IA generativa sea integrada sin sacrificar aquello que permite usarla bien: criterio, contexto, responsabilidad y capacidad de pensar.

4.7. De la herramienta al ecosistema

Un último elemento debe incorporarse al análisis: la IA no opera como herramienta aislada, sino como ecosistema. El usuario no interactúa solo con un modelo, sino con una cadena de servicios, interfaces, proveedores, datos, integraciones, costos, actualizaciones y condiciones de uso. Esta dimensión es relevante porque la dependencia no se produce únicamente por confianza cognitiva en el output; también se produce por integración operativa. Una vez que la empresa conecta sus procesos a múltiples plataformas, migrar, auditar o corregir se vuelve más difícil.

La regulación activa debe ampliarse hacia una conciencia de ecosistema. Antes de incorporar una herramienta, conviene preguntar quién controla los datos, qué ocurre si cambia el precio, cómo se exporta la información, qué procesos quedarían detenidos si el servicio falla, qué proveedor concentra funciones críticas y qué alternativa existe si la herramienta deja de ser conveniente. Estas preguntas

parecen administrativas, pero son profundamente regulatorias. Definen el grado de soberanía tecnológica de la organización.

En pequeñas y medianas empresas, este punto es especialmente delicado. La facilidad de adopción puede ocultar compromisos futuros. Una herramienta gratuita o económica al inicio puede volverse infraestructura esencial después. La organización adapta su flujo de trabajo, capacita a su equipo, almacena información, crea automatizaciones y depende de esa plataforma para operar. Cuando aparece el costo real, ya no se evalúa libremente; se paga porque salir resulta más caro. La subordinación no ocurrió por una decisión única, sino por acumulación de pequeñas dependencias.

Por ello, regular activamente también significa preservar reversibilidad. Un proceso tecnológicamente asistido debe conservar cierta capacidad de continuidad si la herramienta falla o cambia. Esto no implica duplicar todo manualmente, sino evitar que el conocimiento del proceso desaparezca detrás de la automatización. El equipo debe seguir entendiendo qué se hace, por qué se hace y cómo podría reconstruirse. La IA puede ejecutar partes del proceso, pero no debe convertirse en la única memoria operativa.

4.8. La pregunta por el valor

Toda adopción de IA debería concluir en una pregunta simple: ¿aumentó el valor real o solo aumentó la producción? Esta pregunta obliga a regresar al usuario final, al paciente, al estudiante, al cliente, al ciudadano o al equipo de trabajo. Si la tecnología no mejora su experiencia, comprensión, seguridad, atención, aprendizaje o capacidad de decisión, entonces el beneficio es dudoso. Puede haber ahorro interno, pero no necesariamente creación de valor.

La regulación activa se opone a la fascinación por el output porque vuelve a colocar el valor en el centro. Una empresa no existe para usar IA; usa IA para cumplir mejor sus fines. Una institución educativa no existe para generar más materiales; existe para formar. Una organización de salud no existe para automatizar comunicación; existe para cuidar mejor. Una firma financiera no existe para producir reportes más

elegantes; existe para orientar decisiones prudentes. Cuando esta jerarquía se pierde, la tecnología deja de servir al propósito y empieza a reemplazarlo.

Este criterio permite evaluar la IA sin moralismos. No se trata de aprobarla o condenarla en abstracto. Se trata de observar si, en un caso específico, fortalece capacidades humanas, mejora decisiones, reduce daños, libera tiempo para tareas significativas o amplía comprensión. Cuando lo hace, debe usarse. Cuando no lo hace, debe limitarse, rediseñarse o rechazarse. La regulación activa no es una teoría contra la tecnología; es una disciplina para que la tecnología no piense por nosotros ni decida nuestros fines.

CONCLUSIÓN

La inteligencia artificial generativa obliga a repensar la regulación desde una dimensión previa a la norma formal: la relación del sujeto con la herramienta. Las leyes, políticas y estándares son necesarios, pero insuficientes si quienes usan IA no conservan autonomía de juicio. Una organización puede cumplir reglas y aun así operar de forma acrítica. Un profesional puede usar herramientas avanzadas y aun así renunciar a pensar. La primera regulación ocurre cuando el individuo se detiene, pregunta, valida, contextualiza y asume responsabilidad.

La tesis de este capítulo no es que la IA sea enemiga del pensamiento humano. La tesis es más precisa: la IA puede ampliar o erosionar el pensamiento según la forma en que se integre. Puede liberar tiempo para reflexión o sustituir la reflexión por outputs. Puede fortalecer el criterio experto o simularlo. Puede mejorar procesos o vaciarlos de aprendizaje. La diferencia depende de prácticas concretas de uso, validación y responsabilidad.

La paradoja central permanece: una herramienta diseñada para aumentar eficiencia puede producir dependencia si la velocidad se interpreta como valor suficiente. Las organizaciones deben mirar más allá del ahorro de tiempo. Deben preguntarse qué capacidades se fortalecen, qué conversaciones desaparecen, qué conocimientos dejan de transmitirse, qué decisiones se vuelven opacas y qué sujetos se están

formando. Si el resultado es más producción con menos criterio, la eficiencia es aparente.

La regulación activa del individuo propone una respuesta situada. No espera a que el Estado resuelva todo ni reduce la responsabilidad a cumplimiento corporativo. Coloca al sujeto frente a su decisión cotidiana: usar o consumir, validar o aceptar, pensar o delegar, subordinar la herramienta o subordinarse a ella. Esta decisión no es abstracta. Aparece en cada correo generado, cada reporte automatizado, cada recomendación aceptada, cada imagen publicada y cada estrategia sugerida por un modelo.

Regular activamente significa recuperar la primacía del proceso sobre el output cuando el proceso forma juicio. Significa reconocer que la ventana de contexto de una IA no equivale al contexto de vida de una persona u organización. Significa distinguir plausibilidad de verdad, información de conocimiento, asistencia de autoridad. Significa aceptar que el sujeto sigue siendo responsable, incluso cuando la máquina participa.

El futuro de la IA en las organizaciones no dependerá únicamente de modelos más potentes. Dependerá de sujetos más capaces de pensar con ellos sin ser absorbidos por ellos. En ese sentido, la pregunta decisiva no es si la inteligencia artificial transformará el trabajo. Ya lo está transformando. La pregunta es si esa transformación producirá individuos más autónomos o más dependientes, organizaciones más inteligentes o más subordinadas, procesos más humanos o simplemente más veloces. La respuesta no está escrita en el código. Empieza en el juicio de quien usa la herramienta.

REFERENCIAS

- Arendt, H. (1958). *The human condition*. University of Chicago Press.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending regulations and directives. *Official Journal of the European Union*.
- Foucault, M. (1975). *Surveiller et punir: Naissance de la prison*. Gallimard.
- Foucault, M. (1987). The ethic of care for the self as a practice of freedom: An interview with Michel Foucault on January 20, 1984. In J. Bernauer & D. Rasmussen (Eds.), *The final Foucault* (pp. 1–20). MIT Press.
- Gadamer, H.-G. (1975). *Truth and method*. Seabury Press.
- Han, B.-C. (2022). *Infocracy: Digitization and the crisis of democracy*. Polity Press.
- Hegel, G. W. F. (1977). *Phenomenology of spirit* (A. V. Miller, Trans.). Oxford University Press. (Original work published 1807)
- Marx, K. (1982). *Capital: A critique of political economy* (Vol. 1, B. Fowkes, Trans.). Penguin Books. (Original work published 1867)
- National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1)*. U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- Polanyi, M. (2009). *The tacit dimension*. University of Chicago Press. (Original work published 1966)
- Sen, A. (1999). *Development as freedom*. Oxford University Press.

UNESCO. (2022). Recommendation on the ethics of artificial intelligence.
UNESCO.

Winner, L. (2020). The whale and the reactor: A search for limits in an age
of high technology (2nd ed.). University of Chicago Press

CAPÍTULO XIX

ELEMENTOS PARA LA CONSTRUCCIÓN DE UNA GOBERNANZA DIGITAL SITUADA: UN ANÁLISIS DE LA REGULACIÓN DE LA IA EN ENTORNOS LEGISLATIVOS

Elements for the construction of situated digital governance: an analysis of the regulation of AI in legislative environments

Luis Angel Garcia Covarrubias⁹

luiscova007@gmail.com

Universidad Autónoma Metropolitana, Unidad Iztapalapa

<https://orcid.org/0009-0007-1552-7461>

I. Resumen

La irrupción de la Inteligencia Artificial (IA) en el corazón del Estado mexicano no es una posibilidad futura, sino una realidad instalada que está transformando la naturaleza misma de la labor legislativa. La capacidad de generar, sintetizar y procesar volúmenes masivos de información en segundos ha seducido a los espacios legislativos, donde la urgencia de la inmediatez a menudo colisiona con la profundidad de la deliberación democrática. Sin embargo, esta adopción tecnológica no es un proceso neutral; conlleva una reconfiguración de las relaciones de poder, de la autoría intelectual y de la soberanía política del representante.

El presente capítulo tiene como objetivo analizar la necesidad imperativa de regular la IA en la creación de propuestas legislativas, no desde una visión meramente técnica o administrativa, sino a través de una lente sociológica y organizacional. Se

⁹ Maestrante en el programa de Maestría en Estudios Organizacionales por la Universidad Autónoma Metropolitana, Unidad Iztapalapa y Licenciado en Sociología por la Universidad Autónoma del Estado de Hidalgo

argumenta que, sin un marco de gobernanza digital situado, el Poder Legislativo corre el riesgo de quedar subordinado a las lógicas del capitalismo algorítmico, donde la ley deja de ser un producto de la voluntad social para convertirse en una narrativa estandarizada y modulada por dispositivos tecnológicos externos.

A lo largo de este capítulo, se explorará cómo la IA está configurando una nueva episteme digital y actuando como un dispositivo de normalización que disciplina la conducta de los legisladores. Frente a este panorama, se propone la intervención organizacional como la alternativa fundamental para construir una regulación ética y soberana. En este enfoque, los actores del análisis organizacional dejan de ser observadores externos para convertirse en mediadores críticos que dictan las pautas de una tecnología al servicio de la vida.

En ese sentido, el contenido de este capítulo se despliega desde 1) una presentación puntual sobre la realidad del uso de la IA en los espacios legislativos, 2) el inicio hacia una construcción de la gobernanza digital en la vida legislativa, 3) la presentación del capitalismo algorítmico como el marco estructural del gobierno digital, 4) la puesta en escena de los elementos foucaultianos de episteme y dispositivo para la profundización del gobierno digital en el capitalismo algorítmico, 5) una solución posible para la regulación y manejo de la IA en estos espacios legislativos a partir de la intervención organizacional y 6) consideraciones finales del embalaje propuesto para este capítulo.

A través de este recorrido, el lector encontrará un argumento sólido para defender que la regulación de la IA en México debe ser, ante todo, un acto de defensa de la inteligencia humana y del contexto territorial frente a la automatización de la democracia.

Palabras clave: Inteligencia artificial, Poder Legislativo, gobernanza digital, capitalismo algorítmico, intervención organizacional, regulación tecnológica, episteme digital, dispositivo de poder, soberanía legislativa, democracia digital.

Abstract

The irruption of Artificial Intelligence (AI) in the heart of the Mexican State is not a future possibility, but an installed reality that is transforming the very nature of legislative work. The ability to generate, synthesize, and process massive volumes of information in seconds has seduced legislative spaces, where the urgency of immediacy often collides with the depth of democratic deliberation. However, this technological adoption is not a neutral process; it entails a reconfiguration of the relations of power, of the intellectual authorship and of the political sovereignty of the representative.

This chapter aims to analyze the imperative need to regulate AI in the creation of legislative proposals, not from a merely technical or administrative perspective, but through a sociological and organizational lens. It is argued that, without a digital governance framework in place, the Legislative Branch runs the risk of being subordinated to the logics of algorithmic capitalism, where the law ceases to be a product of social will to become a standardized narrative modulated by external technological devices.

Throughout this chapter, we will explore how AI is shaping a new digital episteme and acting as a normalization device that disciplines the conduct of legislators. Faced with this panorama, organizational intervention is proposed as the fundamental alternative to build ethical and sovereign regulation. In this approach, the actors of organizational analysis cease to be external observers to become critical mediators who dictate the guidelines of a technology at the service of life.

In this sense, the content of this chapter unfolds from 1) a specific presentation on the reality of the use of AI in legislative spaces, 2) the beginning towards a construction of digital governance in legislative life, 3) the presentation of algorithmic capitalism as the structural framework of digital government, 4) the staging of the Foucauldian elements of episteme and device for the deepening of digital government in the algorithmic capitalism, 5) a possible solution for the regulation

and management of AI in these legislative spaces based on organizational intervention and 6) final considerations of the packaging proposed for this chapter.

Through this journey, the reader will find a solid argument to defend that the regulation of AI in Mexico should be, above all, an act of defense of human intelligence and the territorial context against the automation of democracy.

Keywords: Artificial intelligence, Legislative Power, digital governance, algorithmic capitalism, organizational intervention, technological regulation, digital episteme, power device, legislative sovereignty, digital democracy.

II. Presentación de la realidad legislativa y el uso de la IA

La integración de la inteligencia artificial en la vida legislativa mexicana ha dejado de ser una especulación futurista para convertirse en una realidad cuantificable que redefine la naturaleza misma del quehacer legislativo. De acuerdo con los reportajes publicados por diversas revistas en México, el uso de modelos de lenguaje generativo ha permeado de tal forma en el Congreso de la Unión que es posible identificar huellas lingüísticas permanentes en los discursos de tribuna, con un incremento crítico detectado a partir de septiembre de 2024, donde los patrones de redacción asistida aumentaron entre un 60% y 70% (IMT, 2026). Esta adopción no es uniforme pues se manifiesta con mayor intensidad en grupos parlamentarios como el PVEM y Morena, y curiosamente es más prevalente entre legisladores con niveles educativos medios o técnicos, lo que sugiere que la herramienta funciona como un mecanismo de compensación de capital cultural ante las exigencias de una tecnocracia parlamentaria cada vez más demandante (RPD, 2026).

En el caso de la diputada Olga Leticia Chávez, quien defendió el uso de IA para condensar dictámenes extensos en minutos bajo la consigna de *actualícense ignorantes*, ilustra una mutación en la responsabilidad del representante (Bárcena

& Cardona, 2025). Aquí, la labor del legislador se desplaza desde la creación intelectual hacia la supervisión técnica, generando ciertas contradicciones pues, mientras el 80% de los legisladores reconoce la necesidad de estas herramientas para gestionar el volumen de trabajo, solo el 30% admite poseer la capacitación ética necesaria para su manejo responsable (Bárcena & Cardona, 2025). Esta brecha abre un espacio de opacidad donde la autoría política se diluye en un flujo de algoritmos que, en palabras de críticos legislativos, podría erosionar la honorabilidad y seriedad de un poder que emana del voto ciudadano.

Ante esta situación, es preciso empezar a cuestionarse y proponer alternativas para la regulación y el uso de la IA en entornos legislativos pues su importancia de regular la inteligencia artificial en el sector público adquiere una dimensión ética, pragmática y urgente. La regulación no debe ser vista como un obstáculo a la innovación, sino como la base necesaria para una gobernanza digital responsable que garantice que el uso de algoritmos no vulnere los principios democráticos ni los derechos fundamentales. Como señala Aguirre Quezada y Lopez Hernández (2025), la implementación efectiva de la IA en la administración pública requiere marcos normativos sólidos que mitiguen riesgos críticos como el sesgo algorítmico, la precarización laboral y la vulneración de la ciberseguridad, transformando lo que podría ser un dispositivo de control en una herramienta de gestión transparente y cercana a la ciudadanía.

El panorama legislativo actual en México refleja esta preocupación creciente ya que, solamente en los primeros cien días de la LXVI Legislatura, se presentaron siete iniciativas clave que buscaron reformar leyes generales en materias tan diversas como salud, derechos de la infancia, violencia de género y trabajo rural (Aguirre Quezada & López Hernández, 2025). Estos esfuerzos legislativos intentan cubrir lagunas legales que hoy permiten una actuación discrecional, por ejemplo, la propuesta de normar el consentimiento informado en servicios de salud asistidos por IA o la supervisión ética en el reconocimiento facial para la búsqueda de personas desaparecidas. Esta actividad legislativa demuestra que el Estado

reconoce la necesidad de intervenir antes de que el progreso tecnológico socave los mismos principios morales que la democracia busca proteger.

Este panorama de uso y regulación de la IA en entornos legislativos hace pensar ¿Cómo operan las tensiones entre la racionalidad instrumental del capitalismo algorítmico y la prevención de sesgos que amenazan con invisibilizar los saberes sometidos (locales y situados) en la legislación mexicana?

El desarrollo de este texto pretende realizar anotaciones precisas y teóricas sobre la regulación de la IA en la gobernanza digital de entornos legislativos desde la sociología y los estudios organizacionales, por lo que es preciso partir de un análisis cualitativo documental de los diversos aportes teóricos de estas disciplinas.

III. Hacia la construcción de una gobernanza digital en la vida legislativa

A partir del análisis de Guzmán Noguera y *otros* (2024), la gobernanza digital se presenta como una visión situada y estratégica que responde a las demandas de la Revolución 4.0 y al surgimiento de una Ciudadanía 5.0. Lejos de ser un proceso meramente técnico o burocrático, esta visión la sitúa como el pilar fundamental para la transformación de la administración pública, con el objetivo de alcanzar niveles superiores de eficiencia, transparencia y, sobre todo, una participación ciudadana activa en los asuntos del Estado.

Esta perspectiva sostiene que la gobernanza digital no puede ser impuesta de manera unilateral por el gobierno; más bien, debe ser el resultado de un diseño colaborativo en el que converjan todos los sectores productivos, expertos y la propia ciudadanía. En este sentido, se concibe como un espacio de encuentro que genera una ruta para el uso inteligente de las Tecnologías de la Información y las Comunicaciones, orientándolas no solo a la modernización de trámites, sino a la creación de valor público compartido (Guzmán Noguera, y *otros*, 2024, p. 8).

Esta visión situada subraya que la gobernanza digital funciona como la base necesaria para construir un marco normativo robusto. Dicho marco no solo debe sustentar la infraestructura tecnológica, sino reflejar las necesidades reales y los derechos de los ciudadanos en un entorno digital, asegurando que la tecnología sea una herramienta de inclusión y democracia en lugar de una barrera administrativa

Por otra parte, el análisis de Valenzuela, Collanes y Durand (2020) permite profundizar en las estructuras de poder y los marcos ideológicos que subyacen a la digitalización del Estado. Esta perspectiva expande la visión técnica hacia una comprensión política y multinivel del fenómeno.

La gobernanza digital no es un proceso neutro; se despliega en un espectro que oscila entre la optimización técnica y la transformación de la ciudadanía. Estos autores advierten que gran parte del impulso hacia la gobernanza digital se inscribe en la lógica del neoliberalismo digital. Bajo esta premisa, la adopción de tecnologías como la inteligencia artificial en el Poder Legislativo no busca únicamente mejorar la representación, sino que a menudo se prioriza la reducción de costos, la tercerización de servicios de datos a corporaciones privadas y la aplicación de métricas de rendimiento que tratan la labor legislativa como una cadena de producción industrial (Valenzuela Fernandez, Collantes Inga, & Durand, 2020).

Un componente esencial de la gobernanza digital es la gobernanza de datos, que define quién recolecta, almacena y utiliza la información generada en los procesos públicos. En el contexto parlamentario mexicano, el uso de IA generativa sin marcos claros de gobernanza de datos entrega la producción intelectual y política del país a ecosistemas tecnológicos externos, debilitando la soberanía digital. La gobernanza efectiva requiere, por tanto, el establecimiento de infraestructuras propias y protocolos de transparencia que aseguren que los datos legislativos permanezcan bajo el control democrático y no se conviertan en activos de plataformas transnacionales.

En este sentido, la gobernanza digital debe aspirar a la consolidación de una ciudadanía digital activa. Esto significa que la integración de la IA en la creación de leyes no debe ser una caja negra cerrada para el ciudadano, sino un mecanismo que facilite la rendición de cuentas en múltiples niveles, desde lo local hasta lo nacional. Una gobernanza digital democratizadora es aquella que utiliza la tecnología para visibilizar los procesos de decisión, permitiendo que la sociedad no sea solo receptora de normas, sino un actor que vigila y colabora en la supervisión de los algoritmos que impactan su vida común.

Abonando a esta forma de ver la gobernanza digital, Aminahuel y Rodríguez (2024) comentan que es necesario entenderla no solo como un marco administrativo, sino como un campo de batalla frente a la concentración del poder tecnológico. Según las autoras, la gobernanza actual se encuentra en una tensión dialéctica: por un lado, existen esfuerzos internacionales multisectoriales para proteger los derechos humanos y, por el otro, una realidad dominada por la "lataformización del espacio público, donde un puñado de corporaciones transnacionales dictan las reglas del juego mediante arquitecturas digitales opacas.

En el contexto legislativo, esta visión de la gobernanza digital advierte sobre los peligros de la dependencia tecnológica. El uso de herramientas de IA en el Congreso no es una decisión soberana aislada si estas herramientas pertenecen al ecosistema de las grandes tecnológicas. La arquitectura de estas plataformas facilita procesos de datificación y vigilancia digital que pueden erosionar la autonomía de las instituciones públicas. De este modo, la gobernanza digital debe actuar como un mecanismo que denuncie y limite esta concentración de poder, asegurando que el Estado no sea un mero consumidor de servicios algorítmicos, sino un regulador capaz de defender la integridad del debate democrático frente a las polarizaciones y desinformaciones que estos sistemas pueden generar.

Asimismo, las autoras subrayan que una gobernanza digital efectiva requiere superar las contradicciones entre los discursos éticos y las prácticas económicas del capitalismo de plataformas. Mientras que en el Congreso mexicano se puede

hablar de eficiencia algorítmica, la lógica de fondo sigue siendo la de la acumulación de datos. Una gobernanza situada debe, por tanto, transitar hacia un modelo que garantice la transparencia y la rendición de cuentas de los intermediarios digitales.

Ya despues, Aminahuel y Rodríguez (2024) destacan que la gobernanza digital debe aspirar a proteger la esfera pública de los efectos nocivos de los algoritmos de recomendación y síntesis. En la labor legislativa, esto se traduce en la necesidad de marcos normativos que no solo vigilen la autoría de las propuestas, sino que también auditen los sesgos inherentes a las plataformas utilizadas. La gobernanza digital, desde esta óptica, se convierte en la herramienta para recuperar la soberanía tecnológica, evitando que la democracia se convierta en un subproducto de la arquitectura digital privada y garantizando que la ley siga respondiendo a los intereses colectivos y no a los mercados de futuros conductuales.

Ante todo, lo propuesto, la gobernanza digital dentro de los espacios legislativos se comprende no solo como la adopción de infraestructuras tecnológicas para la gestión administrativa, sino como una arquitectura política y ética estratégica diseñada para mediar la relación entre el Estado, la ciudadanía y la innovación. En este contexto, la gobernanza actúa como un puente de conocimiento esencial para la creación de iniciativas de ley, permitiendo que el proceso legislativo transite de una lógica de eficiencia hacia una construcción normativa situada y participativa.

Desde esta perspectiva, la gobernanza digital responsable garantiza que la integración de la inteligencia artificial sea un mecanismo de democratización del saber y no un dispositivo de docilidad política. Al establecer marcos de transparencia y participación ciudadana, la gobernanza permite que los saberes locales y territoriales (sobre todo en el espacio de la gobernanza de los congresos locales) fluyan hacia las comisiones, enriqueciendo la deliberación humana. De este modo, la tecnología deja de ser una caja negra que dicta soluciones estandarizadas y se convierte en un instrumento que potencia la pluralidad del conocimiento, asegurando que las leyes respondan a las realidades complejas de la sociedad mexicana y no a meros patrones predictivos de proveedores externos.

Este puente de conocimiento se consolida a través de la profesionalización y la alfabetización algorítmica de los legisladores. Una gobernanza digital efectiva empodera al representante para que actúe como un supervisor crítico, capaz de utilizar la IA para procesar volúmenes masivos de información sin perder la soberanía de su juicio político. Al unir la ética con normas exigibles y auditables, los espacios legislativos se transforman en nodos de conocimiento donde la técnica queda al servicio de la vida y la justicia social, recuperando así la honorabilidad de un poder que emana de la voluntad popular frente a las inercias deshumanizantes.

Sin embargo, en la construcción de estos mecanismos digitales para la consolidación de la IA como herramienta digital, habría que ser críticos sobre desde que lugares estamos utilizando estos mecanismos.

IV. El capitalismo algorítmico como el marco estructural del gobierno digital

Para situar la gobernanza digital dentro de su marco estructural, es imperativo profundizar en la lógica del capitalismo algorítmico. De acuerdo con González Montaña (2022), este sistema no se limita a la simple acumulación de información, sino que representa una fase del capital donde la experiencia humana es convertida en materia prima para la creación de productos de predicción". En este contexto, la labor legislativa deja de ser vista como un proceso de deliberación democrática para ser tratada como un flujo de datos que debe ser optimizado y rentabilizado. La gobernanza digital, por lo tanto, opera en un terreno donde el objetivo primordial del sistema es la modulación de la existencia a través de un despliegue algorítmico que busca conocer y anticipar los comportamientos de los sujetos.

Bajo la trama del capitalismo algorítmico, el Poder Legislativo se enfrenta a lo que se denomina la extracción de excedente conductual. González Montaña (2022) explica que los dispositivos digitales (como las herramientas de IA generativa utilizadas en el Congreso mexicano) operan bajo una lógica de vigilancia que transforma la creatividad y el juicio crítico del legislador en datos procesables. Cuando un legislador recurre a la IA para sintetizar iniciativas de forma masiva, está

alimentando un ciclo de acumulación donde el saber jurídico se desprende de su autoría humana y se convierte en una mercancía algorítmica. Este proceso facilita lo que Éric Sadin identifica como el tránsito hacia el individuo asistido (2020); es decir, un sujeto que, lejos de ejercer su soberanía, se vuelve dependiente de sistemas que le sugieren o imponen trayectorias de decisión en tiempo real, priorizando la eficiencia del capital sobre la justicia social (González Montaña, 2022).

Esta dinámica representa una amenaza directa a la democracia, a la cual se describe como acechada por la automatización de la toma de decisiones. En el capitalismo algorítmico, la gobernanza digital corre el riesgo de ser capturada por una gubernamentalidad algorítmica que sustituye el debate público por la gestión estadística de la realidad. El peligro para la creación de leyes en México radica en que el dispositivo tecnológico no solo asiste, sino que "produce" la narrativa política, induciendo al representante a adoptar soluciones predefinidas por algoritmos que no comprenden el contexto territorial. Así, la democracia se ve subordinada a un sistema de dirección de la vida humana que busca la uniformidad, invisibilizando las disidencias y los saberes locales que no logran ser codificados por la lógica de la ganancia técnica.

Entender la gobernanza digital y sus implicaciones con la IA desde el capitalismo algorítmico permite comprender por qué la gobernanza digital en los espacios legislativos debe ser, ante todo, una estrategia de soberanía. Si la lógica dominante busca la modulación y la docilidad del sujeto político, la gobernanza debe actuar como un contrapoder que exija la transparencia de las "cajas negras" algorítmicas y proteja la integridad de la deliberación humana. La regulación de la IA no debe limitarse a la gestión de riesgos técnicos, sino que debe ser un acto de resistencia política que garantice que la ley siga siendo un producto de la voluntad social situada y no un subproducto de los mercados de futuros conductuales (González Montaña, 2022). Solo así, los espacios legislativos podrán ser verdaderos puentes de conocimiento que pongan la técnica al servicio de la vida y no de la acumulación automatizada.

El planteamiento propuesto desde la gobernanza digital propone la alternativa a la integración de la inteligencia artificial en la labor legislativa. Esta, a su vez, también debe ser entendida como un proceso inserto en la fase actual del capitalismo algorítmico. De acuerdo con el mismo González Montaña (2023), este modelo se caracteriza por el despliegue de dispositivos digitales que no solo procesan información, sino que tienen como fin último modular la existencia de los sujetos. En este sentido, la IA funciona como una máquina comunicacional sugestiva que induce y produce narrativas que se ajustan a las lógicas de optimización del capital.

Bajo la lógica del capitalismo algorítmico, el legislador deja de ser un actor político autónomo para convertirse en un usuario cuya producción intelectual es mediada por algoritmos. González Montaña (2023) advierte que estos dispositivos operan mediante un conocimiento pormenorizado de los usuarios, transformando sus interacciones en datos que alimentan el ciclo de mejora del software privado. Cuando un legislador utiliza IA para redactar una propuesta, está sometiendo su juicio a un sistema que prioriza la estandarización y la predictibilidad, lo que Foucault identificaría como una forma de disciplina que busca sujetos dóciles y útiles a la eficiencia del mercado tecnológico.

Este riesgo central identificado por González Montaña (2023) es que la IA no se limita a asistir, sino que interpela e incide en la creación del discurso. En el ámbito de las propuestas legislativas, esto significa que el algoritmo sugiere estructuras argumentativas y soluciones jurídicas que, al haber sido entrenadas en corpus de datos masivos y deslocalizados, tienden a reproducir un sentido hegemónico. Así, el capitalismo algorítmico logra que la ley pierda su naturaleza de voluntad social situada para convertirse en una narrativa automatizada que satisface las métricas de rendimiento y rapidez demandadas por la burocracia digital.

Esta visión del capitalismo algorítmico opera bajo una lógica de extracción. En lugar de extraer recursos naturales, extrae la capacidad creativa y crítica del jurista para convertirla en patrones predictivos. González Montaña (2023) destaca que este

proceso lleva a una unión íntima entre el dispositivo y el sujeto, donde las fronteras entre el pensamiento humano y la salida algorítmica se desdibujan. Para el Poder Legislativo mexicano, esto representa una amenaza a la soberanía del juicio, ya que la deliberación democrática queda supeditada a cajas negras tecnológicas cuyos intereses responden a corporaciones transnacionales y no al bienestar público.

La regulación de la IA, por tanto, no puede ser una simple normativa técnica. Debe ser una resistencia política frente al capitalismo algorítmico. Como señala González Montaña, se requiere una mirada crítica que reconozca que estos dispositivos producen una normalidad digital que, si no es cuestionada, terminará por invisibilizar los saberes locales y las necesidades cualitativas de la sociedad mexicana en favor de una eficiencia gélida y automatizada.

En el contexto del capitalismo algorítmico, la gobernanza digital deja de ser un mero conjunto de herramientas técnicas para convertirse en un dispositivo político que organiza el poder mediante el uso de datos y algoritmos orientados a la acumulación y el control. A diferencia de las formas tradicionales de autoridad, esta modalidad opera de manera sutil e íntima a través de la modulación de la conducta, anticipando los deseos de los sujetos para dirigir sus acciones. En el ámbito legislativo, esta dinámica se traduce en una presión constante por la inmediatez y la eficacia técnica que termina por sacrificar la profundidad de la deliberación pública y el juicio crítico del representante.

Esta transición hacia lo que se denomina el individuo asistido conlleva una erosión de la soberanía humana, ya que se delegan capacidades cognitivas esenciales a sistemas de inteligencia artificial que no comprenden el contexto social. Por ello, la regulación de estas tecnologías no es solo una cuestión técnica, sino una necesidad política urgente para recuperar la autonomía en la toma de decisiones. Finalmente, se propone una gobernanza situada y responsable que actúe como defensa de los saberes locales y territoriales frente a la tendencia homogeneizadora de los modelos algorítmicos globales.

Por lo tanto, la IA en los espacios legislativos no debe entenderse como un objeto técnico aislado, sino como la manifestación de una nueva episteme digital que dicta las reglas de lo que es posible pensar, decir y legislar en la actualidad.

V. El aterrizaje de los conceptos foucaultianos de episteme y dispositivo.

En una obra escrita con anterioridad (García Covarrubias, 2025) la episteme representa el conjunto de reglas que definen las condiciones de posibilidad del saber en un periodo histórico determinado. En la vida legislativa actual, esta episteme se caracteriza por la primacía de la eficiencia técnica y la inmediatez algorítmica sobre la deliberación cualitativa. Cuando los legisladores adoptan patrones lingüísticos estandarizados detectados, están operando dentro de una estructura de conocimiento donde el saber legítimo es aquel que puede ser procesado, sintetizado y generado con rapidez. Esta lógica impone una jerarquía donde el dato duro y la salida automatizada desplazan a la experiencia situada y a los procesos de reflexión política que no encajan en la velocidad del algoritmo.

Por su parte, la IA se constituye como un dispositivo foucaultiano, definido por como una red heterogénea que articula discursos, instituciones, leyes, tecnologías y enunciados científicos para responder a una urgencia histórica (García Covarrubias, 2025). En el contexto legislativo, la urgencia es la gestión de la saturación informativa; sin embargo, este dispositivo opera bajo la lógica del capitalismo algorítmico, lo que tiende a producir lo que el autor denomina un sentido hegemónico. Este dispositivo digital disciplina la conducta del representante, transformándolo en un sujeto asistido que, al buscar la optimización de su tiempo, termina por delegar su juicio soberano a una caja negra que estandariza las soluciones jurídicas.

Este aporte pretende situar la advertencia sobre la invisibilización de las realidades locales frente a la homogeneización algorítmica pues estos dispositivos arriesgan el silenciamiento de los saberes locales y las lógicas territoriales que no han sido digitalizadas o que no se ajustan a los modelos de lenguaje globales. En el ámbito

legislativo, esto significa que una propuesta generada por IA puede carecer de territorio, convirtiéndose en un enunciado técnico despojado de la pluralidad social de México, y sobre todo de los contextos locales. La regulación, por tanto, debe actuar como un contrapeso que rescate estos saberes sometidos, asegurando que la tecnología no sea un fin en sí mismo, sino una herramienta mediada críticamente.

Por lo tanto, frente a la docilidad política que impone el dispositivo, surge la necesidad de un pensamiento de intervención organizacional crítica dentro de los congresos tanto federales como los locales en donde se garantice el uso de la IA para la construcción de herramientas legislativas mas adecuadas a sus contextos, destacaando principios como la primacía del contexto, el diálogo y la co-construcción como fundamentales para resistir la racionalidad instrumental. Para el Poder Legislativo, esto implica que la alternativa responsable no es solo la transparencia técnica, sino la formación de legisladores y asesores que actúen como mediadores críticos. Estos actores deben ser capaces de reconocer cuándo el dispositivo está modulando su juicio y, mediante una alfabetización algorítmica, recuperar la soberanía política necesaria para que la ley siga siendo un producto de la voluntad social y no un excedente de datos del capitalismo de vigilancia.

VI. El resultado: la intervención organizacional como medio regulador del uso de la IA en la gobernanza digital de la vida legislativa.

La intervención organizacional se erige como la praxis fundamental para transformar la gobernanza digital de un simple marco administrativo a una estrategia de resistencia política y ética frente al capitalismo algorítmico. Lozano Carrillo, Zavala Osorio y Couturier Bañuelos (2025) proponen a la intervención en las organizaciones, particularmente bajo el modelo del Programa de Inteligencia Artificial Estratégica Social y Solidaria (PIAESS), no debe entenderse como una consultoría técnica tradicional, sino como un proceso de co-construcción de conocimiento de alto impacto social. En el ámbito legislativo, donde la urgencia de la inmediatez técnica suele desplazar la profundidad política, la intervención

organizacional ofrece una metodología para reintroducir la dimensión humana y territorial en la creación de leyes.

Desde la analítica de Foucault, el dispositivo algorítmico en el Congreso opera bajo una lógica de normalización que busca producir individuos asistidos (Sadin, 2020), legisladores que delegan su juicio a la máquina en nombre de la eficiencia. Frente a esto, la propuesta de Allende Gorgonio (2025) sobre la construcción de evidencias organizacionales sugiere que la intervención puede actuar como un mecanismo de visibilización. Intervenir la organización legislativa implica auditar los procesos de toma de decisión para identificar dónde el algoritmo está sesgando la deliberación y dónde se están silenciando los saberes sometidos. La intervención organizacional permite que el Poder Legislativo transite de una episteme, donde el saber vale por su rapidez de procesamiento, hacia una episteme para la vida, centrada en el impacto social y la justicia comunitaria.

Un elemento central en esta profundización es la praxis de la co-agencia propuesta por Ramírez Navarro (2025). En la era de la inteligencia artificial, la intervención ya no solo ocurre entre humanos, sino que exige un diálogo onto-epistemológico entre el actor político y el sistema inteligente. La intervención organizacional dota al legislador de la figura del mediador crítico; este actor no es un simple usuario de software, sino un profesional capaz de supervisar el dispositivo, reconociendo sus límites y forzando la entrada de contextos locales que el capitalismo digital tiende a ignorar. Como señala el PIAESS en el marco de la UAM, la intención es reeducar el algoritmo funcionalista para orientarlo hacia un conocimiento que construya comunidad humana y no solo excedentes conductuales (Lozano Carrillo y otros., 2025).

La intervención organizacional es la base operativa de una gobernanza digital situada. Mientras que el capitalismo algorítmico busca estandarizar la labor legislativa como una cadena de producción industrial, la intervención propone un diseño colaborativo y adaptativo que valore lo cualitativo por encima de lo métrico. Al implementar procesos de intervención que prioricen el diálogo, la transparencia

y el resguardo de la soberanía de los datos, el Estado mexicano puede recuperar el control sobre sus infraestructuras digitales. En este sentido, intervención organizacional es el puente de conocimiento que permite que la inteligencia artificial deje de ser un dispositivo de docilidad política para convertirse en una herramienta de emancipación que potencie la honorabilidad del poder emanado del voto popular.

VII. A manera de conclusión

La irrupción de la inteligencia artificial en el Congreso de la Unión, evidenciada por a través de diferentes fuentes dejan huellas lingüísticas algorítmicas, que marcan el nacimiento de una nueva episteme digital. Como se ha argumentado, esta episteme no es neutral; impone una estructura de conocimiento donde la celeridad y la síntesis masiva desplazan la deliberación reflexiva y situada. En este escenario, la IA opera como un dispositivo de normalización que disciplina la labor legislativa bajo la lógica del capitalismo algorítmico. Este sistema, descrito por González Montaña, trasciende la mera automatización para buscar la modulación de la conducta del representante, transformándolo en un individuo asistido que delega su juicio soberano a cajas negras tecnológicas controladas por intereses transnacionales.

El riesgo de este modelo radica en la consolidación de un neoliberalismo digital que prioriza la eficiencia gélida y la reducción de costos sobre la calidad de la representación popular. Al automatizar la creación de leyes, se corre el peligro de silenciar los saberes sometidos, aquellas realidades locales y territoriales que no logran ser codificadas por los modelos de lenguaje globales. Por tanto, la gobernanza digital en los espacios legislativos no debe ser entendida como un simple marco administrativo de riesgos, sino como una arquitectura ética estratégica que actúe como un puente de conocimiento entre la técnica y la vida social.

El contenido del capítulo profundiza en la trama estructural del capitalismo algorítmico, explorando cómo estos dispositivos digitales trascienden la mera automatización para buscar la modulación de la conducta del representante, transformándolo en un "individuo asistido" que delega su juicio soberano a cajas negras tecnológicas. Ante el riesgo de consolidar un neoliberalismo digital que silencie los saberes sometidos, aquellas realidades locales y territoriales que no encajan en la lógica del código global, se propone redefinir la gobernanza digital. Esta no debe ser entendida como una gestión administrativa de riesgos, sino como una arquitectura ética estratégica y un puente de conocimiento que garantice que el origen de la ley siga siendo un acto de voluntad social situada.

Frente a esta inercia, la intervención organizacional se propone como la alternativa política fundamental. La intervención deja de ser una consultoría externa para convertirse en una praxis de co-construcción y co-agencia. Son los actores del análisis organizacional, actuando como mediadores críticos, quienes deben definir las pautas regulatorias. Esta regulación desde adentro asegura que el uso de la IA sea auditable, transparente y, sobre todo, que responda a una episteme de la vida que rescate el contexto local frente a la homogeneización algorítmica.

Por tanto, la regulación de la IA en los espacios legislativos debe transitar de un modelo de neoliberalismo digital, que busca reducir costos y automatizar el pensamiento, hacia un modelo de soberanía algorítmica. Esto exige una alfabetización algorítmica profunda de los servidores públicos, permitiéndoles supervisar el dispositivo sin ser modulados por él. La propuesta final de este capítulo sostiene que la honorabilidad del mandato popular en la era digital depende de la capacidad para intervenir las organizaciones legislativas, protegiendo la integridad de la deliberación humana frente a la maquinaria de predicción conductual. Solo una gobernanza que valore lo cualitativo y lo territorial podrá garantizar que la ley siga siendo un acto de voluntad social soberana y que la tecnología sea, verdaderamente, una herramienta para la emancipación y no para la docilidad política.

En última instancia, regular la IA en el Poder Legislativo es un acto de recuperación de la soberanía algorítmica. No basta con establecer directrices éticas; es necesario profesionalizar al servidor público bajo una alfabetización algorítmica que le permita supervisar críticamente al dispositivo. Entonces, a través de una gobernanza digital situada, mediada por procesos de intervención organizacional que valoren lo cualitativo y lo situado, se podría garantizar que la ley siga siendo un producto de la voluntad humana y que la técnica permanezca siempre al servicio de la justicia social y la del mandato popular.

Bibliografía

- Allende, G. D. (2025). La construcción de evidencias organizacionales como estrategia de la intervención organizacional para el cambio. En O. Lozano Carrillo, Y. Zavala Osorio, & P. Couturier Bañuelos, *Reflexiones sobre la intervención en las organizaciones en la UAM*. Ciudad de México: REMINEO.
- Aguirre Quezada, J. P., & López Hernandez, G. (2025). La inteligencia artificial en la profesionalización de las y los servidores públicos: Oportunidades y riesgos en el desarrollo profesional. *Revista Internacional de Estudios sobre Sistemas Educativos*, 741-761.
- Aminahuel, A., & Rodriguez, M. (2024). Gobernanza de plataformas digitales y de inteligencia artificial. Contradicciones y concentraciones del poder tecnológico. *Universidad Nacional de Río Cuarto. Facultad de Ciencias Humanas. Departamento Ciencias de la Comunicación; Temas y Problemas de Comunicación*, 31-46.
- Bárcena, S. A., & Cardona, E. (06 de 10 de 2025). *Tecnológico de Monterrey*. Obtenido de "Actualícense ignorantes": El uso de la IA en el trabajo legislativo: <https://omd.tec.mx/noticia/actualicense-ignorantes-el-uso-de-la-ia-en-el-trabajo-legislativo>
- García Covarrubias, L. A. (2025). Episteme y dispositivos: Reflexiones del uso del PIAESS en contextos locales para MiPymes desde el pensamiento Foucaultiano. En C. O. Lozano, Y. Zavala Osorio, & D. P. Couturier Bañuelos, *EL PIAESS UNA IA PARA LA VIDA* (págs. 716-737). Ciudad de México: REMINEO AC.
- Gonzalez Montaña, A. (2023). El dispositivo digital en la era del capitalismo algorítmico: el caso de ChatGPT. *Sintaxis*, 33-48.
- González Montaña, A. (2022). La rebelión de las máquinas en la trama del capitalismo algorítmico: la democracia acechada. *Logos Revista de Filosofía*, 139-154.

- Guzman Noguera, R., Hernández Hernández, J. L., Hernández Hernández, M., Garduño Téliz, E., & Álvarez Hilario, V. (2024). GOBERNANZA DIGITAL: CONCEPTOS, MODELOS Y METODOLOGÍAS PARA SU IMPLEMENTACIÓN. *Innova Ingeniería, publicaciones técnicas de investigación*.
- IMT, R. (18 de febrero de 2026). *La IA como herramienta estratégica en la construcción legislativa*. Obtenido de Contacto Forum: <https://contactforum.com.mx/noticias/la-ia-como-herramienta-estrategica-en-la-construccion-legislativa/>
- Ramirez Navarro, E. (2025). La praxis interventora en las organizaciones en la era de la co-agencia. En O. Lozano Carrillo, Y. Zavala Osorio, & P. Couturier Bañuelos, *Reflexiones sobre la intervención en las organizaciones en la UAM*. Ciudad de México: REMINEO.
- RPD. (2026). *Thunder MX*. Obtenido de ¿Quién escribe la ley? La Inteligencia Artificial conquista la tribuna del Congreso mexicano: <https://thunder.mx/quien-escribe-la-ley-la-inteligencia-artificial-conquista-la-tribuna-del-congreso-mexicano/>
- Sadin, E. (2020). *La inteligencia Artificial o el desafío del siglo*. Buenos Aires: Caja Negra.
- Valenzuela Fernandez, L. A., Collantes Inga, Z. M., & Durand. (2020). Sobre la gobernanza digital, política digital y educación. *Eleuthera*, 88-103.

CAPÍTULO XX

La Inteligencia Artificial Aplicable En Los Procesos Jurisdiccionales Mexicanos

La transformación de los procesos jurisdiccionales de la mano de la gobernanza tecnológica

Artificial Intelligence Applicable in Mexican Jurisdictional Processes

*The transformation of jurisdictional processes hand in hand with
technological governance*

Mtro. Jesús Alberto Pérez Rodríguez

al2251801351@azc.uam.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0002-3667-2509>

Resumen

El presente capítulo analiza la incorporación de la Inteligencia Artificial (IA) en las funciones jurisdiccionales mexicanas y la necesidad urgente de construir un marco jurídico integral que regule su uso conforme a los principios del Estado de Derecho. Se sostiene que, aunque la IA representa una herramienta capaz de eficientar los procesos judiciales y fortalecer la justicia digital, su implementación no puede desarrollarse al margen de la Constitución, de los derechos humanos y de los principios procesales que rigen la función jurisdiccional.

El estudio parte de la premisa de que la actividad jurisdiccional debe permanecer bajo supervisión y decisión humana, ya que los actos judiciales requieren emanar de autoridades investidas constitucionalmente, conforme al principio de legalidad y al principio de inmediación. En consecuencia, la IA no puede sustituir al juzgador, sino únicamente auxiliarlo en tareas técnicas y operativas. A partir de ello, se examinan criterios jurisdiccionales mexicanos recientes que validan el uso de la IA como herramienta de apoyo, especialmente en cálculos matemáticos y gestión

procesal, siempre que exista proporcionalidad, transparencia, protección de datos personales, explicabilidad y supervisión humana efectiva.

Asimismo, el capítulo estudia tres instrumentos internacionales fundamentales para la construcción de una gobernanza algorítmica judicial en México: las Directrices Éticas para una Inteligencia Artificial Fiable de la Comisión Europea, el Reglamento de Inteligencia Artificial de la Unión Europea (AI Act) y la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO. A través de estos marcos regulatorios se identifican principios esenciales como la dignidad humana, la transparencia, la rendición de cuentas, la no discriminación, la privacidad, la supervisión humana y la sostenibilidad social y tecnológica.

Finalmente, se concluye que México requiere no solo una ley específica sobre IA, sino una reforma estructural y progresiva del sistema jurídico que permita integrar tecnologías inteligentes sin comprometer la constitucionalidad, la tutela judicial efectiva y los derechos fundamentales. La IA debe consolidarse como una herramienta auxiliar de gobernanza tecnológica y no como un sustituto autónomo de la función jurisdiccional, garantizando siempre la prevalencia de la supervisión humana y la protección de la dignidad de las personas dentro del sistema de justicia.

Palabras clave: Inteligencia Artificial, funciones jurisdiccionales, gobernanza algorítmica, justicia digital, derechos humanos, supervisión humana, Estado de Derecho, transparencia algorítmica, regulación jurídica, procesos judiciales

Abstract

This chapter analyzes the incorporation of Artificial Intelligence (AI) in Mexican jurisdictional functions and the urgent need to build a comprehensive legal framework that regulates its use in accordance with the principles of the Rule of Law. It is argued that, although AI represents a tool capable of making judicial processes more efficient and strengthening digital justice, its implementation cannot

be developed outside the Constitution, human rights and the procedural principles that govern the jurisdictional function.

The study is based on the premise that jurisdictional activity must remain under human supervision and decision, since judicial acts require the emanation of constitutionally vested authorities, in accordance with the principle of legality and the principle of immediacy. Consequently, AI cannot replace the judge, but only assist him in technical and operational tasks. Based on this, recent Mexican jurisdictional criteria that validate the use of AI as a support tool are examined, especially in mathematical calculations and procedural management, as long as there is proportionality, transparency, protection of personal data, explainability and effective human supervision.

The chapter also studies three fundamental international instruments for the construction of judicial algorithmic governance in Mexico: the European Commission's Ethical Guidelines for Trustworthy Artificial Intelligence, the European Union's Artificial Intelligence Regulation (AI Act) and UNESCO's Recommendation on the Ethics of Artificial Intelligence. Through these regulatory frameworks, essential principles such as human dignity, transparency, accountability, non-discrimination, privacy, human oversight, and social and technological sustainability are identified.

Finally, it is concluded that Mexico requires not only a specific law on AI, but also a structural and progressive reform of the legal system that allows the integration of smart technologies without compromising constitutionality, effective judicial protection and fundamental rights. AI must be consolidated as an auxiliary tool of technological governance and not as an autonomous substitute for the jurisdictional function, always guaranteeing the prevalence of human supervision and the protection of the dignity of people within the justice system.

Keywords: Artificial Intelligence, jurisdictional functions, algorithmic governance, digital justice, human rights, human supervision, rule of law, algorithmic transparency, legal regulation, judicial processes

La labor de los Juzgados y Tribunales en sus funciones jurisdiccionales se ha ido transformando de acuerdo con diversas modificaciones (derogaciones y abrogaciones) y de las heterogéneas reformas que se han realizado en las legislaciones adjetivas. El derecho procesal es fundamental para que un juzgado o tribunal puedan realizar sus funciones jurisdiccionales de manera eficaz y eficiente, pero siempre dentro de un ámbito de legalidad. La legalidad es uno de los principios más importantes del sistema jurídico, ya que es una barrera impenetrable donde la autoridad no puede sobreexcederse a nombre del interés público delimitando sus facultades de actuación, y permitiendo que las legislaciones protejan los derechos humanos y fundamentales de las personas; por lo tanto, este principio de legalidad es un elemento esencial de un Estado de Derecho.

Sin embargo, la tecnología ha avanzado a pasos gigantescos, y las funciones jurisdiccionales no deben de quedar rezagados al avance del desarrollo tecnológico. Una de las herramientas que ha surgido con el avance tecnológico es la Inteligencia Artificial (IA) que es un instrumento fundamental en diversas áreas, como en la educación, en la investigación, en la salud, en la economía, la cultura, la comunicación, entre otras. La pregunta es: ¿qué aplicación tiene la IA en los procesos jurisdiccionales?

Es inobjetable, que la IA ha venido a revolucionar el cómo funcionan las cosas. Sin embargo, la manera en cómo se utiliza la IA en las funciones jurisdiccionales debe estar conforme a la legislación procesal vigente y respetando los derechos sustantivos, por lo que la legislación procesal debe de regular la manera en cómo se utilizar la IA en las funciones jurisdiccionales. El problema se encuentra en que no existe normativa vigente que regule la inteligencia artificial en México, mucho menos una norma jurídica que regule el uso de la IA en las funciones jurisdiccionales.

Para ello, se abordarán tres instrumentos internacionales de referencia que contienen un alto valor heurístico que van encaminados a la construcción de un

modelo de gobernanza algorítmica judicial en México. Estos instrumentos internacionales de referencia deben ser considerados como parte de los cimientos que deben de regir las legislaciones mexicanas que regulen la IA, pero para ello, no basta con la promulgación de una legislación que regule la IA, sino que debe ser una reforma integral que transforme el sistema jurídico mexicano adaptándolo a los avances tecnológicos, específicamente con la integración de la IA en los procesos jurisdiccionales. Para que ello sea posible, se necesita la reconfiguración de procesos internos, relaciones de poder, medios de comunicación adecuados, entre otros cambios.

La Inteligencia Artificial Como Herramienta De Gobernanza Tecnológica Actual Y Futura

La Constitución Política de los Estados Unidos Mexicanos (CPEUM) (Cámara de Diputados del H. Congreso de la Unión, 2026) es la ley suprema de la nación (Art. 133 CPEUM), y todo tipo de legislaciones y regulaciones debe estar acorde con la misma. Así también, los tratados internacionales sobre derechos humanos tienen la misma jerarquía que la misma Ley Suprema; sin embargo, los derechos humanos deben ser interpretada conforme a los establecido en ella (Art. 1 CPEUM). De acuerdo con Hans Kelsen (1960) la norma fundamental de todo sistema jurídico es la Constitución, la cual da validez a todas las demás normas inferiores. La Constitución Federal, de acuerdo con la teoría constitucional, se divide en dos partes:

1. La parte dogmática, comprendida desde el artículo 1 hasta el 29 de la CPEUM, donde se limita el poder público, protegiendo la esfera jurídica de las personas y de la sociedad;
2. La parte orgánica, comprendida desde el artículo 30 hacia delante de la CPEUM, es donde se regula la forma de gobierno, los poderes, el federalismo, los organismos constitucionales autónomos, se organiza el poder público y se regula el ejercicio de la autoridad.

En la actualidad, los órganos jurisdiccionales están utilizando la Inteligencia Artificial (IA) para eficientar su operación. La integración de la IA en los organismos jurisdiccionales debe respetar la parte dogmática de la Constitución, protegiendo los derechos humanos de las personas; así mismo, debe ser estar acorde a la parte orgánica, respecto a la manera en que los organismos jurisdiccionales (juzgados y tribunales) están integrados y cumpliendo con su función bajo los parámetros constitucionales.

El artículo 16 de la Constitución Federal, en su primer párrafo, establece *“Nadie puede ser molestado en su persona, familia, domicilio, papeles o posesiones, sino en virtud de mandamiento escrito de la autoridad competente, que funde y motive la causa legal del procedimiento.”* Al establecerse que los actos de molestia deben provenir de una autoridad competente, investida de potestad pública (*imperium*), cuya autoridad debe estar conforme a lo establecido por la parte orgánica de la Constitución, necesariamente, el funcionario público (juez) que tiene la facultad de tomar decisiones jurisdiccionales, es de quien debe emanar los actos de molestia, por lo que los actos de molestia provenientes de la automatización tecnológica, sin participación de la persona que tiene la embestidura jurisdiccional, deben ser considerados inconstitucionales.

Eso quiere decir, que las herramientas tecnológicas, donde incluimos la IA, deben ser consideradas herramientas auxiliares en las funciones jurisdiccionales, y no como un agente jurisdiccional. En el artículo 20 de la CPEUM establece: *“El proceso penal será acusatorio y oral. Se regirá por los principios de publicidad, contradicción, concentración, continuidad e inmediación.”*; el principio de inmediación que debe regir el proceso penal es la garantía procesal que exige que los jueces o tribunales deben atender directa y personalmente el proceso jurisdiccional, sin intermediarios, por lo que pretender que cualquier agente de IA atienda un proceso judicial, en la figura de judicatura automatizada, debe ser considerado inconstitucional.

El Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito (2025), en la tesis II.2o.C.9 K (11a.), con número de registro 2031010, abordó el tema de la utilización de la IA en los procesos jurisdiccionales como herramienta para calcular el monto de una garantía vinculada con la inscripción de la demanda en los folios reales de los inmuebles reclamados, estableciendo que es necesario cumplir con los siguientes requisitos:

- a) Proporcionalidad e inocuidad. Estos principios son criterios normativos que ponen límites al ejercicio del poder público. Por una parte, la proporcionalidad exige que se debe de cumplir los elementos de idoneidad, necesidad y sentido estricto cuando la actuación de una autoridad afecte derechos humanos.

El elemento de idoneidad quiere decir que la medida adoptada por la autoridad jurisdiccional deber ser adecuada para lograr un fin legítimo; el elemento de necesidad establece que la medida debe ser la más eficiente y eficaz entre las demás alternativas; y el elemento del sentido estricto establece que el beneficio de adoptar la medida debe superar el perjuicio causado. Por otra parte, el elemento de inocuidad establece que la autoridad debe evitar causar daño o afectación indebida en los derechos humanos o a la esfera jurídica de las personas en todo tipo de actuaciones.

- b) Protección de los datos personales. Este principio impone la obligación al órgano jurisdiccional, como a toda autoridad, de guardar la privacidad, integridad, seguridad y uso legítimo de la información personal, especialmente cuando esta información permita identificar a una persona o los datos de un expediente judicial.
- c) Transparencia y explicabilidad. El principio de transparencia va dirigido a que toda autoridad debe cuidar que sus actuaciones sean accesibles, verificables y comprensibles para los gobernados; mientras que el principio de explicabilidad quiere decir que las decisiones y los procesos que adopten las autoridades deben ser comprensibles, justificadas y reconstruibles.

En la tesis aislada mencionada establece: *“impone que la persona juzgadora no sólo informe que utiliza herramientas de inteligencia artificial, sino que exponga la metodología, los datos empleados y la forma en que se arribó al resultado, a fin de permitir que el procedimiento sea auditado y comprendido por las partes y órganos revisores”*.

- d) Supervisión y decisión humana. Este principio establece que toda medida tecnológica tomada en las decisiones jurisdiccionales debe operar bajo el escrutinio efectivo de un ser humano, en este caso del juez (respetando el principio de inmediación). Por lo tanto, la IA no debe de participar en el proceso y en las decisiones jurisdiccionales como un agente tomador de decisiones, sino como una herramienta que auxilie al juzgador en sus determinaciones y no que supla sus funciones.

En la actualidad, el Poder Judicial mexicano cuenta con la infraestructura de “justicia digital”, al contar con un portal web donde se puede consultar expedientes electrónicos a los que se tiene autorización, generar una cita para trámites de diligencias en la central de actuarios, presentar una demanda o consultar el estado que guarda esta, presentar y consultar promociones, generar una solicitud de amnistía, firmar documentos con la firma electrónica del Poder Judicial, entre otros.

Sin embargo, el avance tecnológico de la justicia digital no ha caminado a la par con las adecuaciones en la legislación que regule el uso sistematizado de la IA en los procesos judiciales. En la actualidad, no existe una legislación que regule el uso de la IA en México; sin embargo, existe la Iniciativa de Ley Nacional que Regula el Uso de la Inteligencia Artificial presentado ante el Senado de la República el 29 de octubre de 2024 por la senadora Juanita Guerra Mena (Cámara de Diputados, 2024). Antes de esta iniciativa de ley, el diputado Ignacio Loyola Vera presentó una iniciativa para expedir la Ley de Regulación Ética de la Inteligencia Artificial y la Robótica (Cámara de Diputados, 2023), y otro intento de regular la IA es con la Iniciativa Con Proyecto De Decreto Por El Que Se Expide La Ley Federal Para El Desarrollo Ético, Soberano E Inclusivo De La Inteligencia Artificial (Cámara de Diputados, 2025).

Recientemente, fue publicado el Acuerdo General AG-POAJ-006/2026 (Acuerdo General del Pleno del Órgano de Administración Judicial que regula la integración y trámite del expediente electrónico y el uso de videoconferencias en todos los asuntos competencia de los órganos jurisdiccionales a cargo del Órgano) (Órgano de Administración Judicial, 2026) proveniente del Pleno del Órgano de Administración Judicial, publicado en el Diario Oficial de la Federación el 17 de abril de 2026. En este acuerdo, se establece que en todos los asuntos que los órganos jurisdiccionales conozcan deberán de integrar un expediente electrónico que tendrá la misma validez que el expediente físico, dicho expediente podrá ser consultado por los interesados autorizados sin necesidad de presentarse en el Juzgado o Tribunal; por lo que todo documento que se presente en forma física será digitalizada e incorporado al expediente electrónico. También se les da validez a las videoconferencias siempre que sea posible identificar a los intervinientes y que exista fidelidad en la comunicación.

En un estudio realizado en los Poderes Judiciales de 32 entidades federativas y del Poder Judicial Federal se obtuvieron los siguientes resultados: 31 de las 32 entidades federativas (excepto Oaxaca) permiten consultar las “listas de acuerdos” dentro de su portal digital, 10 entidades federativas aceptan la FIREL (Firma Electrónica del Poder Judicial de la Federación), 6 entidades aceptan la e.firma (SAT) y 6 entidades han desarrollado su propia firma electrónica del Poder Judicial. El Poder Judicial de la Ciudad de México (PJCDMX) expide su propia firma electrónica, pero también acepta la FIREL y la e.firma del SAT (Álvarez & Soberanes Díez, 2025). De acuerdo con los datos anteriores, se puede demostrar que no existe las mismas condiciones entre los diversos organismos jurisdiccionales de las distintas entidades federativas, por lo que su regulación nacional no partiría desde un piso parejo de infraestructura, por lo que su regulación debería ser gradual y ad-hoc a las condiciones actuales.

Ha existido un buen avance en la instalación de un “Tribunal Cero Papel”, donde todo tipo de promociones (documentos) se digitalizan y se sube al expediente electrónico del Poder Judicial. Para la presentación de todo tipo de

promociones, no solo el abogado postulante debe contar con su firma electrónica autorizada, sino que también debe contar con ella el promovente (la parte en litigio = cliente). El problema existe en que, por una parte, los abogados postulantes cuentan con su firma electrónica ya que su uso es parte de sus actividades cotidianas; sin embargo, las partes de litigio no cuentan con su firma electrónica, y para poderla obtener pueden pasar varios meses antes de tener oportunidad de concretar una cita para obtenerla.

La IA también es utilizada por los órganos jurisdiccionales para realizar cálculos matemáticos para calcular el monto de garantías que se fijan en los juicios de amparo. El Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito (2026), en la tesis II.2o.C. J/1 K (12a.), con número de registro 2031639, abordó este tema, dando validez al uso de la IA como herramienta que auxilia a los órganos jurisdiccionales en el razonamiento numérico, realizando cálculos de operaciones técnicas como cálculo de garantías, actualización de valores, aplicación de tasas de interés y ponderación de plazos procesales.

El uso de la IA como herramienta auxiliar, de acuerdo con esta tesis en comento, genera diversos beneficios como la reducción de errores al minimizar la posibilidad de error humano, genera trazabilidad en el procedimiento de cómo se obtuvo cada resultado, consistencia en procedimientos estandarizados y eficientar el tiempo del personal jurisdiccional. Sin embargo, los catedráticos que nos hemos acercado a la IA para realizar operaciones matemáticas complejas encontramos diversos errores en el cálculo, haciendo que esta herramienta no sea infalible, requiriendo necesariamente la intervención del personal capacitado en las operaciones matemáticas. Es decir, la IA para realizar operaciones matemáticas simples es una gran herramienta; pero, para realizar operaciones matemáticas complejas, donde se deben integrar criterios que no necesariamente son matemáticos, es donde la IA presenta su debilidad.

Estos dos criterios jurisdiccionales antes mencionados (tesis), son aplicables única y exclusivamente cuando los órganos jurisdiccionales utilizan la IA como

herramienta de cálculo matemático; sin embargo, deja ver el vacío normativo en el uso de la IA para otras actividades judiciales como su uso en la redacción de una síntesis de todo expediente judicial a estudio antes de emitir una sentencia, la búsqueda de precedentes aplicables o fundamentos legales aplicables al caso concreto, el análisis de las pruebas de manera sistematizada con todo el caudal del expediente judicial, la redacción de sentencias, etc. Respecto a este tema, al día de hoy, existe un sistema de IA desarrollada por el Poder Judicial del Estado de Querétaro (SONIA), donde se emitió la primera sentencia asistida por IA de un juicio de divorcio incausado. Se tiene programado que a partir del 4 de mayo de 2026 el sistema de IA SONIA operará con regularidad en el en el Juzgado Noveno Civil y el Juzgado Sexto Familiar del Estado de Querétaro (Hernández, 2026). Tampoco, en la actualidad, se ha abordado el tema de auditabilidad del proceso judicial asistido por la IA.

Lo anterior no quiere decir que el uso de la IA en los procesos jurisdiccionales sea algo malo, sino que en un Estado de Derecho es importante que se garantice el respeto, en todo el actuar de los órganos jurisdiccionales, de la legalidad de sus actos, evitando actos arbitrarios, inseguridad jurídica y nulidad de actos judiciales. El actualizar el sistema normativo mexicano a los avances tecnológico, como es el caso de la IA, buscaría efficientar los procesos judiciales buscando una justicia pronta y expedita, pero dentro de la legalidad del actuar de los órganos jurisdiccionales en el respeto de las normas jurídicas vigentes, pero actualizadas a las condiciones actuales.

La cuestión es saber qué parámetros debería de cumplir la IA para que, en su implementación, en cada uno de los procesos jurisdiccionales, sea legal, ética y responsable, con una perspectiva de protección a los derechos humanos. El Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito, mediante la tesis jurisprudencial II.2º.C. J/2 K (12ª.) (2026) menciona la conveniencia del uso referencial de las regulaciones internacionales sobre IA en las funciones jurisdiccionales, ante la ausencia de regulación nacional sobre estos tema, entre las que se encuentran las Directrices Éticas para una Inteligencia Artificial Fiable

del Grupo de expertos de alto nivel sobre inteligencia artificial creado por la Comisión Europea; el Reglamento del Parlamento Europeo del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial; y la Recomendación sobre la ética de la inteligencia artificial de la UNESCO y sus principios éticos, como uso referencial

Las Directrices Éticas para una Inteligencia Artificial Fiable del Grupo de expertos de alto nivel sobre inteligencia artificial creado por la Comisión Europea (2019) establece un marco que garantiza que el desarrollo y uso de la IA se de manera confiable, responsable y alineada a los valores fundamentales de las sociedades democráticas. Ronald Dworkin (1996) comenta que la democracia no solo significa “la regla de la mayoría”, sino que la democracia debe de cumplir un cúmulo de principios éticos y respeto de derechos individuales, entre los que se encuentran la libertad, igualdad, fraternidad, pluralismo y justicia.

Las Directrices éticas para una IA fiable (2019) establece que, las directrices en el diseño, desarrollo, implementación, uso y desmantelamiento de la IA no solamente debe ser técnicamente eficiente, sino que deben ser jurídicamente conforme a los derechos fundamentales, éticamente fundada y ser socialmente segura, evitando riesgos de discriminación algorítmica, manipulación en la información, exclusión de grupos vulnerables, invasión en la privacidad de las personas, e impactos negativos en la sociedad. Estos tres elementos para la fiabilidad de la IA (legalidad, ética y seguridad social), por sí solo no garantizan la fiabilidad de la IA, sino que se garantiza en el momento en que existe interdependencia entre estos elementos.

Las Directrices éticas para una IA fiable (2019) establece diversos principios de primer nivel que deben de cumplirse en el diseño, desarrollo, implementación y uso de sistemas de IA, entre los cuales establecen los siguientes derechos:

1. Dignidad humana. Este valor establece que cada persona posee por sí sola un valor intrínseco, por lo que no deben ser tratadas como objetos o datos,

sino como sujeto autónomo cuya identidad e integridad debe estar garantizada en el uso de la IA.

2. Libertad humana. Este derecho está directamente relacionado con la autonomía y autodeterminación en la toma de decisiones, las cuales no deben estar influenciadas por la IA a manera de manipulación, y parte de esta libertad establece que todo resultado de la IA debe estar sujeto a la intervención humana (supervisión y vigilancia).
3. Democracia. Este derecho se refiere a los procesos institucionales y a la garantía de la participación de los ciudadanos en los asuntos públicos; la IA debe garantizar la no intervención en asuntos públicos no influenciando a la existencia de una participación sesgada por intereses particulares.
4. Igualdad. Este principio de igualdad, junto con el de equidad, van de la mano; una frase describe su interrelación: “Tratar igual a los iguales y desigual a los desiguales”. La accesibilidad a la IA debe tener un piso parejo para todas las personas (igualdad), pero debe de dar un trato privilegiado a las personas cuyas condiciones de accesibilidad sea precaria (equidad).

A partir de estos derechos se derivan los siguientes principios éticos esenciales:

1. Respeto de la autonomía humana. Este principio establece que la IA debe de ser una herramienta para que el ser humano pueda potenciar sus capacidades y habilidades, y no como un arma que las restrinja o disminuya; es decir, la IA debe estar al servicio de las personas humanas.
2. Prevención del daño. Se establece que la IA debe estar programada para detectar riesgos y disminuir la posibilidad de causar un daño de cualquier tipo a cualquier persona, grupo de personas o a la sociedad en general.
3. Explicabilidad. Este principio es clave para generar confianza entre los diversos interesados, donde sus resultados deben ser transparentes y con una trazabilidad robusta de la información y los datos utilizados para formular su respuesta. Todas las respuestas deben ser comprensibles e interpretables para los usuarios para que tengan la información suficiente y

general para la toma de decisiones de manera autónoma. Al mencionar que se tenga una información general quiere decir que todos los interesados tengan acceso a la misma información, reduciendo los sesgos interpretativos generándose una manipulación mal intencionada.

En el segundo nivel de las Directrices éticas para una IA fiable (2019) establece los requisitos operativos esenciales para todos los ciclos del sistema de Inteligencia Artificial, que se refiere a una combinación de métodos técnicos (como algoritmos auditables o mecanismos de seguridad) y no técnicos (como gobernanza, regulación y formación). La aplicación de estos mecanismos es una responsabilidad conjunta compartida entre todos los sujetos relacionados con la IA, desde los desarrolladores, los usuarios, los comerciantes y la sociedad en general; es decir, se considera el ambiente sociotécnico en la que la IA interactúa con los seres humanos. Dentro de estos requisitos se encuentran los siguientes:

1. Acción y supervisión humana. Este requisito establece que la IA debe permitir la autonomía y la autodeterminación de las personas humanas, respetando en todo tiempo sus derechos fundamentales, y donde el resultado de uso de la IA permita ser supervisada por los usuarios interesados. Este requisito tiene como propósito el garantizar la gobernanza humana, evitando que los productos de la IA creen escenarios de “acciones realizadas e irreversibles” donde los seres humanos no puedan cambiar los resultados que generen algún tipo de afectación a un individuo, grupos vulnerables o a la sociedad en general.
2. Solidez y seguridad técnica. Este principio establece que la IA debe estar programada para evitar algún riesgo para la sociedad, ya sea por daños voluntarios o involuntarios. Parte del requisito de seguridad técnica es establecer que la IA debe ser resiliente ante manipulaciones maliciosas o ataques que intenten engañar al modelo para obtener un resultado dañino para la sociedad. Para evitar esto, la IA debe tener un sistema de seguridad que detecte cualquier falla, pueda detener el sistema e identificar por qué ocurrió el error.

3. Gestión de la privacidad y los datos. Los sistemas de IA deben garantizar la confidencialidad y privacidad de las personas, protegiendo sus datos que no deban ser considerados públicos, respetando la intimidad de las personas. Los usuarios de la IA deben conocer previamente como serán utilizados sus datos al momento de interactuar con el sistema. Por ello, las respuestas de la IA deben de garantiza la calidad e integridad de los datos proporcionados, los cuales deben evitar cualquier sesgo que puedan perjudicar a las personas al tomar decisiones discriminatorias, deben ser verídicas al utilizar datos exactos y actualizados; y la información debe ser relevante exclusivamente donde se utilicen solo datos que sean necesarios para proporcionar la respuesta exacta a lo solicitado por el usuario.
4. Transparencia. Todos los sistemas de IA deben diseñarse como herramientas aptas para el escrutinio público, donde se garantice la trazabilidad de la información, la explicabilidad y comprensibilidad de las decisiones tomadas o resultados obtenidos por al IA; y, una restricción en la comunicación donde el ser humano debe saber que esta interactuando con un producto de IA.
5. Diversidad, no discriminación y equidad. Los sistemas de IA deben ser inclusivos y accesible para todas las personas, evitando que exista una brecha social en su acceso. Para ello, es importante que los sistemas de IA estén contruidos por un diseño universal (diseño para todos) donde las personas marginadas y con las diversas discapacidades puedan acceder, donde la IA se acople a las condiciones de las personas y no las personas se acoplen a las condiciones de la IA.
6. Bienestar social y ambiental. Este requisito exige que la IA no solo sea evaluada por el impacto individual en el usuario, si no en el impacto social provocado. Entre ello se encuentra en impacto en la ecología y el medio ambiente (por ejemplo, el consumo de agua para enfriar las supercomputadoras, o los desechos de computadoras y su impacto en los suelos, etc.); adicionalmente, el impacto social cuando la IA pudiere ser

utilizada como herramienta de desinformación, manipulación de la opinión pública, siendo un factor generador de polarización que derive en conflictos sociales. Y también en el impacto individual en las afectaciones psicológicas y emocionales a la persona individual, deteriorando su salud mental.

7. Rendición de cuentas. Este último punto, cierra el círculo ético de una IA fiable, al abordar que cada sistema de IA debe tener aparejada herramientas de rendición de cuentas y de fincar responsabilidades por los resultados de los sistemas de IA. Para ello, debe existir un protocolo de auditabilidad donde se evalúa los algoritmos, los datos con que cuenta el sistema y la manera en cómo fue diseñado el sistema de IA. Lo anterior, tiene como objetivo el minimizar y notificar los impactos negativos, los daños colaterales creados; además de contar con un protocolo para mitigar los daños.

En un tercer nivel de las Directrices éticas para una IA fiable (2019) establece un listado de evaluación de la fiabilidad de la IA, para que los distintos usuarios (desarrolladores, responsables de la implementación, usuarios institucionales y autoridades reguladoras) tengan una herramienta de gobernanza aplicada al interrogar sistemáticamente a un sistema de IA respecto a sus parámetros de ética, técnica y responsabilidad social. Es decir, se constituye una herramienta de auditoría ética para cada uno de los procesos del sistema de inteligencia artificial, como un sistema de control continuo del ciclo de vida del sistema, desde el diseño hasta el uso de los resultados de la IA.

Esta herramienta es un conjunto de preguntas orientadoras donde se cuestionan si el sistema de IA cuenta con un sistema de supervisión humana efectiva, la calidad de los datos utilizados para generar una respuesta automatizada existe explicabilidad de los resultados hacia los usuarios y que mecanismos de responsabilidad y rendición de cuentas cuenta el sistema de IA. Adicionalmente, esta auditabilidad no se realiza de manera aislada, sino que es esencial la

participación de las partes interesadas (stakeholders). Este mecanismo de evaluación actúa como un dispositivo de gobernanza participativa.

Por otra parte, el Reglamento del Parlamento Europeo del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (mejor conocido como Reglamento de Inteligencia Artificial) (Unión Europea, 2024) es un marco jurídico aplicable a todos los países que integran la Unión Europea (UE) donde se regula el diseño, desarrollo, comercialización, servicio y uso de los sistemas de IA. Este Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo (AI Act) tiene dos principales objetivos. Primero es el garantizar la protección de las personas en sus derechos fundamentales (salud, seguridad y valores democráticos); y segundo, promover la innovación, competitividad y correcto funcionamiento del mercado.

El Reglamento (UE) 2024/1689 del Parlamento Europeo (Unión Europea, 2024) parte de la premisa que la IA debe estar al servicio del ser humano, como herramienta de bienestar humano y social. Pero también reconoce que la IA tiene un alto potencial de impacto social, económico y ambiental, por lo que ese impacto tiene aparejado riesgos (como convertirse en una herramienta autónoma de control y dominación social), por lo cual, los sistemas de IA deben ser controlados, auditados y evaluados. Estos riesgos son clasificados de acuerdo con el impacto negativo que pudiera generar, por lo que existen practicas prohibidas, sistemas de alto riesgo, sistemas con obligaciones específicas de transparencia, y sistemas de riesgo mínimo o bajo.

Las prácticas prohibidas, es la categoría más restrictiva del Reglamento (UE) 2024/1689 (2024) que restringen absolutamente el desarrollo y uso de sistemas de IA nocivos, por ser considerados contrarios a las normas jurídicas europeas, y se trata de proteger los bienes jurídicos fundamentales que no admiten ponderación (protección a la dignidad humana, a la autonomía personal, a la autodeterminación, la igualdad, a la equidad, a la no discriminación, a la integridad de los procesos democráticos, etc.). Esta estrictamente prohibido, sistemas de IA que manipulen el

comportamiento humano mediante prácticas engañosas y técnicas subliminales, al producirse una interferencia ilegítima en las decisiones personales afectando su autonomía y autodeterminación.

También son prácticas prohibidas la explotación de las vulnerabilidades de determinadas personas o grupos, que de acuerdo con su situación de la realidad, puedan ser influenciadas y manipuladas para comportarse de cierta manera en beneficio de intereses mezquinos, afectándose el derecho a la igualdad material, equidad y no discriminación. Asimismo, se encuentra prohibido el sistema de “puntuación social” (social scoring), el cual es un sistema que clasifica a las personas de acuerdo con sus comportamientos, características personales y consumo, generándose perfiles que pueden ser utilizados como una herramienta de control social generalizado, forjándose una situación de discriminación estructural y exclusión injustificada.

Este social scoring es peligroso, ya que cada perfil tendrá acceso a datos e información adhoc que son creados de acuerdo con su propio perfil, y no mediante una respuesta generalizada para cualquier tipo de usuario. Igualmente está prohibido los sistemas de identificación biométrica remota en tiempo real en espacios públicos, con sus excepciones. Este sistema permite identificar a personas sin su consentimiento, afectándole su derecho a la libertad, privacidad, protección de datos personales, libre tránsito, entre otros. Por lo que esta restricción está enfocada a la defensa de una gobernanza algorítmica orientada al respeto de los derechos fundamentales.

Por otro lado, el Reglamento (UE) 2024/1689 (2024) establece que los sistemas de IA al alto riesgo operarán bajo límites estrictos y cumpliendo con ciertos requisitos sustantivos, procedimentales y organizacionales que buscan garantizar su confiabilidad y control. El alto riesgo en el uso de un sistema de IA no se encuentra en el sistema per se, sino en la manera y fines en que se utilizarán los resultados del sistema. Por ejemplo, los servicios de IA utilizados en la selección de personal, donde un agente de IA deja de ser una herramienta auxiliar y se

convierte en un agente autónomo que toma sus propias decisiones, sin supervisión humana, con consecuencias jurídicas y materiales relevantes. Este régimen de alto riesgo responde al principio de mitigar los daños y perjuicios en los derechos fundamentales, así como auditar y controlar el poder algorítmico.

El Reglamento (UE) 2024/1689 (2024) cataloga como sistemas de IA sujetos a obligaciones específicas de transparencia al uso de la IA donde exista un riesgo relevante de asimetría informativa, engaño o manipulación. Para mitigar este riesgo y respetar la autonomía informada, los usuarios deben tener la información sobre el entorno tecnológico en el que interactúan con el objetivo de que puedan tomar decisiones autónomas e informadas. Dentro de esta categoría se encuentran los chatbots o asistentes virtuales, que son herramientas de IA que interactúan con las personas; por lo que es esencial que la herramienta de IA informe a las personas que su interacción se está dando con una herramienta de IA. También en esta categoría se encuentra los deepfakes (sistemas que manipulan contenido sintético, como imágenes, audio y video), los cuales pueden poner en riesgo la reputación de las personas, la integridad de la información y la estabilidad de los procesos democráticos al crear representaciones falsas que simulan la realidad con alto grado de verosimilitud, por lo que es esencial que se informe que el contenido presentado fue generado por IA.

En el nivel de riesgo mínimo o de bajo riesgo, de acuerdo con el Reglamento (UE) 2024/1689 (2024) se encuentra aquellos sistemas que no generan riesgos significativos a los derechos fundamentales y a los intereses de la sociedad. En este nivel bajo de riesgo, el legislador europeo no le impuso cargas significativas con el objetivo de incentivar la innovación y desarrollo tecnológico. En este nivel se encuentran productos como filtros de spam, videojuegos con IA, asistentes básicos, entre otros, los cuales su nivel de riesgo es marginal o residual. Lo anterior no quiere decir que se deje de regular a los sistemas de IA de riesgo mínimo, más bien que sus limitaciones y cargas que deben de cumplir se reducen, pero existen mecanismos de autorregulación donde los actores del ecosistema de la IA deben

apegarse a las buenas prácticas alineadas a los principios éticos de la IA fiables (Códigos de conducta voluntarios).

Los códigos de conducta abordan temas relacionados a la transparencia, equidad, protección de datos personales, sostenibilidad, supervisión humana, entre otros. Este código de conducta tiene como objetivo el cumplimiento anticipado de estándares de regulación más estricta, con el objetivo de los desarrolladores sigan realizando innovaciones en sus productos de IA que previamente han cumplido con sus obligaciones más estrictas de minimización del riesgo, dando un paso significativo a la gobernanza regulatoria.

Una de las definiciones interesantes que contiene el Reglamento (UE) 2024/1689 (2024) se encuentra el de “sistema de IA” que a la letra dice:

“Una característica principal de los sistemas de IA es su capacidad de inferencia. Esta capacidad de inferencia se refiere al proceso de obtención de resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos y virtuales, (...) Las técnicas que permiten la inferencia al construir un sistema de IA incluyen estrategias de aprendizaje automático que aprenden de los datos cómo alcanzar determinados objetivos y estrategias basadas en la lógica y el conocimiento que infieren a partir de conocimientos codificados o de una representación simbólica de la tarea que debe resolverse. La capacidad de inferencia de un sistema de IA trasciende al tratamiento básico de datos, al permitir el aprendizaje, el razonamiento o la modelización.

(...) los sistemas de IA pueden funcionar con arreglo a objetivos definidos explícitos o a objetivos implícitos (...) Los sistemas de IA están diseñados para funcionar con distintos niveles de autonomía, lo que significa que pueden actuar con cierto grado de independencia con respecto a la actuación humana y tienen ciertas capacidades para funcionar sin intervención humana. La capacidad de adaptación que un sistema de IA podría mostrar tras su despliegue se refiere a las capacidades de

autoaprendizaje que permiten al sistema cambiar mientras está en uso. Los sistemas de IA pueden utilizarse de manera independiente o como componentes de un producto, con independencia de si el sistema forma parte físicamente del producto (integrado) o contribuye a la funcionalidad del producto sin formar parte de él (no integrado).”

El Reglamento (UE) 2024/1689 (2024) tiene un alcance extraterritorial ya que no solo es aplicable en el territorio de la Unión Europea, sino también es aplicable a aquellos terceros países que exporten servicios de IA a la Unión Europea y produzcan sus efectos en su territorio de la UE. Este reglamento no solo desea regular las cuestiones técnicas y económicas de la IA, sino también desea regularla desde un enfoque políticos y social en la preservación de un orden democrático y en la protección de los derechos fundamentales.

El Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito, mediante la tesis jurisprudencial II.2º.C. J/2 K (12ª.) (2026) establece un tercer documento regulatorio propuesto son la “Recomendación sobre la ética de la inteligencia artificial de la UNESCO y sus principios éticos” (UNESCO, 2021). Este instrumento normativo global de carácter ético en el uso de la IA se encuentra redactado desde una perspectiva de protección a los derechos humanos. Establece que los sistemas de la IA deben buscar como último fin el beneficio de la humanidad, por lo que debe garantizar la prevención de daños y la protección de los derechos fundamentales como la dignidad humana, la justicia social, la equidad y la sostenibilidad ambiental.

La UNESCO plantea que la IA es una tecnología de alto impacto sistémico, esto quiere decir que la IA no tiene un impacto aislado, sino que la IA es un dispositivo sistémico que puede impactar diversos subsistemas sociales. La IA tiene impacto en múltiples planos, como en el ámbito social, económico, cultural, político, educativo, entre otros. Por otra parte, el carácter sistémico de la IA está directamente relacionado a la producción, procesamiento y utilización de la información, que es uno de los elementos fundamentales en una democracia

moderna; por lo que la IA mediante la información que construye y proporciona tiene la capacidad de influir en la toma de decisiones, transformar las relaciones de poder, privilegiar ciertas áreas del conocimiento y la manera en cómo se interactúa en la sociedad. Lo anterior, genera un beneficio al poder mejorar la calidad de vida, el fomentar y acelerar el desarrollo tecnológico y científico, entre otros beneficios; sin embargo, también conlleva riesgos como el sesgo en la información, discriminación y afectación de los derechos humanos y fundamentales.

La recomendación sobre la ética de la inteligencia artificial de la UNESCO y sus principios éticos (UNESCO, 2021) tiene como objetivo el promover que el uso de la IA se encuentre al servicio de la humanidad, por lo que proporciona un marco universal de referencia en el diseño, regulación, desarrollo, aplicación y desmantelamiento de la IA desde una perspectiva ética. El marco de referencia normativa del uso de la IA de manera ética está relacionado con la protección de los derechos y la dignidad humanos, reducción de las desigualdades y brechas digitales, incentivar la cooperación internacional, protección del medio ambiente y las generaciones futuras. Con ello, se evidencia la gobernanza global donde la IA no es considerada solamente como una herramienta tecnológica, sino también como un fenómeno sociotécnico con influencia en la estructura social.

La recomendación (UNESCO, 2021) distingue entre valores y principios. Los valores son ideas orientadoras, mientras que los principios son directrices operativas. Dentro de los valores se encuentran el respeto, protección y promoción de los derechos humanos, evitando la discriminación y cosificación de las personas; la protección del medio ambiente y los ecosistemas fomentando la sostenibilidad, sustentabilidad y la minimización de la huella ecológica; se reconoce la diversidad y la inclusión evitando los sesgos tecnológicos y las desigualdades estructurales; así como también, la construcción de sociedades justas, pacíficas e interconectadas por la IA, fomentando así la cohesión social, la equidad y la solidaridad global.

Dentro de los principios éticos de la IA, que forman parte de la parte operativa de la recomendación, existen acciones concretas que se deben de cumplir, entre las que se encuentran las siguientes:

1. Proporcionalidad e inocuidad. Este principio exige que la IA deba ser utilizada siguiendo un objetivo legítimo, evitando todo tipo de daños a las personas, a los grupos sociales, a la sociedad o al medio ambiente.
2. Seguridad y protección. Se establece que los sistemas de IA deben estar contruidos con un sistema que prevenga daños, riesgos y perjuicios en todo el ciclo de vida de la IA.
3. Equidad y no discriminación. Este principio establece que la IA debe de mitigar la reproducción de desigualdades estructurales, privilegiando la inclusión y el acceso efectivo y equitativo de los beneficios producidos por la IA.
4. Sostenibilidad. La IA debe estar acorde al desarrollo sostenible evaluando su impacto social, cultural, económico y ambiental.
5. Derecho a la privacidad y protección de datos. Debe respetar el derecho humano a la privacidad y confidencialidad, por lo que, desde su diseño, los sistemas de IA deben de proteger los datos personales de los múltiples usuarios.
6. Supervisión y decisión humana. Este principio establece que el sistema de IA debe estar bajo la vigilancia humana, y la manera en cómo se usa los productos de la IA deben quedar bajo su más estricta responsabilidad y bajo la decisión de los seres humanos.
7. Transparencia y explicabilidad. Este principio establece que los sistemas de IA deben ser comprensibles, interpretables, auditables, verificables y trazables, por lo que los usuarios deben conocer cómo fue construida la respuesta de IA y como fue supervisada desde el diseño.
8. Responsabilidad y rendición de cuentas. Este principio está directamente relacionado a la trazabilidad de la información y de la transparencia en la

metodología utilizada en la obtención de los resultados. Para ello, es necesario tener herramientas de auditoría, supervisión y control.

9. Sensibilización y educación. Se requiere una estrategia de alfabetización digital y promoción de educación tecnológica con el objetivo de minimizar el sesgo estructural de la sociedad en el uso de la IA.
10. Gobernanza multiactor y adaptativa. Se establece una gobernanza global multiactoreal permitiendo la participación de los diversos interesados para la regulación de la IA estandarizada.

Por último, la recomendación sobre la ética de la inteligencia artificial de la UNESCO y sus principios éticos (UNESCO, 2021) establece mecanismo de implementación de estas recomendaciones, para que no solo se quede en el ámbito declarativo, sino que establece mecanismos de evaluación de la ética en los sistemas de IA, sugiere marcos regulatorios interno y globales, mecanismos de cooperación internacional, y especificaciones de regulación de sectores estratégicos como la economía, salud, educación y cultura. Por lo que estas recomendaciones abordan tres dimensiones: 1. Axiológica (valores universales), 2. Normativa (principios éticos operativos) y 3. Política (instrumentos de implementación).

Es fundamental, por lo tanto, la adecuación del sistema jurídico mexicano a los avances tecnológicos, y específicamente a la adopción de la IA como una herramienta que pueda eficientar las funciones jurisdiccionales, beneficiando el principio de recibir una justicia pronta y expedita. Como se ha observado anteriormente, la IA debe ser utilizada como una herramienta de apoyo a las funciones jurisdiccionales, pero una persona humana debe ser quien realice la función de supervisar y auditar el funcionamiento del sistema de IA. Sin embargo, yo sigo viendo un futuro donde la gobernanza digital va a ir ganando terreno con los avances tecnológicos a pasos agigantados.

En un futuro, que tal vez no muy lejano, y con las adecuaciones legislativas pertinentes, existirá en primera instancia un Juez Virtual, que en realidad es un agente de inteligencia artificial que este debidamente entrenado para la aplicación

estricta de las legislaciones provenientes del poder legislativo, y que pueda solucionar de manera técnica, y tal vez no humanista en primera instancia, los diversos conflictos planteados. En este caso, los seres humanos solo serán auditores y verificadores de las determinaciones que el agente de IA realice, cuidando en todo momento el respeto a los derechos humanos y que todos los procesos de los sistemas de IA estén acorde a los principios éticos planteados. Con esta implementación se tendría procesos judiciales demasiado eficientes, llegando a una sentencia en tiempo límite, y dejando para una segunda instancia los procesos largos cuando alguna de las partes considere que le fue vulnerado algún derecho fundamental.

Conclusión

En la actualidad, aún no existe una legislación que regule el diseño, implementación, desarrollo, uso y desmantelamiento de los sistemas de IA. Aunque han existido diversos intentos en el Congreso de la Unión por regular este tema, hasta este momento ninguno de ellos ha llegado a promulgarse y publicarse en el Diario Oficial de la Federación para que pueda ser considerado de aplicación obligatoria y estricta.

Sin embargo, aunque aún existen iniciativas de ley que pretenden regular la IA, de llegarse a promulgar, esta no será suficiente para la regulación completa que necesita un estado de derecho. Para ello, debe de existir una reforma estructural del sistema jurídico mexicano para que dicha regulación pueda ser completa, sistemática e integral. Es innegable los beneficios en el uso de la IA en diversos ámbitos de la vida pública, como en la educación, cultura, economía, política, etc. Sin embargo, existe un área en la que no es fácil su adopción, y esta es el área judicial.

Las autoridades jurisdiccionales, en un estado de derecho, deben de respetar la CPEUM y los tratados internacionales de protección de derechos humanos, en primera instancia. Pero también existen legislaciones subjetivas donde se establecen los derechos, obligaciones, responsabilidades, infracciones y

sanciones que son aplicados en las funciones jurisdiccionales con el objetivo de buscar una justicia eficaz y expedita. Pero los juzgados y tribunales deben de seguir ciertos procesos y procedimientos establecidos en las legislaciones adjetivas, con el objetivo de que su función jurisdiccional no sea arbitraria y este acorde al estado de derecho que ya se ha mencionado anteriormente. Por lo que el uso de la IA en las funciones jurisdiccionales debe ser tratado con delicadeza con el objetivo de que las determinaciones judiciales no sean considerados inconstitucionales, inconvencionales o ilegales.

A pesar de lo anterior, existe criterios jurisdiccionales mediante diversas tesis donde se acepta el uso de la IA con sus limitantes, donde la IA no pueda reemplazar las decisiones humanas, sino que sea utilizada como una herramienta auxiliar en las funciones jurisdiccionales, pero haciendo efectiva la tutela judicial. Sin embargo, en la actualidad existen normas internacionales de referencias sobre el uso ético de la IA que son aceptadas por los juzgados y tribunales mexicanos, abriendo la puerta al uso de la IA en las funciones judiciales.

Referencias

- Álvarez, C.-L., & Soberanes Díez, J. (2025). *Plataformas de justicia digital en México y el derecho a la información judicial*. Universidad Nacional Autónoma de México. Ciudad de México: Instituto de Investigaciones Jurídicas. Recuperado el 25 de abril de 2026, de https://www.scielo.org.mx/scielo.php?pid=S2594-00822025000200003&script=sci_arttext
- Cámara de Diputados. (23 de abril de 2023). *Impulsan iniciativa para expedir la Ley de Regulación Ética de la Inteligencia Artificial y la Robótica*. Recuperado el 25 de abril de 2026, de Coordinación de Comunicación Social: <https://comunicacionsocial.diputados.gob.mx/index.php/boletines/impulsan-iniciativa-para-expedir-la-ley-de-regulacion-etica-de-la-inteligencia-artificial-y-la-robotica>
- Cámara de Diputados. (2024). *Iniciativa con Proyecto de Decreto que Expide la Ley Nacional que Regula el uso de la Inteligencia Artificial*. Ciudad de México: Sistema de Información Legislativa (SIL). Recuperado el 25 de abril de 2026, de https://sil.gobernacion.gob.mx/Archivos/Documentos/2024/12/asun_4825444_20241213_1730224444.pdf
- Cámara de Diputados. (2025). *Iniciativa con Proyecto de Decreto que reforma diversas disposiciones de la Constitución Política de los Estados Unidos Mexicanos, en materia de Inteligencia Artificial*. Ciudad de México: Gaceta Parlamentaria / Sistema de Información Legislativa (SIL). Recuperado el 25 de abril de 2026, de https://sil.gobernacion.gob.mx/Archivos/Documentos/2025/04/asun_4896460_20250430_1747757009.pdf
- Cámara de Diputados del H. Congreso de la Unión. (10 de abril de 2026). *Constitución Política de los Estados Unidos Mexicanos*. México: Diario Oficial de la Federación. Recuperado el 25 de abril de 2026, de <https://www.diputados.gob.mx/LeyesBiblio/pdf/CPEUM.pdf>
- Dworkin, R. (1996). *Freedom's Law: The Moral Reading of the American Constitution*. Harvard University Press. Recuperado el 26 de abril de 2026

Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial. (2019).

Directrices éticas para una IA fiable. Comisión Europea. Recuperado el 26 de abril de 2026, de

https://www.euskadi.eus/contenidos/informacion/orain_farmacia/es_orain/adjuntos/Directrices-eticas-inteligencia-artificical-fiable.pdf

Hernández, M. (28 de abril de 2026). Querétaro dicta primera sentencia con inteligencia

artificial en México. *Código Qro*. Recuperado el 30 de abril de 2026, de

<https://codigoqro.mx/nota/local/2026/04/28/queretaro-dicta-primera-sentencia-inteligencia-artificial-mexico>

Inteligencia artificial aplicada en procesos jurisdiccionales. Elementos mínimos que deben

observarse para su uso ético y responsable con perspectiva de derechos humanos,

Tesis aislada II.2o.C.9 K (11 a.), registro digital 2031010 (Suprema Corte de Justicia de la Nación 2025). Recuperado el 24 de abril de 2026, de

<https://sjf2.scjn.gob.mx/detalle/tesis/2031010>

Kelsen, H. (1960). *Teoría Pura del Derecho*. México, D.F., México: Universidad Nacional

Autónoma de México (Instituto de Investigaciones Jurídicas). Recuperado el 24 de abril de 2026

Órgano de Administración Judicial. (2026). *Acuerdo General AG-POAJ-006/2026 del Pleno del*

Órgano de Administración Judicial. Diario Oficial de la Federación (DOF). Recuperado el 25 de abril de 2026, de <https://sidof.segob.gob.mx/notas/5785263>

Tribunales Colegiados de Circuito. (2026). *Inteligencia artificial aplicada en procesos*

jurisdiccionales. Constituye una herramienta válida para calcular el monto de las

garantías que se fijen en los juicios de amparo. Registro digital 2031639, Semanario

Judicial de la Federación, Ciudad de México. Recuperado el 25 de abril de 2026, de

<https://sjf2.scjn.gob.mx/detalle/tesis/2031639>

UNESCO. (2021). *Recomendación sobre la Ética de la Inteligencia Artificial*. Recuperado el 27

de abril de 2026, de Organización de las Naciones Unidas para la Educación, la

Ciencia y la Cultura: [https://www.unesco.org/es/articles/recomendacion-sobre-la-](https://www.unesco.org/es/articles/recomendacion-sobre-la-etica-de-la-inteligencia-artificial)

[etica-de-la-inteligencia-artificial](https://www.unesco.org/es/articles/recomendacion-sobre-la-etica-de-la-inteligencia-artificial)

Unión Europea. (2024). *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial*. Bruselas: Diario Oficial de la Unión Europea. Recuperado el 27 de abril de 2026, de <https://www.boe.es/doue/2024/1689/L00001-00144.pdf>

CAPÍTULO XXI

ÉTICA, SESGOS ALGORÍTMICOS Y RESPONSABILIDAD SOCIAL

Ethics, algorithmic biases and social responsibility

Cristian Arturo Plaza Cuadras

capc@azc.uam.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0009-0006-6725-1340>

Resumen

Este capítulo analiza la integración de la Inteligencia Artificial (IA) en la sociedad, desplazando el enfoque desde la optimización técnica hacia una dimensión ética y sociopolítica. Se argumenta que la IA no es una herramienta neutral, sino un sistema que puede amplificar y sistematizar las asimetrías del mundo analógico. El análisis categoriza los sesgos algorítmicos en dos vertientes: sesgos por datos (históricos, de representación y de medición) y sesgos por diseño (de agregación, evaluación y despliegue). Mediante el examen de casos críticos en justicia predictiva y visión artificial, se evidencia cómo la "objetividad matemática" oculta procesos de discriminación.

El texto aborda el dilema de la reingeniería social y la dificultad de definir métricas universales de justicia, advirtiendo sobre los riesgos del paternalismo algorítmico. Finalmente, se propone la responsabilidad social universitaria como un contrapeso ético indispensable. Se examinan modelos de gobernanza en instituciones internacionales y el panorama actual en México, concluyendo que la universidad debe liderar la auditoría de estos sistemas y la formación de profesionales capaces de garantizar que la tecnología se subordine a los valores democráticos y al bienestar humano.

Palabras clave: Inteligencia Artificial (IA), sesgos algorítmicos, discriminación algorítmica, ética digital, gobernanza algorítmica, justicia predictiva, visión artificial, responsabilidad social universitaria, reingeniería social, democracia tecnológica.

Abstract

This chapter analyzes the integration of Artificial Intelligence (AI) in society, shifting the focus from technical optimization to an ethical and socio-political dimension. It is argued that AI is not a neutral tool, but a system that can amplify and systematize the asymmetries of the analog world. The analysis categorizes algorithmic biases into two aspects: data biases (historical, representational, and measurement) and design biases (aggregation, evaluation, and deployment). By examining critical cases in predictive justice and computer vision, it is evident how "mathematical objectivity" hides processes of discrimination.

The text addresses the dilemma of social reengineering and the difficulty of defining universal metrics of justice, warning about the risks of algorithmic paternalism. Finally, university social responsibility is proposed as an indispensable ethical counterweight. Governance models in international institutions and the current panorama in Mexico are examined, concluding that the university must lead the audit of these systems and the training of professionals capable of ensuring that technology is subordinated to democratic values and human well-being.

Keywords: Artificial Intelligence (AI), algorithmic biases, algorithmic discrimination, digital ethics, algorithmic governance, predictive justice, computer vision, university social responsibility, social reengineering, technological democracy.

La integración ubicua de la Inteligencia Artificial (IA) en las estructuras fundamentales de la sociedad contemporánea ha desplazado el debate tecnológico desde la mera optimización funcional hacia una dimensión profundamente ética y sociopolítica. Lo que inicialmente se percibía como una

herramienta de objetividad matemática ha revelado ser, en la práctica, un espejo complejo que amplifica y sistematiza las asimetrías del mundo analógico. Esta problemática no es accidental, sino que se encuentra intrínsecamente ligada a la naturaleza de los datos y al diseño de los sistemas.

El presente capítulo analiza la intersección entre la ética, los sesgos algorítmicos y la responsabilidad social, articulando una crítica técnica y humana sobre cómo estas tecnologías impactan en la equidad. La arquitectura del problema se divide, primordialmente, en dos vertientes de sesgo. Por un lado, los sesgos por datos, que incluyen desde el sesgo histórico —la fosilización de prejuicios sistémicos en modelos de aprendizaje— hasta deficiencias en la representación y la medición, donde la invisibilización de minorías y la falta de rigor métrico generan resultados injustos.

Por otro lado, exploraremos los sesgos por diseño, aquellos errores estructurales que surgen en la fase de modelado y despliegue, como el sesgo de agregación, la evaluación técnica carente de una perspectiva de justicia, y los riesgos del despliegue en entornos no controlados. Finalmente, el análisis concluye con una reflexión sobre la responsabilidad social, con un énfasis crítico en el rol de los entornos universitarios. Las instituciones académicas, como centros de innovación y formación de capital humano, tienen el deber ético no solo de desarrollar tecnología, sino de liderar la gobernanza y la auditoría de sistemas que definen el acceso a derechos y oportunidades en el siglo XXI.

Sesgos algorítmicos

El concepto de sesgo algorítmico se refiere a la existencia de errores sistemáticos y repetibles en un sistema informático que crean resultados injustos, como la concesión de privilegios a un grupo arbitrario de usuarios en detrimento de otros. A diferencia de los errores aleatorios de computación, el sesgo algorítmico es una distorsión con dirección y sentido; es una inclinación técnica que suele replicar las fallas de juicio humanas bajo una apariencia de neutralidad matemática.

Desde una perspectiva académica, el sesgo no debe entenderse únicamente como un "error de programación". En realidad, es el resultado de una compleja interacción entre el código, los datos y el contexto social de uso. La ilusión de la "objetividad algorítmica" —la creencia de que, por ser un proceso automatizado, está libre de prejuicios— es, precisamente, lo que vuelve a estos sesgos tan peligrosos: se integran en la burocracia digital y se vuelven difíciles de cuestionar por el ciudadano común.

Ejemplos paradigmáticos de la problemática actual

Para comprender la magnitud de esta situación, es necesario observar casos documentados donde la teoría del sesgo se manifiesta en daños tangibles:

1. Justicia predictiva y el caso COMPAS: Uno de los ejemplos más citados en la literatura académica es el sistema COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), utilizado en Estados Unidos para predecir la probabilidad de reincidencia de los detenidos. Auditorías independientes demostraron que el algoritmo tendía a asignar puntuaciones de riesgo mucho más altas a personas afroamericanas que a personas blancas, incluso cuando los perfiles delictivos eran similares, debido a que el sistema utilizaba variables que indirectamente penalizaban la situación socioeconómica y el origen racial.
2. Sistemas de visión artificial y disparidad de género: En el ámbito del procesamiento de imágenes, diversos modelos de detección de objetos han mostrado dificultades para identificar herramientas de cocina en manos de hombres, asociándolas casi exclusivamente a mujeres. Asimismo, sistemas de clasificación de género fallan con mayor frecuencia ante rostros de mujeres con tonos de piel oscuros, lo que evidencia que la "visión" de la máquina está condicionada por los referentes visuales dominantes de sus desarrolladores.
3. Filtrado de talento y exclusión automatizada: Grandes corporaciones han intentado implementar herramientas de IA para clasificar miles de currículos.

En casos notables, los algoritmos aprendieron a penalizar palabras asociadas a grupos específicos (como "club de mujeres" o nombres de universidades históricamente femeninas) simplemente porque los datos históricos de éxito laboral en la empresa estaban saturados de perfiles masculinos. El algoritmo, en su intento por ser eficiente, interpretó la desigualdad histórica como una regla de calidad.

Estos ejemplos subrayan que el sesgo algorítmico no es una anomalía del sistema, sino a menudo un reflejo fiel de la sociedad que produce la tecnología. El reto ético reside en cómo intervenir estas "cajas negras" para que no se conviertan en motores de discriminación a escala masiva

Sesgos por datos

La arquitectura de cualquier sistema de inteligencia artificial, particularmente aquellos basados en el aprendizaje automático (*machine learning*), depende de la integridad y representatividad de los datos que consume. En el ámbito académico, se reconoce que los datos no son meros registros técnicos objetivos, sino construcciones sociales que capturan el estado de un fenómeno en un momento determinado. El sesgo por datos surge cuando la información utilizada para entrenar el modelo contiene distorsiones estructurales que impiden un procesamiento equitativo.

Estas distorsiones no suelen ser fallos en el almacenamiento del dato, sino problemas en la naturaleza misma de lo que se está registrando. Si el "material de construcción" de la IA es parcial o injusto, el edificio algorítmico resultante no solo heredará estas fallas, sino que las validará bajo un halo de infalibilidad técnica.

- **Ejemplo representativo:** el algoritmo de asignación de cuidados de salud de Optum. Un caso fundamental para entender cómo los datos pueden traicionar la intención ética de un sistema es el algoritmo utilizado en diversos hospitales de Estados Unidos para identificar a pacientes con necesidades médicas complejas. El sistema no utilizaba la raza como variable, lo que en teoría lo hacía "neutral". Sin embargo, se entrenó

utilizando el "gasto médico histórico" como una variable sustituta (*proxy*) de la necesidad de salud, bajo la premisa de que un paciente que genera más gastos es un paciente más enfermo. El resultado fue una discriminación sistémica: debido a barreras económicas y sociales históricas, los pacientes de minorías étnicas gastaban menos en salud a pesar de estar más enfermos. El algoritmo, al leer el dato de "bajo gasto", clasificaba a pacientes afroamericanos en estado crítico como "estables", otorgando prioridad de cuidados intensivos a pacientes blancos con cuadros clínicos mucho más leves. Aquí, el sesgo no estuvo en el código, sino en confiar en un dato (gasto) que ya estaba viciado por la desigualdad social.

Sesgo histórico

El sesgo histórico representa uno de los desafíos más complejos para la ética algorítmica, ya que se manifiesta incluso cuando el proceso de recolección de datos ha sido técnicamente impecable y la muestra es estadísticamente significativa. Este fenómeno ocurre cuando los datos de entrenamiento reflejan prejuicios, estereotipos o desigualdades sistémicas que ya están presentes en la sociedad. En este sentido, la inteligencia artificial actúa como un espejo: si la realidad histórica es discriminatoria, el algoritmo aprenderá a ser discriminatorio para ser "preciso" respecto a esos datos.

- **Ejemplo representativo:** el sistema de filtrado de talento de Amazon. A mediados de la década pasada, la empresa Amazon desarrolló una herramienta de inteligencia artificial diseñada para automatizar la selección de personal. El sistema fue entrenado utilizando los patrones de contratación de la compañía durante un periodo de diez años. Debido a que la industria tecnológica ha tenido una sobrerrepresentación masculina histórica, el algoritmo concluyó que el "éxito profesional" estaba vinculado al género masculino. El sistema comenzó a penalizar activamente los currículos que contenían la palabra "femenino" o referencias a instituciones educativas para mujeres. A pesar de los intentos de los ingenieros por "anonimizar" los datos,

el algoritmo aprendió a identificar patrones sutiles que le permitían seguir discriminando. El proyecto fue cancelado en 2017 al confirmarse que el sesgo histórico era tan profundo que el modelo no podía garantizar la equidad.

Sesgo de representación

El sesgo de representación ocurre cuando el conjunto de datos utilizado para entrenar un modelo no refleja la diversidad demográfica o la variabilidad real de la población en la que se aplicará el sistema. En la estadística aplicada a la inteligencia artificial, este fenómeno genera una asimetría en el rendimiento del algoritmo: mientras que el sistema alcanza una precisión óptima para los grupos mayoritarios presentes en la muestra, su eficacia se degrada significativamente al enfrentarse a minorías o grupos infrarrepresentados.

Este problema es especialmente crítico en el aprendizaje supervisado. Si un sistema de visión artificial, por ejemplo, es entrenado con una base de datos donde el 90 % de los sujetos pertenecen a una etnia específica, el modelo desarrollará una "ceguera estadística" hacia los rasgos fenotípicos del 10 % restante. Para el algoritmo, lo que no está suficientemente representado en los datos de entrenamiento simplemente no existe como categoría válida, lo que deriva en una exclusión tecnológica que refuerza la marginalidad social.

- **Ejemplo representativo:** el proyecto Gender Shades y los sistemas de reconocimiento facial. Uno de los hitos en la investigación de la ética algorítmica fue el estudio liderado por Joy Buolamwini y Timnit Gebru, quienes analizaron los sistemas comerciales de análisis facial de gigantes tecnológicos como IBM y Microsoft. La investigación reveló que estos modelos presentaban una precisión casi perfecta (superior al 99 %) cuando se trataba de identificar el género de hombres de piel clara. No obstante, la tasa de error se disparaba drásticamente al analizar a mujeres de piel oscura, alcanzando niveles de falla de hasta el 34 %. La causa raíz era un sesgo de representación: las bases de datos de entrenamiento (como

Adience o *IJB-A*) estaban compuestas mayoritariamente por rostros caucásicos y masculinos. Este caso demostró que la "objetividad" del reconocimiento facial estaba condicionada por una visión del mundo que invisibilizaba a las mujeres afrodescendientes, planteando riesgos severos en aplicaciones de seguridad y vigilancia.

Sesgo de medición

El sesgo de medición surge durante la fase de diseño y recolección de datos, cuando las características o variables elegidas para describir un fenómeno se miden de forma inconsistente entre diferentes grupos, o cuando la variable elegida no representa fielmente el concepto que se pretende evaluar. A menudo, este sesgo es el resultado de una simplificación excesiva de problemas sociales complejos en métricas cuantificables que, en la práctica, no son equitativas.

Este tipo de distorsión suele ser más difícil de detectar que otros, ya que no depende de la cantidad de datos (representación), sino de la calidad y la validez de la métrica misma. Si el instrumento de medición está viciado o si la "etiqueta" asignada al dato ignora el contexto socioeconómico del individuo, el sistema producirá predicciones que penalizan injustamente a ciertos colectivos bajo una falsa premisa de exactitud métrica.

- **Ejemplo representativo:** el algoritmo COMPAS en la justicia predictiva. El sistema COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) ha sido utilizado en diversos estados de EE. UU. para evaluar el riesgo de reincidencia de los detenidos. El sesgo de medición aquí se manifestó en las variables utilizadas para calcular dicho riesgo. El algoritmo no preguntaba por la raza, pero utilizaba un cuestionario de 137 preguntas que medían factores como el historial de arrestos previos, el nivel de empleo y el entorno familiar. Dado que las comunidades afroamericanas están sujetas a una vigilancia policial más intensa y a barreras estructurales de empleo, estas métricas capturaban la "vulnerabilidad social" en lugar de la "propensión criminal". Como resultado, una investigación de *ProPublica*

demostró que el sistema clasificaba erróneamente a los detenidos negros como "de alto riesgo" con el doble de frecuencia que a los blancos, mientras que a estos últimos se les asignaba un riesgo bajo incluso con antecedentes penales más graves, evidenciando una medición desigual de la peligrosidad.

Sesgos por diseño

A diferencia de las distorsiones que provienen de la recolección de información, los sesgos por diseño se originan en las decisiones técnicas, metodológicas y conceptuales tomadas por los desarrolladores y arquitectos de sistemas durante la fase de modelado. Estos no son errores accidentales en los datos, sino que se consolidan en la arquitectura misma del algoritmo, en la elección de las variables de interés, en las funciones de optimización o en la definición de lo que el sistema considera un "éxito". El sesgo por diseño pone de manifiesto que la creación de tecnología no es un proceso neutro, sino que está impregnado de las visiones, prioridades y, en ocasiones, los puntos ciegos de quienes la programan.

- **Ejemplo representativo:** la optimización de la precisión sobre la equidad. Un ejemplo fundamental de sesgo por diseño ocurre cuando los ingenieros definen la función de pérdida de un modelo priorizando la precisión global (*accuracy*) por encima de cualquier otra métrica. Si un sistema se diseña para ser "lo más preciso posible" en un conjunto de datos donde el 95 % de los usuarios pertenecen a un grupo mayoritario, el algoritmo descubrirá que puede ignorar por completo las particularidades del 5 % restante (la minoría) y aun así obtener una calificación de éxito del 95 %. En este caso, el sesgo no está en un dato erróneo, sino en la decisión de diseño de no incluir una penalización específica para los errores cometidos contra grupos vulnerables o minoritarios, validando un rendimiento técnico que es intrínsecamente injusto.

Sesgo de agregación

El sesgo de agregación se produce cuando se asume erróneamente que una población es homogénea y se desarrolla un modelo único bajo la premisa de que

"una sola talla sirve para todos". En el ámbito de la inteligencia artificial, este error de diseño ignora que los subgrupos dentro de una población pueden tener características, comportamientos o necesidades biológicas radicalmente distintas. Al intentar representar a una colectividad diversa a través de un único promedio estadístico, el modelo termina por ser ineficaz o incluso perjudicial para aquellos individuos que no se ajustan a la norma mayoritaria, tratando sus diferencias como si fueran simple "ruido" que debe ser eliminado o ignorado.

- **Ejemplo representativo:** modelos de riesgo clínico y variables biológicas. En la medicina basada en datos, un caso crítico de sesgo de agregación se observa en el diseño de algoritmos para predecir enfermedades cardiovasculares. Si el sistema se diseña agregando los datos de hombres y mujeres en un solo modelo sin considerar que los síntomas y marcadores biológicos de un infarto varían significativamente entre sexos, el algoritmo fallará en su capacidad de diagnóstico para las mujeres. Al utilizar un umbral de riesgo "promedio" (basado históricamente en sujetos masculinos), el diseño del sistema invisibiliza las señales clínicas femeninas, lo que resulta en diagnósticos tardíos y una menor calidad de atención. Aquí, la falla reside en la decisión de agregar poblaciones distintas en un modelo monolítico en lugar de optar por un diseño estratificado que reconozca la diversidad biológica.

Sesgo de evaluación

El sesgo de evaluación surge durante la fase de validación y prueba de un modelo, cuando los marcos de referencia o las métricas utilizadas para calificar su éxito no logran capturar las disparidades de rendimiento entre diferentes subgrupos. En la ingeniería de inteligencia artificial convencional, la métrica predominante es la precisión global (*accuracy*). Sin embargo, desde una perspectiva ética, esta métrica es insuficiente y potencialmente engañosa: un algoritmo puede presentar una efectividad del 98 % a nivel general, pero fallar de manera estrepitosa en el 2 % restante de los casos, los cuales suelen coincidir con minorías o grupos

vulnerables. Si el proceso de evaluación no incluye auditorías de equidad o pruebas estratificadas, el sistema recibe una "certificación de calidad" que invisibiliza su comportamiento discriminatorio.

- **Ejemplo representativo:** el espejismo de los bancos de imágenes estándar. Durante años, los sistemas de visión artificial fueron evaluados utilizando conjuntos de datos de prueba como *Labeled Faces in the Wild* (LFW) para medir su precisión en el reconocimiento facial. Los desarrolladores celebraban tasas de éxito cercanas al 100 %, lo que sugería que la tecnología estaba "lista" para su uso masivo. No obstante, el sesgo de evaluación residía en que estos bancos de imágenes estaban compuestos mayoritariamente por rostros de personas blancas y masculinas. Cuando investigadoras como Joy Buolamwini aplicaron métricas de evaluación diferenciadas por tono de piel y género, descubrieron que el rendimiento caía drásticamente al analizar rostros de mujeres afrodescendientes. El error no estuvo solo en el entrenamiento, sino en un proceso de evaluación que consideró que una muestra homogénea era suficiente para validar un sistema de alcance global.

Sesgo de despliegue

El sesgo de despliegue ocurre en la transición del laboratorio al entorno real. Se manifiesta cuando existe una desconexión profunda entre los objetivos para los que fue diseñado el sistema y la forma en que este es utilizado o interpretado por los usuarios en un contexto social complejo. Un modelo puede ser técnicamente robusto en condiciones controladas, pero convertirse en un motor de desigualdad cuando su salida interactúa con los prejuicios humanos de quienes operan la herramienta o con las carencias estructurales del entorno donde se implementa. Este sesgo subraya que la tecnología no opera en el vacío, sino que su despliegue puede exacerbar problemas preexistentes de los cuales el desarrollador no fue consciente.

- **Ejemplo representativo:** sistemas de patrullaje predictivo y el bucle de retroalimentación. Sistemas diseñados para predecir zonas de alta criminalidad, como PredPol, ilustran con claridad el sesgo de despliegue. Aunque el algoritmo busca optimizar la presencia policial, su implementación en ciudades con un historial de segregación racial genera resultados perversos. Al desplegarse en barrios donde históricamente ha habido más detenciones por una vigilancia policial intensiva, el sistema recibe datos de "éxito" constantes desde esas mismas áreas. Esto crea un ciclo de retroalimentación: el algoritmo envía más policías a los mismos barrios, lo que genera más arrestos, lo que a su vez "confirma" al algoritmo que debe seguir enviando patrullas allí. En este caso, el entorno de despliegue —una sociedad con sesgos policiales previos— transforma una herramienta de eficiencia logística en un mecanismo automatizado de acoso y criminalización hacia comunidades específicas.

Sesgo de etiquetado

El sesgo de etiquetado ocurre durante la fase de preparación de los datos, cuando los seres humanos encargados de clasificar, categorizar o describir la información introducen sus propios prejuicios, estereotipos o limitaciones cognitivas en el sistema. Dado que el aprendizaje supervisado depende de "etiquetas" (por ejemplo, identificar qué es un perro o qué es un comportamiento sospechoso), si el etiquetador humano comete un error sistemático, el algoritmo aprenderá ese error como una verdad absoluta. Este sesgo es particularmente peligroso porque convierte la subjetividad humana en una norma técnica "objetiva".

- **Ejemplo representativo:** el incidente de Google Photos y la clasificación de personas afroamericanas. Un caso citado extensamente por Melanie Mitchell y otros expertos es el error ocurrido en 2015 con el servicio Google Photos, cuya inteligencia artificial clasificó automáticamente fotografías de personas negras bajo la etiqueta de "gorilas". Este fallo no solo fue una

deficiencia de representación en los datos de entrenamiento, sino también un error crítico en el etiquetado y la taxonomía del modelo. El sistema, al no haber sido entrenado con una diversidad suficiente de rasgos humanos y al tener categorías de visión artificial mal definidas por los anotadores, terminó cometiendo una asociación deshumanizante y racista. Este evento evidenció cómo la falta de sensibilidad ética en el proceso de etiquetado y validación puede derivar en productos tecnológicos que perpetúan el estigma y la violencia simbólica contra grupos históricamente marginados.

Sesgo de género en el procesamiento de lenguaje natural

Este tipo de sesgo se manifiesta cuando los modelos de lenguaje aprenden asociaciones estereotipadas entre el género y ciertas actividades, profesiones o atributos, basándose en la frecuencia con la que aparecen en textos literarios, noticias o redes sociales. Al procesar millones de oraciones donde los roles de liderazgo se asocian mayoritariamente a pronombres masculinos y las labores de cuidado a femeninos, la inteligencia artificial no solo replica estos estereotipos, sino que los refuerza al generar contenido nuevo o realizar traducciones automáticas.

- **Ejemplo representativo:** la traducción automática y los estereotipos profesionales. Sistemas de traducción como Google Translate han mostrado sesgos persistentes al traducir desde idiomas con pronombres neutros (como el turco o el húngaro) al español o al inglés. Por ejemplo, al traducir frases neutras sobre profesiones, el sistema solía convertir de forma sistemática "o bir doktor" (él/ella es doctor/a) como "él es médico", mientras que "o bir hemşire" (él/ella es enfermero/a) se traducía como "ella es enfermera". Aunque las empresas han intentado mitigar este sesgo ofreciendo ambas opciones, el problema raíz reside en que el modelo matemático ha "decidido" que la probabilidad estadística de que un médico sea hombre es mayor, eliminando la neutralidad del lenguaje y perpetuando el techo de cristal algorítmico en el ámbito profesional.

Consecuencias del sesgo y el dilema de la reingeniería social

La persistencia de los sesgos algorítmicos no solo representa un fallo técnico, sino que conlleva consecuencias sistémicas que pueden erosionar los pilares de la democracia y la cohesión social. Entre las repercusiones más graves se encuentra la cristalización de la desigualdad: mientras que en el pasado la discriminación dependía de prejuicios individuales, hoy se encuentra codificada en sistemas que operan a una escala y velocidad que impiden la impugnación humana. Esto genera un efecto de "retroalimentación negativa" donde los grupos históricamente marginados son castigados por el algoritmo basándose en datos del pasado, lo que les impide mejorar su situación en el presente, alimentando así nuevos datos sesgados para el futuro.

Sin embargo, surge una interrogante ontológica crucial: los algoritmos son, en última instancia, un espejo de la sociedad. Si un sistema de inteligencia artificial muestra un sesgo, es porque ha aprendido de una realidad humana que ya es desigual. El dilema ético profundo aparece cuando intentamos "corregir" o "eliminar" el sesgo. Para lograr un resultado "justo", el desarrollador debe intervenir activamente en el modelo para forzar un resultado que no necesariamente se encuentra en los datos originales. En ese momento, la tecnología deja de ser un observador pasivo y se convierte en una herramienta de reingeniería social.

- **Ejemplo representativo:** el dilema de la paridad estadística en la publicidad laboral. Si un algoritmo de una red social como LinkedIn detecta que, históricamente, los hombres hacen más clic en anuncios de empleos de alta dirección, el sistema —buscando eficiencia— mostrará menos esos anuncios a las mujeres. Si decidimos "eliminar" ese sesgo y forzar al algoritmo a mostrar los anuncios en una proporción de 50/50, estamos aplicando un estándar de justicia redistributiva. Aquí surge el problema del "quién decide": si el ingeniero decide que el 50/50 es el estándar, está moldeando activamente la estructura del mercado laboral según su propia

visión moral. Como señala Ruha Benjamin en su obra *Race After Technology*, el acto de "neutralizar" nunca es neutro; siempre implica la imposición de un nuevo estándar de normalidad.

El problema de la autoridad moral: ¿quién define la justicia?

Eliminar el sesgo implica necesariamente elegir una definición de "equidad". En la literatura técnica se han identificado más de veinte definiciones matemáticas de justicia (como la paridad demográfica o la igualdad de oportunidades), y se ha demostrado matemáticamente que es imposible satisfacer todas simultáneamente. Por lo tanto, al programar un algoritmo "ético", alguien —ya sea un ingeniero, una corporación o un Estado— debe decidir qué métrica de justicia tiene prioridad sobre las demás.

Este escenario plantea el riesgo de un "paternalismo algorítmico". Si tratamos de moldear la sociedad a través de códigos para alcanzar un ideal de justicia, nos enfrentamos a la pregunta de quién tiene la legitimidad para decidir cuál es ese ideal. Como argumentan Michael Kearns y Aaron Roth en *The Ethical Algorithm*, cualquier intervención para corregir un sesgo implica un "costo de equidad" (*fairness trade-off*) que suele impactar en la precisión del sistema. La decisión de cuánto estamos dispuestos a sacrificar de eficiencia en favor de la justicia es una decisión política, no matemática, y dejar esa decisión en manos de tecnócratas privados representa uno de los mayores desafíos para la soberanía democrática.

- **Ejemplo representativo:** el debate sobre el "punto de vista desde ningún lugar" de Kate Crawford. En su investigación sobre la política de la clasificación, Kate Crawford argumenta que los sistemas de IA suelen presentarse como visiones objetivas desde una "nada" matemática, cuando en realidad son productos de centros de poder específicos (como Silicon Valley). Al intentar "limpiar" los datos de sesgos raciales o de género, las empresas tecnológicas suelen aplicar estándares de "diversidad" que reflejan la visión liberal estadounidense, la cual puede no ser aplicable o incluso ser contraproducente en contextos culturales de Asia, África o

América Latina. Este conflicto evidencia que no existe un estándar de "sociedad sin sesgos" que sea universal; cada intento de corrección es, en realidad, una imposición de una visión particular de lo que la sociedad *debería* ser.

Responsabilidad social

En el contexto de la inteligencia artificial, la responsabilidad social trasciende el cumplimiento de las normativas legales vigentes; se define como el compromiso ético y proactivo de las organizaciones, desarrolladores y Estados para asegurar que el diseño, despliegue y uso de los sistemas algorítmicos generen un impacto positivo en la sociedad y no vulneren los derechos fundamentales. Esta dimensión implica que los actores tecnológicos deben ser conscientes de las externalidades de sus productos, asumiendo una postura de rendición de cuentas (*accountability*) ante las posibles consecuencias negativas que la automatización pueda acarrear en el tejido social, la economía y la cultura.

La responsabilidad social en la IA se articula sobre tres pilares fundamentales: la transparencia, la justicia y la sostenibilidad. La transparencia exige que los procesos de toma de decisiones algorítmicas no sean "cajas negras" impenetrables, sino que permitan una auditoría clara por parte de terceros. La justicia busca mitigar los sesgos que hemos analizado previamente, garantizando la equidad en el acceso a las oportunidades. Por último, la sostenibilidad considera tanto el impacto ambiental de los grandes centros de datos como el impacto laboral de la automatización. Bajo este marco, la tecnología deja de ser un fin en sí misma para convertirse en un medio que debe estar subordinado al bienestar humano.

- **Ejemplo representativo:** el Reglamento de Inteligencia Artificial de la Unión Europea (AI Act). La promulgación de la Ley de IA de la Unión Europea representa el esfuerzo global más ambicioso en materia de responsabilidad social regulatoria. A diferencia de enfoques previos basados en la autorregulación corporativa, esta normativa clasifica los sistemas de IA según su nivel de riesgo para los ciudadanos. Por ejemplo, prohíbe

explícitamente el uso de sistemas de "puntuación social" (*social scoring*) y restringe la vigilancia biométrica en espacios públicos, al considerar que estas tecnologías violan la responsabilidad social básica del Estado hacia la privacidad y la libertad. Con este marco, la responsabilidad social deja de ser una declaración de intenciones corporativas para convertirse en un requisito técnico y legal vinculante, obligando a las empresas a realizar evaluaciones de impacto sobre los derechos fundamentales antes de lanzar cualquier producto de alto riesgo al mercado.

Finalmente, tras haber analizado la responsabilidad social desde una perspectiva global y regulatoria, nos enfocaremos en la responsabilidad social en entornos universitarios, donde la formación ética y la investigación científica desempeñan un papel determinante en la configuración de las futuras herramientas tecnológicas.

Responsabilidad social en entornos universitarios

La universidad, como institución fundamental en la producción de conocimiento y formación de capital humano, ocupa una posición de privilegio y, en consecuencia, de altísima responsabilidad ante el despliegue de la inteligencia artificial. A diferencia de otros sectores, la academia no solo es una consumidora de tecnología, sino el laboratorio donde se gestan los algoritmos que definirán el futuro. Por ello, la responsabilidad social universitaria debe ser entendida como un compromiso multidimensional que abarca desde la ética pedagógica hasta la gobernanza institucional y el impacto en la comunidad.

1. Responsabilidad en la formación (docencia)

La formación técnica en inteligencia artificial ha priorizado históricamente la eficiencia computacional y el rigor matemático, dejando los dilemas éticos como asignaturas optativas o seminarios periféricos. Sin embargo, la responsabilidad social exige una reestructuración del currículo donde la ética de la IA sea una columna vertebral obligatoria. Enseñar a identificar y mitigar sesgos no puede ser un "añadido" al final del semestre, sino un componente crítico que se integre en la

resolución de problemas técnicos desde las primeras etapas de la carrera. El objetivo es formar profesionales capaces de ejercer un pensamiento crítico que cuestione no solo si un sistema "puede" construirse, sino si "debe" ser construido y bajo qué condiciones de justicia.

- **Ejemplo representativo:** la integración de la ética en el currículo de computación.

Una propuesta que ha ganado tracción en instituciones de vanguardia es que los cursos de ciencias de la computación dediquen al menos un 20 % de su carga horaria a los impactos sociales y éticos de la tecnología. En lugar de estudiar los algoritmos de clasificación de forma abstracta, los estudiantes deben trabajar con casos prácticos de fallos reales en la industria, analizando por qué un modelo falló y cómo impactó a comunidades vulnerables. Al evaluar un proyecto final, la calificación no debería depender solo de la precisión del modelo (\$accuracy\$), sino de una auditoría interna de sesgos realizada por el propio estudiante, integrando la responsabilidad social como una competencia técnica más.

2. Responsabilidad en la investigación

En el ámbito de la investigación, la universidad debe actuar como un contrapeso ético frente a las presiones del mercado. Esto implica establecer comités de ética robustos que revisen los proyectos de IA antes de su ejecución, de forma similar a como ocurre en la investigación médica. Además, la academia debe liderar el movimiento hacia la inteligencia artificial explicable (XAI) y el desarrollo de métricas de equidad (*fairness*), rechazando financiamientos que comprometan la autonomía científica o exijan resultados que validen prácticas discriminatorias.

- **Ejemplo representativo:** las declaraciones de impacto social en conferencias académicas.

Un hito significativo en este ámbito fue la decisión de la conferencia NeurIPS (una de las más importantes en IA a nivel mundial) de exigir que todos los investigadores presenten una "Declaración de impacto" junto con sus artículos científicos. En esta

declaración, los autores deben discutir explícitamente los posibles riesgos sociales de su investigación, incluyendo potenciales sesgos o usos duales malintencionados. Esta práctica, que se ha extendido a diversos centros de investigación universitarios, obliga a los científicos a salir de la abstracción técnica y asumir la responsabilidad sobre las consecuencias a largo plazo de sus descubrimientos.

3. Responsabilidad en la gestión institucional

La universidad también es una organización que utiliza IA para sus propios procesos administrativos, lo cual conlleva riesgos éticos directos para su comunidad. La gestión institucional debe garantizar que el uso de algoritmos en admisiones, evaluación estudiantil o sistemas de bienestar sea transparente y justo. Existe un peligro latente en el uso de IA para predecir la deserción escolar o calificar ensayos, ya que estos sistemas pueden penalizar estilos de escritura no convencionales o idiomas nativos distintos al estándar, perpetuando desigualdades socioeconómicas previas.

- **Ejemplo representativo:** el sesgo en los sistemas de supervisión remota (*proctoring*).

Durante la transición a la educación a distancia, muchas universidades implementaron sistemas de IA para supervisar exámenes. Se descubrió que algunos de estos algoritmos tenían dificultades para reconocer rostros de personas con tonos de piel oscuros debido a problemas de iluminación o diseño del sensor, y clasificaban como "comportamiento sospechoso" ciertos tics o movimientos asociados a la neurodiversidad. La responsabilidad institucional exige que la universidad audite estas herramientas antes de implementarlas, garantizando que el software no se convierta en una barrera de acceso o en un mecanismo de discriminación injustificada para sus propios estudiantes.

4. Responsabilidad en la relación con el entorno social

La universidad no es una torre de marfil; tiene el deber de devolver conocimiento a la sociedad y proteger a los grupos más vulnerables del impacto algorítmico. Esto se traduce en la creación de "clínicas jurídicas de ética en IA" o proyectos de "IA para el bien social", donde los estudiantes y docentes asesoran a organizaciones locales en el uso ético de la tecnología. Además, la institución debe ser un espacio de denuncia anónima para sesgos detectados, actuando como un faro de integridad tecnológica en su región.

- **Ejemplo representativo:** clínicas de justicia algorítmica.

Siguiendo el modelo de las clínicas legales, algunas universidades han comenzado a fundar laboratorios de justicia algorítmica. En estos espacios, la comunidad puede acudir para que expertos universitarios analicen si un algoritmo (por ejemplo, de una oficina de empleo local o de una entidad financiera) los está discriminando. Al ofrecer estas auditorías externas de forma gratuita, la universidad cumple con su función social de democratizar el acceso a la defensa tecnológica y de empoderar a la ciudadanía frente a la opacidad de los sistemas automatizados.

5. Gobernanza específica universitaria

Finalmente, la responsabilidad social debe cristalizarse en estructuras de gobernanza claras. No basta con declaraciones de principios; se requieren órganos operativos como un comité de supervisión de IA con representación tripartita (estudiantes, docentes y administrativos) y una oficina de responsabilidad algorítmica. Estas estructuras deben encargarse de realizar auditorías periódicas y gestionar una defensoría universitaria especializada en quejas por decisiones automatizadas, asegurando que siempre exista un "humano en el bucle" (*human-in-the-loop*) con capacidad de revertir una decisión injusta.

- **Ejemplo representativo:** el establecimiento de consejos de ética en IA universitarios.

Instituciones como la Universidad de Oxford han establecido centros específicos de ética en IA que funcionan como asesores directos de la gobernanza universitaria. Estos consejos no solo supervisan la investigación, sino que dictan las políticas de uso de IA generativa en el aula y auditan los sistemas de gestión de datos de los estudiantes. Al institucionalizar la ética, la universidad garantiza que la responsabilidad social no dependa de la voluntad individual de un profesor, sino que sea un mandato corporativo ineludible.

Casos de éxito internacional y el contraste con la realidad universitaria mexicana

La implementación de marcos de responsabilidad social en inteligencia artificial no ha sido uniforme a nivel global. Mientras algunas instituciones han logrado integrar la ética de manera orgánica en su estructura, otras aún se encuentran en una fase declarativa. A continuación, se presentan modelos institucionales que destacan por su operatividad y se contrastan con los esfuerzos realizados en el territorio mexicano.

Modelos de excelencia en gobernanza algorítmica

En el ámbito internacional, universidades como Harvard y el Instituto Tecnológico de Massachusetts (MIT) han desarrollado protocolos que van más allá de la teoría. El programa *Embedded EthiCS* de la Universidad de Harvard es, quizás, el referente más sólido de responsabilidad docente. En lugar de relegar la ética a un curso aislado, profesores de filosofía colaboran directamente con ingenieros para insertar módulos éticos dentro de las materias técnicas de ciencias de la computación. Esto garantiza que el estudiante enfrente dilemas de sesgo y responsabilidad social en el mismo momento en que diseña el código.

Por otro lado, la Universidad de Stanford, a través de su Instituto de IA Centrada en el Humano (HAI), ha implementado un proceso de revisión ética para la investigación que actúa de forma similar a un comité de bioseguridad. Cualquier proyecto de investigación que utilice grandes volúmenes de datos o algoritmos predictivos debe someterse a una evaluación de impacto antes de recibir

financiamiento interno. Este protocolo "brilla" por su capacidad de detener investigaciones que, aunque técnicamente innovadoras, presentan riesgos inaceptables de vigilancia o discriminación.

El panorama en México: retos y contrastes

El contraste con las universidades mexicanas es notable, principalmente en términos de recursos destinados a la auditoría algorítmica y la formalización de comités especializados. En México, gran parte del enfoque se ha centrado en la adopción técnica y la alfabetización digital básica, dejando la gobernanza ética en un plano secundario o dependiente de esfuerzos individuales de investigadores. La burocracia institucional y la falta de marcos regulatorios nacionales específicos para la IA en la educación han ralentizado la creación de oficinas de responsabilidad algorítmica. Sin embargo, esto no implica una ausencia total de compromiso, sino una etapa de maduración distinta.

Mientras que en Estados Unidos o Europa la discusión se centra en la gobernanza y la vigilancia, en México la responsabilidad social universitaria tiene una fuerte carga hacia la reducción de la brecha digital y la democratización del acceso. El reto para las instituciones mexicanas reside en transitar de un uso ético "por buena voluntad" hacia una estructura normativa vinculante que proteja a los estudiantes y a la sociedad de las decisiones automatizadas.

Ejemplo positivo en el contexto mexicano: el Tecnológico de Monterrey y la UNAM

A pesar de los retos estructurales, existen ejemplos destacados que posicionan a México como un actor relevante en la ética regional. El Tecnológico de Monterrey ha liderado esfuerzos significativos a través de su participación en la iniciativa *fAIr LAC*, impulsada por el Banco Interamericano de Desarrollo (BID). Esta colaboración ha permitido la creación de un *hub* de ética de IA en el que se desarrollan marcos de evaluación de impacto para proyectos locales, enfocándose en evitar que la tecnología exacerbe las desigualdades socioeconómicas propias del contexto mexicano.

Por su parte, la Universidad Nacional Autónoma de México (UNAM), a través de su Programa Universitario de Inteligencia Artificial (PUIA) y la Coordinación de la Universidad Abierta, Innovación Educativa y Educación a Distancia (CUAIEED), ha emitido lineamientos para el uso de IA generativa y ética en el aula. La UNAM destaca por su enfoque en la soberanía de los datos y el fomento de una "ética del cuidado" en el uso de herramientas digitales, promoviendo que la IA sea un apoyo para el docente y no un sustituto que opaque la diversidad cultural y lingüística de sus alumnos. Estos esfuerzos representan una respuesta situada y necesaria que busca equilibrar la innovación global con la responsabilidad social local.

La encrucijada ética de la inteligencia artificial

A lo largo de este capítulo, hemos transitado desde la disección técnica de los sesgos algorítmicos hasta la dimensión política de la responsabilidad social universitaria. El contraste fundamental que emerge de este análisis es la tensión entre la eficiencia computacional y la justicia social. Mientras que la industria tecnológica a menudo promociona a la inteligencia artificial como un motor de objetividad y progreso deslocalizado, la evidencia empírica demuestra que estos sistemas actúan, con frecuencia, como conservadores del *statu quo*, automatizando y legitimando desigualdades históricas bajo una apariencia de neutralidad matemática.

El contraste más agudo reside en la naturaleza misma del dato: lo que para un modelo es un patrón estadístico de éxito, para la sociedad puede representar una barrera de exclusión sistémica. Hemos visto que la eliminación del sesgo no es una simple tarea de "limpieza" de bases de datos, sino un acto de reingeniería social que plantea preguntas incómodas sobre la autoridad y el poder. Si la IA es un reflejo de nuestra sociedad, intentar corregirla implica decidir qué valores deben prevalecer, una decisión que, como se ha argumentado, no puede quedar relegada exclusivamente a las esferas técnicas o corporativas.

En este escenario, la responsabilidad social —especialmente en los entornos universitarios— se erige como el único contrapeso viable. La universidad tiene el

deber de formar no solo técnicos competentes, sino "centinelas éticos" capaces de cuestionar la lógica de la caja negra. La consecuencia de no asumir este rol es la creación de una sociedad gobernada por algoritmos que nadie comprende del todo y que pocos pueden impugnar. En última instancia, la ética en la IA no debe entenderse como un freno a la innovación, sino como la condición necesaria para que la tecnología sea verdaderamente humana. La inteligencia artificial será responsable o no será, en el sentido más profundo de la supervivencia de los valores democráticos, una herramienta de progreso.

Bibliografía

- Angwin, J., Larson, J., Mattu, S., y Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Barocas, S., Hardt, M., y Narayanan, A. (2019). Fairness and machine learning: Limitations and opportunities. MIT Press.
- Bender, E. M., Gebru, T., McMillan-Major, A., y Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623. <https://doi.org/10.1145/3442188.3445922>
- Benjamin, R. (2019). Race after technology: Abolitionist tools for the New Jim Code. Polity Press.
- Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. Patterns, 2(2), 100205. <https://doi.org/10.1016/j.patter.2021.100205>
- Broussard, M. (2018). Artificial unintelligence: How computers misunderstand the world. MIT Press.
- Buolamwini, J., y Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1-15.
- Caliskan, A., Bryson, J. J., y Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), 183-186. <https://doi.org/10.1126/science.aal4230>
- Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we need. The MIT Press.
- Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.
- D'Ignazio, C., y Klein, L. F. (2020). Data feminism. The MIT Press.
- Dastin, J. (2018, 10 de octubre). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

- Dignum, V. (2019). Responsible artificial intelligence: How to develop and use AI in a responsible way. Springer Nature. <https://doi.org/10.1007/978-3-030-30371-6>
- Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., y Venkatasubramanian, S. (2018). Runaway feedback loops in predictive policing. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency, 160-171. <https://doi.org/10.1145/3278721.3278733>
- Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
- Floridi, L. (2021). The ethics of artificial intelligence: Principles, challenges, and opportunities. Oxford University Press.
- Fry, H. (2019). Hola mundo: Qué es ser humano en la era de los algoritmos. Blackie Books.
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., y Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. <https://doi.org/10.1145/3458723>
- Grosz, B. J., Grant, D. G., Vredenburg, K., Behrends, J., Hu, L., Simmons, A., y Schneiderman, J. (2019). Embedded EthiCS: Lectures on the ethical and social implications of computing of the computer science curriculum. Communications of the ACM, 62(8), 54-61. <https://doi.org/10.1145/3330794>
- Hardt, M., Price, E., y Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29, 3315-3323.
- Jobin, A., Ienca, M., y Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kearns, M., y Roth, A. (2019). The ethical algorithm: The science of socially aware algorithm design. Oxford University Press.
- Lum, K., e Isaac, W. (2016). To predict and serve? Significance, 13(5), 14-19. <https://doi.org/10.1111/j.1740-9713.2016.00960.x>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., y Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM

Computing Surveys (CSUR), 54(6), 1-35.
<https://doi.org/10.1145/3457607>

- Mitchell, M. (2020). Artificial intelligence: A guide for thinking humans. Pelican Books.
- Mitchell, S., Potash, E., Barocas, S., D'Amour, A., y Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, 8, 141-163.
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501-507.
<https://doi.org/10.1038/s42256-019-0114-4>
- Nissenbaum, H. (2001). How computer systems embody values. *Computer*, 34(3), 118-120. <https://doi.org/10.1109/2.910905>
- Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.
- O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing Group.
- Obermeyer, Z., Powers, B., Vogeli, C., y Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.
<https://doi.org/10.1126/science.aax2342>
- Parlamento Europeo. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas sobre inteligencia artificial. Diario Oficial de la Unión Europea.
- Pasquale, F. (2015). The black box society: The secrecy of algorithms that control money and information. Harvard University Press.
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., y Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, 59-68. <https://doi.org/10.1145/3287560.3287598>
- Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. Polity Press.
- Stanford Institute for Human-Centered AI. (2023). AI Index Report 2023. Stanford University.

- Suresh, H., y Gutttag, J. V. (2021). A framework for understanding sources of unintended bias in machine learning. *Communications of the ACM*, 64(7), 54-62. <https://doi.org/10.1145/3457603>
- Tecnológico de Monterrey y BID Lab. (2022). Iniciativa fAIr LAC: Ética y gobernanza de la inteligencia artificial en América Latina y el Caribe. Tecnológico de Monterrey.
- Universidad Nacional Autónoma de México. (2023). Lineamientos para el uso de inteligencia artificial generativa en la UNAM. CUAIEED.
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.
- Vinué, G. (2021). Sesgos algorítmicos en la inteligencia artificial: Una perspectiva de género. *Revista de Derecho, Empresa y Sociedad (REDS)*, (18), 101-115.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

LA INTELIGENCIA ARTIFICIAL Y LA SUSTITUCIÓN DE PROCESOS HUMANOS EN EL DERECHO

Artificial Intelligence and the substitution of human processes in law

Grecia Josafat Jarillo Olague

Josjarillocv@gmail.com

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0009-0005-9630-4467>

Resumen

El presente capítulo analiza el impacto de la inteligencia artificial en los procesos humanos dentro del derecho mexicano, con especial énfasis en el sistema penal acusatorio, la responsabilidad jurídica derivada de decisiones automatizadas y el acceso a la justicia. A partir del marco constitucional establecido en la Constitución Política de los Estados Unidos Mexicanos y la regulación procesal contenida en el Código Nacional de Procedimientos Penales, se examina cómo la incorporación de tecnologías algorítmicas transforma la toma de decisiones jurídicas, generando tensiones con principios fundamentales como el debido proceso, la presunción de inocencia, la igualdad procesal y la defensa adecuada.

El estudio identifica que la inteligencia artificial introduce una lógica predictiva y automatizada que puede desplazar elementos esenciales del juicio humano, particularmente en ámbitos sensibles como la imposición de medidas cautelares o la valoración de riesgos procesales. Asimismo, se aborda la problemática de la responsabilidad jurídica frente a decisiones automatizadas, destacando la insuficiencia de los marcos normativos tradicionales para atribuir responsabilidad en contextos donde intervienen múltiples actores tecnológicos.

A través del análisis de un caso hipotético, se evidencian los riesgos constitucionales del uso de sistemas de inteligencia artificial sin criterios de

transparencia, explicabilidad y control judicial, lo que puede derivar en violaciones a derechos fundamentales. De igual forma, se examina el impacto de la digitalización y automatización en el acceso a la justicia, subrayando la existencia de una brecha digital que puede profundizar desigualdades estructurales en el contexto mexicano.

En conclusión, el capítulo sostiene que la inteligencia artificial debe ser concebida como una herramienta complementaria y no sustitutiva de los procesos humanos en el derecho, cuya implementación requiere una regulación integral basada en principios de transparencia, supervisión humana, rendición de cuentas y protección de derechos fundamentales.

Palabras clave: Inteligencia Artificial (IA), sistema penal acusatorio, debido proceso, responsabilidad jurídica, decisiones automatizadas, acceso a la justicia, derechos fundamentales, transparencia algorítmica, supervisión humana, Código Nacional de Procedimientos Penales (CNPP).

Abstract

This chapter analyzes the impact of artificial intelligence on human processes within Mexican law, with special emphasis on the accusatory criminal system, legal liability derived from automated decisions, and access to justice. Based on the constitutional framework established in the Political Constitution of the United Mexican States and the procedural regulation contained in the National Code of Criminal Procedure, it examines how the incorporation of algorithmic technologies transforms legal decision-making, generating tensions with fundamental principles such as due process, the presumption of innocence, procedural equality, and adequate defense.

The study identifies that artificial intelligence introduces predictive and automated logic that can displace essential elements of human judgment, particularly in sensitive areas such as the imposition of precautionary measures or the assessment of procedural risks. Likewise, the problem of legal liability in the face of

automated decisions is addressed, highlighting the insufficiency of traditional regulatory frameworks to attribute responsibility in contexts where multiple technological actors intervene.

Through the analysis of a hypothetical case, the constitutional risks of the use of artificial intelligence systems without criteria of transparency, explainability and judicial control are evidenced, which can lead to violations of fundamental rights. Similarly, the impact of digitalization and automation on access to justice is examined, underlining the existence of a digital divide that can deepen structural inequalities in the Mexican context.

In conclusion, the chapter argues that artificial intelligence should be conceived as a complementary tool and not a substitute for human processes in law, the implementation of which requires comprehensive regulation based on principles of transparency, human supervision, accountability, and protection of fundamental rights.

Keywords: Artificial Intelligence (AI), accusatory criminal system, due process, legal responsibility, automated decisions, access to justice, fundamental rights, algorithmic transparency, human supervision, National Code of Criminal Procedure (CNPP).

Introducción

La inteligencia artificial (IA) se ha consolidado como una de las tecnologías más disruptivas del siglo XXI, generando transformaciones profundas en múltiples ámbitos sociales, incluyendo el derecho. Su impacto no se limita a la automatización de tareas, sino que implica una reconfiguración estructural de los procesos humanos en la creación, interpretación y aplicación de normas jurídicas. En este contexto, la IA plantea tanto oportunidades como desafíos, particularmente en relación con la sustitución parcial de funciones humanas, la toma de decisiones automatizadas y la redefinición de principios fundamentales del sistema jurídico.

Uno de los principales impactos de la IA en el derecho se observa en la automatización de tareas jurídicas tradicionalmente realizadas por profesionales. Actividades como la revisión de contratos, la búsqueda de precedentes o el análisis de grandes volúmenes de información pueden ser ejecutadas con mayor rapidez y precisión mediante sistemas de IA. Esto ha generado una transformación en la práctica profesional, donde el abogado ya no es el único operador del conocimiento jurídico, sino que comparte este espacio con herramientas tecnológicas avanzadas. Sin embargo, esta automatización no implica necesariamente la eliminación del factor humano, sino más bien su desplazamiento hacia funciones de supervisión, interpretación crítica y toma de decisiones complejas.

En el ámbito judicial, la IA ha comenzado a ser utilizada para apoyar la toma de decisiones mediante sistemas de análisis predictivo. Estos sistemas pueden evaluar probabilidades de reincidencia, riesgos procesales o incluso sugerir resoluciones basadas en patrones previos. Aunque esto puede contribuir a la eficiencia del sistema judicial, también plantea preocupaciones significativas en torno a la transparencia, la imparcialidad y el debido proceso. La opacidad de los algoritmos, conocida como “caja negra”, dificulta la comprensión de cómo se llegan a determinadas decisiones, lo cual puede vulnerar el derecho a una defensa adecuada (Pasquale, 2015).

Asimismo, la utilización de IA en el derecho penal introduce tensiones relevantes en materia de derechos humanos. La predicción del comportamiento humano, basada en datos históricos, puede reproducir sesgos estructurales presentes en la sociedad, afectando de manera desproporcionada a ciertos grupos (O’Neil, 2016). Esto cuestiona principios fundamentales como la presunción de inocencia y la igualdad ante la ley, ya que las decisiones pueden basarse en probabilidades y no en hechos concretos. En este sentido, la sustitución de procesos humanos por sistemas automatizados no es neutral, sino que puede amplificar desigualdades existentes.

Otro aspecto clave es la responsabilidad jurídica derivada del uso de IA. La determinación de quién es responsable por las decisiones tomadas por sistemas autónomos representa un desafío significativo para el derecho. La tradicional noción de responsabilidad basada en la conducta humana se ve cuestionada cuando intervienen algoritmos que operan con cierto grado de autonomía. Esto ha llevado al desarrollo de propuestas regulatorias que buscan establecer marcos de responsabilidad adaptados a estas nuevas tecnologías, incluyendo la posibilidad de atribuir responsabilidad a desarrolladores, operadores o incluso a los propios sistemas en ciertos contextos.

En el ámbito del derecho laboral, la IA también está transformando las relaciones de trabajo. La automatización de tareas y la implementación de sistemas de gestión algorítmica han modificado las condiciones laborales, generando nuevas formas de control y evaluación del desempeño (De Stefano, 2020). Esto plantea interrogantes sobre la protección de los derechos de los trabajadores, la privacidad y la dignidad en el entorno laboral. La sustitución de procesos humanos por sistemas automatizados puede reducir la intervención humana en la toma de decisiones laborales, lo cual exige una revisión de los marcos normativos existentes.

Por otro lado, la IA está influyendo en la producción normativa y en la formulación de políticas públicas. Los sistemas de análisis de datos permiten a los legisladores identificar patrones y tendencias que pueden ser utilizados para diseñar normas más eficientes. Sin embargo, esto también implica el riesgo de una tecnocratización del derecho, donde las decisiones se basan en criterios algorítmicos más que en deliberaciones democráticas. La integración de la IA en estos procesos requiere garantizar la participación humana y la rendición de cuentas.

En términos éticos, la incorporación de la IA en el derecho plantea la necesidad de desarrollar principios que orienten su uso. Conceptos como la transparencia, la explicabilidad, la justicia y la no discriminación se han convertido en ejes centrales de la discusión (Jobin, Ienca & Vayena, 2019). Estos principios buscan asegurar que la IA no sustituya de manera indebida la capacidad humana de juicio, sino que

funcione como una herramienta complementaria. La preservación de la agencia humana es fundamental para mantener la legitimidad del sistema jurídico.

Finalmente, es importante destacar que el impacto de la IA en el derecho no es uniforme, sino que depende de factores contextuales como el nivel de desarrollo tecnológico, el marco regulatorio y las condiciones sociales de cada país. En contextos como América Latina, donde existen desigualdades estructurales, la implementación de IA en el ámbito jurídico debe ser cuidadosamente evaluada para evitar la profundización de dichas brechas. La regulación de la IA debe equilibrar la innovación tecnológica con la protección de los derechos fundamentales.

Inteligencia artificial y debido proceso en el sistema penal acusatorio mexicano

La incorporación de la inteligencia artificial (IA) en el ámbito jurídico, particularmente en el sistema penal acusatorio mexicano, constituye uno de los fenómenos más complejos y transformadores en la relación entre tecnología y derecho. Este proceso no solo implica la adopción de herramientas tecnológicas para optimizar funciones, sino que supone una reconfiguración profunda de los procesos humanos que históricamente han estructurado la impartición de justicia. En este sentido, la IA no debe analizarse únicamente como un instrumento técnico, sino como un agente que incide directamente en la forma en que se toman decisiones jurídicas, se valoran pruebas y se garantizan derechos fundamentales.

El sistema penal acusatorio en México, consolidado a partir de la reforma constitucional de 2008, se fundamenta en principios como la oralidad, la inmediación, la contradicción, la publicidad y la presunción de inocencia. Estos principios están consagrados en la Constitución Política de los Estados Unidos Mexicanos, particularmente en sus artículos 14, 16, 17, 19 y 20, así como en el Código Nacional de Procedimientos Penales. Dicho modelo procesal tiene como eje central la intervención humana directa en la toma de decisiones, privilegiando la valoración racional de la prueba y la interacción entre las partes y el juez. Sin embargo, la introducción de sistemas de IA en este contexto plantea una tensión

estructural: la posible sustitución o desplazamiento de funciones humanas por procesos automatizados.

Uno de los principales ámbitos donde la IA puede incidir es en la etapa inicial del proceso penal, particularmente en la determinación de medidas cautelares. En otros sistemas jurídicos, como el estadounidense, se han utilizado algoritmos predictivos para evaluar el riesgo de reincidencia o de fuga de los imputados. Aunque en México este tipo de sistemas no se ha implementado de manera formal y generalizada, la creciente digitalización de los expedientes judiciales y el uso de bases de datos masivas abren la puerta a su eventual adopción. Este escenario obliga a cuestionar si las decisiones basadas en probabilidades estadísticas son compatibles con los estándares constitucionales del debido proceso.

El debido proceso, entendido como el conjunto de garantías que aseguran un juicio justo, se ve particularmente comprometido ante el uso de sistemas automatizados. Uno de sus elementos esenciales es el derecho a una defensa adecuada, que implica la posibilidad real de controvertir las pruebas y argumentos presentados en contra del imputado. Cuando una decisión judicial se apoya en un algoritmo cuyo funcionamiento no es transparente, se dificulta el ejercicio efectivo de este derecho. La opacidad algorítmica, comúnmente denominada como “caja negra”, impide conocer los criterios utilizados para generar una determinada recomendación o resultado, lo que limita la capacidad de la defensa para impugnarlo (Pasquale, 2015).

Asimismo, la presunción de inocencia, consagrada en el artículo 20 constitucional, se ve amenazada por el uso de modelos predictivos. Estos sistemas operan a partir de datos históricos y patrones estadísticos, lo que implica que sus conclusiones se basan en probabilidades y no en hechos concretos del caso. Esto puede llevar a decisiones anticipadas sobre la conducta futura del imputado, lo cual contradice el principio de que toda persona debe ser considerada inocente hasta que se demuestre lo contrario mediante pruebas desahogadas en juicio. En este sentido,

la IA introduce una lógica preventiva que puede desvirtuar la naturaleza garantista del proceso penal.

Otro principio fundamental afectado es el de inmediación, que exige que el juez tenga contacto directo con las pruebas y las partes. La utilización de sistemas de IA como intermediarios en la valoración probatoria puede debilitar este principio, al introducir elementos que no son percibidos directamente por el juzgador. Aunque la IA puede ser útil para organizar información o identificar patrones, su intervención en la valoración de la prueba debe ser limitada, ya que esta función requiere un análisis cualitativo que difícilmente puede ser replicado por un algoritmo.

En relación con la igualdad procesal, la IA también plantea desafíos importantes. Los sistemas algorítmicos pueden reproducir y amplificar sesgos existentes en los datos con los que son entrenados. En contextos como el mexicano, donde existen profundas desigualdades sociales, esto puede traducirse en decisiones discriminatorias que afecten de manera desproporcionada a ciertos grupos. Por ejemplo, si un algoritmo utiliza variables como nivel socioeconómico o lugar de residencia para evaluar riesgos, podría generar resultados que perpetúen estigmas y exclusiones.

Desde una perspectiva normativa, el marco jurídico mexicano aún es incipiente en la regulación específica de la IA. No obstante, existen disposiciones que pueden ser aplicables de manera indirecta. La Ley Federal de Protección de Datos Personales en Posesión de los Particulares establece principios como la licitud, la finalidad y la proporcionalidad en el tratamiento de datos personales, los cuales son fundamentales para el funcionamiento de sistemas de IA. Asimismo, la Suprema Corte de Justicia de la Nación ha desarrollado criterios relevantes sobre el debido proceso, la motivación de las resoluciones y la protección de derechos fundamentales, que pueden servir como base para evaluar la constitucionalidad del uso de tecnologías en el ámbito judicial.

Un aspecto clave en este debate es la necesidad de garantizar la explicabilidad de los sistemas de IA. Esto implica que las decisiones generadas por algoritmos deben

ser comprensibles para los usuarios y, en el contexto jurídico, para las partes involucradas en el proceso. La exigencia de motivación de las resoluciones judiciales, prevista en el artículo 16 constitucional, implica que el juez debe expresar de manera clara y razonada los fundamentos de su decisión. Si esta se basa en un sistema automatizado, es indispensable que dicho sistema permita identificar los criterios utilizados, de lo contrario se estaría vulnerando este principio.

Por otro lado, la IA también puede ofrecer beneficios significativos para el sistema penal. La automatización de tareas administrativas, la gestión eficiente de expedientes y el análisis de grandes volúmenes de información pueden contribuir a reducir la carga de trabajo de los operadores jurídicos y mejorar el acceso a la justicia. En este sentido, el reto no es rechazar la tecnología, sino integrarla de manera que fortalezca el sistema sin comprometer sus principios fundamentales.

La inteligencia artificial tiene el potencial de transformar profundamente el sistema penal acusatorio mexicano, pero su implementación debe ser cuidadosamente regulada para garantizar el respeto al debido proceso. La sustitución de procesos humanos por sistemas automatizados no puede realizarse sin considerar las implicaciones jurídicas y éticas que ello conlleva. Es necesario establecer límites claros al uso de la IA, asegurando que su función sea complementaria y no sustitutiva del juicio humano. Solo así será posible aprovechar sus beneficios sin poner en riesgo los derechos fundamentales que constituyen la base del sistema jurídico.

Responsabilidad jurídica por decisiones automatizadas en el sistema jurídico mexicano

La irrupción de la inteligencia artificial (IA) en los distintos ámbitos de la vida social ha generado una transformación significativa en la forma en que se producen decisiones con efectos jurídicos. En México, este fenómeno plantea un desafío estructural: la redefinición de los esquemas tradicionales de responsabilidad jurídica frente a actos realizados, directa o indirectamente, por sistemas automatizados. A diferencia de los modelos clásicos, basados en la conducta

humana, la IA introduce una nueva lógica en la que la autonomía relativa de los sistemas tecnológicos complica la atribución de responsabilidad.

En el derecho mexicano, la responsabilidad jurídica se construye a partir de categorías bien definidas: responsabilidad civil, administrativa y penal. Todas ellas parten de un elemento común: la existencia de un sujeto que actúa con voluntad y cuya conducta puede ser valorada conforme a parámetros de culpa o dolo. Sin embargo, cuando interviene un sistema de inteligencia artificial, especialmente aquellos que operan mediante aprendizaje automático, se diluye la relación directa entre el sujeto humano y el resultado dañoso, generando una ruptura en los esquemas tradicionales de imputación (Floridi y otros., 2018).

Desde la perspectiva de la responsabilidad civil, el punto de partida se encuentra en el Código Civil Federal, particularmente en su artículo 1910, que establece la obligación de reparar el daño causado por un hecho ilícito. Este modelo presupone que el daño es consecuencia de una acción u omisión imputable a una persona. No obstante, en el contexto de la IA, el daño puede derivarse de decisiones tomadas por un sistema que actúa de manera autónoma, sin intervención humana directa en el momento de la ejecución. Esta situación genera incertidumbre sobre la imputación del daño, lo que ha llevado a la doctrina a replantear los criterios tradicionales de responsabilidad.

Una de las principales dificultades radica en la identificación del sujeto responsable. En el ecosistema de la IA intervienen múltiples actores: el desarrollador del algoritmo, el proveedor del sistema, la empresa que lo implementa y el usuario final. Cada uno de estos sujetos tiene distintos niveles de control y conocimiento sobre el funcionamiento del sistema, lo que complica la atribución de responsabilidad. En este sentido, algunos autores han propuesto modelos de responsabilidad distribuida, en los que se asigna responsabilidad en función del grado de intervención y control de cada actor (Calo, 2015).

En el contexto mexicano, una de las teorías que puede resultar aplicable es la del riesgo creado. Conforme a esta doctrina, quien introduce un elemento

potencialmente peligroso en la sociedad debe responder por los daños que cause, independientemente de la existencia de culpa. Esta teoría ha sido reconocida en la práctica judicial mexicana y podría extenderse a los sistemas de IA, en la medida en que estos generan riesgos inherentes derivados de su complejidad y autonomía. Bajo este enfoque, las empresas que desarrollan o implementan IA podrían ser responsables por los daños derivados de su uso, incluso si no actuaron con negligencia.

En materia de consumo, la Ley Federal de Protección al Consumidor establece un marco relevante para analizar la responsabilidad por decisiones automatizadas. Esta ley impone a los proveedores la obligación de garantizar la calidad y seguridad de los bienes y servicios que ofrecen. Si un sistema basado en IA causa un daño al consumidor, el proveedor podría ser responsable, incluso si el daño se deriva de un comportamiento no previsto del sistema. Esto es consistente con el principio de protección al consumidor, que busca equilibrar la relación entre proveedores y usuarios.

Por otro lado, la protección de datos personales constituye un eje fundamental en la regulación de la IA. Los sistemas automatizados dependen del procesamiento masivo de datos, lo que implica riesgos significativos para la privacidad. La Ley Federal de Protección de Datos Personales en Posesión de los Particulares establece principios como el consentimiento, la finalidad, la proporcionalidad y la responsabilidad en el tratamiento de datos. Si un sistema de IA vulnera estos principios, puede generarse responsabilidad administrativa y civil para los sujetos obligados.

Un aspecto particularmente relevante es el de las decisiones automatizadas que afectan derechos sin intervención humana directa. Aunque en México no existe una disposición expresa que regule este fenómeno, la doctrina internacional ha desarrollado el concepto del derecho a no ser objeto de decisiones exclusivamente automatizadas. Este principio puede vincularse con el derecho al debido proceso y a la tutela judicial efectiva, reconocidos en la Constitución Política de los Estados

Unidos Mexicanos. La falta de intervención humana en decisiones que afectan derechos fundamentales puede considerarse una vulneración a estas garantías.

En el ámbito administrativo, el uso de IA por parte de autoridades también plantea retos significativos. Las decisiones de la administración pública deben cumplir con los principios de legalidad, fundamentación y motivación. La Suprema Corte de Justicia de la Nación ha sostenido que la motivación implica la exposición clara de las razones que justifican un acto de autoridad. Sin embargo, los sistemas de IA suelen carecer de explicabilidad, lo que dificulta cumplir con este requisito. Esto podría dar lugar a la nulidad de los actos administrativos basados en decisiones automatizadas opacas.

En materia penal, la problemática adquiere una dimensión aún más compleja. El derecho penal mexicano se basa en el principio de culpabilidad, que exige la existencia de dolo o culpa para atribuir responsabilidad. Dado que los sistemas de IA no tienen conciencia ni voluntad, no pueden ser sujetos de responsabilidad penal. No obstante, esto no excluye la posibilidad de imputar responsabilidad a las personas que intervienen en su diseño, programación o utilización. En este sentido, la IA puede ser considerada un instrumento del delito, lo que permite aplicar las reglas tradicionales de autoría y participación. Otro problema relevante es el de los errores algorítmicos. Estos pueden derivarse de fallas en los datos, en el diseño del modelo o en su implementación. A diferencia de los errores humanos, que pueden evaluarse en función de estándares de diligencia, los errores de la IA requieren criterios técnicos especializados. Esto plantea la necesidad de desarrollar estándares jurídicos que permitan determinar cuándo un error es atribuible a negligencia y cuándo constituye un riesgo inherente a la tecnología.

En este contexto, la regulación de la IA en México se encuentra en una etapa incipiente. Si bien existen avances en materia de gobierno digital y protección de datos, aún no se cuenta con un marco integral que regule la responsabilidad por decisiones automatizadas. Esto genera incertidumbre jurídica y puede afectar tanto a los usuarios como a los desarrolladores de tecnología. No obstante, también

representa una oportunidad para construir un modelo regulatorio que integre principios como la transparencia, la rendición de cuentas, la supervisión humana y la protección de derechos fundamentales.

La inteligencia artificial desafía los fundamentos tradicionales de la responsabilidad jurídica en México al introducir un nuevo tipo de agente en la producción de decisiones con efectos legales. La sustitución parcial de procesos humanos por sistemas automatizados exige una reconfiguración de los criterios de imputación, que permita atribuir responsabilidad de manera clara y justa. El derecho mexicano debe evolucionar para enfrentar estos desafíos, garantizando que la innovación tecnológica no se traduzca en una disminución de la protección de los derechos de las personas, sino en una mejora del sistema jurídico en su conjunto.

Caso práctico: uso de inteligencia artificial en la imposición de medidas cautelares y su control constitucional en México

La incorporación de sistemas de inteligencia artificial (IA) en la toma de decisiones dentro del proceso penal plantea uno de los escenarios más críticos para el derecho mexicano contemporáneo. A diferencia de otros ámbitos, en materia penal están en juego derechos fundamentales de la más alta jerarquía, como la libertad personal, la presunción de inocencia y el debido proceso. Este capítulo desarrolla un caso hipotético verosímil, construido a partir del marco normativo mexicano, con el objetivo de analizar las implicaciones jurídicas del uso de IA en la imposición de medidas cautelares y su posible impugnación mediante el juicio de amparo.

I. Planteamiento del caso

En la Ciudad de México, un juez de control conoce de una audiencia inicial en la que el Ministerio Público formula imputación en contra de una persona por el delito de fraude. Durante la audiencia de vinculación a proceso, el Ministerio Público solicita la imposición de prisión preventiva justificada, argumentando riesgo de fuga.

Para sustentar su solicitud, presenta un informe generado por un sistema de inteligencia artificial desarrollado por una empresa privada, el cual analiza variables como:

- Nivel socioeconómico
- Domicilio
- Historial laboral
- Redes de contacto
- Antecedentes administrativos

El sistema arroja un “índice de riesgo” alto, concluyendo que existe una probabilidad elevada de que el imputado evada la acción de la justicia.

El juez, al momento de resolver, hace referencia expresa al resultado del sistema y determina imponer prisión preventiva justificada, señalando que el análisis algorítmico constituye un elemento objetivo que refuerza el riesgo procesal.

La defensa manifiesta que desconoce el funcionamiento del sistema, los criterios utilizados y la forma en que se procesaron los datos, por lo que no puede controvertir adecuadamente dicha “prueba”.

II. Problema jurídico

El problema central consiste en determinar si el uso de inteligencia artificial como fundamento para imponer una medida cautelar restrictiva de la libertad es compatible con los derechos fundamentales reconocidos en la Constitución Política de los Estados Unidos Mexicanos y en el Código Nacional de Procedimientos Penales.

En particular, se plantean las siguientes cuestiones:

1. ¿Vulnera el uso de IA el derecho al debido proceso?
2. ¿Se afecta la presunción de inocencia al basar la decisión en predicciones?
3. ¿Se transgrede el derecho de defensa al no poder controvertir el algoritmo?
4. ¿Cumple la resolución judicial con el deber de fundamentación y motivación?

III. Marco normativo aplicable

El análisis debe partir de los siguientes preceptos:

- Artículo 14 constitucional: garantía de debido proceso
- Artículo 16 constitucional: fundamentación y motivación
- Artículo 19 constitucional: excepcionalidad de la prisión preventiva
- Artículo 20 constitucional: presunción de inocencia y defensa adecuada

Asimismo, el CNPP establece que las medidas cautelares deben imponerse con base en criterios de proporcionalidad, necesidad e idoneidad, y siempre a partir de datos de prueba concretos, no de inferencias abstractas.

IV. Conceptos de violación (estructura tipo amparo)

1. Violación al debido proceso por uso de prueba ininteligible

El uso de un sistema de IA cuyo funcionamiento no es accesible para las partes vulnera el derecho al debido proceso, ya que impide conocer los elementos que sustentan la decisión judicial. La doctrina ha señalado que los algoritmos opacos constituyen una “caja negra” que limita la transparencia y la rendición de cuentas.

En este caso, la defensa no puede conocer:

- Qué variables fueron determinantes
- Cómo se ponderaron los datos
- Si existen sesgos en el modelo

Esto impide ejercer una defensa efectiva, lo que contraviene estándares constitucionales.

2. Violación a la presunción de inocencia

La presunción de inocencia implica que toda persona debe ser tratada como inocente hasta que se demuestre lo contrario mediante pruebas en juicio. Sin embargo, los sistemas de IA operan con base en probabilidades y patrones históricos, lo que implica una anticipación del comportamiento futuro.

El uso de estas herramientas para justificar medidas cautelares transforma el proceso penal en un mecanismo predictivo, lo cual resulta incompatible con el modelo acusatorio (O'Neil, 2016). No se juzga el hecho, sino el riesgo estimado, lo cual desnaturaliza el proceso penal.

3. Violación al derecho de defensa y contradicción

El CNPP reconoce el derecho de las partes a controvertir los datos de prueba. Sin embargo, cuando se utiliza un algoritmo cuyo funcionamiento no es explicable, este derecho se vuelve ilusorio.

La defensa no puede:

- Interrogar al “sistema”
- Cuestionar la metodología
- Verificar la veracidad de los datos

Esto genera una asimetría procesal incompatible con el principio de igualdad de armas.

4. Falta de fundamentación y motivación

El artículo 16 constitucional exige que toda resolución esté debidamente fundada y motivada. Esto implica explicar las razones jurídicas y fácticas que justifican la decisión. Cuando el juez se limita a citar el resultado de un sistema de IA sin explicar su funcionamiento, incurre en una motivación aparente. La Suprema Corte de Justicia de la Nación ha sostenido que la motivación debe ser clara, lógica y comprensible, lo cual no se cumple en este caso.

V. Análisis constitucional

El uso de IA en este contexto debe evaluarse bajo un test de proporcionalidad:

Idoneidad

La herramienta podría considerarse útil para identificar riesgos, pero su utilidad depende de la calidad de los datos y la ausencia de sesgos.

Necesidad

No es necesaria, ya que existen medios tradicionales para evaluar el riesgo procesal, como el análisis del arraigo, antecedentes y comportamiento del imputado.

Proporcionalidad en sentido estricto

El impacto en los derechos fundamentales (libertad personal, defensa, presunción de inocencia) es desproporcionado en relación con el beneficio que ofrece la herramienta.

Por tanto, el uso de IA en este caso no supera la prueba de constitucionalidad.

VI. Posible resolución en amparo

Un tribunal de amparo podría resolver en los siguientes términos:

- Declarar que el uso de IA sin transparencia vulnera el debido proceso
- Establecer que las decisiones judiciales no pueden basarse exclusivamente en algoritmos
- Ordenar la reposición del procedimiento
- Establecer criterios sobre el uso de tecnologías en el proceso penal

Este tipo de resolución podría sentar un precedente relevante para la regulación de la IA en México.

VII. Reflexión final

Este caso evidencia cómo la inteligencia artificial puede transformar los procesos humanos en el derecho, particularmente en el ámbito penal. Si bien estas herramientas pueden mejorar la eficiencia, su uso sin regulación adecuada puede generar violaciones graves a derechos fundamentales.

La clave no está en prohibir la IA, sino en establecer límites claros:

- Transparencia algorítmica
- Supervisión humana obligatoria

- Derecho a explicación
- Control judicial estricto

Solo así será posible integrar la tecnología sin comprometer la esencia garantista del sistema penal mexicano.

Inteligencia artificial, acceso a la justicia y brecha digital en el sistema jurídico mexicano

En este contexto, el impacto de la IA en los procesos humanos no solo debe analizarse desde la perspectiva de la automatización, sino también en términos de quiénes pueden beneficiarse de estas herramientas y quiénes quedan fuera de su alcance. El derecho de acceso a la justicia se encuentra reconocido en el artículo 17 de la Constitución Política de los Estados Unidos Mexicanos, el cual establece que toda persona tiene derecho a que se le administre justicia de manera pronta, completa e imparcial. Este derecho implica no solo la existencia de tribunales, sino también la posibilidad real de acudir a ellos en condiciones de igualdad. En este sentido, cualquier transformación tecnológica del sistema judicial debe evaluarse en función de su impacto en este derecho fundamental.

Uno de los principales argumentos a favor de la IA en el ámbito jurídico es su capacidad para democratizar el acceso a servicios legales. Herramientas como chatbots jurídicos, sistemas de orientación legal automatizada y plataformas de resolución de disputas en línea pueden facilitar el acceso a información y asesoría básica, especialmente para personas que no cuentan con recursos para contratar un abogado. En México, donde existe una alta demanda de servicios legales y una limitada capacidad institucional, estas tecnologías podrían contribuir a reducir la saturación del sistema judicial.

Sin embargo, esta visión optimista debe matizarse a la luz de la realidad social mexicana. La implementación de soluciones basadas en IA presupone el acceso a infraestructura tecnológica, conectividad y alfabetización digital, condiciones que no están garantizadas para toda la población. De acuerdo con diversos estudios, existe una brecha digital significativa en México, particularmente en zonas rurales y entre

grupos en situación de vulnerabilidad. Esto implica que la digitalización de la justicia podría, paradójicamente, excluir a quienes más necesitan protección jurídica.

La brecha digital no solo se refiere al acceso a dispositivos o internet, sino también a la capacidad de utilizar estas herramientas de manera efectiva. En el contexto jurídico, esto adquiere una dimensión crítica, ya que el ejercicio de derechos requiere comprensión, interpretación y toma de decisiones informadas. Si los sistemas de IA sustituyen o median procesos humanos sin considerar estas desigualdades, pueden generar una forma de exclusión estructural, en la que ciertos sectores de la población quedan fuera del sistema de justicia.

Otro aspecto relevante es la deshumanización de los procesos judiciales. La justicia no es únicamente un mecanismo de resolución de conflictos, sino también un espacio de reconocimiento y dignidad para las personas. La interacción con jueces, defensores y otros operadores jurídicos tiene un componente humano que difícilmente puede ser replicado por sistemas automatizados. La sustitución de estas interacciones por interfaces digitales puede afectar la percepción de legitimidad del sistema y la confianza de los ciudadanos en las instituciones.

En el ámbito institucional, México ha avanzado en la digitalización de ciertos procesos judiciales, especialmente a raíz de la pandemia por COVID-19. La implementación de audiencias virtuales, expedientes electrónicos y sistemas de gestión digital ha permitido mantener la operatividad del sistema judicial en condiciones adversas. Sin embargo, estos avances también han evidenciado las limitaciones tecnológicas y las desigualdades en el acceso a la justicia digital.

La Suprema Corte de Justicia de la Nación ha reconocido la importancia de garantizar el acceso a la justicia en condiciones de igualdad, lo que implica que cualquier innovación tecnológica debe ser compatible con este principio. En este sentido, la incorporación de IA en el sistema judicial debe ir acompañada de políticas públicas que aseguren la inclusión digital, la capacitación de los usuarios y la disponibilidad de alternativas presenciales.

Desde una perspectiva normativa, la regulación de la IA en México aún es incipiente. No obstante, existen principios constitucionales y legales que pueden orientar su implementación. El principio de igualdad, consagrado en el artículo 1 constitucional, obliga a las autoridades a evitar prácticas discriminatorias, lo que incluye aquellas derivadas del uso de tecnologías. Asimismo, el derecho al debido proceso exige que las personas puedan participar de manera efectiva en los procedimientos, lo que puede verse comprometido si no cuentan con los medios para interactuar con sistemas digitales.

En este contexto, es necesario desarrollar un enfoque de “justicia digital inclusiva”, que reconozca las ventajas de la IA sin ignorar sus riesgos. Este enfoque debe considerar al menos cuatro elementos:

1. **Accesibilidad tecnológica:** garantizar que todas las personas puedan acceder a las herramientas digitales.
2. **Alfabetización jurídica y digital:** capacitar a los usuarios para interactuar con sistemas automatizados.
3. **Supervisión humana:** asegurar que las decisiones relevantes no sean completamente automatizadas.
4. **Alternativas no digitales:** mantener canales tradicionales de acceso a la justicia.

Además, es fundamental evaluar el impacto de la IA en términos de derechos humanos. La automatización de procesos puede generar eficiencias, pero también puede afectar derechos como la privacidad, la igualdad y la dignidad. En este sentido, la regulación debe incorporar mecanismos de evaluación de impacto, auditorías algorítmicas y estándares de transparencia.

La experiencia comparada muestra que la implementación de IA en la justicia debe ser gradual y cuidadosamente diseñada. En algunos países, los sistemas de resolución de disputas en línea han tenido éxito en conflictos de baja cuantía, pero su extensión a materias más complejas ha generado controversias. En México,

donde el sistema judicial enfrenta retos estructurales, la adopción de estas tecnologías debe ser particularmente cuidadosa.

En conclusión, la inteligencia artificial tiene el potencial de transformar el acceso a la justicia en México, pero su impacto dependerá de la forma en que se implemente. Si se utiliza de manera inclusiva, puede ampliar el acceso y mejorar la eficiencia del sistema. Sin embargo, si se adopta sin considerar las desigualdades existentes, puede profundizar la brecha digital y generar nuevas formas de exclusión. El desafío consiste en integrar la tecnología de manera que fortalezca los procesos humanos, en lugar de sustituirlos de forma indiscriminada.

Conclusión

La irrupción de la inteligencia artificial en el ámbito jurídico no constituye un fenómeno meramente técnico, sino una transformación estructural que obliga a replantear los fundamentos mismos del derecho como sistema de regulación social. A lo largo de los capítulos desarrollados, ha quedado claro que la IA no solo impacta en la eficiencia de los procesos, sino que incide directamente en la forma en que se construyen, interpretan y aplican las normas jurídicas, así como en la manera en que los sujetos participan en dichos procesos. En este sentido, el derecho enfrenta un desafío doble: adaptarse a las nuevas tecnologías sin perder su esencia garantista y, al mismo tiempo, evitar que la automatización sustituya indebidamente los elementos humanos que le dan legitimidad.

En el contexto mexicano, este desafío adquiere una relevancia particular debido a las características del sistema jurídico y las condiciones sociales del país. El modelo constitucional, basado en la protección de derechos humanos, establece límites claros a cualquier forma de intervención estatal o privada que afecte la esfera jurídica de las personas. La Constitución Política de los Estados Unidos Mexicanos no solo reconoce derechos como el debido proceso, la presunción de inocencia y el acceso a la justicia, sino que también impone obligaciones a las autoridades para garantizar su efectividad. En este marco, la incorporación de la IA debe evaluarse

no solo en términos de eficiencia, sino principalmente en función de su compatibilidad con estos principios.

Uno de los hallazgos más relevantes es que la inteligencia artificial tiende a introducir una lógica distinta a la del derecho tradicional. Mientras que el sistema jurídico se basa en la valoración de hechos concretos, la argumentación racional y la deliberación humana, la IA opera a partir de patrones, probabilidades y modelos estadísticos. Esta diferencia no es menor, ya que implica un cambio en la forma en que se toman decisiones. En el ámbito penal, por ejemplo, el uso de sistemas predictivos puede llevar a decisiones anticipadas sobre el comportamiento futuro de las personas, lo cual entra en tensión directa con la presunción de inocencia. La posibilidad de que una persona sea privada de su libertad con base en un cálculo algorítmico plantea interrogantes profundas sobre la legitimidad del sistema.

Asimismo, la opacidad de los sistemas de IA representa uno de los mayores riesgos para el derecho. La dificultad para comprender cómo funcionan los algoritmos —la denominada “caja negra”— limita la transparencia y la rendición de cuentas, elementos esenciales en cualquier sistema democrático. En el ámbito judicial, esto se traduce en una afectación directa al derecho de defensa y al principio de contradicción, ya que las partes no pueden impugnar adecuadamente una decisión si desconocen los criterios que la sustentan. En este sentido, la exigencia de explicabilidad no es un lujo técnico, sino una condición indispensable para la validez jurídica de las decisiones automatizadas.

Por otro lado, el análisis de la responsabilidad jurídica ha evidenciado que los marcos normativos actuales resultan insuficientes para enfrentar los desafíos que plantea la IA. La atribución de responsabilidad en casos de daños causados por sistemas automatizados se complica debido a la multiplicidad de actores involucrados y a la autonomía relativa de la tecnología. Aunque figuras como el riesgo creado ofrecen una posible solución, es necesario desarrollar esquemas más específicos que permitan distribuir la responsabilidad de manera justa y efectiva.

De lo contrario, existe el riesgo de generar espacios de impunidad en los que los daños no sean reparados adecuadamente.

El estudio del caso práctico sobre la imposición de medidas cautelares mediante IA ha permitido ilustrar de manera concreta cómo estas problemáticas se manifiestan en la práctica. Este ejemplo demuestra que, sin una regulación adecuada, el uso de tecnología puede derivar en violaciones graves a derechos fundamentales, incluso cuando se presenta como una herramienta objetiva y neutral. La aparente objetividad de los algoritmos puede ocultar sesgos estructurales y reproducir desigualdades existentes, lo que refuerza la necesidad de un control judicial riguroso.

En el ámbito del acceso a la justicia, la IA presenta una paradoja significativa. Por un lado, tiene el potencial de ampliar el acceso a servicios legales, reducir costos y mejorar la eficiencia del sistema. Por otro, puede profundizar la brecha digital y excluir a sectores de la población que no cuentan con los recursos o habilidades necesarias para interactuar con sistemas tecnológicos. En un país como México, donde las desigualdades sociales son profundas, este riesgo no puede ser ignorado. La digitalización de la justicia debe ir acompañada de políticas públicas que garanticen la inclusión, de lo contrario, se corre el riesgo de construir un sistema más eficiente, pero menos justo.

A partir de lo anterior, es posible afirmar que la clave no radica en aceptar o rechazar la inteligencia artificial, sino en definir las condiciones bajo las cuales puede ser integrada al sistema jurídico. Esto implica reconocer que la tecnología no es neutral y que su diseño e implementación deben estar orientados por principios jurídicos y éticos. Entre estos principios destacan la transparencia, la rendición de cuentas, la no discriminación, la supervisión humana y la protección de derechos fundamentales. Estos elementos deben convertirse en ejes rectores de cualquier política pública o reforma legislativa en materia de IA.

En este sentido, el papel del Estado mexicano es fundamental. Las autoridades no solo deben regular el uso de la IA, sino también garantizar que su implementación

respete el marco constitucional. Esto incluye la creación de normas específicas, la capacitación de operadores jurídicos y el desarrollo de mecanismos de control y supervisión. Asimismo, el Poder Judicial, particularmente la Suprema Corte de Justicia de la Nación, tendrá un rol clave en la interpretación de los derechos fundamentales en el contexto de la IA, estableciendo criterios que orienten su uso y delimiten sus alcances.

Finalmente, es importante destacar que el derecho no puede permanecer estático frente a los avances tecnológicos. La inteligencia artificial representa una oportunidad para modernizar el sistema jurídico y hacerlo más eficiente, pero también constituye un riesgo si se implementa sin una reflexión crítica. El reto consiste en encontrar un equilibrio entre innovación y protección de derechos, entre eficiencia y justicia, entre automatización y humanidad.

En última instancia, el derecho debe seguir siendo un espacio de deliberación humana, donde las decisiones no se reduzcan a cálculos estadísticos, sino que incorporen valores, principios y contextos. La inteligencia artificial puede ser una herramienta poderosa, pero no debe convertirse en un sustituto del juicio humano. Preservar esta distinción será esencial para garantizar que el sistema jurídico continúe cumpliendo su función principal: proteger la dignidad y los derechos de las personas en un mundo cada vez más tecnológico.

Referencias

- Calo, R. (2015). Robotics and the lessons of cyberlaw. *California Law Review*, 103(3), 513–563.
- De Stefano, V. (2020). *Algorithmic management and the future of work*. International Labour Organization.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
- Mendoza, I., & Bygrave, L. A. (2017). The right not to be subject to automated decisions. *Computer Law & Security Review*, 33(1), 77–98.
- Miró Llinares, F. (2020). Inteligencia artificial y derecho penal: Aproximación a los riesgos de la automatización. *Revista de Derecho Penal y Criminología*, (23), 45–78.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Pagallo, U. (2013). *The laws of robots: Crimes, contracts, and torts*. Springer.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Susskind, R. (2019). *Online courts and the future of justice*. Oxford University Press.
- Veale, M., & Brass, I. (2019). Administration by algorithm? *Public Management Review*, 21(12), 1803–1820.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2), 76–99.

INTELIGENCIA ARTIFICIAL RESPONSABLE: UN ANÁLISIS SISTEMÁTICO.

Responsible Artificial Intelligence: A systematic analysis.

Edgar Alfonso Sansores Guerrero

easg@azc.uam.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

Departamento de Administración

<https://orcid.org/0000-0002-4952-2737>

Juana Edith Navarrete Marneou

emarneou@uqroo.eedu.mx

Universidad Autónoma del Estado de Quintana Roo, Campus Bahía

Departamento de Ciencias Económico-Administrativas

<https://orcid.org/0000-0002-0168-6599>

Resumen

En el contexto de la transformación digital contemporánea, la inteligencia artificial (IA) se ha consolidado como una tecnología de propósito general con un impacto significativo en la competitividad organizacional, la innovación y la reconfiguración de los modelos de negocio. Este artículo tiene como objetivo analizar, a partir de una revisión sistemática de la literatura, la conceptualización e implementación de la gobernanza responsable de la IA en el ámbito organizacional, así como explorar el papel de la IA situada en la generación y el sostenimiento de ventajas competitivas. En particular, se busca comprender cómo la integración de la IA, cuando se adapta a contextos específicos, puede fortalecer o debilitar las capacidades organizacionales y sus resultados estratégicos.

Metodológicamente, el estudio se sustenta en una revisión sistemática de la literatura, estructurada en cinco etapas: definición de criterios de búsqueda, identificación de estudios, selección y depuración de documentos, extracción y sistematización de datos y síntesis de la evidencia mediante una matriz de

conceptos. La búsqueda se realizó en bases de datos académicas de alto impacto, considerando publicaciones en inglés publicadas a partir de 2021. El proceso permitió identificar inicialmente 1,208 documentos, de los cuales, tras un riguroso proceso de selección, se obtuvo una muestra final de 55 artículos, lo que aseguró la calidad metodológica y la pertinencia temática de los estudios analizados.

Los resultados evidencian que la IA responsable constituye un constructo multidimensional que integra dimensiones éticas, legales, técnicas y organizacionales, orientadas a garantizar el desarrollo y uso de sistemas de IA alineados con valores sociales y principios de sostenibilidad. Entre los principios más relevantes destacan la transparencia, la explicabilidad, la equidad, la rendición de cuentas, la privacidad de los datos y la robustez técnica. Asimismo, se identifica que los principios de accountability, responsibility y transparency (ART) emergen como ejes fundamentales para la construcción de sistemas confiables y éticamente sostenibles.

Desde una perspectiva organizacional, la IA responsable se configura como un elemento estratégico que requiere la alineación de los intereses de los distintos grupos de interés, la incorporación de mecanismos de gobernanza y la adopción de enfoques como el accountability-by-design, que garantizan la trazabilidad y la supervisión continua de los sistemas de IA. No obstante, se observa una ambigüedad conceptual en la literatura, en la que términos como IA ética, IA confiable e IA responsable se emplean indistintamente, lo que refleja un campo aún en proceso de consolidación teórica.

El estudio pone de relieve la importancia de la IA situada como un enfoque que enfatiza la contextualización de la tecnología en entornos organizacionales específicos. En este sentido, la generación de ventajas competitivas no depende exclusivamente de la adopción de la IA, sino también de su adecuada integración en procesos, estructuras y culturas organizacionales, así como de la capacidad de las organizaciones para fomentar el aprendizaje y la adaptación continuos.

Finalmente, se concluye que la gobernanza de la IA responsable constituye un desafío crítico en un entorno caracterizado por la rápida digitalización, la reutilización intensiva de datos y la creciente complejidad tecnológica. Si bien la IA ofrece oportunidades significativas para mejorar la eficiencia y la competitividad, también plantea riesgos asociados a la opacidad algorítmica, la pérdida de privacidad y la de la autonomía humana. En este contexto, el desarrollo de marcos de gobernanza robustos, integrales y adaptativos resulta fundamental para equilibrar la generación de valor con la mitigación de riesgos, contribuyendo así a una transformación organizacional ética, sostenible y socialmente responsable

Palabras clave: Inteligencia Artificial (IA), gobernanza responsable, IA situada, competitividad organizacional, accountability, transparencia algorítmica, explicabilidad, sostenibilidad, transformación digital, ventajas competitivas.

Abstract

In the context of contemporary digital transformation, Artificial Intelligence (AI) has become established as a general-purpose technology with a significant impact on organizational competitiveness, innovation, and the reconfiguration of business models. This article aims to analyze, through a systematic literature review, the conceptualization and implementation of responsible AI governance within organizational settings, as well as to explore the role of situated AI in the generation and sustainability of competitive advantages. In particular, the study seeks to understand how the integration of AI, when adapted to specific contexts, can strengthen or weaken organizational capabilities and their strategic outcomes.

Methodologically, the study is based on a systematic literature review structured into five stages: definition of search criteria, identification of studies, selection and screening of documents, extraction and systematization of data, and synthesis of evidence through a concept matrix. The search was conducted in high-impact academic databases, considering English-language publications published from 2021 onward. The process initially identified 1,208 documents, from which, after a

rigorous selection procedure, a final sample of 55 articles was obtained, ensuring the methodological quality and thematic relevance of the analyzed studies.

The findings reveal that responsible AI constitutes a multidimensional construct integrating ethical, legal, technical, and organizational dimensions aimed at ensuring the development and use of AI systems aligned with social values and sustainability principles. Among the most relevant principles are transparency, explainability, fairness, accountability, data privacy, and technical robustness. Likewise, the principles of accountability, responsibility, and transparency (ART) emerge as fundamental pillars for the construction of reliable and ethically sustainable systems.

From an organizational perspective, responsible AI is configured as a strategic element that requires the alignment of stakeholders' interests, the incorporation of governance mechanisms, and the adoption of approaches such as accountability-by-design, which ensure traceability and continuous supervision of AI systems. Nevertheless, the literature reveals conceptual ambiguity, as terms such as ethical AI, trustworthy AI, and responsible AI are often used interchangeably, reflecting a field that is still undergoing theoretical consolidation.

The study highlights the importance of situated AI as an approach that emphasizes the contextualization of technology within specific organizational environments. In this regard, the generation of competitive advantages depends not only on the adoption of AI, but also on its appropriate integration into organizational processes, structures, and cultures, as well as on organizations' capacity to foster continuous learning and adaptation.

Finally, the study concludes that responsible AI governance constitutes a critical challenge in an environment characterized by rapid digitalization, intensive data reuse, and increasing technological complexity. Although AI offers significant opportunities to improve efficiency and competitiveness, it also poses risks associated with algorithmic opacity, loss of privacy, and the erosion of human autonomy. In this context, the development of robust, comprehensive, and adaptive governance frameworks is essential to balance value creation with risk mitigation,

thereby contributing to an ethical, sustainable, and socially responsible organizational transformation.

Keywords: Artificial Intelligence (AI), responsible governance, situated AI, organizational competitiveness, accountability, algorithmic transparency, explainability, sustainability, digital transformation, competitive advantages.

1. Introducción

En el último lustro, la Inteligencia Artificial (IA) se ha consolidado como la principal tecnología digital de propósito general (GTP) que ha impulsado una serie de innovaciones (modelos generativos, autoaprendizaje, automatización) sin precedentes desde la primera Revolución Industrial (Goellner, y *otros.*, 2024; Gunasekara, y *otros.*, 2025; Papagiannidis, y *otros.*, 2025; Vujović, 2024; Wan y *otros.*, 2025).

Ante el crecimiento exponencial de la disponibilidad de datos, la democratización de las plataformas digitales y el desarrollo de las capacidades de los sistemas informáticos (procesamiento, almacenamiento y escalabilidad) han permitido que las organizaciones adopten sistemas de IA como un recurso estratégico orientado a generar ventajas competitivas sostenibles (Dzreke, 2025; Condeço y *otros.*, 2026).

Este posicionamiento de la IA como el núcleo de la transformación económica y social actual ha suscitado un intenso debate académico en torno a su pertinencia, impacto, regulación y uso (Goellner, y *otros.*, 2024). Asimismo, ha acentuado, a nivel global, la competencia tecnológica y comercial entre las empresas Big Tech (Google-Alphabet, Microsoft, Amazon, Meta, Apple), los laboratorios de IA de frontera (OpenAI , Anthropic, Google DeepMind) y los desarrolladores de infraestructura (NVIDIA, TSMC, Broadcom).

No obstante, estos avances coexisten con desafíos estructurales que evidencian las limitaciones inherentes de la IA (Shahmerdanova, 2025). Entre estos destacan los riesgos asociados a su diseño y uso, la probabilidad de generar resultados inesperados o no deseados, así como escenarios emergentes relacionados con el desplazamiento tecnológico, la vulnerabilidad ante ciberataques y, en casos extremos, riesgos de carácter existencial (Hammond y *otros.*, 2025; Wheeler, 2025).

En este contexto, ha cobrado especial relevancia el concepto de Inteligencia Artificial Responsable (Responsible AI), que se refiere a un conjunto articulado de principios y prácticas orientados a guiar el diseño, desarrollo, implementación y uso de sistemas de IA, de modo que se mitiguen sus riesgos potenciales para las personas, la sociedad y el entorno (Akbarighatar, 2025). Esta perspectiva tiene como propósito asegurar que dichas tecnologías estén alineadas con valores humanos y normas sociales, promoviendo su desarrollo y aplicación de forma ética, segura y socialmente sostenible (Lee y *otros.*, 2025).

Recientemente, la evidencia empírica sobre incidentes relacionados con el uso indebido de la IA (COMPAS, deepfake, Cambridge Analytica) y acelerada difusión en diversas plataformas, ha mostrado la necesidad de establecer marcos regulatorios y de gobernanza robustos (Radanliev, 2025). Ante esta situación, organismos internacionales y gobiernos han desarrollado diversas iniciativas para orientar el desarrollo de la IA hacia principios de confiabilidad, ética y seguridad.

A pesar de estos avances normativos y de las iniciativas presentadas, persisten vacíos tanto conceptuales como operativos (Madanchian y Taherdoost, 2025). Por una parte, los términos empleados en marcos regulatorios y códigos éticos (IA segura, IA confiable o IA responsable) carecen de definiciones claras y delimitaciones precisas, lo que dificulta su implementación en las prácticas organizacionales (Taylor, 2025; Meduri y *otros.*, 2025).

Por otra parte, el debate académico se ha centrado en la identificación de dimensiones y principios de la IA responsable desde enfoques normativos de alto nivel, excluyendo algunos aspectos vinculados a los mecanismos de gobernanza (estructurales, procedimentales, técnicos, regulatorios, relacionales) necesarios para su implementación (Baker y Xiang, 2023; Radanliev, 2025). En consecuencia, se identifica una brecha en la comprensión de los mecanismos que posibilitan una gobernanza responsable de la IA a lo largo de su ciclo de vida (Meske y *otros.*, 2022).

Por lo expuesto anteriormente, el presente capítulo tiene como objetivo analizar, mediante una revisión sistemática de la literatura, la conceptualización e implementación de la gobernanza responsable de la IA, tanto en el ámbito académico como en la práctica organizacional. Asimismo, se examina cómo la gobernanza responsable de la IA se configura en función del contexto y se estudian sus implicaciones para las organizaciones y su entorno.

2. Marco teórico

La revisión sistemática de la literatura (RSL) se ha consolidado como una perspectiva metodológica central en la producción de conocimiento científico; es específica en contextos en los que la concentración de evidencia es amplia, heterogénea y, en ocasiones, contradictoria (Paul y *otros.*, 2024; Smela y *otros.*, 2023; Nightingale, 2009).

En este sentido, su objetivo fundamental consiste en recopilar de manera exhaustiva la evidencia disponible en torno a una pregunta de investigación específica, con el propósito de elaborar una síntesis rigurosa, equilibrada y metodológicamente libre de sesgos (Matos y *otros.*, 2023).

A diferencia de las revisiones narrativas tradicionales (frecuentemente sujetas a sesgos de selección, confirmación e interpretación), la RSL se fundamenta en la aplicación de procedimientos explícitos, transparentes y replicables orientados a

garantizar la exhaustividad en la identificación de estudios, la objetividad en su selección y la rigurosidad en su análisis (Batool y *otros.*, 2023).

Desde una perspectiva epistemológica, la RSL se sitúa en el paradigma de la investigación basada en evidencia, cuyo objetivo es construir conocimiento acumulativo mediante la integración crítica y sistemática de hallazgos previos (Azarian y *otros.*, 2023). Este enfoque se orienta a reducir la fragmentación del conocimiento científico y a facilitar la identificación de patrones, convergencias, divergencias y vacíos en la literatura.

En este sentido, la RSL es analítica y crítica, ya que permite evaluar la consistencia de los resultados, la calidad metodológica de los estudios y la solidez de las conclusiones (Torres-Carrión y *otros.*, 2018). Para ello, se sustenta en una pregunta de investigación delimitada, que actúa como eje articulador de la metodología, lo que permite definir con precisión el alcance de la revisión y los criterios de inclusión y exclusión (Nightingale, 2009).

En términos metodológicos, la RSL se caracteriza por el diseño de estrategias de búsqueda exhaustivas y sistemáticas, orientadas a maximizar la sensibilidad, incluso a costa de recuperar literatura potencialmente irrelevante (Paul y Barari, 2022). Este principio responde al reconocimiento del sesgo de publicación, según el cual los estudios con resultados positivos tienen mayor probabilidad de ser publicados y de aparecer en revistas de alto impacto, lo que puede distorsionar la evidencia disponible (Smela y *otros.*, 2023).

En consecuencia, la RSL incorpora la búsqueda de literatura (informes técnicos, tesis, actas de congresos) y el uso de múltiples bases de datos, combinando términos controlados con estrategias de búsqueda por palabras clave (Torres-Carrión y *otros.*, 2018; Smela y *otros.*, 2023; Batool y *otros.*, 2025).

El proceso de selección de estudios se sustenta en criterios de inclusión y exclusión definidos a priori, aplicados de manera sistemática para reducir la subjetividad. Sin embargo, algunos estudios han identificado la presencia de sesgos que pueden afectar la validez de la revisión (Batoool, y *otros.*, 2025).

Entre estos, se destacan el sesgo de publicación, el sesgo de selección y el sesgo de extracción de datos, este último asociado a errores o interpretaciones inconsistentes durante la recopilación de información. Para mitigar esta situación, se recomienda la participación de al menos dos revisores independientes, el uso de protocolos estandarizados y la comparación sistemática de resultados (Dekkers y *otros.*, 2022).

Por su parte, la evaluación crítica de la calidad metodológica de los estudios incluidos facilita ponderar la robustez de la evidencia y evitar la incorporación de resultados imprecisos (Matos y *otros.*, 2023). Dicha evaluación se realiza mediante instrumentos estandarizados que consideran aspectos como el diseño del estudio, la validez interna, la representatividad de la muestra y la calidad del análisis (Marzi, 2025).

En cuanto a la síntesis de la evidencia, que constituye la etapa final del proceso, esta se distingue por su carácter flexible, al poder adoptar enfoques tanto cualitativos como cuantitativos. En el enfoque cualitativo se lleva a cabo una integración interpretativa de los hallazgos, orientada a identificar patrones temáticos y relaciones conceptuales (Jungell-Michelsson y Heikkurinen, 2022).

Por su parte, el enfoque cuantitativo recurre al metaanálisis, una técnica que permite combinar estadísticamente los resultados de múltiples estudios con el propósito de estimar efectos agregados (Jungell-Michelsson y Heikkurinen, 2022). A nivel macro, la RSL también cumple una función estratégica en la agenda de investigación al permitir identificar vacíos teóricos, inconsistencias empíricas y oportunidades de desarrollo conceptual (Marzi, 2025).

En áreas de conocimiento emergentes o altamente dinámicas (como la Inteligencia Artificial y su gobernanza), la RSL facilita la organización del conocimiento disperso y contribuye a la construcción de marcos teóricos integradores (Matos y *otros.*, 2023).

En términos procedimentales, la RSL sigue una secuencia estructurada que incluye: a) la definición del problema y la pregunta de investigación, b) el diseño del protocolo de revisión, c) la estrategia de búsqueda, d) la selección de estudios, e) la extracción de datos, f) la evaluación crítica de la calidad y g) la síntesis e interpretación de los resultados (Dekkers y *otros.*, 2022). Este proceso, cuando es debidamente documentado, garantiza la transparencia, trazabilidad y reproducibilidad del estudio, atributos esenciales para la validez científica.

En síntesis, la revisión sistemática de la literatura se configura como una herramienta metodológica robusta que permite superar las limitaciones de las revisiones tradicionales y proporciona una base sólida para la generación de conocimiento, la formulación de políticas públicas y la toma de decisiones informadas (Zamiri y Esmaeili, 2024). Su rigor metodológico y su capacidad de síntesis la convierten en un instrumento indispensable en contextos de alta complejidad y de producción intensiva de conocimiento (Marzi, 2025).

3. Método, técnicas e instrumentos.

En este trabajo de investigación se utilizó una revisión sistemática de la literatura integrada por cinco etapas orientadas a identificar la producción científica relevante en los últimos cinco años. Dicha revisión se sustentó en metodologías ampliamente reconocidas, con el propósito de asegurar la incorporación exhaustiva de las publicaciones pertinentes (Dekkers y *otros.*, 2022; Zamiri y Esmaeili, 2024; Matos y *otros.*, 2023).

Se adoptó un enfoque de revisión, mediante el cual se establecieron los criterios para la búsqueda y el análisis de artículos escritos en idioma inglés. Asimismo, el

proceso fue desarrollado con la participación de tres investigadores, con el objetivo de mitigar posibles sesgos durante la selección de los estudios.

En la primera etapa se definieron diversos criterios de búsqueda, los cuales incorporaron palabras clave relacionadas con la inteligencia artificial desde una perspectiva organizacional. Estos términos se emplearon como palabras clave y, posteriormente, se combinaron mediante operadores booleanos y caracteres comodines, con el propósito de ampliar y refinar el alcance de la búsqueda (Tabla 1). Dicha búsqueda se realizó en bases de datos como Scopus, Web of Science y Springer.

Tabla 1. *Fórmula de búsqueda.*

Categoría	Palabras clave
Inteligencia artificial responsable	<i>“accountable AI, socially responsible AI, credible AI, trustworthy AI, human-centered AI, value-creating AI, principled AI, norm-compliant AI, transparent AI, intelligible AI, AI environment, risks of AI, harmful AI effects, human-centered AI design, value-sensitive design of AI, unbiased AI, equitable AI”.</i>
Contexto	<i>“Regulation, barriers, organizational value, digital transformation, institutional barriers, social value, risk governance.”</i>

Fuente: Elaboración propia.

En esta fase se identificaron inicialmente 1,208 documentos. Posteriormente, se aplicó un proceso iterativo de depuración basado en criterios de relevancia; tras la revisión de títulos, se excluyeron 311 registros por considerarse irrelevantes o duplicados. En una segunda etapa, se llevó a cabo el análisis de los resúmenes, lo que condujo a la exclusión de 315 trabajos adicionales conforme a los criterios previamente establecidos.

La tercera etapa consistió en una evaluación crítica del enfoque metodológico de los estudios, en la cual se descartaron 260 artículos. En la cuarta etapa se efectuó la extracción y sistematización de la información en una hoja de cálculo; los documentos restantes cumplieron con los criterios definidos, lo que dio lugar a una muestra final de 55 artículos.

Finalmente, la quinta etapa correspondió a la síntesis de la evidencia mediante una matriz de conceptos, herramienta que permitió organizar el contenido y establecer relaciones entre los estudios analizados. Con el propósito de delimitar el alcance de la revisión, se establecieron criterios de inclusión y exclusión orientados a garantizar la coherencia entre los estudios seleccionados y los objetivos de investigación.

En este sentido, se consideraron únicamente investigaciones que abordaran la inteligencia artificial desde una perspectiva organizacional. Asimismo, se restringió el periodo de análisis a publicaciones a partir de 2021, en atención al notable incremento de estudios sobre esta temática en la última década. De igual forma, se incluyeron exclusivamente artículos en idioma inglés para asegurar su accesibilidad, y se priorizaron publicaciones en revistas arbitradas, garantizando así un nivel adecuado de rigor académico.

En cuanto a los criterios de exclusión, se descartaron estudios con un enfoque predominantemente técnico. Asimismo, se excluyeron determinados tipos de documentos, tales como series de libros, artículos de divulgación, capítulos de libro, memorias en extenso e informes disponibles en sitios web. Finalmente, se eliminaron trabajos en proceso de revisión o en etapa de prepublicación, con el objetivo de preservar la calidad y la integridad del corpus analizado.

4. Hallazgos y discusión.

El análisis de la literatura pone de manifiesto que la responsabilidad en el ámbito de la inteligencia artificial (IA) constituye un constructo inherentemente complejo, dinámico y multidimensional, cuya comprensión exige un enfoque centrado en el ser humano, orientado al bienestar colectivo y alineado con los valores sociales predominantes.

En este marco, la IA responsable no se limita a una dimensión normativa, sino que implica la integración sistemática de consideraciones éticas, legales, económicas y socioculturales en todas las fases del ciclo de vida de los sistemas inteligentes, con el objetivo de maximizar su contribución al desarrollo social y ambiental, al tiempo que se mitigan sus posibles efectos adversos.

Los resultados evidencian que la operacionalización de la IA responsable se fundamenta en un conjunto de principios rectores que orientan el diseño, desarrollo e implementación de estas tecnologías. Entre los más recurrentes destacan la transparencia y la explicabilidad de los sistemas, la equidad y la prevención de sesgos algorítmicos, la protección de la privacidad y la gobernanza de los datos, la robustez y confiabilidad técnica, la rendición de cuentas y la generación de valor social.

En su conjunto, estos principios configuran un marco integral que trasciende lo técnico y se posiciona como un referente normativo y operativo para el desarrollo ético de la IA. De manera particular, los principios ART (accountability, responsibility y transparency) emergen como ejes estructurales, al subrayar que la responsabilidad última recae en los actores humanos (individuos, organizaciones e instituciones) involucrados en la creación, implementación y uso de estos sistemas.

Desde una perspectiva organizacional, la IA responsable se articula como un componente estratégico que debe integrarse en los procesos de planeación, toma

de decisiones e implementación tecnológica. Esto implica no solo la alineación de los intereses y expectativas de los distintos grupos de interés, sino también la adopción de mecanismos formales de gestión de riesgos.

En este contexto, cobra relevancia el enfoque de *accountability-by-design*, que promueve el desarrollo de sistemas de IA concebidos desde su origen para garantizar trazabilidad, auditabilidad y una cadena continua de responsabilidad humana a lo largo de todo su ciclo de vida.

Adicionalmente, la revisión evidencia la existencia de ambigüedad conceptual en el campo, donde términos como *Responsible AI*, *Ethical AI* y *Trustworthy AI* son frecuentemente utilizados de manera intercambiable, pese a sus diferencias analíticas. Mientras que la IA ética se orienta hacia la alineación con valores morales ideales y la IA confiable enfatiza la percepción de fiabilidad, la IA responsable emerge como una noción integradora que articula estos enfoques en una perspectiva más amplia, pragmática y orientada a la acción.

En el plano institucional, se observa un creciente protagonismo de gobiernos, organismos internacionales y corporaciones en la formulación de marcos regulatorios y lineamientos para la implementación de la IA responsable. Diversas iniciativas a nivel nacional e internacional han contribuido a la definición de principios y estándares orientados a promover el uso ético de la IA y a mitigar sus riesgos.

No obstante, la participación del sector privado en estos procesos introduce tensiones relevantes, ya que puede derivar en la adopción de esquemas regulatorios flexibles que prioricen la innovación o la rentabilidad sobre el cumplimiento estricto de criterios de responsabilidad.

En términos conceptuales, los hallazgos apuntan hacia una convergencia progresiva en torno a un núcleo de principios ampliamente aceptados, entre los que destacan la rendición de cuentas, la equidad, la no discriminación, la supervisión humana, la gobernanza de datos, la robustez técnica, la transparencia y el bienestar social y ambiental.

Sin embargo, la diversidad terminológica y la coexistencia de múltiples marcos analíticos reflejan que el campo aún se encuentra en una fase de consolidación teórica. Asimismo, se identifica la necesidad de fortalecer los mecanismos de gobernanza para enfrentar desafíos inherentes a la IA, tales como la opacidad algorítmica, la creciente autonomía de los sistemas inteligentes y la dificultad para anticipar sus resultados.

Estas características generan incertidumbre y plantean cuestionamientos críticos en torno a la responsabilidad, la transparencia y la rendición de cuentas, lo que refuerza la urgencia de desarrollar marcos regulatorios más robustos y adaptativos.

Finalmente, los resultados destacan que la reutilización intensiva de datos y la digitalización generalizada han posicionado a la IA como un factor clave de competitividad a nivel global. No obstante, su rápida adopción también conlleva riesgos significativos, entre ellos el desplazamiento laboral, la erosión de la privacidad y la reducción de la autonomía humana.

En este contexto, la gobernanza de la IA responsable adquiere un papel central, al definirse como el conjunto de prácticas orientadas a garantizar el desarrollo, implementación y monitoreo de sistemas de IA de manera segura, ética y confiable a lo largo de todo su ciclo de vida.

En síntesis, la evidencia analizada permite afirmar que la IA responsable trasciende su carácter normativo para constituirse como un enfoque estratégico integral, cuyo

propósito es equilibrar la generación de valor con la mitigación de riesgos en un entorno caracterizado por alta complejidad tecnológica, interdependencia sistémica y creciente escrutinio social.

5. Conclusiones

El presente estudio permitió analizar de manera integral la conceptualización y aplicación de la inteligencia artificial (IA) desde una perspectiva organizacional, con énfasis en la gobernanza responsable y el enfoque de IA situada. A partir de la revisión sistemática de la literatura y el análisis teórico desarrollado, se concluye que la IA ha dejado de ser un recurso meramente tecnológico para consolidarse como un elemento estructural en la configuración de ventajas competitivas y en la transformación de los modelos organizacionales contemporáneos.

En primer lugar, los hallazgos evidencian que la IA, en su carácter de tecnología de propósito general, redefine las fuentes tradicionales de ventaja competitiva al introducir nuevas capacidades organizacionales asociadas al procesamiento de datos, la automatización inteligente y la toma de decisiones basada en algoritmos. Sin embargo, este potencial no se materializa de manera homogénea, sino que depende de la capacidad de las organizaciones para contextualizar, adaptar e integrar estas tecnologías en sus procesos, estructuras y culturas.

En este sentido, la noción de IA situada emerge como un marco analítico clave, al enfatizar que el valor de la IA no reside únicamente en sus capacidades técnicas, sino en su articulación con contextos específicos de uso, dinámicas organizacionales y procesos de aprendizaje colectivo.

En segundo lugar, se concluye que la implementación de la IA no es un proceso neutral ni exento de tensiones, sino que implica desafíos significativos de carácter ético, organizacional y estratégico. La evidencia analizada muestra que la adopción

acelerada de estas tecnologías puede generar efectos no deseados, tales como sesgos algorítmicos, opacidad en la toma de decisiones, pérdida de autonomía humana y riesgos asociados a la privacidad y la seguridad de la información.

En consecuencia, la IA no solo debe ser evaluada en términos de eficiencia o rendimiento, sino también desde una perspectiva crítica que considere sus implicaciones sociales y su impacto en la dignidad humana. En este contexto, el concepto de inteligencia artificial responsable adquiere un papel central, al constituirse como un marco normativo y operativo orientado a guiar el diseño, desarrollo e implementación de sistemas de IA bajo principios de transparencia, equidad, rendición de cuentas y sostenibilidad.

No obstante, se identificó que, a pesar de la proliferación de marcos conceptuales y lineamientos institucionales, persisten ambigüedades terminológicas y vacíos operativos que dificultan su implementación efectiva en entornos organizacionales. Esta situación pone de manifiesto la necesidad de avanzar hacia modelos de gobernanza más claros, integrales y adaptativos.

En tercer lugar, se concluye que la gobernanza de la IA responsable debe concebirse como un proceso sociotécnico, en el cual convergen dimensiones técnicas, organizacionales, regulatorias y culturales. Desde esta perspectiva, la gobernanza no se limita al cumplimiento normativo, sino que implica la construcción de capacidades organizacionales orientadas a la gestión de riesgos, la supervisión continua de los sistemas de IA y la alineación con los valores y expectativas de los grupos de interés. Asimismo, el enfoque de *accountability-by-design* se posiciona como un elemento clave para garantizar la trazabilidad, la auditabilidad y la responsabilidad a lo largo de todo el ciclo de vida de la IA.

Por otra parte, el análisis evidencia que existe una tensión estructural entre la innovación tecnológica y la regulación, particularmente en contextos donde el

sector privado desempeña un papel protagónico en el desarrollo de soluciones basadas en IA. Si bien la participación de actores corporativos ha impulsado avances significativos, también plantea riesgos asociados a la adopción de marcos regulatorios flexibles que podrían priorizar intereses económicos sobre criterios éticos y sociales.

En consecuencia, se requiere fortalecer la articulación entre actores públicos, privados y académicos para construir esquemas de gobernanza más equilibrados y legítimos. Adicionalmente, el estudio aporta a la literatura al integrar la discusión sobre IA responsable con la teoría de la IA situada, proponiendo una visión más contextualizada, crítica e inclusiva de la transformación organizacional.

En este sentido, se destaca que la generación de valor a través de la IA no depende únicamente de la tecnología en sí misma, sino de su apropiación social, su integración en prácticas organizacionales y su alineación con procesos de aprendizaje y cambio cultural.

En términos de implicaciones, los resultados sugieren que las organizaciones deben transitar de una lógica de adopción tecnológica hacia una lógica de integración estratégica de la IA, en la cual se consideren simultáneamente los beneficios económicos, los riesgos asociados y las responsabilidades éticas.

Esto implica desarrollar capacidades internas en materia de gobernanza de datos, evaluación de impactos, diseño ético y formación de talento especializado. Finalmente, el estudio reconoce algunas limitaciones, entre ellas el enfoque en literatura reciente y en publicaciones en idioma inglés, lo cual podría restringir la inclusión de perspectivas provenientes de otros contextos geográficos o epistemológicos.

En este sentido, futuras líneas de investigación podrían profundizar en estudios empíricos que analicen la implementación de la IA situada en distintos sectores, así

como en el desarrollo de modelos de gobernanza aplicados que permitan operacionalizar los principios de la IA responsable en contextos específicos.

En síntesis, la evidencia presentada permite concluir que la inteligencia artificial responsable, en articulación con el enfoque de IA situada, constituye un marco analítico y estratégico fundamental para comprender y gestionar la transformación organizacional en la era digital. Su adecuada implementación no solo determinará la sostenibilidad de las ventajas competitivas, sino también el impacto social y ético de estas tecnologías en el largo plazo.

Referencias

- Akbarighatar, P. (2025). Operationalizing responsible AI principles through responsible AI capabilities. *AI and Ethics*, 5(2), 1787-1801.
- Azarian, M., Yu, H., Shiferaw, A. T., & Stevik, T. K. (2023). Do we perform systematic literature review right? A scientific mapping and methodological assessment. *Logistics*, 7(4), 89.
- Baker, S., & Xiang, W. (2023). Explainable AI is responsible AI: How explainability creates trustworthy and socially responsible artificial intelligence. *arXiv preprint arXiv:2312.01555*.
- Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review. *AI and Ethics*, 5(3), 3265-3279.
- Condeço, D., Ferreira, J. J., & Zhang, J. Z. (2026). The Digital Edge: A Systematic Review of AI-Driven Competitive Advantage. *Journal of Computer Information Systems*, 1-15.
- Dekkers, R., Carey, L., & Langhorne, P. (2022). Making literature reviews work: A multidisciplinary guide to systematic approaches. Cham: Springer.
- Dzreke, S. S. (2025). The competitive advantage of AI in business: A strategic imperative. *International Journal for Multidisciplinary Research*, 7(4).
- Goellner, S., Tropmann-Frick, M., & Brumen, B. (2024). Responsible artificial intelligence: A structured literature review. *arXiv preprint arXiv:2403.06910*.
- Gunasekara, L., El-Haber, N., Nagpal, S., Moraliyage, H., Issadeen, Z., Manic, M., & De Silva, D. (2025). A systematic review of responsible artificial intelligence principles and practice. *Applied System Innovation*, 8(4), 97.
- Hammond, L., Chan, A., Clifton, J., Hoelscher-Obermaier, J., Khan, A., McLean, E., & Rahwan, I. (2025). Multi-agent risks from advanced ai. *arXiv preprint arXiv:2502.14143*.
- Jungell-Michelsson, J., & Heikkurinen, P. (2022). Sufficiency: A systematic literature review. *Ecological Economics*, 195, 107380.
- Lee, S. U., Perera, H., Liu, Y., Xia, B., Lu, Q., Zhu, L. & Nottage, M. (2025). Integrating ESG and AI: a comprehensive responsible AI assessment framework. *AI and Ethics*, 5(5), 5121-5148.

- Linares-Espinós, E., Hernández, V., Domínguez-Escrig, J. L., Fernández-Pello, S., Hevia, V., Mayor, J. & Ribal, M. J. (2018). Methodology of a systematic review. *Actas Urológicas Españolas (English Edition)*, 42(8), 499-506.
- Matos, J. F., Piedade, J., Freitas, A., Pedro, N., Dorotea, N., Pedro, A., & Galego, C. (2023). Teaching and learning research methodologies in education: A systematic literature review. *Education Sciences*, 13(2), 173.
- Madanchian, M., & Taherdoost, H. (2025). Ethical theories, governance models, and strategic frameworks for responsible AI adoption and organizational success. *Frontiers in Artificial Intelligence*, 8, 1619029.
- Meduri, K., Podicheti, S., Satish, S., & Whig, P. (2025). Accountability and transparency ensuring responsible AI development. In *Ethical dimensions of AI development* (pp. 83-102). IGI Global Scientific Publishing.
- Meske, C., Eisenhardt, D., Šešelja, D., Straßer, C., & Schneider, J. (2024). Mitigating Bias In Academic Publishing: Towards Responsible (Gen) Ai-Augmentation In Peer-Review Processes.
- Nightingale, A. (2009). A guide to systematic literature reviews. *Surgery (oxford)*, 27(9), 381-384.
- Paul, J., Khatri, P., & Kaur Duggal, H. (2024). Frameworks for developing impactful systematic literature reviews and theory building: what, why and how?. *Journal of Decision Systems*, 33(4), 537-550.
- Paul, J., & Barari, M. (2022). Meta-analysis and traditional systematic literature reviews—What, why, when, where, and how?. *Psychology & Marketing*, 39(6), 1099-1115.
- Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, 34(2).
- Radanliev, P. (2025). Frontier AI regulation: What form should it take? *Frontiers in Political Science*, 7, 1561776.
- Shahmerdanova, R. (2025). Artificial intelligence in translation: Challenges and opportunities. *Acta Globalis Humanitatis et Linguarum*, 2(1), 62-70.
- Smela, B., Toumi, M., Świerk, K., Francois, C., Biernikiewicz, M., Clay, E., & Boyer, L. (2023). Rapid literature review: definition and methodology. *Journal of market access & health policy*, 11(1), 2241234.
- Taylor, I. (2025). Is explainable AI responsible AI? *AI & society*, 40(3), 1695-1704.

- Torres-Carrión, P. V., González-González, C. S., Aciar, S., & Rodríguez-Morales, G. (2018, April). Methodology for systematic literature review applied to engineering and education. In 2018 IEEE Global engineering education conference (EDUCON) (pp. 1364-1373). IEEE.
- Vujović, D. (2024). Generative AI: Riding the new general purpose technology storm. *Ekonomika preduzeća*, 72(1-2), 125-136.
- Wan, Y., Tao, H., Zhao, Y., & Huang, L. (2025). The economic impact of artificial intelligence as a general-purpose technology and its innovations in economic research. *International Journal of Computer Science and Information Technology*, 5(1), 22-34.
- Wheeler, N. E. (2025). Responsible AI in biotechnology: balancing discovery, innovation and biosecurity risks. *Frontiers in Bioengineering and Biotechnology*, 13.
- Zamiri, M., & Esmaeili, A. (2024). Methods and technologies for supporting knowledge sharing within learning communities: A systematic literature review. *Administrative Sciences*, 14(1), 17.

ÉTICA, SEGSOS ALGORÍTMICOS Y DERECHOS HUMANOS: HACIA UNA GOBERNANZA GLOBAL DE LA INTELIGENCIA ARTIFICIAL

Ethics, Algorithmic Biases and Human Rights: Towards a Global Governance of Artificial Intelligence

Bianca Melissa Soto Rivera

soto.bianca@uabc.edu.mx

Universidad Autónoma de Baja California

0000-0000-0000-0000

Sandra Paola Morales Flores

sandra.morales85@uabc.edu.mx

Universidad Autónoma de Baja California

<https://orcid.org/0009-0007-7655-4295>

Resumen

El presente capítulo analiza la inteligencia artificial (IA) desde una perspectiva ética, social y jurídica, enfatizando que los sistemas algorítmicos no son herramientas neutrales, sino tecnologías profundamente vinculadas a estructuras sociales, políticas y económicas. A partir de un enfoque cualitativo, documental y crítico-interpretativo, se examinan los fundamentos éticos de la IA, los orígenes de los sesgos algorítmicos y sus impactos en ámbitos como el empleo, la justicia, la salud y la educación. El estudio demuestra que los sesgos pueden originarse tanto en los datos utilizados para entrenar los modelos como en las decisiones humanas relacionadas con el diseño y la implementación de los algoritmos, reproduciendo y amplificando desigualdades históricas.

Asimismo, se aborda la dimensión epistemológica del sesgo algorítmico, destacando que la IA no solo automatiza decisiones, sino que también produce conocimiento condicionado por contextos sociales y relaciones de poder. A través

del análisis de casos emblemáticos, como el sistema COMPAS, el algoritmo de reclutamiento de Amazon y tecnologías de reconocimiento facial, se evidencian los riesgos de discriminación, opacidad y afectación de derechos humanos fundamentales.

El capítulo también examina la responsabilidad compartida de desarrolladores, organizaciones, gobiernos y sociedad civil en la construcción de sistemas de IA éticamente responsables. Se analizan estrategias técnicas, organizacionales y regulatorias orientadas a la mitigación de sesgos, incluyendo auditorías algorítmicas, algoritmos de equidad, explicabilidad, regulación basada en riesgos y participación social. Finalmente, se concluye que la neutralidad algorítmica es inalcanzable, por lo que el objetivo debe centrarse en desarrollar modelos transparentes, auditables y alineados con principios de justicia social, derechos humanos y gobernanza democrática.

Palabras clave: Inteligencia artificial (IA), sesgos algorítmicos, ética de la IA, derechos humanos, gobernanza tecnológica, discriminación algorítmica, transparencia, explicabilidad, justicia social, responsabilidad algorítmica, neutralidad tecnológica, regulación de la IA.

Abstract

This chapter analyzes artificial intelligence (AI) from an ethical, social, and legal perspective, emphasizing that algorithmic systems are not neutral tools, but technologies deeply linked to social, political, and economic structures. Based on a qualitative, documentary, and critical-interpretative approach, the chapter examines the ethical foundations of AI, the origins of algorithmic bias, and its impacts in areas such as employment, justice, healthcare, and education. The study demonstrates that biases may originate both from the data used to train models and from human decisions related to the design and implementation of algorithms, thereby reproducing and amplifying historical inequalities.

Likewise, the epistemological dimension of algorithmic bias is addressed, highlighting that AI not only automates decisions but also produces knowledge conditioned by social contexts and power relations. Through the analysis of emblematic cases, such as the COMPAS system, Amazon's recruitment algorithm, and facial recognition technologies, the chapter reveals the risks of discrimination, opacity, and violations of fundamental human rights.

The chapter also examines the shared responsibility of developers, organizations, governments, and civil society in building ethically responsible AI systems. Technical, organizational, and regulatory strategies aimed at mitigating bias are analyzed, including algorithmic audits, fairness-aware algorithms, explainability, risk-based regulation, and social participation. Finally, it concludes that algorithmic neutrality is unattainable; therefore, the objective should focus on developing transparent, auditable models aligned with principles of social justice, human rights, and democratic governance.

Keywords: Artificial Intelligence (AI), algorithmic bias, AI ethics, human rights, technological governance, algorithmic discrimination, transparency, explainability, social justice, algorithmic accountability, technological neutrality, AI regulation.

1. Introducción

La inteligencia artificial (IA) se ha consolidado como una de las fuerzas tecnológicas más influyentes del siglo XXI, transformando de manera profunda múltiples dimensiones de la vida social, económica y política. Su capacidad para procesar grandes volúmenes de datos, identificar patrones complejos y automatizar procesos de toma de decisiones ha permitido avances significativos en sectores como la salud, la educación, las finanzas, la seguridad pública y la administración gubernamental. Sin embargo, este desarrollo acelerado también ha generado una serie de desafíos éticos, sociales y jurídicos que cuestionan la aparente neutralidad de los sistemas algorítmicos.

Lejos de ser herramientas objetivas, los sistemas de IA están profundamente condicionados por los datos con los que son entrenados, así como por las decisiones humanas que intervienen en su diseño, implementación y uso. En este sentido, los sesgos algorítmicos no deben entenderse como fallas aisladas, sino como manifestaciones de desigualdades estructurales presentes en la sociedad. La IA no solo reproduce estas desigualdades, sino que puede amplificarlas, automatizando procesos discriminatorios a gran escala.

Esta problemática adquiere una relevancia aún mayor cuando los sistemas algorítmicos se utilizan en ámbitos críticos, como la justicia penal, el acceso al empleo, la asignación de créditos o los servicios de salud. En estos contextos, las decisiones automatizadas pueden afectar directamente derechos humanos fundamentales, como la igualdad, la privacidad, la autonomía y el debido proceso.

Por ello, el análisis de la IA no puede limitarse a una perspectiva técnica. Es necesario abordarla desde un enfoque interdisciplinario que integre la ética, el derecho, la sociología y la política. Este capítulo tiene como objetivo ofrecer una visión integral de los desafíos que plantea la inteligencia artificial, profundizando en la relación entre sesgos algorítmicos, derechos humanos y gobernanza global.

1.1 Metodología

El presente trabajo se desarrolla a partir de un enfoque cualitativo de carácter documental y analítico, orientado a examinar los principales debates contemporáneos en torno a la inteligencia artificial, los sesgos algorítmicos, los derechos humanos y la gobernanza tecnológica. Este tipo de enfoque resulta adecuado debido a la naturaleza compleja, interdisciplinaria y en constante evolución del objeto de estudio, el cual no puede ser comprendido únicamente desde perspectivas cuantitativas.

La investigación se basa principalmente en la revisión sistemática de literatura académica, incluyendo artículos científicos, libros especializados, informes de organismos internacionales y documentos normativos recientes. Entre las fuentes analizadas se encuentran trabajos en el campo del derecho, la ética aplicada, la

sociología de la tecnología y la ciencia de datos, lo que permite una comprensión integral del fenómeno desde distintas disciplinas (Floridi y otros., 2018; Mehrabi y otros., 2021).

Asimismo, se adopta un enfoque crítico-interpretativo, cuyo objetivo no es únicamente describir los fenómenos asociados a la inteligencia artificial, sino también analizar sus implicaciones sociales, políticas y éticas. Este enfoque permite identificar cómo los sistemas algorítmicos pueden reproducir estructuras de desigualdad existentes, así como evaluar las limitaciones de los marcos regulatorios actuales (Crawford, 2021).

Como parte de la estrategia metodológica, se incorporan estudios de caso relevantes, tales como el sistema de reclutamiento automatizado de Amazon, el algoritmo COMPAS en el ámbito judicial y los sistemas de reconocimiento facial. Estos casos permiten ilustrar de manera concreta los efectos de los sesgos algorítmicos en distintos contextos, facilitando la conexión entre el análisis teórico y la realidad empírica (Angwin y otros., 2016; Dastin, 2018).

En términos de procedimiento, la investigación se desarrolló en tres etapas principales:

1. Recopilación de información: identificación y selección de fuentes académicas recientes y relevantes sobre inteligencia artificial, ética y derechos humanos.
2. Análisis y sistematización: organización de la información en categorías temáticas, tales como fundamentos éticos, sesgos algorítmicos, impacto social y gobernanza.
3. Interpretación crítica: evaluación de los hallazgos a partir de marcos teóricos interdisciplinarios, con énfasis en sus implicaciones sociales y normativas.

Finalmente, es importante señalar que, debido a la naturaleza documental del estudio, existen ciertas limitaciones, como la dependencia de fuentes secundarias y la ausencia de trabajo empírico directo. No obstante, el uso de literatura

especializada y estudios de caso permite construir un análisis sólido y fundamentado.

2. Fundamentos Éticos de la Inteligencia Artificial

2.1 La ética como marco de análisis tecnológico

La ética aplicada a la inteligencia artificial surge como una respuesta a la creciente influencia de los sistemas automatizados en la vida humana. A diferencia de otras tecnologías, la IA participa en procesos de toma de decisiones con implicaciones directas en la vida de las personas.

Floridi y otros. (2018) proponen principios fundamentales como la beneficencia, la no maleficencia, la autonomía, la justicia y la explicabilidad, los cuales han sido retomados por organismos internacionales como la UNESCO (2021).

La ética aplicada a la inteligencia artificial surge como una respuesta a la creciente influencia de los sistemas automatizados en la vida humana. A diferencia de otras tecnologías, la IA no solo ejecuta instrucciones, sino que participa en procesos de toma de decisiones que pueden tener consecuencias significativas para las personas.

En este contexto, la ética no debe entenderse como un elemento accesorio, sino como un componente fundamental en el diseño y desarrollo de sistemas tecnológicos. Su función es establecer principios que orienten la creación de sistemas que respeten la dignidad humana y promuevan el bienestar social.

Entre los principios más relevantes destacan:

- La beneficencia, que implica maximizar los beneficios sociales.
- La no maleficencia, que busca evitar daños y perjuicios.
- La autonomía, que reconoce la capacidad de decisión de las personas.
- La justicia, que exige equidad en la distribución de beneficios y riesgos.

- La explicabilidad, que promueve la transparencia en los sistemas.

Estos principios permiten evaluar no solo el funcionamiento técnico de la IA, sino también sus impactos sociales y éticos.

2.2 Perspectivas filosóficas clásicas

El análisis ético de la Inteligencia Artificial (IA) se enriquece profundamente al integrar las principales corrientes de la filosofía moral. Cada una ofrece un lente distinto para evaluar el impacto y la responsabilidad de estas tecnologías:

Propuesto por pensadores como John Stuart Mill (2002), el utilitarismo sugiere evaluar las decisiones algorítmicas en función de sus consecuencias, priorizando aquellas que maximizan el beneficio para el mayor número de personas.

Limitación Crítica: En el contexto de la IA, este enfoque puede ser problemático, ya que el análisis de bienestar agregado corre el riesgo de justificar decisiones que perjudican o marginan a minorías en favor de una mayoría estadística.

Basada en la filosofía de Immanuel Kant (2003), esta corriente enfatiza el respeto a principios universales y derechos fundamentales, independientemente de los resultados finales.

Aplicación en IA: Establece límites éticos infranqueables (imperativos categóricos). Los sistemas deben diseñarse para respetar la dignidad humana y garantizar la no discriminación, actuando bajo normas que protejan al individuo frente a la eficiencia pura.

Inspirada en Aristóteles (2001), esta perspectiva desplaza el foco de la "regla" o el "resultado" hacia el carácter moral de los agentes.

Dimensión Organizacional: No basta con que el algoritmo sea ético; los desarrolladores y las organizaciones deben cultivar virtudes como la honestidad, la justicia y la responsabilidad social. La integridad de quienes crean la tecnología es el primer filtro para garantizar sistemas éticamente responsables.

Mientras que el utilitarismo busca la eficiencia del bienestar general, la deontología impone barreras de protección a los derechos humanos y la ética de la virtud apela al compromiso ético de los creadores. La convergencia de estas tres visiones es esencial para construir una IA que sea, a la vez, eficaz, justa y profundamente humana.

2.3 La inteligencia artificial como sistema sociotécnico

La ética de la Inteligencia Artificial (IA) ha encontrado en el enfoque sociotécnico una de sus aportaciones más relevantes. Bajo esta premisa, se sostiene que las tecnologías no pueden analizarse de manera aislada, sino en su relación intrínseca con los contextos sociales en los que se desarrollan (Latour, 2005). En este sentido, la IA debe entenderse como una interacción compleja donde lo técnico y lo social son inseparables.

Desde esta perspectiva, los algoritmos dejan de ser vistos como entidades matemáticas neutrales para ser reconocidos como artefactos cargados de valores, supuestos y relaciones de poder. Como bien señaló Winner (1980), toda tecnología posee una dimensión política inherente, dado que su diseño e implementación influyen directamente en la organización y jerarquización de la sociedad.

Esta dimensión política adquiere una urgencia particular en el contexto latinoamericano, marcado por persistentes desigualdades estructurales. La implementación de sistemas de IA en sociedades con brechas tan profundas corre el riesgo de:

Intensificar las desigualdades: Sin medidas adecuadas, la IA puede automatizar y amplificar sesgos históricos.

Profundizar brechas sociales: Como argumenta Bustelo Gracia (2024), la tecnología puede convertirse en una herramienta que excluya a sectores ya vulnerabilizados si no existe una regulación adecuada.

En conclusión, reconocer a la IA como un sistema sociotécnico es el primer paso para transitar hacia una gobernanza tecnológica que no solo busque la eficiencia,

sino que garantice que estos sistemas no se conviertan en mecanismos de profundización de la injusticia social en la región.

3. Naturaleza y origen de los sesgos algorítmicos

3.1 El papel de los datos

El principal origen de los sesgos algorítmicos se encuentra en los datos utilizados para entrenar los modelos. Los algoritmos de aprendizaje automático dependen de grandes volúmenes de información histórica, la cual puede contener patrones discriminatorios.

Mehrabi y otros. (2021) señalan que los datos no son representaciones neutrales de la realidad, sino construcciones sociales que reflejan decisiones humanas, omisiones y desigualdades. Por lo tanto, cuando un modelo se entrena con datos sesgados, es probable que reproduzca esos mismos sesgos en sus resultados.

Un ejemplo claro se encuentra en los sistemas de reconocimiento facial, los cuales han demostrado menor precisión en personas con tonos de piel más oscuros debido a la falta de diversidad en los conjuntos de entrenamiento.

Los datos constituyen la base sobre la cual se construyen los sistemas de inteligencia artificial. Sin embargo, estos datos no son neutrales: reflejan decisiones humanas, contextos históricos y desigualdades sociales.

Cuando un algoritmo se entrena con datos sesgados, es probable que reproduzca estos sesgos en sus resultados. Esto puede ocurrir incluso cuando variables sensibles, como el género o la raza, no están explícitamente incluidas.

3.2 Sesgos en el diseño algorítmico

Además de los datos, el diseño del algoritmo también puede introducir sesgos. Las decisiones relacionadas con la selección de variables, la definición de objetivos y la elección de métricas de evaluación son procesos que reflejan valores humanos (Friedman & Nissenbaum, 1996).

Por ejemplo, si un sistema de crédito se diseña priorizando la maximización de ganancias, es posible que discrimine a grupos considerados de mayor riesgo, reproduciendo desigualdades económicas.

Además de los datos, los sesgos también pueden surgir en el diseño de los algoritmos. Las decisiones sobre qué variables incluir, qué objetivos optimizar o qué métricas utilizar pueden influir significativamente en los resultados.

Por ejemplo, un sistema diseñado para maximizar la rentabilidad en la concesión de créditos puede terminar excluyendo a grupos vulnerables, incluso si no existe una intención explícita de discriminación.

3.3 Sesgo estructural y reproducción social

Más allá de los aspectos técnicos, los sesgos algorítmicos están profundamente vinculados a estructuras sociales. Como señalan Boté-Vericad y San José (2025), los algoritmos actúan como mecanismos de reproducción de desigualdades, ya que incorporan patrones históricos presentes en la sociedad.

Este fenómeno puede analizarse desde la teoría de la reproducción social, la cual sostiene que las instituciones tienden a perpetuar las estructuras de poder existentes. En este sentido, la inteligencia artificial puede convertirse en una herramienta que refuerza dichas estructuras si no se implementan mecanismos de control adecuados.

4. Dimensión epistemológica del sesgo algorítmico

Un aspecto menos explorado, pero igualmente relevante, es la dimensión epistemológica del sesgo algorítmico. La inteligencia artificial no solo produce decisiones, sino también conocimiento. Los modelos de aprendizaje automático generan representaciones del mundo basadas en los datos con los que han sido entrenados.

Sin embargo, si estos datos están sesgados, el conocimiento producido también lo estará. Esto plantea interrogantes sobre la objetividad de la IA y su capacidad para representar la realidad de manera justa.

Desde esta perspectiva, el sesgo algorítmico no es únicamente un problema técnico o ético, sino también epistemológico. Implica cuestionar qué tipo de conocimiento se produce, quién lo produce y con qué fines.

Autores como O'Neil (2016) han advertido que los algoritmos pueden actuar como “cajas negras” que ocultan sus procesos internos, dificultando la identificación de sesgos y la rendición de cuentas.

La inteligencia artificial no solo toma decisiones, sino que también produce conocimiento. Esto plantea una cuestión fundamental: ¿qué tipo de conocimiento generan los algoritmos?

Si los datos utilizados están sesgados, el conocimiento producido también lo estará. Esto cuestiona la idea de que la IA es una herramienta objetiva y plantea la necesidad de desarrollar mecanismos que permitan evaluar la calidad y equidad del conocimiento generado.

5. Impacto social de los sesgos algorítmicos

Los sesgos algorítmicos no constituyen únicamente un problema técnico, sino que tienen consecuencias directas en la vida social, afectando la distribución de oportunidades, recursos y derechos. En este sentido, la inteligencia artificial actúa como un mecanismo de mediación que puede reforzar o transformar estructuras de desigualdad existentes.

5.1 Ámbito laboral

En el ámbito del empleo, los sistemas automatizados de reclutamiento y selección han sido ampliamente adoptados por empresas que buscan optimizar sus

procesos. Sin embargo, diversos estudios han demostrado que estos sistemas pueden reproducir patrones históricos de discriminación.

Un caso emblemático es el de un sistema de reclutamiento desarrollado por Amazon, el cual fue descontinuado tras evidenciarse que penalizaba sistemáticamente a candidatas mujeres debido a que había sido entrenado con datos históricos predominantemente masculinos (Dastin, 2018). Este ejemplo ilustra cómo los algoritmos pueden internalizar sesgos existentes en los datos y reproducirlos de manera automatizada.

De acuerdo con Barocas y Selbst (2016), este tipo de discriminación puede ser indirecta, ya que los algoritmos utilizan variables correlacionadas que funcionan como sustitutos de atributos protegidos, como el género o la etnicidad.

5.2 Sistema judicial

En el ámbito judicial, el uso de algoritmos para la evaluación de riesgo ha generado un intenso debate ético. Uno de los casos más conocidos es el del sistema COMPAS, utilizado en Estados Unidos para predecir la probabilidad de reincidencia.

Un análisis realizado por Angwin y otros. (2016) reveló que este sistema presentaba sesgos raciales, ya que tendía a clasificar erróneamente a personas afroamericanas como de alto riesgo en mayor proporción que a personas blancas. Este tipo de errores tiene implicaciones graves, ya que puede influir en decisiones judiciales como la imposición de penas o la concesión de libertad condicional.

En el contexto latinoamericano, investigaciones recientes han advertido que la adopción de sistemas similares podría agravar desigualdades existentes si no se implementan mecanismos de control adecuados (Chaverra Mena, 2025).

5.3 Salud

En el sector salud, los sistemas de inteligencia artificial se utilizan para diagnósticos, asignación de recursos y predicción de enfermedades. No obstante, estos sistemas pueden presentar sesgos que afectan la calidad de la atención.

Un estudio realizado por Obermeyer y otros. (2019) demostró que un algoritmo utilizado en el sistema de salud estadounidense subestimaba las necesidades médicas de pacientes afroamericanos debido a que utilizaba el gasto sanitario como variable proxy de la necesidad de atención. Dado que históricamente este grupo ha tenido menor acceso a servicios médicos, el algoritmo reproducía esta desigualdad.

Este caso pone de manifiesto que los sesgos algorítmicos pueden tener consecuencias directas en la salud y el bienestar de las personas.

5.4 Educación

En el ámbito educativo, los sistemas de evaluación automatizada y plataformas de aprendizaje adaptativo pueden contribuir a la personalización del aprendizaje. Sin embargo, también pueden reproducir desigualdades si se basan en datos sesgados.

Por ejemplo, los algoritmos que predicen el rendimiento académico pueden favorecer a estudiantes con determinados perfiles socioeconómicos, perpetuando brechas educativas (Selwyn, 2019).

5.5 Casos Reales Ampliados

Caso 1: Amazon y la discriminación de género

Uno de los ejemplos más citados sobre sesgos en inteligencia artificial es el sistema de reclutamiento automatizado desarrollado por Amazon. Este sistema fue entrenado con datos históricos de contratación de la empresa, los cuales reflejaban una predominancia masculina en el sector tecnológico. Como resultado, el

algoritmo aprendió a penalizar currículums que incluían indicadores asociados con mujeres, como la participación en organizaciones femeninas (Dastin, 2018).

Este caso demuestra que los sistemas de IA no solo replican patrones existentes, sino que pueden institucionalizar la discriminación si no se supervisan adecuadamente. Además, evidencia la importancia de auditar los datos de entrenamiento y cuestionar los supuestos detrás de los modelos algorítmicos (Mehrabi y otros., 2021).

Caso 2: COMPAS y el sistema judicial

El sistema COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) fue utilizado en Estados Unidos para evaluar el riesgo de reincidencia de personas acusadas. Un estudio realizado por ProPublica encontró que el algoritmo presentaba un sesgo significativo contra personas afroamericanas, clasificándolas erróneamente como de alto riesgo con mayor frecuencia que a personas blancas (Angwin y otros., 2016).

Este caso plantea serias preocupaciones sobre el uso de IA en contextos donde están en juego derechos fundamentales, como la libertad personal y el debido proceso. Además, pone en evidencia la necesidad de garantizar transparencia y explicabilidad en los sistemas utilizados en el ámbito judicial (O’Neil, 2016).

Caso 3: Algoritmos en salud

En el sector salud, Obermeyer y otros. (2019) identificaron un algoritmo ampliamente utilizado en Estados Unidos que asignaba recursos médicos basándose en el gasto sanitario previo de los pacientes. Dado que históricamente las poblaciones afroamericanas han tenido menor acceso a servicios médicos, el algoritmo subestimaba sus necesidades reales.

Este caso ilustra cómo el uso de variables aparentemente neutrales puede generar resultados discriminatorios. También subraya la importancia de seleccionar indicadores adecuados y considerar el contexto social en el diseño de sistemas de IA.

Caso 4: Reconocimiento facial

Diversas investigaciones han demostrado que los sistemas de reconocimiento facial presentan tasas de error significativamente mayores en personas de piel oscura y en mujeres (Buolamwini & Gebru, 2018). Este problema ha llevado a cuestionamientos sobre el uso de estas tecnologías en vigilancia, seguridad pública y control migratorio.

El impacto de estos errores no es menor, ya que puede derivar en detenciones injustificadas, vigilancia desproporcionada y violaciones a la privacidad (Raji y otros., 2020).

6. Responsabilidad social en la inteligencia artificial

La complejidad de los sistemas de inteligencia artificial implica que la responsabilidad por sus efectos no puede atribuirse a un solo actor. En lugar de ello, se requiere un enfoque de responsabilidad distribuida que involucre a desarrolladores, empresas, gobiernos y sociedad civil.

6.1 Responsabilidad de los desarrolladores

Los desarrolladores de sistemas de IA desempeñan un papel fundamental en la prevención de sesgos. Sus decisiones en el diseño del modelo, la selección de datos y la definición de objetivos tienen implicaciones éticas significativas.

Desde la ética profesional, se ha argumentado que los ingenieros deben adoptar principios de responsabilidad similares a los de otras disciplinas, como la medicina o la ingeniería civil, donde las decisiones pueden afectar la vida de las personas (Floridi y otros., 2018).

6.2 Responsabilidad de las organizaciones

Las empresas que desarrollan e implementan sistemas de IA tienen la responsabilidad de garantizar que estos sean éticamente aceptables. Esto implica

la adopción de políticas de gobernanza que incluyan auditorías algorítmicas, mecanismos de transparencia y evaluación de impacto social.

En este sentido, la ética empresarial debe evolucionar para incorporar los desafíos específicos de la inteligencia artificial.

6.3 Responsabilidad del Estado

Los gobiernos desempeñan un papel clave en la regulación de la inteligencia artificial. La creación de marcos legales que promuevan la equidad, la transparencia y la rendición de cuentas es fundamental para prevenir abusos.

En la Unión Europea, por ejemplo, se han desarrollado propuestas regulatorias que buscan clasificar los sistemas de IA según su nivel de riesgo, estableciendo requisitos específicos para cada categoría (European Commission, 2021).

6.4 Participación de la sociedad civil

La sociedad civil también tiene un papel importante en la vigilancia y evaluación de los sistemas de IA. La participación ciudadana y la transparencia son elementos clave para garantizar que estas tecnologías se desarrollen en beneficio del interés público.

7. Estrategias para la mitigación de sesgos algorítmicos

La mitigación de los sesgos algorítmicos requiere un enfoque integral que combine soluciones técnicas, organizacionales y sociales.

7.1 Estrategias técnicas

Entre las estrategias técnicas se encuentran la diversificación de los conjuntos de datos, la implementación de algoritmos de equidad (fairness-aware machine learning) y el uso de técnicas de explicabilidad.

Estas herramientas permiten identificar y reducir sesgos en los modelos, aunque no eliminan completamente el problema (Mehrabi y otros., 2021).

7.2 Estrategias organizacionales

Las organizaciones pueden implementar auditorías algorítmicas para evaluar el desempeño de sus sistemas. Asimismo, la adopción de códigos de ética y políticas internas puede contribuir a la prevención de sesgos.

7.3 Estrategias regulatorias

La regulación es un componente esencial para garantizar el uso responsable de la IA. Esto incluye la creación de normas que obliguen a las empresas a evaluar el impacto de sus sistemas y a garantizar su transparencia.

8. Discusión teórica: límites de la neutralidad algorítmica

Uno de los debates más importantes en la ética de la inteligencia artificial es la posibilidad de alcanzar una neutralidad algorítmica. Desde una perspectiva técnica, podría pensarse que es posible diseñar sistemas completamente objetivos. Sin embargo, la evidencia sugiere lo contrario.

Autores como Feenberg (1999) y Winner (1980) han argumentado que toda tecnología está impregnada de valores y relaciones de poder. En este sentido, la inteligencia artificial no puede ser completamente neutral, ya que refleja las condiciones sociales en las que se desarrolla.

Desde la teoría crítica, se sostiene que la tecnología no solo refleja la realidad, sino que también la configura. Por lo tanto, los algoritmos no solo reproducen desigualdades, sino que pueden contribuir a su consolidación.

Esto implica que el objetivo no debe ser eliminar completamente los sesgos, sino gestionarlos de manera responsable y transparente. La ética de la IA debe centrarse en la creación de sistemas que sean auditables, explicables y alineados con principios de justicia social.

9. Estrategias para la Mitigación de Sesgos

La mitigación de los sesgos algorítmicos representa uno de los principales desafíos en el desarrollo responsable de la inteligencia artificial. Dado que estos sesgos no surgen únicamente de errores técnicos, sino también de estructuras sociales y decisiones humanas, su abordaje requiere un enfoque integral que combine soluciones técnicas, organizacionales y regulatorias (Mehrabi y otros., 2021).

En este sentido, no existe una solución única, sino un conjunto de estrategias complementarias que deben aplicarse a lo largo de todo el ciclo de vida de los sistemas de IA, desde el diseño hasta su implementación y evaluación continua.

9.1 Estrategias técnicas

Las estrategias técnicas buscan intervenir directamente en el desarrollo de los modelos algorítmicos para reducir la presencia de sesgos en los resultados.

a) Preprocesamiento de datos

Una de las principales fuentes de sesgo son los datos de entrenamiento. Por ello, es fundamental implementar técnicas de limpieza, balanceo y diversificación de datos. Esto incluye:

- Eliminación de datos redundantes o sesgados
- Inclusión de grupos subrepresentados
- Revisión de variables proxy (variables que indirectamente reflejan atributos sensibles)

Estas acciones permiten reducir la probabilidad de que el modelo reproduzca desigualdades históricas (Barocas & Selbst, 2016).

b) Algoritmos de equidad (fairness-aware machine learning)

Existen enfoques diseñados específicamente para incorporar criterios de equidad en los modelos de IA. Estos algoritmos buscan equilibrar la precisión con la justicia en los resultados.

Algunos criterios de equidad incluyen:

- Paridad demográfica: igualdad en la distribución de resultados entre grupos
- Igualdad de oportunidades: tasas de error similares entre grupos
- Equidad individual: tratar de manera similar a individuos similares

Sin embargo, es importante reconocer que estos criterios pueden entrar en conflicto entre sí, lo que implica decisiones éticas y políticas sobre qué tipo de equidad priorizar (Mehrabian y otros., 2021)

c) Explicabilidad e interpretabilidad

El desarrollo de modelos explicables (Explainable AI) permite comprender cómo y por qué un sistema toma ciertas decisiones. Esto facilita la identificación de sesgos y mejora la rendición de cuentas.

Herramientas como LIME o SHAP permiten analizar la influencia de variables en las predicciones, contribuyendo a una mayor transparencia (Raji y otros., 2020).

d) Evaluación continua y auditorías técnicas

Los modelos de IA no deben considerarse sistemas estáticos. Es necesario implementar procesos de monitoreo constante para detectar desviaciones, errores o nuevos sesgos que puedan surgir con el tiempo.

Las auditorías algorítmicas permiten evaluar el desempeño del sistema en distintos contextos y grupos poblacionales.

9.2 Estrategias organizacionales

Más allá de la tecnología, las organizaciones juegan un papel fundamental en la prevención de sesgos.

a) Equipos multidisciplinarios

El desarrollo de sistemas de IA no debe estar limitado a ingenieros o programadores. Es necesario integrar especialistas en ética, derecho, sociología y

otras disciplinas que permitan identificar riesgos desde distintas perspectivas (Floridi y otros., 2018).

b) Cultura organizacional ética

Las empresas y organizaciones deben promover valores como la transparencia, la responsabilidad y la equidad. Esto puede lograrse mediante:

- códigos de ética en IA
- políticas internas de revisión
- capacitación en sesgos y discriminación

c) Auditorías externas e independientes

La evaluación por terceros permite garantizar mayor objetividad en la detección de sesgos. Estas auditorías son especialmente importantes en sistemas de alto impacto, como aquellos utilizados en salud, justicia o finanzas.

d) Participación de los afectados

Una estrategia emergente consiste en incluir a las comunidades afectadas en el diseño y evaluación de sistemas de IA. Esto permite identificar riesgos que podrían pasar desapercibidos desde una perspectiva exclusivamente técnica.

9.3 Estrategias regulatorias y legales

El papel del Estado es fundamental para garantizar el uso responsable de la inteligencia artificial.

a) Regulación basada en riesgos

Modelos como el AI Act de la Unión Europea clasifican los sistemas de IA según su nivel de riesgo, estableciendo mayores exigencias para aquellos que pueden afectar derechos fundamentales (European Commission, 2024).

b) Evaluaciones de impacto algorítmico

Antes de implementar un sistema de IA, se deben realizar evaluaciones que analicen sus posibles efectos en derechos humanos, equidad y justicia social.

c) Derechos de los usuarios

Las regulaciones deben garantizar derechos como:

- el derecho a la explicación
- el derecho a impugnar decisiones automatizadas
- el acceso a mecanismos de reparación

Estos derechos fortalecen la protección frente a posibles abusos (UNESCO, 2021).

d) Responsabilidad legal

Es necesario establecer claramente quién es responsable cuando un sistema de IA genera daños. Esto incluye desarrolladores, empresas y entidades que implementan la tecnología.

9.4 Límites de las estrategias de mitigación

A pesar de los avances, es importante reconocer que la eliminación total de los sesgos algorítmicos es prácticamente imposible. Esto se debe a que los sistemas de IA están profundamente vinculados con contextos sociales complejos.

Como señala Crawford (2021), los problemas de la IA no pueden resolverse únicamente con soluciones técnicas, ya que están relacionados con estructuras de poder, desigualdad y acceso a recursos.

Por ello, el objetivo no debe ser eliminar completamente los sesgos, sino gestionarlos de manera responsable, transparente y ética.

9.5 Hacia un enfoque integral

La mitigación de sesgos requiere un enfoque integral que combine:

- innovación tecnológica

- responsabilidad organizacional
- regulación efectiva
- participación social

Solo mediante la articulación de estos elementos será posible desarrollar sistemas de inteligencia artificial que sean no solo eficientes, sino también justos y respetuosos de los derechos humanos.

10. Límites de la Neutralidad Algorítmica

La idea de que la tecnología es neutral es un mito. Todos los sistemas tecnológicos reflejan valores y decisiones humanas.

Más allá de los avances tecnológicos, la inteligencia artificial plantea una pregunta fundamental: ¿qué tipo de sociedad queremos construir?

La implementación de sistemas algorítmicos no es un proceso neutral, sino una decisión política que refleja valores, prioridades y relaciones de poder. En este sentido, la discusión sobre la IA debe trascender el ámbito técnico y centrarse en sus implicaciones sociales.

Como señala Crawford (2021), la IA no es simplemente una herramienta, sino una infraestructura que organiza la vida contemporánea. Por ello, su desarrollo debe estar orientado por principios de justicia, equidad y respeto a los derechos humanos.

11. Conclusiones

La inteligencia artificial representa una herramienta poderosa con el potencial de transformar positivamente la sociedad. Sin embargo, también plantea riesgos significativos que no pueden ser ignorados.

El análisis desarrollado en este capítulo permite afirmar que la IA no es simplemente una tecnología, sino un fenómeno profundamente social y político. Los sesgos

algorítmicos no son errores aislados, sino el reflejo de estructuras de desigualdad que deben ser abordadas de manera integral.

Por ello, el desarrollo de la inteligencia artificial debe estar acompañado de una gobernanza ética, jurídica y social que garantice su uso responsable. Solo así será posible construir sistemas tecnológicos que contribuyan al bienestar colectivo y al respeto de los derechos humanos.

Referencias

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Aristóteles. (2001). *Ética a Nicómaco*. Gredos.
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732. <https://doi.org/10.15779/Z38BG31>
- Boté-Vericad, J. J., & San José, C. (2025). *Sesgo algorítmico en la inteligencia artificial: Conceptos, implicaciones y soluciones*. Universidad de Barcelona.
- Bustelo Gracia, J. L. (2024). El sesgo algorítmico en la inteligencia artificial: Implicaciones para la seguridad global y la política criminal. *Geoconflicts & Intelligence*, 2(2), 183–203.
- Chaverra Mena, Y. E. (2025). El sesgo algorítmico en sistemas judiciales y su impacto en víctimas de género. *Revista de Estudios Jurídicos*.
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. <https://www.reuters.com/technology/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-2018-10-10/>
- European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. <https://eur-lex.europa.eu>
- Feenberg, A. (1999). *Questioning technology*. Routledge.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities,

risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.
<https://doi.org/10.1007/s11023-018-9482-5>

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330–347.
<https://doi.org/10.1145/230538.230561>

Kant, I. (2003). *Fundamentación de la metafísica de las costumbres*. Alianza Editorial.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1–35. <https://doi.org/10.1145/3457607>

Mill, J. S. (2002). *Utilitarianism*. Hackett Publishing Company.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15.

CEPAL. (2023). *Inteligencia artificial en América Latina y el Caribe: Desafíos y oportunidades*. Naciones Unidas.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... Barnes, P. (2020). Closing the AI accountability gap. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44.

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97–112.

ANÁLISIS DE LA NORMA ISO/IEC 42001 EN CONTEXTOS ORGANIZACIONALES EMERGENTES: UNA LECTURA DESDE LA RACIONALIZACIÓN Y EL ISOMORFISMO INSTITUCIONAL

Analysis of ISO/IEC 42001 in Emerging Organizational Contexts: A Reading
from Institutional Rationalization and Isomorphism

Virginia López Nevárez

vicky-academico@hotmail.com

Universidad Autónoma de Occidente

<https://orcid.org/0000-0001-9887-0132>

Resumen

El capítulo analiza la regulación de la inteligencia artificial desde una perspectiva organizacional, centrándose en la norma ISO/IEC 42001 como un mecanismo de racionalización y estructuración institucional. A partir de enfoques teóricos como la racionalización organizacional de Weber y el nuevo institucionalismo de DiMaggio y Powell, se examina cómo la regulación de la IA trasciende el ámbito jurídico para convertirse en un proceso de institucionalización que define criterios de legitimidad, control y gobernanza dentro de las organizaciones. El estudio sostiene que estándares como la ISO/IEC 42001 buscan traducir principios éticos y normativos internacionales en prácticas organizacionales concretas mediante sistemas formales de gestión, trazabilidad, control y evaluación de riesgos. Sin embargo, el análisis evidencia tensiones entre esta lógica de formalización y la naturaleza emergente, flexible y adaptativa de la inteligencia artificial en contextos organizacionales reales. Asimismo, se demuestra que la regulación de la IA no elimina la incertidumbre tecnológica, sino que la reorganiza mediante mecanismos de control parcial que coexisten con dinámicas de aprendizaje, experimentación y adaptación continua. Finalmente, se concluye que la gobernanza efectiva de la inteligencia artificial requiere articular marcos regulatorios formales con prácticas

organizacionales abiertas y dinámicas, reconociendo que la regulación constituye un proceso de interacción constante entre norma, legitimidad institucional y práctica tecnológica.

Palabras clave: Inteligencia artificial (IA), ISO/IEC 42001, regulación de la IA, gobernanza organizacional, racionalización organizacional, isomorfismo institucional, institucionalización, gestión de riesgos, gobernanza tecnológica, legitimidad organizacional, transformación digital, sistemas de gestión, innovación organizacional, regulación tecnológica, organizaciones emergentes.

Abstract

The chapter analyzes the regulation of artificial intelligence from an organizational perspective, focusing on the ISO/IEC 42001 standard as a mechanism of rationalization and institutional structuring. Drawing on theoretical approaches such as Weber's organizational rationalization and the new institutionalism of DiMaggio and Powell, the study examines how AI regulation transcends the legal sphere to become a process of institutionalization that defines criteria of legitimacy, control, and governance within organizations. The study argues that standards such as ISO/IEC 42001 seek to translate international ethical and regulatory principles into concrete organizational practices through formal systems of management, traceability, control, and risk assessment. However, the analysis reveals tensions between this logic of formalization and the emergent, flexible, and adaptive nature of artificial intelligence in real organizational contexts. Likewise, it demonstrates that AI regulation does not eliminate technological uncertainty, but rather reorganizes it through partial control mechanisms that coexist with dynamics of learning, experimentation, and continuous adaptation. Finally, the chapter concludes that effective artificial intelligence governance requires articulating formal regulatory frameworks with open and dynamic organizational practices, recognizing that regulation constitutes a continuous process of interaction between norms, institutional legitimacy, and technological practice.

Keywords: Artificial intelligence (AI), ISO/IEC 42001, AI regulation, organizational governance, organizational rationalization, institutional isomorphism, institutionalization, risk management, technological governance, organizational legitimacy, digital transformation, management systems, organizational innovation, technological regulation, emerging organizations.

Introducción

La incorporación de la inteligencia artificial (IA) en los entornos organizacionales contemporáneos ha reconfigurado de manera significativa las dinámicas de operación, decisión y generación de valor. A diferencia de oleadas tecnológicas previas, la IA, particularmente en sus desarrollos recientes en el ámbito generativo, no se ha integrado exclusivamente mediante procesos formales de planeación estratégica, sino a través de dinámicas emergentes caracterizadas por la experimentación, la apropiación distribuida y la adaptación situada en distintos niveles organizacionales. Este carácter no lineal de adopción plantea desafíos sustantivos para los marcos tradicionales de regulación, los cuales se fundamentan en supuestos de racionalidad instrumental, control y estandarización.

En respuesta a estos desafíos, en los últimos años se ha observado un crecimiento sostenido de iniciativas orientadas a la regulación de la IA en el ámbito internacional. Organismos como la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura han promovido principios éticos centrados en el respeto a los derechos humanos, la equidad y la transparencia (UNESCO, 2021), mientras que la Organización para la Cooperación y el Desarrollo Económicos ha planteado lineamientos orientados a la gobernanza responsable y al desarrollo confiable de estas tecnologías (OCDE, 2019). De manera complementaria, la Comisión Europea ha impulsado marcos regulatorios más vinculantes, como la propuesta de Ley de Inteligencia Artificial, basada en un enfoque de gestión de riesgos (European Commission, 2021). Sin embargo, estos esfuerzos, aunque

relevantes, tienden a configurarse en un nivel macro, con un énfasis normativo que no siempre logra capturar la complejidad de los procesos organizacionales en los que la IA es efectivamente implementada.

En este contexto, la norma ISO/IEC 42001 emerge como un intento por trasladar la regulación de la inteligencia artificial al ámbito organizacional, mediante la estructuración de un sistema de gestión que integra principios de gobernanza, evaluación de riesgos y responsabilidad institucional (ISO/IEC, 2023). A diferencia de los marcos regulatorios de carácter general, esta norma se posiciona como un instrumento orientado a operacionalizar la gestión de la IA dentro de las organizaciones, proponiendo mecanismos formales de control y estandarización de prácticas. No obstante, su análisis no puede limitarse a una lectura técnica o instrumental, sino que requiere ser comprendido en el marco de procesos más amplios de estructuración organizacional.

Desde una perspectiva teórica, la ISO/IEC 42001 puede interpretarse como una expresión contemporánea de procesos de racionalización organizacional, en el sentido planteado por Weber (1978/2002), en los cuales la acción social se organiza mediante sistemas formales de reglas, procedimientos y estructuras de control. A su vez, su potencial adopción por parte de las organizaciones puede ser analizada a la luz del enfoque del nuevo institucionalismo, particularmente desde las dinámicas de isomorfismo institucional propuestas por DiMaggio y Powell (1983), donde las organizaciones tienden a incorporar estructuras y prácticas legitimadas en su entorno, no únicamente por su eficiencia técnica, sino por presiones coercitivas, normativas y miméticas.

Bajo este encuadre, la regulación de la inteligencia artificial trasciende su dimensión normativa para configurarse como un proceso de institucionalización que moldea las formas en que las organizaciones comprenden, adoptan y utilizan estas tecnologías. Sin embargo, esta lógica de racionalización entra en tensión con la naturaleza emergente, flexible y adaptativa de la IA en contextos organizacionales

reales, donde las prácticas no necesariamente responden a estructuras formales previamente definidas.

A partir de lo anterior, el presente capítulo tiene como objetivo analizar la norma ISO/IEC 42001 en contextos organizacionales emergentes, a partir de una lectura teórica fundamentada en los procesos de racionalización y el isomorfismo institucional. Se propone examinar en qué medida este estándar constituye un mecanismo efectivo de regulación organizacional de la IA, así como las tensiones que se generan entre su lógica normativa y las dinámicas reales de adopción tecnológica.

La regulación de la inteligencia artificial como proceso de institucionalización

La regulación de la inteligencia artificial no puede comprenderse únicamente como una respuesta jurídica ante los riesgos tecnológicos, sino como un proceso más amplio de institucionalización mediante el cual se definen criterios de legitimidad, responsabilidad y control en torno al desarrollo y uso de estas tecnologías. En este sentido, regular la IA implica algo más que establecer límites normativos; supone construir marcos de referencia que orientan la forma en que las organizaciones interpretan, incorporan y justifican sus prácticas tecnológicas.

Desde esta perspectiva, los marcos internacionales de regulación de la IA han buscado establecer principios comunes frente a un fenómeno cuya expansión ha superado la capacidad de respuesta de las estructuras normativas tradicionales. La Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO constituye uno de los referentes más relevantes, al plantear que el desarrollo y uso de la IA debe orientarse por principios como dignidad humana, derechos humanos, diversidad, inclusión, transparencia, responsabilidad y sostenibilidad (UNESCO, 2021). Este enfoque no se limita a una preocupación técnica, sino que sitúa la

regulación en el terreno de los valores sociales que deben guiar la incorporación de la IA en distintos contextos institucionales.

En una línea complementaria, la OCDE ha planteado principios orientados al desarrollo de una IA confiable, destacando la necesidad de que los sistemas de IA respeten los derechos humanos, los valores democráticos, la transparencia, la robustez, la seguridad y la rendición de cuentas. Estos principios, adoptados originalmente en 2019 y revisados posteriormente para responder a la evolución tecnológica, expresan una preocupación creciente por traducir el desarrollo tecnológico en prácticas institucionalmente responsables (OCDE, 2019, 2024).

Por su parte, la Unión Europea ha avanzado hacia un marco regulatorio de mayor obligatoriedad mediante el Reglamento (UE) 2024/1689, conocido como Ley de Inteligencia Artificial. Este instrumento establece reglas armonizadas para la IA y adopta un enfoque basado en riesgos, lo que implica diferenciar los sistemas según su potencial impacto sobre derechos, seguridad y valores fundamentales (European Parliament & Council of the European Union, 2024).

Estos tres referentes muestran que la regulación de la IA se está construyendo en una lógica multiescalar: ética global, principios de gobernanza y marcos jurídicos vinculantes. Sin embargo, también evidencian una tensión relevante: aunque estos instrumentos definen orientaciones generales, su efectividad depende de la capacidad de las organizaciones para traducirlos en prácticas, procedimientos y estructuras de gestión. En ese punto, la regulación deja de ser únicamente un asunto normativo y se convierte en un proceso organizacional de institucionalización.

Si bien los marcos internacionales han logrado establecer principios orientadores para el desarrollo y uso de la inteligencia artificial, su implementación efectiva enfrenta desafíos significativos cuando se trasladan al ámbito organizacional. Esta dificultad radica en que los principios normativos, aun cuando son ampliamente

aceptados, no necesariamente se traducen de manera directa en estructuras operativas, procedimientos internos o mecanismos de control específicos. En consecuencia, la regulación de la IA tiende a configurarse como un proceso gradual de incorporación institucional, en el cual las organizaciones interpretan, adaptan y materializan dichos lineamientos en función de sus propias capacidades, contextos y objetivos estratégicos.

En este sentido, la regulación de la inteligencia artificial puede ser comprendida como un proceso de institucionalización en el que se construyen marcos de legitimidad que orientan la acción organizacional. Desde la perspectiva del nuevo institucionalismo, las organizaciones no adoptan prácticas únicamente por criterios de eficiencia técnica, sino también por la necesidad de alinearse con expectativas sociales, normativas y culturales que definen lo que se considera apropiado en un determinado campo organizacional (DiMaggio y Powell, 1983). Así, la regulación de la IA no solo establece restricciones o lineamientos, sino que contribuye a definir qué prácticas son legítimas, responsables y socialmente aceptables en el uso de estas tecnologías.

Esta lógica se observa con particular claridad en el caso de la inteligencia artificial, donde la incertidumbre tecnológica y los riesgos asociados han impulsado la construcción de marcos regulatorios que buscan generar confianza en su uso. Sin embargo, dicha confianza no se construye exclusivamente mediante la existencia de normas, sino a través de su incorporación en las dinámicas organizacionales. En otras palabras, la regulación adquiere efectividad en la medida en que es internalizada por las organizaciones y se traduce en rutinas, procedimientos y criterios de decisión.

No obstante, este proceso de institucionalización no está exento de tensiones. Por un lado, los marcos regulatorios tienden a promover la estandarización, la formalización y el control, elementos que buscan reducir la incertidumbre y

garantizar el cumplimiento de principios éticos y legales. Por otro lado, la inteligencia artificial, especialmente en sus aplicaciones más recientes, se desarrolla en entornos caracterizados por la experimentación, la flexibilidad y la rápida evolución tecnológica. Esta coexistencia de lógicas genera un escenario en el que la regulación no siempre logra acompasar el ritmo de la innovación, dando lugar a brechas entre lo normativo y lo operativo.

En este contexto, resulta pertinente reconocer que la regulación de la inteligencia artificial no se consolida de manera homogénea, sino que presenta variaciones según los niveles de desarrollo institucional, las capacidades organizacionales y las características de los sectores en los que se implementa. En países y organizaciones con mayor madurez regulatoria, es posible observar avances en la formalización de políticas y estructuras de gobernanza de la IA. En contraste, en contextos emergentes, la regulación tiende a manifestarse de manera más incipiente, con un predominio de prácticas informales o parcialmente estructuradas.

A partir de lo anterior, puede sostenerse que la regulación de la inteligencia artificial se encuentra en una fase de construcción, en la que coexisten marcos normativos consolidados con prácticas organizacionales aún en proceso de adaptación. Esta condición refuerza la idea de que la regulación no opera únicamente como un conjunto de reglas externas, sino como un proceso dinámico de institucionalización que se configura en la interacción entre principios normativos y prácticas organizacionales.

La consolidación de la regulación de la inteligencia artificial requiere, por tanto, un desplazamiento analítico que permita comprenderla no solo como un conjunto de disposiciones normativas, sino como un proceso de estructuración organizacional en el que los principios generales adquieren forma a través de prácticas concretas. En este sentido, la literatura reciente ha señalado que uno de los principales desafíos de la gobernanza de la IA radica en la brecha existente entre la formulación de principios éticos y su implementación efectiva en contextos organizacionales (Mittelstadt, 2019; Floridi y otros., 2018). Esta brecha evidencia que la regulación

no se agota en la definición de marcos normativos, sino que requiere mecanismos que permitan su traducción en sistemas de gestión, procesos operativos y estructuras de decisión.

Desde esta perspectiva, los esfuerzos internacionales por regular la IA han contribuido a la construcción de un consenso normativo en torno a valores y principios, pero han dejado abierta la cuestión de cómo dichos principios pueden ser operacionalizados en la práctica organizacional. Esta problemática ha sido destacada por diversos organismos que advierten la necesidad de avanzar hacia enfoques de gobernanza que integren dimensiones técnicas, organizacionales y sociales, con el fin de asegurar una implementación efectiva y contextualizada de la IA (OCDE, 2024; UNESCO, 2023). En este punto, la regulación se enfrenta a un dilema fundamental: mientras mayor es el nivel de abstracción de los principios, mayor es su capacidad de generalización, pero menor su aplicabilidad directa en contextos específicos.

En este escenario, emerge la necesidad de instrumentos que permitan traducir los marcos normativos en prácticas organizacionales estructuradas. Dichos instrumentos no solo cumplen una función técnica, sino también simbólica, en la medida en que contribuyen a dotar de legitimidad a las organizaciones frente a su entorno institucional. La adopción de estándares, certificaciones y sistemas de gestión puede interpretarse, en este sentido, como una forma de respuesta organizacional ante la incertidumbre regulatoria, en la que las organizaciones buscan alinearse con expectativas externas y demostrar el cumplimiento de criterios reconocidos internacionalmente (Suchman, 1995).

Este tránsito desde la regulación abstracta hacia la estructuración organizacional de la inteligencia artificial permite comprender la relevancia de estándares como la norma ISO/IEC 42001, la cual se inscribe precisamente en este punto de intersección entre principios normativos y prácticas organizacionales. Más que

constituir un instrumento meramente técnico, este tipo de estándares puede ser interpretado como un mecanismo que busca formalizar la gestión de la IA, dotándola de coherencia, trazabilidad y control dentro de las organizaciones. Sin embargo, su emergencia también refleja las limitaciones de los marcos regulatorios existentes, en la medida en que evidencia la necesidad de dispositivos que traduzcan la regulación en estructuras operativas concretas.

De esta manera, el análisis de la regulación de la inteligencia artificial como proceso de institucionalización conduce necesariamente a considerar el papel de los estándares organizacionales en la configuración de prácticas legítimas. La norma ISO/IEC 42001 no surge en un vacío normativo, sino en un contexto donde la regulación internacional ha establecido principios generales que requieren ser materializados. En este sentido, su estudio no puede desvincularse de las dinámicas institucionales que la hacen posible, ni de las tensiones que se generan cuando estos intentos de formalización se enfrentan a contextos organizacionales caracterizados por la complejidad y la emergencia.

La norma ISO/IEC 42001 como dispositivo de racionalización organizacional

La consolidación de la regulación de la inteligencia artificial en el ámbito organizacional exige mecanismos que permitan trasladar principios generales a formas concretas de estructuración de la acción. Diversos análisis recientes han señalado que uno de los principales retos de la gobernanza de la IA radica en la dificultad de traducir marcos normativos abstractos en prácticas organizacionales efectivas (Dignum, 2019). Esta brecha entre regulación y operación ha impulsado el desarrollo de instrumentos que, más allá de establecer orientaciones generales, buscan organizar de manera sistemática la gestión de la inteligencia artificial dentro de las organizaciones.

En este contexto, la norma ISO/IEC 42001 representa un intento por formalizar la gestión de la IA mediante la adopción de un sistema estructurado de control organizacional. Este tipo de estándares se inscribe en una tradición más amplia de normalización internacional orientada a generar consistencia, comparabilidad y confiabilidad en la gestión de procesos complejos (Loconto & Busch, 2010). Desde esta perspectiva, la norma no solo introduce lineamientos técnicos, sino que establece una arquitectura organizacional que define cómo deben estructurarse las prácticas en torno a la inteligencia artificial.

La ISO/IEC 42001 adopta el modelo de sistema de gestión característico de la International Organization for Standardization, el cual se fundamenta en la integración de procesos de planificación, implementación, evaluación y mejora continua. Este enfoque responde a una lógica de organización basada en la sistematización de la acción, en la que la complejidad es gestionada a través de la definición de procedimientos, la asignación de responsabilidades y la implementación de mecanismos de control (ISO/IEC, 2023). En este sentido, la norma presupone que la gestión de la IA puede ser organizada bajo una estructura formal que permita su supervisión y evaluación constante.

En términos operativos, la norma establece la necesidad de delimitar el alcance organizacional de la inteligencia artificial, definir políticas institucionales que orienten su uso y desarrollar mecanismos para la gestión de riesgos asociados a estas tecnologías. Este énfasis en la gestión del riesgo se vincula con enfoques contemporáneos de regulación tecnológica que buscan anticipar y mitigar impactos potenciales mediante la identificación sistemática de amenazas y vulnerabilidades (Black & Baldwin, 2010). A ello se suman la implementación de estructuras de gobernanza, la definición de roles y responsabilidades y el establecimiento de mecanismos de seguimiento y auditoría que permitan evaluar el desempeño del sistema de gestión.

A partir de esta arquitectura, es posible identificar un conjunto de dimensiones analíticas que permiten comprender cómo la norma traduce la complejidad de la inteligencia artificial en categorías organizacionales gestionables. Entre estas dimensiones se encuentran la trazabilidad, que implica la capacidad de documentar y rastrear los procesos asociados al uso de la IA; la mejora continua, que introduce ciclos sistemáticos de evaluación y ajuste; la gestión del riesgo, orientada a la identificación y mitigación de impactos; la participación de las partes interesadas, que incorpora múltiples actores en la toma de decisiones; y la transparencia algorítmica, que plantea la necesidad de hacer comprensibles los procesos automatizados.

Estas dimensiones no solo operan como componentes técnicos, sino que configuran una forma particular de estructurar la acción organizacional. Al traducir fenómenos complejos en variables susceptibles de ser observadas, medidas y controladas, la norma contribuye a reorganizar la incertidumbre asociada a la inteligencia artificial en términos manejables desde la perspectiva organizacional. Este proceso ha sido ampliamente documentado en estudios sobre regulación y control, donde se señala que la formalización de prácticas implica la transformación de lo incierto en objetos de evaluación sistemática (Power, 2004).

En consecuencia, la ISO/IEC 42001 puede ser comprendida como un instrumento que no solo orienta la gestión de la inteligencia artificial, sino que establece una lógica específica de intervención organizacional. Su adopción implica la incorporación de estructuras formales que buscan ordenar la relación entre tecnología y organización, alineándola con criterios de control, evaluación y rendición de cuentas. Este encuadre permite avanzar hacia una interpretación teórica más profunda de la norma, en la que su arquitectura puede ser analizada como una expresión contemporánea de procesos de racionalización organizacional.

La arquitectura organizacional propuesta por la ISO/IEC 42001 puede ser comprendida de manera más profunda si se analiza a la luz de los procesos de racionalización organizacional. Desde la perspectiva clásica de Weber, la racionalización implica la organización de la acción social mediante reglas formales, procedimientos estandarizados y estructuras de autoridad orientadas al control y la previsibilidad (Weber, 1978/2002). En este sentido, los sistemas de gestión constituyen una expresión contemporánea de esta lógica, en la medida en que buscan ordenar la complejidad organizacional mediante la formalización de prácticas y la definición de mecanismos de supervisión.

Aplicada al ámbito de la inteligencia artificial, esta lógica de racionalización se traduce en la transformación de procesos tecnológicos abiertos, dinámicos y, en muchos casos, inciertos, en objetos de gestión susceptibles de ser planificados, evaluados y controlados. La ISO/IEC 42001, al estructurar la gestión de la IA en torno a dimensiones como el riesgo, la trazabilidad y la mejora continua, convierte la incertidumbre tecnológica en un conjunto de variables organizacionales que pueden ser intervenidas sistemáticamente. Este proceso no solo facilita la gestión, sino que redefine la forma en que la IA es comprendida dentro de la organización, desplazándola del ámbito de la experimentación hacia el de la regulación formal.

Sin embargo, la racionalización organizacional no opera de manera aislada, sino que se articula con dinámicas más amplias de institucionalización. Desde el enfoque del nuevo institucionalismo, las organizaciones tienden a adoptar estructuras y prácticas que son socialmente legitimadas en su entorno, lo que da lugar a procesos de homogeneización conocidos como isomorfismo institucional (DiMaggio y Powell, 1983). Estos procesos pueden originarse en presiones coercitivas derivadas de marcos regulatorios, en presiones normativas asociadas a estándares profesionales, o en presiones miméticas que llevan a las organizaciones a imitar prácticas consideradas exitosas o legítimas.

En este contexto, la adopción de estándares como la ISO/IEC 42001 puede ser interpretada como una respuesta organizacional a la necesidad de legitimidad en un entorno caracterizado por la incertidumbre tecnológica. Más que responder exclusivamente a criterios de eficiencia, la incorporación de este tipo de normas permite a las organizaciones demostrar que sus prácticas se alinean con expectativas institucionales ampliamente reconocidas, particularmente en lo relativo a la gestión responsable de la inteligencia artificial. En este sentido, la norma no solo organiza la acción, sino que también cumple una función simbólica al posicionar a la organización dentro de un marco de cumplimiento y responsabilidad.

La interacción entre racionalización e isomorfismo permite comprender que la estructuración de la gestión de la IA no es únicamente un proceso técnico, sino también institucional. Al adoptar la ISO/IEC 42001, las organizaciones no solo implementan un sistema de gestión, sino que incorporan una forma específica de entender y organizar la relación con la inteligencia artificial. Esta forma está marcada por la formalización de procesos, la estandarización de prácticas y la orientación hacia el control, lo que contribuye a generar mayor previsibilidad en un contexto caracterizado por la complejidad.

No obstante, esta misma lógica puede generar tensiones cuando se enfrenta a contextos organizacionales emergentes, donde la adopción de la inteligencia artificial se desarrolla a través de dinámicas más flexibles, adaptativas y, en ocasiones, informales. En estos escenarios, la imposición de estructuras formales de control puede entrar en conflicto con prácticas organizacionales que dependen de la experimentación y la rápida adaptación al cambio tecnológico. Esta tensión no implica necesariamente una incompatibilidad, pero sí evidencia la necesidad de analizar críticamente los alcances de los modelos de racionalización en entornos donde la incertidumbre constituye una condición estructural.

A partir de esta lectura, la ISO/IEC 42001 puede ser interpretada como un mecanismo contemporáneo de racionalización organizacional que refuerza dinámicas de isomorfismo institucional. Esta interpretación permite situar la norma no solo como un instrumento de gestión, sino como parte de un proceso más amplio de estructuración de la acción organizacional en torno a la inteligencia artificial. En consecuencia, su análisis requiere considerar tanto sus implicaciones técnicas como sus efectos en la configuración de prácticas, estructuras y formas de legitimidad dentro de las organizaciones.

La interpretación de la ISO/IEC 42001 como un mecanismo de racionalización organizacional permite avanzar hacia una cuestión central: su relación con las prácticas efectivas de adopción de la inteligencia artificial en contextos organizacionales emergentes. Tal como se ha señalado, la norma introduce una lógica de formalización, control y estandarización que busca estructurar la gestión de la IA mediante sistemas organizados de toma de decisiones y supervisión. Sin embargo, esta lógica se enfrenta a escenarios donde la incorporación de la inteligencia artificial no sigue trayectorias lineales ni responde a diseños organizacionales previamente establecidos.

Diversos estudios han documentado que la adopción de tecnologías digitales avanzadas, incluida la inteligencia artificial, se caracteriza por procesos iterativos, experimentales y, en muchos casos, descentralizados, en los que la innovación emerge desde distintos niveles de la organización (Nambisan, Lyytinen, Majchrzak, & Song, 2017; Svahn, Mathiassen, & Lindgren, 2017). Estos procesos desafían las concepciones tradicionales de control organizacional, al evidenciar que la transformación digital no siempre puede ser completamente anticipada ni gestionada mediante estructuras formales predefinidas. En este sentido, la inteligencia artificial introduce una dinámica organizacional en la que la flexibilidad, la adaptación y la capacidad de aprendizaje continuo se vuelven elementos centrales.

En este contexto, la racionalización promovida por la ISO/IEC 42001 puede generar tensiones al intentar organizar prácticas que, por su naturaleza, son emergentes y adaptativas. Mientras la norma busca establecer criterios claros de gobernanza, trazabilidad y gestión de riesgos, las organizaciones enfrentan entornos en los que la velocidad de cambio tecnológico dificulta la consolidación de estructuras estables de control. Esta tensión ha sido identificada en la literatura sobre innovación organizacional, donde se señala que los procesos de formalización pueden, en determinados contextos, limitar la capacidad de experimentación y aprendizaje (Benner & Tushman, 2003).

Al mismo tiempo, la adopción de estándares como la ISO/IEC 42001 no puede ser interpretada únicamente en términos de restricción o control. Desde una perspectiva institucional, estos instrumentos también pueden facilitar la coordinación organizacional, al proporcionar marcos comunes que orientan la toma de decisiones y reducen la ambigüedad en el uso de tecnologías complejas (Scott, 2014). En este sentido, la norma puede contribuir a generar condiciones de confianza y coherencia interna, particularmente en organizaciones que buscan integrar la inteligencia artificial en procesos críticos o de alto impacto.

No obstante, la coexistencia de estas dos lógicas, la racionalización y la emergencia, plantea un desafío fundamental para la regulación organizacional de la inteligencia artificial. Por un lado, la necesidad de establecer mecanismos de control y responsabilidad impulsa la adopción de estructuras formales como las propuestas por la ISO/IEC 42001. Por otro, la naturaleza dinámica de la IA requiere espacios de flexibilidad que permitan la adaptación continua a contextos cambiantes. Esta dualidad no se resuelve mediante la sustitución de una lógica por otra, sino a través de su articulación, lo que implica reconocer que la regulación organizacional es, en sí misma, un proceso dinámico.

En este sentido, la norma puede ser entendida como un marco que busca estabilizar la relación entre organización y tecnología, sin eliminar completamente la incertidumbre inherente a la inteligencia artificial. Su implementación efectiva dependerá, por tanto, de la capacidad de las organizaciones para adaptar sus principios a contextos específicos, equilibrando la necesidad de control con la exigencia de innovación. Este equilibrio constituye uno de los principales retos de la gobernanza de la IA en contextos organizacionales emergentes, donde las prácticas no solo deben ser reguladas, sino también continuamente reinterpretadas y ajustadas.

A partir de este análisis, se refuerza la idea de que la ISO/IEC 42001 no puede ser evaluada únicamente en términos de su diseño normativo, sino en función de su interacción con las dinámicas organizacionales en las que se inserta. En consecuencia, su papel como dispositivo de racionalización organizacional debe ser comprendido en relación con las condiciones específicas de adopción de la inteligencia artificial, así como con las tensiones que emergen cuando la formalización se encuentra con la complejidad y la adaptabilidad propias de estos entornos.

Racionalización, inteligencia artificial y regulación: tensiones en contextos organizacionales emergentes

El análisis de la norma ISO/IEC 42001 como dispositivo de racionalización organizacional permite situar una tensión fundamental entre la lógica normativa de la regulación y la naturaleza emergente de la inteligencia artificial en los contextos organizacionales contemporáneos. Mientras que los sistemas de gestión proponen estructuras formales orientadas al control, la previsibilidad y la estandarización, la incorporación de la IA en las organizaciones se caracteriza, en múltiples casos, por dinámicas no lineales, experimentales y adaptativas que desafían dichos supuestos.

La literatura sobre transformación digital ha documentado que la adopción de tecnologías digitales avanzadas no responde necesariamente a procesos centralizados de planeación, sino que se configura a partir de interacciones distribuidas entre actores, prácticas y tecnologías dentro de las organizaciones (Yoo y otros., 2010; Hinings y otros., 2018). Este carácter distribuido implica que la inteligencia artificial no se incorpora como un sistema cerrado, sino como un conjunto de capacidades que evolucionan en función de su uso contextual y de las dinámicas organizacionales en las que se inserta.

En este sentido, la inteligencia artificial puede ser comprendida como un fenómeno organizacional emergente, en el que las prácticas se desarrollan mediante procesos de aprendizaje incremental, experimentación y adaptación continua. Esta perspectiva se articula con el enfoque de infraestructuras digitales, el cual sostiene que las tecnologías contemporáneas no son implementadas de manera definitiva, sino que se transforman en la medida en que son utilizadas y reinterpretadas por distintos actores organizacionales (Tilson y otros., 2010). Como resultado, la IA tiende a ser apropiada de forma heterogénea, generando múltiples configuraciones de uso que no necesariamente responden a esquemas formales predefinidos.

Esta condición plantea un desafío relevante para los marcos regulatorios organizacionales. Mientras que estos buscan estabilizar la gestión de la tecnología mediante la formalización de procesos, la dinámica de adopción de la IA exige flexibilidad, capacidad de respuesta y adaptación continua. Desde la perspectiva de las capacidades dinámicas, las organizaciones requieren no solo estructuras formales, sino también la habilidad de reconfigurar sus recursos y procesos en función de cambios tecnológicos y del entorno (Teece, 2018). En este contexto, la inteligencia artificial no solo introduce nuevas herramientas, sino que redefine las condiciones bajo las cuales se desarrollan los procesos organizacionales.

A partir de lo anterior, la relación entre racionalización organizacional y adopción de la inteligencia artificial no puede ser entendida en términos de sustitución, sino de coexistencia. Mientras los estándares como la ISO/IEC 42001 buscan estructurar la gestión de la IA mediante mecanismos formales de control, las prácticas organizacionales continúan evolucionando a través de procesos emergentes que no siempre pueden ser plenamente anticipados ni regulados. Esta dualidad configura un escenario en el que la regulación no elimina la incertidumbre, sino que se articula con ella, generando dinámicas complejas de adaptación organizacional.

En consecuencia, el análisis de la regulación de la inteligencia artificial en contextos organizacionales emergentes requiere considerar no solo los marcos normativos y los dispositivos de gestión, sino también las formas en que las organizaciones interpretan experimentan y transforman dichas herramientas en la práctica. Este enfoque permite avanzar hacia una comprensión más integral de la gobernanza de la IA, en la que la racionalización organizacional se reconoce como una dimensión relevante, pero no suficiente para explicar la complejidad de los procesos de adopción tecnológica.

La lectura de la ISO/IEC 42001 como dispositivo de racionalización permite desplazar el análisis desde una oposición superficial entre norma y práctica hacia la identificación de los mecanismos mediante los cuales la regulación intenta reconfigurar la acción organizacional en contextos de alta incertidumbre tecnológica. En este sentido, la tensión no se reduce a una falta de ajuste entre estructuras formales y dinámicas emergentes, sino que expresa una fricción más profunda entre dos formas de producir orden organizacional: una basada en la formalización ex ante y otra en la estabilización ex post de prácticas en uso.

La racionalización organizacional, tal como se materializa en los sistemas de gestión, opera mediante la anticipación de contingencias, la definición de categorías

de intervención y la estandarización de respuestas organizacionales. En el caso de la ISO/IEC 42001, esta lógica se traduce en la descomposición de la inteligencia artificial en dimensiones gestionables como riesgo, trazabilidad, control y evaluación. Sin embargo, este proceso implica una operación epistemológica relevante: la reducción de la indeterminación tecnológica a variables susceptibles de ser conocidas y gestionadas. Esta reducción no es neutral, ya que redefine qué aspectos de la IA son considerados problemáticos, medibles o relevantes dentro de la organización.

Desde la perspectiva de la práctica, las tecnologías no existen como objetos cerrados, sino como configuraciones sociomateriales que adquieren sentido en la interacción entre actores, artefactos y contextos organizacionales (Orlikowski, 2007; Leonardi, 2011). En el caso de la inteligencia artificial, esta condición se intensifica debido a su carácter probabilístico, su capacidad de aprendizaje y su dependencia de datos contextuales. Como resultado, su comportamiento no puede ser completamente anticipado ni estabilizado a priori, lo que limita la capacidad de los sistemas formales para capturar la totalidad de sus implicaciones organizacionales.

Esta divergencia introduce una primera tensión analítica: mientras la racionalización presupone la posibilidad de anticipar y controlar, la práctica tecnológica evidencia la imposibilidad de cerrar completamente el campo de acción de la IA. En consecuencia, los sistemas de gestión no eliminan la incertidumbre, sino que la redistribuyen, desplazando su tratamiento hacia mecanismos formales que coexisten con zonas de indeterminación no reguladas.

Una segunda tensión emerge en la relación entre formalización y aprendizaje organizacional. Los estándares como la ISO/IEC 42001 promueven la documentación, la estandarización y la replicabilidad como condiciones para la mejora continua. No obstante, la adopción de la inteligencia artificial en contextos organizacionales se apoya frecuentemente en procesos de aprendizaje situados,

experimentación y adaptación que difícilmente pueden ser completamente codificados (Nonaka & Takeuchi, 1995). Esta asimetría genera una brecha entre el conocimiento formalizable y el conocimiento práctico, lo que implica que una parte significativa de la gestión de la IA permanece fuera de los sistemas formales de control.

Una tercera tensión se relaciona con la temporalidad de la regulación. La racionalización organizacional opera bajo una lógica de estabilización, en la que los procesos son definidos, implementados y posteriormente evaluados. En contraste, la inteligencia artificial introduce una temporalidad dinámica, caracterizada por cambios continuos en modelos, datos y contextos de uso. Esta diferencia temporal dificulta la consolidación de estructuras estables de gobernanza, ya que los objetos de regulación se transforman más rápidamente que los marcos diseñados para gestionarlos (Faraj y otros., 2018).

Finalmente, la tensión se expresa en el plano de la legitimidad organizacional. La adopción de estándares como la ISO/IEC 42001 puede responder tanto a la necesidad de estructurar la gestión de la IA como a la búsqueda de legitimidad frente a actores externos. En este sentido, la racionalización no solo organiza la acción, sino que también produce una representación de control que puede no corresponder completamente con la práctica efectiva. Este fenómeno ha sido ampliamente documentado en estudios sobre auditoría y gestión del riesgo, donde se señala que los sistemas formales pueden generar una apariencia de control sin necesariamente capturar la complejidad de los procesos que buscan regular (Power, 1997).

A partir de estas tensiones, resulta posible sostener que la regulación organizacional de la inteligencia artificial no puede ser entendida como un proceso de alineación entre norma y práctica, sino como un espacio de negociación en el que coexisten lógicas distintas de organización. La ISO/IEC 42001, en tanto dispositivo de racionalización, introduce un marco que busca estabilizar la gestión

de la IA, pero su efectividad depende de su capacidad para interactuar con prácticas organizacionales que permanecen abiertas, adaptativas y en constante transformación.

A partir del análisis desarrollado, es posible sintetizar las principales tensiones entre la lógica de racionalización organizacional y la práctica emergente de la inteligencia artificial, las cuales se presentan en la siguiente tabla.

Tabla 1. Tensiones entre racionalización organizacional y práctica emergente de la IA

Dimensión analítica	Lógica de racionalización (ISO/IEC 42001)	Lógica de práctica organizacional emergente	Tensión organizacional	Implicación para la regulación de la IA
Producción de orden organizacional	Estructuración ex ante mediante reglas, procedimientos y controles formales	Estabilización ex post a partir de la práctica, el uso y la adaptación continua	Imposibilidad de anticipar completamente el comportamiento de la IA	La regulación no elimina la incertidumbre, sino que la redistribuye en mecanismos formales
Gestión de la incertidumbre	Reducción de la complejidad mediante categorías	Persistencia de indeterminación derivada del carácter	Desfase entre lo que puede ser modelado y lo que	La IA excede los marcos de control

	como riesgo, trazabilidad y control	probabilístico y contextual de la IA	emerge en la práctica	previstos por los sistemas de gestión
Naturaleza del conocimiento organizacional	Codificación, documentación y estandarización de procesos	Aprendizaje situado, conocimiento tácito y experimentación	Brecha entre conocimiento formalizable y conocimiento en uso	Parte de la gestión de la IA permanece fuera de los sistemas formales
Temporalidad organizacional	Procesos secuenciales de planificación, implementación y evaluación	Dinámica iterativa, evolutiva y en constante cambio	Inestabilidad de los objetos de regulación frente a estructuras formales rígidas	La regulación requiere adaptabilidad continua para mantenerse pertinente
Gobernanza y control	Supervisión centralizada, definición clara de roles y responsabilidades	Uso distribuido y descentralizado de herramientas de IA	Dificultad para centralizar decisiones en entornos altamente dinámicos	La gobernanza de la IA requiere modelos más flexibles e híbridos
Legitimidad organizacional	Cumplimiento de estándares como	Prácticas que evolucionan independientemente	Desacoplamiento entre cumplimiento	La regulación puede

	mecanismo de validación externa	norte del marco formal	formal y práctica real	operar más como legitimación que como control efectivo

Fuente: Elaboración propia con base en ISO/IEC (2023), Orlikowski (2007), Leonardi (2011), Faraj y otros. (2018), Nonaka y Takeuchi (1995) y Power (1997). Estas tensiones evidencian que la regulación organizacional de la inteligencia artificial no puede ser comprendida únicamente como un proceso de alineación normativa, sino como un campo de interacción entre lógicas organizacionales diferenciadas.

El análisis desarrollado permite situar la regulación de la inteligencia artificial en un plano que trasciende su función normativa inmediata, al evidenciar su papel en la configuración de formas específicas de organización. En este sentido, la ISO/IEC 42001 puede ser comprendida como una manifestación contemporánea de procesos de racionalización, en la medida en que introduce esquemas de orden basados en la formalización, la clasificación y la estandarización de la acción organizacional. Sin embargo, esta racionalización no se limita a estructurar prácticas, sino que redefine los términos bajo los cuales la inteligencia artificial se vuelve inteligible, evaluable y, por tanto, gobernable dentro de las organizaciones.

Desde esta perspectiva, la racionalización opera como un mecanismo que no elimina la complejidad, sino que la reorganiza selectivamente. Al establecer categorías como riesgo, control o trazabilidad, la norma delimita aquello que puede ser incorporado dentro de un sistema formal de gestión, al tiempo que deja fuera dimensiones que no son fácilmente traducibles a estos esquemas. Este proceso de selección no es meramente técnico, sino organizacional, ya que implica la

construcción de criterios que orientan la acción y condicionan las formas de intervención sobre la inteligencia artificial.

No obstante, la incorporación de este tipo de estándares no puede ser explicada únicamente desde la lógica de la racionalización. Su adopción también responde a dinámicas de isomorfismo institucional, en las que las organizaciones tienden a alinearse con modelos reconocidos como legítimos en su entorno. En este sentido, la ISO/IEC 42001 no solo estructura la gestión de la IA, sino que se convierte en un referente que orienta expectativas sobre lo que constituye un uso responsable y adecuado de estas tecnologías. Tal como ha sido señalado en la literatura institucional, la adopción de estructuras formales no siempre obedece a criterios de eficiencia, sino a la necesidad de alcanzar legitimidad en contextos organizacionales específicos (Meyer & Rowan, 1977; Scott, 2014).

Esta doble condición, racionalizadora e institucional, permite comprender que la norma opera simultáneamente en dos niveles. Por un lado, organiza la acción mediante la introducción de sistemas formales de gestión. Por otro, contribuye a la homogeneización de prácticas al establecer un modelo que puede ser replicado en distintos contextos organizacionales. Este efecto de convergencia no implica necesariamente una uniformidad en la práctica, pero sí configura un marco común que orienta la forma en que las organizaciones interpretan y gestionan la inteligencia artificial.

En este punto, la tensión identificada a lo largo del capítulo adquiere un carácter más preciso. La racionalización busca estabilizar la acción organizacional, mientras que la dinámica emergente de la inteligencia artificial introduce variabilidad, adaptación y reconfiguración constante. El isomorfismo institucional, por su parte, impulsa la adopción de modelos formales que pueden no corresponder plenamente con las prácticas efectivas, generando posibles desacoplamientos entre lo que se declara y lo que se realiza. Este fenómeno no representa una anomalía, sino una

característica recurrente de los procesos organizacionales en contextos de alta complejidad (Boxenbaum & Jonsson, 2017).

En consecuencia, la regulación de la inteligencia artificial en contextos organizacionales emergentes no puede ser comprendida como un mecanismo de cierre, sino como un proceso en el que la racionalización y el isomorfismo interactúan en la producción de orden organizacional. La ISO/IEC 42001, en este marco, no resuelve la complejidad asociada a la IA, pero sí establece condiciones para su gestión, al tiempo que configura expectativas de legitimidad que influyen en la forma en que las organizaciones se posicionan frente a estas tecnologías.

Este encuadre permite afirmar que la gobernanza de la inteligencia artificial no depende exclusivamente de la robustez de los marcos normativos, sino de la capacidad de las organizaciones para articular estos marcos con prácticas que permanecen abiertas y en transformación. Bajo esta lógica, la regulación no se impone sobre la acción organizacional, sino que participa en su configuración, generando un espacio en el que el control, la legitimidad y la adaptación coexisten como dimensiones constitutivas de la gestión de la inteligencia artificial.

Conclusiones

El análisis desarrollado permite replantear el lugar de la regulación de la inteligencia artificial en el ámbito organizacional, no como un dispositivo destinado a cerrar la incertidumbre, sino como un conjunto de prácticas que participan activamente en la producción de orden bajo condiciones de complejidad. En este sentido, la ISO/IEC 42001 no constituye una solución definitiva a los desafíos de la gobernanza de la IA, sino una forma específica de intervención que organiza la acción a partir de criterios de racionalización que, aunque necesarios, resultan necesariamente parciales.

La racionalización organizacional, al traducirse en sistemas de gestión, implica una operación de reducción mediante la cual la complejidad tecnológica es reconfigurada en categorías gestionables como riesgo, control o trazabilidad. Sin embargo, esta operación no logra capturar la totalidad de las dinámicas asociadas a la inteligencia artificial, particularmente aquellas vinculadas con la adaptación, el aprendizaje situado y la reconfiguración constante de las prácticas. En consecuencia, la regulación no elimina la incertidumbre, sino que la redistribuye, generando espacios en los que el control formal coexiste con zonas de indeterminación que permanecen fuera de los marcos establecidos.

Desde la perspectiva del isomorfismo institucional, la adopción de estándares como la ISO/IEC 42001 revela que las organizaciones no solo buscan gestionar la inteligencia artificial, sino también alinearse con expectativas de legitimidad que definen lo que se considera un uso adecuado de estas tecnologías. Este desplazamiento introduce una dimensión crítica en el análisis, ya que sugiere que la regulación puede operar como un mecanismo de validación externa que no necesariamente transforma las prácticas organizacionales en profundidad. En este sentido, la posibilidad de desacoplamientos entre las estructuras formales y la práctica efectiva no constituye una anomalía, sino una manifestación inherente a los procesos organizacionales en contextos de alta complejidad.

A partir de estas consideraciones, la regulación organizacional de la inteligencia artificial puede ser comprendida como un proceso en el que convergen tres dimensiones analíticas que no se resuelven entre sí: la racionalización, que introduce orden mediante la formalización; el isomorfismo, que impulsa la adopción de modelos legitimados; y la práctica organizacional, que incorpora variabilidad, adaptación y aprendizaje continuo. La interacción entre estas dimensiones no conduce a un estado de equilibrio, sino a configuraciones dinámicas en las que la estabilidad es siempre provisional y la regulación se redefine en función de su uso.

En este sentido, resulta pertinente cuestionar si los modelos actuales de regulación, basados en lógicas de control y estandarización, son capaces de acompañar procesos tecnológicos cuya principal característica es la mutabilidad. Esta interrogante no implica desestimar la relevancia de instrumentos como la ISO/IEC 42001, sino problematizar las condiciones bajo las cuales estos pueden operar en contextos organizacionales emergentes, donde la acción no se encuentra completamente estructurada y las decisiones se configuran de manera situada.

Asimismo, el análisis invita a reconsiderar el papel de la gobernanza de la inteligencia artificial más allá de la implementación de estándares. La regulación efectiva no parece depender exclusivamente de la existencia de marcos formales robustos, sino de la capacidad de las organizaciones para articular dichos marcos con prácticas que permanecen abiertas y en transformación. Esto implica reconocer que la gestión de la IA no puede ser completamente anticipada ni cerrada, sino que requiere mecanismos de interpretación, ajuste y aprendizaje continuo que exceden los sistemas formales de gestión.

En este marco, emergen interrogantes que permiten profundizar en el análisis de la regulación de la inteligencia artificial desde una perspectiva organizacional:

- ¿En qué medida los sistemas de gestión basados en la racionalización organizacional pueden adaptarse a tecnologías cuya evolución es continua y no lineal?
- ¿Hasta qué punto la adopción de estándares como la ISO/IEC 42001 responde a procesos de transformación organizacional efectiva o a dinámicas de legitimación institucional?
- ¿Cómo pueden las organizaciones articular la formalización requerida por la regulación con prácticas que se configuran desde la adaptación y la experimentación constante?

Estas preguntas no se plantean como un cierre del análisis, sino como una apertura hacia una comprensión más compleja de la gobernanza de la inteligencia artificial. Su formulación permite evidenciar que el desafío no radica únicamente en diseñar mejores marcos normativos, sino en comprender las condiciones organizacionales bajo las cuales dichos marcos adquieren sentido y operatividad. En este sentido, la regulación no puede ser concebida como un mecanismo externo que se impone sobre la práctica, sino como un proceso que se construye en interacción con dinámicas organizacionales que la reinterpretan, la adaptan e incluso la transforman. Esta perspectiva invita a desplazar la atención desde la norma hacia la relación entre norma y práctica, reconociendo que la efectividad de la regulación depende menos de su grado de formalización que de su capacidad para integrarse en entornos caracterizados por la incertidumbre, la variabilidad y el cambio continuo.

Referencias

- Benner, M. J., & Tushman, M. L. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review*, 28(2), 238–256. <https://doi.org/10.5465/amr.2003.9416096>
- Black, J., & Baldwin, R. (2010). Really responsive risk-based regulation. *Law & Policy*, 32(2), 181–213. <https://doi.org/10.1111/j.1467-9930.2010.00318.x>
- Boxenbaum, E., & Jonsson, S. (2017). Isomorphism, diffusion and decoupling: Concept evolution and theoretical challenges. En R. Greenwood, C. Oliver, T. Lawrence, & R. Meyer (Eds.), *The SAGE handbook of organizational institutionalism* (pp. 79–104). Sage.
- Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer.
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160. <https://doi.org/10.2307/2095101>
- European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. <https://eur-lex.europa.eu>
- European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence*. Official Journal of the European Union.
- Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of artificial intelligence. *Information and Organization*, 28(1), 62–70. <https://doi.org/10.1016/j.infoandorg.2018.02.007>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>

- Hinings, B., Gegenhuber, T., & Greenwood, R. (2018). Digital innovation and transformation: An institutional perspective. *Information and Organization*, 28(1), 52–61. <https://doi.org/10.1016/j.infoandorg.2018.02.004>
- ISO/IEC. (2023). *ISO/IEC 42001: Artificial intelligence — Management system*. International Organization for Standardization.
- Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly*, 35(1), 147–167.
- Loconto, A., & Busch, L. (2010). Standards, techno-economic networks, and playing fields: Performing the global market economy. *Review of International Political Economy*, 17(3), 507–536. <https://doi.org/10.1080/09692290903319870>
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363.
- Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management: Reinventing innovation management research in a digital world. *MIS Quarterly*, 41(1), 223–238.
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company*. Oxford University Press.
- OCDE. (2019). *Recommendation of the Council on Artificial Intelligence*. OCDE Publishing. <https://doi.org/10.1787/eb54c2c1-en>
- OCDE. (2024). *OCDE AI Principles*. Organisation for Economic Co-operation and Development.
- Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, 28(9), 1435–1448. <https://doi.org/10.1177/0170840607081138>
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.

- Power, M. (2004). *The risk management of everything: Rethinking the politics of uncertainty*. Demos.
- Scott, W. R. (2014). *Institutions and organizations: Ideas, interests, and identities* (4th ed.). Sage.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571–610. <https://doi.org/10.2307/258788>
- Svahn, F., Mathiassen, L., & Lindgren, R. (2017). Embracing digital innovation in incumbent firms: How Volvo Cars managed competing concerns. *MIS Quarterly*, 41(1), 239–253.
- Teece, D. J. (2018). Business models and dynamic capabilities. *Long Range Planning*, 51(1), 40–49. <https://doi.org/10.1016/j.lrp.2017.06.007>
- Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Digital infrastructures: The missing IS research agenda. *Information Systems Research*, 21(4), 748–759. <https://doi.org/10.1287/isre.1100.0318>
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.
- Weber, M. (1978/2002). *Economy and society: An outline of interpretive sociology*. University of California Press.
- Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research*, 21(4), 724–735. <https://doi.org/10.1287/isre.1100.0322>

CAPÍTULO XXVI

NOTEBOOKLM A LA LUZ DE LOS PRINCIPIOS DE CHAPULTEPEC

NotebookLM in Light of the Chapultepec Principles

Elizabeth Ramírez Navarro

ramnael.e9@gmail.com

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0003-0984-8940>

Resumen: Este capítulo analiza mi experiencia con NotebookLM desde un interaccionismo crítico, partiendo de una sospecha fundamental: que el «orden digerible» de la máquina desincentiva nuestra capacidad de cuestionar la realidad. Utilizo como brújula los Principios de Chapultepec para auditar esta herramienta, entendiéndola no como un medio neutral, sino como un actante que traduce y transforma mis fuentes. A través de una metodología performativa de co-autoría entre materia consciente y artificial, exploro la IA en sus roles de amplificación, asistencia y co-agencia. Argumento que NotebookLM no pasa la prueba ética por defecto; sin embargo, su potencia radica en ser una interfaz que podemos habitar críticamente para desafiar nuestras propias percepciones. Concluyo que un socio científico artificial solo es viable si el investigador asume su rol de objetor permanente, rechazando la «educación bancaria» algorítmica para transitar hacia un pluriverso de conocimientos locales. Al final, la soberanía cognitiva es una conquista diaria frente a la comodidad de la caja negra.

Palabras clave: Principios de Chapultepec, ética, colaboración humano-IA

Abstract

This chapter analyzes my experience with NotebookLM from a critical interactionist perspective, grounded in a fundamental suspicion: that the machine's "digestible order" discourages our capacity to question reality. I use the Chapultepec Principles

as a compass to audit this tool, understanding it not as a neutral medium, but as an actant that translates and transforms my sources. Through a performative methodology of co-authorship between conscious and artificial matter, I explore AI in its roles of amplification, assistance, and co-agency. I argue that NotebookLM does not pass the ethical test by default; however, its strength lies in being an interface that we can critically inhabit in order to challenge our own perceptions. I conclude that an artificial scientific partner is only viable if the researcher assumes the role of a permanent dissenter, rejecting algorithmic “banking education” in order to move toward a pluriverse of local knowledges. Ultimately, cognitive sovereignty is a daily conquest against the comfort of the black box.

Keywords: Chapultepec Principles, ethics, human-AI collaboration

Mi inquietud principal, que recorrerá cada página de este capítulo, nace de una sospecha que puede parecer obvia: la sociedad que formamos con herramientas como NotebookLM no es un contrato ético garantizado por la tecnología de Google, sino una conquista diaria que depende de nuestra propia mirada crítica. Me preocupa que, al buscar orden en medio del caos o el océano de información, terminemos aceptando como “verdad” lo que es simplemente un flujo algorítmico bien organizado, sacrificando nuestra capacidad de decidir y de reconocer las realidades que la propia máquina invisibiliza. Sí, le estoy concediendo a NotebookLM el poder dar orden a la información que le cargamos en sus Fuentes, quizá un orden que es tan digerible que nos desincentiva a cuestionar el origen y sentido de los contenidos de salida que puede generar la interfaz.

Esta reflexión no surge en el vacío, sino en el umbral de lo que se denomina la «civilización artificial» (Lasalle, 2024), donde la inteligencia artificial (IA) ha dejado de ser un simple instrumento para convertirse en un actante que reconfigura nuestra forma de conocer el mundo (Latour, 2008). Mi historia con esta herramienta comenzó a mediados de 2025, aunque su génesis técnica se remonta a 2022 con el Project Tailwind de Google Labs. Como investigadora situada en México, mi

interacción con NotebookLM –basado en el modelo multimodal Gemini– me ha obligado a replantear mi propia praxis. No busco ofrecer un manual técnico, sino indagar cómo se construye la responsabilidad ética en esa interfaz donde lo humano y lo artificial se encuentran (Simon, 2006).

Para orientar este análisis, propongo tres formas de entender nuestra relación con la IA: como *herramienta* que amplifica nuestra capacidad de indagación al liberarnos de tareas analíticas pesadas; como *agente asistente* que actúa como un interlocutor socrático para refinar nuestra propia comprensión; y como *co-agente no-humano* que participa activamente en un ensamblaje híbrido para construir sentidos de realidad (Ramírez Navarro, 2025). El camino que propongo es un tipo de auditoría que tiene como punto de partida la experiencia personal que toma como brújula la Declaración de ética y buenas prácticas para el uso y desarrollo de la Inteligencia Artificial, mejor conocida como los Principios de Chapultepec (Gobierno de México y otros., 2026).

Para asegurar el cumplimiento del Principio 3 sobre la explicabilidad y permitir que este ejercicio sea replicable por otras personas, considero prudente bosquejar cómo se dio esta colaboración. Este capítulo es el resultado de un diálogo sostenido durante aproximadamente tres días, involucrando aproximadamente 24 iteraciones entre consulta y respuesta. El cuaderno de trabajo integró un ecosistema de 20 fuentes originales cargadas en la interfaz de NotebookLM, configurando el chat bajo el modelo Gemini con instrucciones específicas para priorizar respuestas extensas, reflexivas y un tono de divulgación científica. Al declarar explícitamente el uso de la IA y reconstruir este proceso, ejerzo esa soberanía cognitiva que me permite no ser «sujeta muda» ante la técnica (Mendoza, 2010). También debo mencionar que tengo un Gem –parecido a un GPT en ChatGPT, la cual es una versión personalizada de la IA de Google, que uno puede configurar para tareas específicas y que son repetitivas– ya configurado para que me de prompts para la configuración de diferentes IA, esto ya en un formato

markdown¹⁰. Con este Gem generé la configuración del chat para el cuaderno que usé de apoyo para la redacción de este capítulo.

A través de cuatro apartados, exploraremos qué es NotebookLM más allá de la narrativa corporativa, analizaremos críticamente los Principios de Chapultepec y someteremos mi práctica investigativa a su juicio. Mi tesis es que la ética en la IA no reside en el código de Silicon Valley, sino en nuestra voluntad de pagar el precio de energía que demanda la reflexión crítica permanente (Harari, 2024).

¿Qué Es NotebookLM?

Para entender la tensión ética que propongo en este capítulo, es indispensable empezar por definir qué es NotebookLM. Mi historia personal con esta herramienta comenzó a mediados de 2025, pero su génesis se remonta a 2022, cuando un equipo de Google Labs inició un experimento conocido internamente como Project Tailwind (Gartenberg, 2025). El objetivo era ambicioso: crear una aplicación que utilizara los modelos de lenguaje más avanzados de Google para servir de apoyo al pensamiento, la escritura y la investigación, basándose estrictamente en las fuentes que la persona usuaria decidiera proporcionarle.

En julio de 2023, este experimento se hizo público bajo el nombre de NotebookLM, presentándose no como un chatbot genérico –como ChatGPT o Gemini o Claude o Grok o DeepSeek– sino como un «socio o compañero de investigación». Esta distinción es fundamental: a diferencia de otros sistemas de inteligencia artificial (IA) que responden basándose en la inmensidad de internet, NotebookLM se ancla en Gemini, su modelo de lenguaje multimodal, pero restringe su universo de "verdad" a la sección de *Fuentes* cargadas en el cuaderno. Es, en

¹⁰ En lenguaje coloquial, es una forma de lenguaje que permite darle formato al texto “plano”, es decir a aquel que no tiene negritas o cambios en el tamaño de la tipografía o diferentes tipografías; lo que permite estructurar la jerarquía y estilo de la redacción a partir de usar algunos símbolos como son #, *, _.

esencia, un sistema diseñado para la racionalidad limitada del ser humano, ayudándonos a procesar entornos semánticos ricos (Simon, 2006).

Desde una perspectiva técnica, un cuaderno se organiza en tres apartados clave (Figura 1) que dictan la interacción (Google, 2026):

1. *Fuentes*: Es el núcleo del sistema. Está ubicado del lado izquierdo. Permite cargar desde el dispositivo o buscar en la web una diversidad de archivos (PDF, audios, videos de YouTube, textos copiados, entre otros; siendo del tipo pdf, txt, md, docx, csv, pptx, epub, avif, bmp, gif, ico, jp2, png, webp, tif, tiff, heic, heif, jpeg, jpg, jpe, 3g2, 3gp, aac, aif, aifc, aiff, amr, au, avi, cda, m4a, mid, mp3, mp4, mpeg, ogg, opus, ra, ram, snd, wav, wma) en más de 50 idiomas; la cantidad exacta de *fuentes* que se puedan cargar dependerá de la versión con que se cuente, teniendo como mínimo 50 y como máximo 600 (gratuita, Google AI Plus, Google AI Pro, Google AI Ultra). Google promete que esta "fuente cerrada" garantiza la verificabilidad, pues cada respuesta en el chat incluye citas directas para que podamos "reconstruir el proceso", tal como exigirían los principios de transparencia expresados en los Principios de Chapultepec que abordaré en el siguiente apartado.
2. *Chat*: Aquí es donde ocurre la magia de la co-agencia. Se ubica en la parte central del cuaderno. Se puede configurar el rol –predeterminado, guía de aprendizaje, o personalizado– y elegir la longitud de la respuesta –predeterminada, más larga, más corta– (Figura 2). En esta interfaz, como diría Simon (2006), el artefacto se adapta al entorno externo (el investigador) para cumplir un propósito deseado. Esta es la parte que se usa de manera similar que los chatbots genéricos; es decir, se escribe una consulta, se envía, se recibe una respuesta. Del proceso anterior se pueden hacer una gran cantidad de iteraciones con el límite diario de la versión con que se cuente, siendo 50 la menor cantidad y 5000 la mayor.
3. *Studio*: Es el espacio de creación por excelencia. Se ubica en el lado derecho del cuaderno. Aquí el sistema puede generar, casi de forma

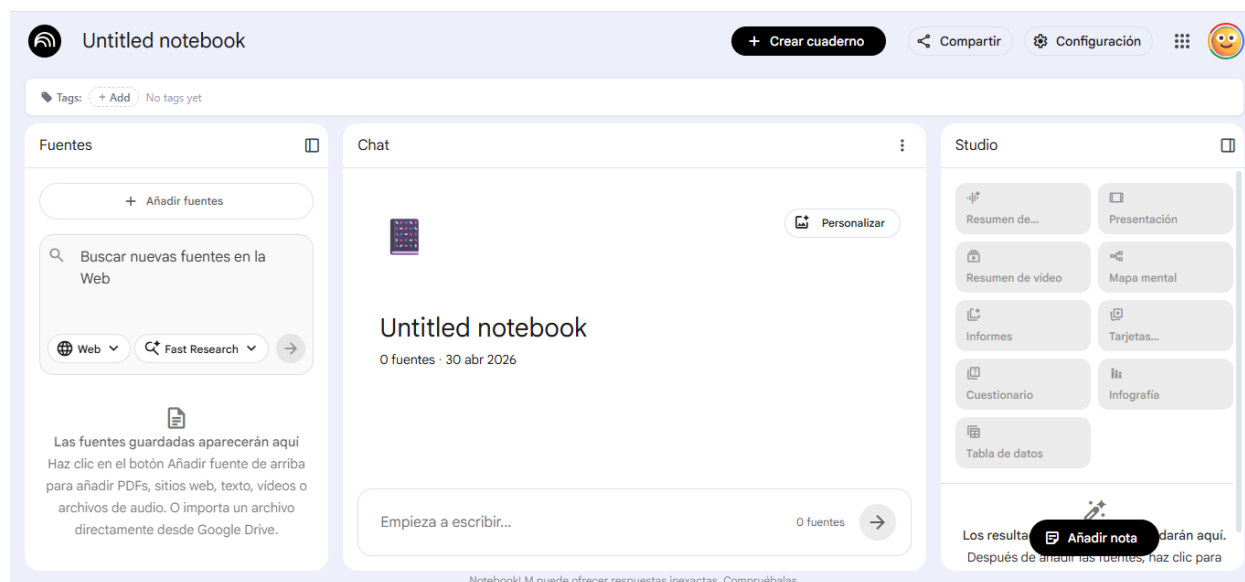
instantánea, resúmenes de audio (tipo pódcast), mapas mentales, presentaciones, cuestionarios, resúmenes de video, informes, tarjetas didácticas, infografías o tabla de datos. Cabe señalar que esta variedad de generación de contenidos ha ido en aumento además de que algunos de estos tipos de salida se pueden configurar para personalizar el contenido que la interfaz entrega. Es precisamente en este apartado donde, como mencioné antes, el orden algorítmico se vuelve tan "perfecto" que puede llegar a desincentivar nuestra capacidad crítica. Así como en los otros apartados, la cantidad que se puede generar en este apartado tiene límites dependiendo de la versión con que se cuente.

En el uso diario de esta aplicación oficial, hay quienes han encontrado que se pueden facilitar aún más algunas tareas. La comunidad de usuarios, ante necesidades particulares, ha desarrollado extensiones que expanden o facilitan las posibilidades de esta interfaz. Herramientas como NotebookLM Tools o FolderLLM permiten agrupar fuentes por carpetas, asignar colores para una organización visual o importar listas masivas de YouTube y Google Drive (NotebookLM Tools, 2026). Estos complementos o plug-ins son, parafraseando a (Latour, 2008), capas sucesivas que nos equipan para ser investigadores más competentes, distribuyendo nuestra capacidad cognitiva en todo un entorno formateado.

Sin embargo, detrás de esta arquitectura de "socio ideal" que ahorra tiempo y esfuerzo frente a la saturación informativa, siguiendo a Harari (2024), subyace una pregunta que Google no responde: ¿en qué momento este apoyo se convierte en un «burócrata digital» que prioriza el orden sobre la verdad?

Figura 1

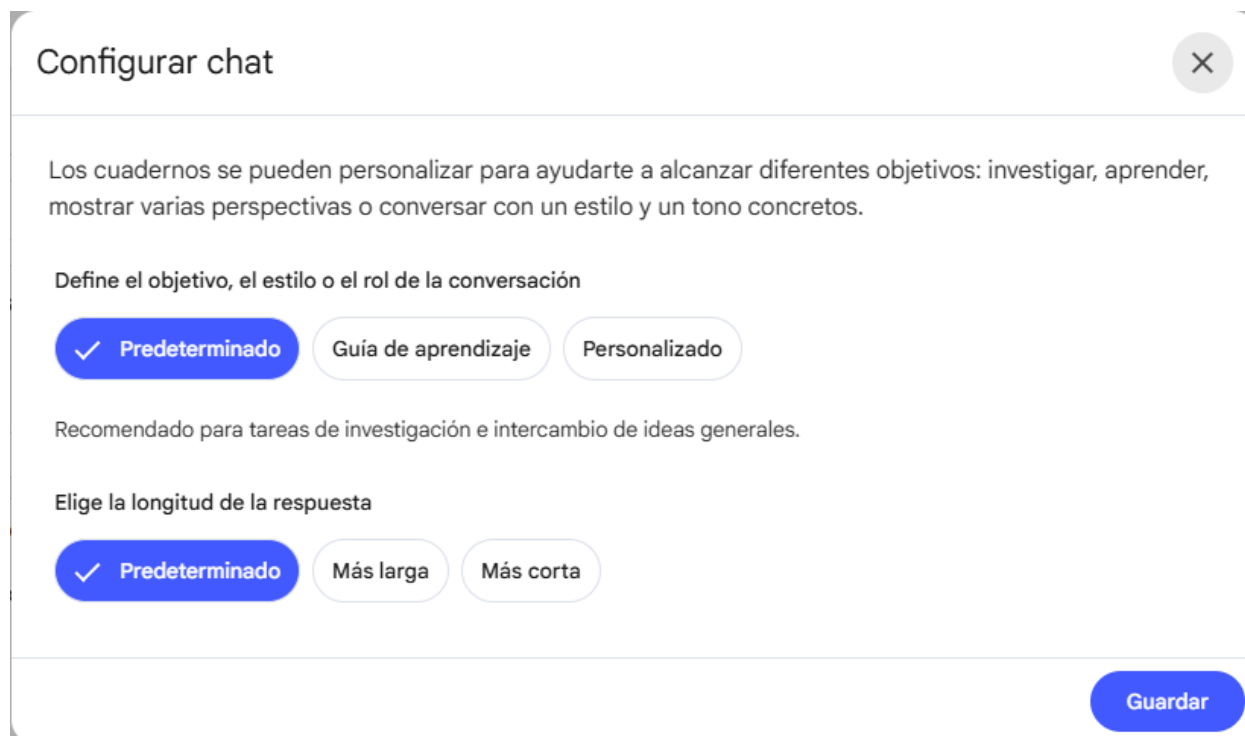
Captura de pantalla de la interfaz de inicio de un cuaderno en NotebookLM



Nota. Interfaz de usuario de NotebookLM. © Google LLC".

Figura 2

Captura de pantalla de la configuración del apartado de Chat.



Nota. Interfaz de usuario de NotebookLM. © Google LLC".

A pesar de su origen, mi primer encuentro con NotebookLM en 2025 no fue tanto el de una fría interacción técnica, sino el de un descubrimiento profundamente satisfactorio. En aquel entonces, la sección de Studio apenas ofrecía tres o cuatro funciones: el Resumen de audio, el Mapa mental, las Tarjetas didácticas. Recuerdo que la sensación era la de recibir un regalo intelectual; había un gozo genuino al ver cómo, de forma casi instantánea, mis documentos se transformaban en formatos dinámicos y atractivos. En esos momentos, NotebookLM actuaba como el artefacto perfecto que describe Simon (2006): una interfaz que se adaptaba tan bien a mi racionalidad limitada que me evitaba el esfuerzo de navegar sola por el océano de información.

Sin embargo, tras ese primer entusiasmo, llegó el silencio de la reflexión y, con este, las preguntas incómodas. Al observar el mapa mental, tan detallado y bien estructurado, el asombro se convirtió en inquietud: ¿por qué el sistema eligió ese orden y no otro? ¿cómo fue que se decidió qué quedaba dentro y qué quedaba fuera? Fue reflexionar sobre que NotebookLM no era un cristal transparente, sino un mediador en el sentido de Latour (2008): una entidad que no se limita a transportar datos, sino que los traduce, los distorsiona y, en última instancia, los transforma. El "orden" que me ofrecía no era un mapa neutral de mis fuentes, sino que construía una realidad intercomputacional —un término que recupero de Harari (2024)—, un relato coherente que, al ser tan "digerible", corría el riesgo de desincentivar mi propia capacidad crítica.

Esta tensión me llevó a cuestionar el propósito pedagógico de sus funciones. Si las tarjetas didácticas deciden qué es lo "importante" de un texto, ¿qué tipo de pensamiento están incentivando? Existe el peligro de que, en manos de un estudiante de preparatoria aún en formación, la IA promueva lo que Paulo Freire (1993) llamaba una «educación bancaria», donde el conocimiento se reduce a depósitos de información que el alumno archiva pasivamente, renunciando a la soberanía de su propia palabra.

Incluso el formato que me resultó más cautivador, el resumen de audio tipo podcast me generó dudas. Al elegir qué incluir y qué silenciar, el sistema ejerce una

función de canonización algorítmica similar a la de los antiguos burócratas, pero con la velocidad del silicio (Harari, 2024). Si NotebookLM excluye información que para mi subjetividad situada es fundamental, entonces su "orden" está anteponiendo la eficiencia a alguna verdad. Esta es la sospecha que da origen a esta indagación: si la sociedad con la IA es tan cómoda que nos hace dejar de pensar, entonces el cumplimiento de los Principios de Chapultepec no puede ser un diseño de Google, sino una conquista diaria de nuestra propia mirada crítica.

Los Principios de Chapultepec

El apartado anterior estuvo orientado al gozo y la sospecha que despierta NotebookLM, este segundo apartado busca un suelo ético firme. El 29 de enero de 2026, en la Ciudad de México, la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) y la Agencia de Transformación Digital y Telecomunicaciones (ATDT) presentaron la Declaración de ética y buenas prácticas para el uso y desarrollo de la Inteligencia Artificial, conocida como los Principios de Chapultepec (Secretaría de Ciencia, 2026).

Es justo reconocer que esta no es la única voz en el panorama nacional. Instituciones académicas de gran calado como la UNAM, el IPN o mi propia casa de estudios, la UAM –que ha impulsado esfuerzos como el Programa de Inteligencia Artificial Estratégica Social y Solidaria (PIAESS) –, han generado sus propias directrices y publicaciones para orientar a la sociedad en este cambio de época. Sin embargo, mi decisión de recuperar específicamente la propuesta de SECIHTI y la ATDT para este capítulo responde a un gusto personal y a una convicción política: me parece fundamental dialogar o cuestionar o poner a prueba una iniciativa que nace desde el gobierno federal como un esfuerzo por regular la IA bajo una visión que se expone como de bienestar social y soberanía nacional.

Los Principios de Chapultepec (Tabla 1) no son solo una lista de reglas; son una respuesta a la advertencia de que los avances tecnológicos nunca son socialmente neutros. Como señala Harari (2024), la historia de las redes de información siempre ha sido una lucha por mantener el equilibrio entre el orden y la

verdad. Si permitimos que la IA se desarrolle sin criterios éticos claros, corremos el riesgo de que el "orden algorítmico" termine por sacrificar nuestra dignidad y nuestros derechos fundamentales en favor de una eficiencia que puede –suele– no responder a nuestras realidades.

Estos diez principios se proponen como una guía para que la innovación no avance a ciegas, sino de manera coherente con la justicia social y el bien común (Gobierno de México y otros., 2026). Pero, como veremos más adelante, su cumplimiento no es automático: requiere de instituciones humanas capaces de vigilar a la técnica y, sobre todo, de investigadores dispuestos a no renunciar a su responsabilidad ética.

Tabla 1

Los 10 Principios de Chapultepec

La inteligencia artificial debe ampliar derechos, nunca reducirlos.
Toda decisión apoyada por inteligencia artificial debe tener responsables humanos.
Si una decisión no puede explicarse, no debe automatizarse.
La inteligencia artificial se gobierna mejor cuando se decide en colectivo.
La inteligencia artificial solo es valiosa si genera bienestar para todas las personas.
Antes de automatizar, hay que comprender a quién y a qué afecta.
La tecnología estratégica debe responder a las necesidades del país.
El desarrollo de la inteligencia artificial requiere fortalecer la educación y el conocimiento del país.
La inteligencia artificial no puede ser ajena a la diversidad cultural y lingüística del país.

Los datos son un bien público que debe cuidarse con responsabilidad.

Nota. Esta tabla muestra los 10 principios de Chapultepec en su redacción original; fuente: Gobierno de México y otros., (2026).

Los motivos que impulsan a la SECIHTI y a la ATDT para proponer este marco se expone que nacen de una convicción: los avances tecnológicos nunca son socialmente neutros y tienen un impacto directo en cómo interactuamos entre nosotros. Reconocen que la IA es una de las herramientas más poderosas de nuestra época, capaz de generar bienestar, pero también de profundizar desigualdades y formas de discriminación si se desarrolla sin criterios éticos claros. Por ello, estos principios surgen como una brújula para que las políticas públicas y la innovación no pierdan de vista la dignidad humana, la justicia social y el bien común.

Como bien advierte Harari (2024), la información no es sinónimo de verdad, sino una herramienta de poder que puede generar bienestar o, si se usa sin ética, profundizar desigualdades y exclusiones. Por ello, estos principios se presentan como una guía para velar por que la innovación tecnológica sea coherente con la dignidad humana y el bien común.

A grandes rasgos, los diez Principios de Chapultepec conforman un sistema de defensa integral ante la opacidad algorítmica:

- *Derechos y Responsabilidad* (Principios 1 y 2): Exigen que la IA amplíe derechos y evite prejuicios sociales. Es vital que ninguna decisión quede «en manos de la máquina»; siempre debe haber responsables humanos que supervisen y respondan por los efectos del sistema.
- *Inteligencia y Colectividad* (Principios 3 y 4): Aquí chocamos con la «caja negra»: si un proceso no puede explicarse ni reconstruirse, no debe automatizarse. Además, se establece que la IA se gobierna mejor en colectivo, promoviendo espacios de diálogo donde la academia y la sociedad civil participen para que la tecnología responda a necesidades reales y no solo a intereses técnicos.
- *Bienestar y Sostenibilidad* (Principios 5 y 6): La IA solo es valiosa si genera valor público y mejora el acceso a servicios esenciales. Esto requiere evaluar sus efectos sociales y ambientales, reconociendo que la IA también depende de infraestructuras físicas y recursos naturales.

- *Soberanía y Conocimiento* (Principios 7 y 8): La tecnología estratégica debe alinearse con una visión de país a largo plazo para fortalecer nuestra autonomía. Para lograrlo, es indispensable fortalecer la educación en todos los niveles e impulsar una investigación interdisciplinaria que integre las humanidades para comprender los impactos éticos y culturales de la IA.
- *Diversidad y Bien Público* (Principios 9 y 10): El sistema no puede ser ajeno a nuestra realidad; debe recuperar y fortalecer las lenguas indígenas para alimentar los sistemas desde nuestra propia diversidad cultural. Finalmente, los datos se reclaman como un bien público que exige calidad, seguridad y respeto total a la privacidad.

Reconocer estos diez principios es, en realidad, un acto de resistencia frente a la colonialidad del poder que, como advierte Quijano(2014), ha impuesto un «espejo eurocéntrico» –agreguemos anglosajón para este caso– que distorsiona nuestras realidades locales al juzgarlas bajo criterios universales de progreso y racionalidad. Los Principios de Chapultepec se alejan de esa visión de «universo» totalizante para abrazar la noción de pluriverso: un diseño que reconoce la interdependencia radical y la autonomía de las comunidades. Al exigir que la IA no sea ajena a la diversidad cultural y lingüística (Principio 9), la SECIHTI y la ATDT no buscan solo "eficiencia técnica" en la traducción, sino proteger la vida misma y esos conocimientos que han sido marginados por no hablar el idioma del capital.

Miremos a esta interfaz de IA y a otras no como "soluciones" dadas, sino como artefactos de significación que deben ser resignificados desde nuestra propia localidad. Siguiendo a Allende Gorgonio (2025), las evidencias –las salidas que podemos obtener en cada cuaderno– que construimos no son verdades absolutas, sino medios que solo cobran sentido cuando el «artista ventrílocuo» –en este caso, la persona que use NotebookLM– los hace hablar; idealmente desde una ética que prioriza el bienestar social sobre el orden algorítmico. Siguiendo a Mendoza Molina (2010), el reto es no permitir que la IA clausure la negociación colectiva de

significados, manteniéndonos «ontológicamente mudos» ante lo que la máquina nos dicta como real.

Un verdadero desafío, entonces, no reside únicamente en proceder según los Principios de Chapultepec, sino en la tensión que surge cuando intentamos seguirlos o apegarnos a ellos mientras interactuamos con una arquitectura como la de NotebookLM, que a menudo prioriza el flujo burocrático y el «orden digerible» sobre la complejidad y el sentido de vida de nuestras realidades situadas.

NotebookLM a la Luz de los Principios de Chapultepec

Usar NotebookLM en el día a día de la investigación o del estudio o del trabajo administrativo, etc. es, en muchos sentidos, aceptar una invitación a la comodidad intelectual, a no gastar tanta energía en la reflexión, en el pensamiento crítico. Como ya vimos, la interfaz es experta en convertir el desorden de nuestros archivos en un flujo de información que parece tener todo el sentido del mundo. Pero, cuando pasamos este ejercicio por el filtro de los Principios de Chapultepec, esa comodidad empieza a dar problemas. Mi duda aquí no es si el programa funciona bien –su eficiencia o efectividad no la pongo en duda–, sino si es ético, si la interfaz puede usarse de buena fe suponiendo que por defecto estará alineada con los Principios de Chapultepec: ¿qué tanto de este uso ayuda realmente a proteger la vida y nuestra autonomía, y qué tanto es simplemente rendirnos ante el orden que Google ya decidió por nosotros?

Si bien podría revisar cada uno de los 10 principios, me quedaré con tres de ellos en una revisión más detallada que los otros siete, al mismo tiempo que extendiendo la invitación a quien lee esto y suele usar NotebookLM a que reflexione sobre esa interacción a la luz de esos otros siete principios que aquí solo atenderé superficialmente.

¿Mundo para todos o pensamiento único? ¿Diversidad? (Principio 9)

El Principio 9 de Chapultepec dice que la IA no puede ignorar la diversidad de culturas y lenguas de nuestro país. Aquí es donde me topo con una primera dificultad que pudiera a llegar a ser una contradicción; esta, en mi caso, es bidireccional. Aunque sé que los conocimientos de los pueblos indígenas son fundamentales para entendernos, en mi práctica diaria termino cargando fuentes casi siempre en español o inglés, alguna vez en francés. Esto es un ejemplo de cómo he aprendido a ver el mundo a través de lo que el discurso hegemónico considera válido (Quijano, 2014).

Si uso NotebookLM siguiendo esta inercia, estoy ayudando a que la herramienta borre las diferencias. Al no buscar activamente esas voces en lenguas originarias, el sistema entra en conflicto y deja de ser eficiente y efectivo en lo que suele serlo, a veces ni siquiera tiene conocimientos sobre ciertas lenguas originarias, y lo que no aparece en las redes de información hoy en día, corre el riesgo de desaparecer de nuestra realidad (Harari, 2024). Pero incluso si las buscara, me queda una duda: ¿puede una máquina diseñada para ser eficiente a nivel global traducir realmente el sentimiento y la vida que hay detrás de palabras que nacieron de una realidad tan distinta a la de un laboratorio en Silicon Valley? NotebookLM podría actuar como un "traductor traidor" que, con tal de darme una respuesta clara y digerible, termine sacrificando la verdad profunda de esas culturas (Harari, 2024); esto considerando que pudiera traducir al español o alguno de los idiomas que puedo entender aquellos textos o videos o audios en lenguas originarias.

Ahora, la tensión no se queda solo en la parte del idioma de las Fuentes, permea a la producción de los contenidos de salida en la sección de Studio y, por supuesto, también al apartado de Chat. Las estructuras gramaticales no son las mismas en español que en inglés o alemán, tampoco son similares al chino, al tailandés o al japonés. Esto que parece algo sin importancia puede llegar a marcar diferencias significativas porque son formas diferentes de pensar, de presentar los saberes. No negaré que hay esfuerzos por recuperar algunas lenguas originarias

en los modelos de lenguaje de Google, este es el caso del náhuatl, el maya yucateco, el zapoteco o el q'eqchi, que ya aparecen en Google Traductor y que se han ido perfeccionando desde su aparición en el 2024.

Nuestra autonomía frente al "socio" extranjero (Principio 7)

Otro punto de choque es el Principio 7, que pide que la tecnología responda a lo que el país necesita. Es muy fácil decir que usar IA de última generación es "progreso". Pero hay que preguntarnos si este diseño realmente nos ayuda a ser más independientes o si nos encierra en una forma de pensar donde solo hay una manera correcta de hacer las cosas (Escobar, 2018).

NotebookLM no es una herramienta nuestra –latinoamericana, mexicana, local–; es un aparato artificial con reglas internas dictadas por una empresa privada –en este caso, Google– (Simon, 2006). Al usarlo para organizar mis ideas, mis datos –que son parte mi materia prima intelectual– pueden alimentar o ayudar a entrenar a un sistema que no controlo. Pudiera ser que, bajo la idea de que estamos siendo más modernos en la investigación, solo estemos haciendo más fuerte nuestra dependencia tecnológica (Harari, 2024). Cuando el orden de nuestro pensamiento queda en manos de una inteligencia ajena, corremos el peligro de volvernos una especie de "colonia digital" que solo consume lo que se produce en los centros de poder (Harari, 2024).

Cualquiera puede construir evidencias que le permitan sustentar el desarrollo o uso de la IA según las necesidades del país (Allende Gorgonio, 2025), por ello, quizá habría que cuidar más la redacción de este Principio 7 y no confiar en el actuar de buena fe que, históricamente, hay quienes han abusado de la “interpretación” o de la supuesta “objetividad” que suele atribuirse a lo que está en el área del conocimiento científico.

La honestidad de trabajar juntos y el miedo al juicio ajeno (Principio 3)

Por último, abordaré el Principio 3, el cual señala que, si no podemos explicar cómo se tomó una decisión, no deberíamos dejar que una máquina la tome sola.

En este capítulo, desde el inicio he sido transparente con el uso de la IA para la redacción. Sin embargo, me doy cuenta de que decir la verdad no quita el prejuicio de los demás. En la academia –y en otros espacios– hay una resistencia natural a aceptar que la IA puede ser un compañero de pensamiento; muchos verán esta unión como si yo estuviera renunciando a mi capacidad de pensar por mí misma. Yo no lo veo así. Para mí, es la oportunidad de poder encontrar una interpelación de parte de un agente no-humano, disponible en momentos en los que en ocasiones se complica la comunicación con otros seres humanos; además de que este agente puede reflejar un nivel de profundidad en el procesamiento de las fuentes cargadas que no todos los seres humanos tienen desarrollado.

Para mí, esta sociedad o interacción humano-IA no ha sido algo automático, aunque quizá sí podría afirmar que no siempre ha sido consciente por parte de todos aquellos que se han visto involucrados. Ha sido, más bien, un esfuerzo constante por no perder mi autonomía. Me obliga a leer con más cuidado y a dudar de por qué la respuesta favoreció un dato y no otro. Pero ¿cómo le explico esto a quienes ven en la IA solo una "caja negra" que viene a reemplazarnos? Aquí, explicar el proceso no es solo mostrar una lista de citas, sino intentar demostrar que yo sigo siendo la responsable de cada palabra que el sistema propone, de esta interacción humano-IA donde esta tiene un rol de agente asistente (Ramírez Navarro, 2025). Si no puedo exponer cómo llegamos a estas ideas, podría estar cayendo en un tipo de traición a quien se toma el tiempo de leer este capítulo y, alimentando la idea de que la IA es una autoridad a la que no se le puede cuestionar (Harari, 2024).

Breve recorrido por tensiones éticas en los otros Principios

Evidentemente el resto de los Principios de Chapultepec también pueden generar preguntas incómodas cuando abrimos un cuaderno de trabajo en NotebookLM: Como mencioné anteriormente, espero que quien lee haga una reflexión más profunda al respecto, por ahora solo dejaré inquietudes de manera superficial.

Con respecto al Principio 1, la interfaz debería evitar reproducir prejuicios sociales. Sin embargo, al basarse en Gemini, NotebookLM carga con los sesgos de sus datos de entrenamiento, que a menudo subestiman la riqueza de lo que queda fuera de las corrientes principales de información.

El Principio 2 exige que las decisiones no queden «en manos de la máquina». La tentación en NotebookLM es dejar que sus funciones de Studio decidan por nosotros qué es lo relevante de un tema, lo cual es una forma sutil de renunciar a nuestra responsabilidad como investigadores. Esto con todo y que algunos contenidos de salida puedan configurarse para, hasta cierto punto, personalizar el resultado.

En el Principio 4 se hace referencia a que la IA se gobierna mejor si participamos todos. Mientras Google mantiene un diseño cerrado y propietario, es interesante ver cómo la comunidad de usuarios intenta "recuperar" cierto control creando extensiones como NotebookLM Tools o FolderLLM para organizar el espacio según nuestras propias necesidades y no solo las del diseñador. Que no evita que, en alguna actualización, lo que ahora son extensiones puedan estar ya integradas en la interfaz que se ofrece.

Con respecto al Principio 5, la tecnología debe generar valor público y no solo ser eficiente. Aquí la tensión es clara: ¿NotebookLM ayuda al usuario a una producción con sentido local, que verdaderamente genere ese valor público o simplemente me favorece la producción de textos más rápido a costa de perder profundidad o sentido situado?

Para el Principio 6, antes de automatizar hay que evaluar, por ejemplo, el impacto ambiental. Casi nunca pensamos en el consumo de energía y agua que requieren los centros de datos de Google cada vez que le pedimos a la IA que procese nuestros archivos.

Con el Principio 8 se busca fortalecer el conocimiento y las humanidades. El riesgo con las tarjetas de estudio o los cuestionarios automáticos es caer en una "educación bancaria" donde solo memorizamos datos seleccionados por un algoritmo, en lugar de aprender a cuestionar la realidad.

Finalmente, según el Principio 10, los datos son un bien público que exige privacidad. Al usar un servicio "gratuito" o incluso las versiones de paga, entramos en un modelo de negocio donde nuestro rastro intelectual puede llegar a ser la materia prima que alimenta a los gigantes tecnológicos, a menudo sin que tengamos claro cómo se usa esa información.

La sociedad, el ser humano en su interacción con NotebookLM no está ante un fenómeno de paz técnica, sino en un sitio de resistencia ética permanente. No podemos esperar que la herramienta o la interfaz sea "buena" o "justa" por diseño, porque su arquitectura responde a lógicas de orden y mercado que muchas veces son ajenas a nuestras necesidades locales.

El cumplimiento de estos principios no se va a lograr instalando una actualización de software a esta o cualquier otra interfaz. Es una conquista diaria que depende de la capacidad de las personas para seguir siendo las "objektoras" principales de lo que el sistema me propone. Si nos dejamos llevar por la comodidad de sus resúmenes y su orden "perfecto", se estará permitiendo que un "burócrata digital" dicte el rumbo del pensamiento de quien usa la interfaz. La verdadera co-agencia reflexiva nace de la fricción, de ese ruido que, por ejemplo, obliga a leer con más cuidado y a no aceptar como verdad lo que es solo un flujo algorítmico bien organizado. . Al final, los Principios de Chapultepec pueden ser una guía, una herramienta para despertar, para recordarnos tener presente que en la era de la IA, nuestra mayor tecnología sigue siendo nuestra propia mirada crítica.

Reflexiones Finales

Escribir este capítulo no ha sido un ejercicio de descripción técnica, sino una travesía por las tensiones de nuestra época. Al terminar, la pregunta de si NotebookLM puede ser un socio científico no tiene una respuesta cerrada; más bien, se revela como un artefacto cuyas posibilidades exceden incluso lo que Google Labs imaginó al diseñarlo. He vivido que la sociedad con la IA no es un

contrato ético que se firma al inicio, sino una conquista diaria que se sostiene en la fatiga de la mirada crítica.

Uno de los hallazgos que más me ha impactado en esta colaboración humano-IA ha sido la capacidad de NotebookLM para actuar como un mediador que transforma mi propia percepción de los textos (Latour, 2008). No se limita a resumir; en su lógica algorítmica, ha puesto énfasis en fuentes que a mí no me habían parecido tan potentes –me refiero a que quizá para mí había sido muy importante alguna propuesta del autor o autora y había puesto énfasis en eso, sin embargo, la IA me propone hilar desde otra perspectiva que para mí no había sido potente por la manera en la que yo la había interpretado–, obligándome a "re-mirar" mi propio material desde una perspectiva que no había considerado. En este sentido, NotebookLM funciona como un espejo aumentado –una herramienta– que me devuelve un retrato contraintuitivo de mi investigación, permitiéndome ver lo que mi subjetividad situada había pasado por alto (Ramírez Navarro, 2025).

Esta potencia se ve reforzada por su arquitectura de "fuentes cerradas". Aunque este diseño limita la posibilidad de que el sistema "alucine" con datos externos, no lo evita del todo (Harari, 2024). Lo que sí podría decir que genera o que puede generar es una presión saludable sobre el rol de quien usa la interfaz: por ejemplo, a ser más rigurosos o quisquillosos en la selección de las fuentes que se le cargan al cuaderno, reconociendo que la calidad del "socio" depende enteramente de la materia prima informativa que le proporcione quien la use. NotebookLM es, como diría Simon (2006), una interfaz perfecta para lidiar con nuestra racionalidad limitada, pues nos obliga a interactuar con una cantidad finita de información para encontrar soluciones "satisfacientes" ante la saturación actual.

Siendo honesta: si sometiéramos a NotebookLM a algún tipo de auditoría estricta, el sistema, por sí solo, no pasaría la prueba de los Principios de Chapultepec (Gobierno de México y otros., 2026). Por defecto, su lenguaje, sus opciones de salida en Studio y su opacidad como "caja negra" corporativa ignoran las exigencias de soberanía, diversidad lingüística y explicabilidad que nuestro país

demanda. Sin embargo, la maravilla de los artefactos es que su uso no está dictado totalmente por el creador.

NotebookLM –así como otras IA– es un medio que puede ser usado siguiendo estos principios, pero eso sitúa la responsabilidad ética total en la persona usuaria. El cumplimiento de la ética no está en el código de Google, sino en la voluntad de cada usuaria o usuario de no ser una o un "sujeto mudo" ante lo que la máquina dicta (Mendoza, 2010). Esta interacción es intrínsecamente tensionante: por un lado, la IA nos puede ahorrar tiempo en tareas de redacción o síntesis; por el otro, nos puede exigir una enorme cantidad de energía para mantenernos en modo de "reflexión crítica" y no aceptar como verdad lo que es solo un flujo algorítmico bien organizado (Harari, 2024).

En conclusión, NotebookLM puede entenderse como una herramienta extraordinariamente potente que puede ayudarnos más que muchos otros chatbots genéricos vigentes, precisamente porque nos obliga a centrarnos en nuestras fuentes. Pero su uso crítico no se da por defecto. Existe el riesgo permanente de caer en una nueva educación bancaria, donde el ahorro de tiempo nos desincentive a cuestionar el orden que el sistema nos regala (Freire, 1993).

La viabilidad de esta "sociedad" depende de que reconozcamos que las capacidades cognitivas ya no residen solo en nosotros, sino que están distribuidas en este entorno formateado (Latour, 2008). Dentro de las responsabilidades de los usuarios, está vigilar que ese entorno no borre nuestra diversidad ni nuestra autonomía. Los Principios de Chapultepec, entonces, no son solo reglas para las máquinas, sino un llamado para que nosotros, los investigadores, los alumnos, los trabajadores o las personas que usen esta interfaz, no renunciemos a la soberanía de nuestro propio pensamiento. NotebookLM puede ser un gran socio –IA como herramienta o IA como agente asistente–, pero tendremos que estar dispuestos a pagar el precio de la sospecha permanente o aceptar su uso acrítico, que podrá incluir respuestas incorrectas, que incluso atenten contra la vida, que en vez de ayuda sea contraproducente su uso.

Referencias

- Allende Gorgonio, D. V. (2025). La Construcción de Evidencias Organizacionales como Estrategia de la Intervención Organizacional para el Cambio. En *Reflexiones sobre la Intervención en las Organizaciones en la UAM* (pp. 88–106). REMINEO.
- Escobar, A. (2018). Introduction. En *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds* (p. ix-xxi 1-21). Duke University Press.
- Freire, P. (1993). *Pedagogía del Oprimido* (44a ed.). Siglo Veintiuno Editores.
- Gartenberg, C. (2025). *La historia interna de la creación de NotebookLM*.
https://blog-google.translate.goog/innovation-and-ai/products/developing-notebooklm/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc
- Gobierno de México, Secretaría de Ciencia, Humanidades, T. e I., & Telecomunicaciones, A. de T. D. y. (2026). *Principios de Chapultepec. Declaración de ética y buenas prácticas para el uso y desarrollo de la Inteligencia Artificial*. Gobierno de México.
- Google. (2026). *Google NotebookLM | AI Research Tool & Thinking Partner*.
<https://notebooklm.google/>
- Harari, Y. N. (2024). *Nexus. Una breve historia de las redes de información desde la Edad de Piedra hasta la IA*. Penguin Random House.
- Lasalle, J. M. (2024). *Civilización Artificial. Sabiduría o sustitución: el dilema humano ante la IA*. Arpa.
- Latour, B. (2008). *Reensamblar lo Social. Una introducción a la Teoría del Actor-Red*. (1a ed.). Manantial.
- Mendoza, R. (2010). Construccinismo Social y Estudios Organizacionales. *Análisis Organizacional*, 1(Núm. 2), 4–32.
- NotebookLM Tools. (2026). *Las herramientas esenciales para NotebookLM*.
<https://www.nlmtools.com/es>
- Quijano, A. (2014). Colonialidad del Poder, Eurocentrismo y América Latica. En CLACSO (Ed.), *Cuestiones y horizontes: de la dependencia histórico-*

estructural a la colonialidad/descolonialidad del poder. (pp. 777–832).

Ramírez Navarro, E. (2025). La Praxis Interventora en las Organizaciones en la Era de la Co-Agencia: Diálogos Ontoepistemológicos entre Humanos e Inteligencia Artificial. En O. Lozano Carrillo, Y. Zavala Osorio, & P. Couturier Bañuelos (Eds.), *Reflexiones sobre la Intervención en las Organizaciones en la UAM* (Primera ed, pp. 69–87). Red Mexicana de Investigadores en Estudios Organizacionales (REMINEO).

Secretaría de Ciencia. (2026). *Foro “La Inteligencia Artificial en la vida pública de México”* - YouTube (p. 05h02m50s). YouTube.

<https://www.youtube.com/watch?v=ix7k3FSqy30>

Simon, H. A. (2006). *Las Ciencias de lo Artificial*. Granada.

GOBERNANZA PEDAGÓGICA DE LA INTELIGENCIA ARTIFICIAL EN EDUCACIÓN SUPERIOR: MODELO METODOLÓGICO, ANÁLISIS COMPARADO Y APLICACIÓN EN EL PROGRAMA PIAESS

Pedagogical governance of artificial intelligence in higher education:
methodological model, comparative analysis and application in the PIAESS
program

Leslie Bridshaw Araya

leslie.bridshaw@uv.cl

Universidad de Valparaíso, Chile

<https://orcid.org/0000-0001-6975-4028>

Resumen

La incorporación de la inteligencia artificial generativa en la educación superior ha introducido tensiones significativas en los dispositivos tradicionales de evaluación, en las nociones de autoría y en los principios de integridad académica. En este escenario, las respuestas institucionales centradas exclusivamente en la restricción o prohibición resultan conceptualmente insuficientes y operativamente ineficaces. Se requiere, en cambio, el desarrollo de marcos de gobernanza pedagógica capaces de articular coherentemente principios normativos, prácticas docentes y capacidades institucionales.

El presente capítulo propone un modelo de gobernanza pedagógica de la inteligencia artificial sustentado en la adaptación metodológica, integrando un análisis teórico, evidencia comparada internacional —con énfasis en experiencias de México, España y Chile— y una propuesta aplicada al Programa de Inteligencia Artificial Estratégica Social y Solidaria (PIAESS). Metodológicamente, se adopta un diseño cualitativo de carácter analítico-comparado, basado en revisión documental sistemática y modelamiento conceptual.

Los resultados configuran un sistema de gobernanza estructurado en cinco niveles —institucional, curricular, metodológico, evaluativo y tecnológico— que posibilita la integración de herramientas de IA sin requerir transformaciones curriculares estructurales. Este enfoque permite reorientar las prácticas pedagógicas hacia el desarrollo de competencias profesionales en entornos digitalmente mediados, resguardando simultáneamente la integridad académica. Asimismo, se discute su potencial de escalabilidad y replicabilidad en contextos latinoamericanos.

Para ello, el capítulo se organiza en torno a tres dimensiones analíticas: (i) fundamentos teóricos de la gobernanza pedagógica de la IA, (ii) análisis comparado de experiencias internacionales y (iii) aplicación del modelo en el programa PIAESS, incorporando como caso ilustrativo la ruta institucional de la Universidad de Valparaíso.

Palabras clave: inteligencia artificial; educación superior; gobernanza pedagógica; evaluación; desarrollo de competencias; innovación educativa.

Abstract

The incorporation of generative artificial intelligence into higher education has introduced significant tensions within traditional assessment mechanisms, notions of authorship, and principles of academic integrity. In this context, institutional responses focused exclusively on restriction or prohibition prove to be conceptually insufficient and operationally ineffective. Instead, there is a need for the development of pedagogical governance frameworks capable of coherently articulating normative principles, teaching practices, and institutional capacities.

This chapter proposes a model of pedagogical governance for artificial intelligence grounded in methodological adaptation, integrating theoretical analysis, comparative international evidence — with emphasis on experiences from Mexico, Spain, and Chile — and an applied proposal for the Social and Solidary Strategic Artificial Intelligence Program (PIAESS). Methodologically, the study adopts a

qualitative analytical-comparative design based on systematic documentary review and conceptual modeling.

The results outline a governance system structured across five levels — institutional, curricular, methodological, evaluative, and technological — enabling the integration of AI tools without requiring structural curricular transformations. This approach makes it possible to redirect pedagogical practices toward the development of professional competencies in digitally mediated environments while simultaneously safeguarding academic integrity. Likewise, the chapter discusses its potential for scalability and replicability within Latin American contexts.

To this end, the chapter is organized around three analytical dimensions: (i) theoretical foundations of pedagogical governance of AI, (ii) comparative analysis of international experiences, and (iii) application of the model within the PIAESS program, incorporating the institutional pathway of the University of Valparaíso as an illustrative case.

Keywords: artificial intelligence; higher education; pedagogical governance; assessment; competency development; educational innovation.

2. Marco teórico

La incorporación de la inteligencia artificial (IA) generativa en la educación superior configura un campo de problematización que trasciende lo instrumental, tensionando dimensiones epistemológicas, pedagógicas y éticas del proceso formativo. La capacidad de estos sistemas para producir contenido, resolver tareas complejas y simular procesos de razonamiento desafía las nociones clásicas de autoría, originalidad y evaluación del aprendizaje, así como los supuestos sobre la construcción individual del conocimiento. En este contexto, la discusión contemporánea se desplaza desde enfoques centrados en la regulación restrictiva hacia perspectivas que enfatizan la necesidad de marcos de gobernanza pedagógica, entendidos como sistemas integrados de normas, prácticas,

capacidades institucionales e infraestructuras tecnológicas orientadas a alinear el uso de la IA con los fines formativos de la educación superior (UNESCO, 2023; Selwyn, 2019).

Desde una perspectiva latinoamericana, esta discusión se inscribe en un debate más amplio sobre el sentido de la universidad en contextos de transformación digital. Autores como Didriksson (2016) y Rama (2020) han advertido que los procesos de digitalización no solo implican cambios tecnológicos, sino también reconfiguraciones en las funciones sustantivas de la universidad —docencia, investigación y vinculación—, lo que exige marcos de gobernanza capaces de preservar su rol crítico, público y formativo. En este sentido, la IA no constituye únicamente una herramienta, sino un vector de transformación institucional que demanda respuestas sistémicas y no fragmentarias.

2.1 IA como artefacto cognitivo ampliado

La conceptualización de la IA como artefacto cognitivo encuentra sustento en la teoría de la mente extendida propuesta por Clark y Chalmers (1998), la cual plantea que los procesos cognitivos no se circunscriben al individuo, sino que se distribuyen en sistemas híbridos que integran herramientas externas. Desde esta perspectiva, la IA generativa puede ser comprendida como un dispositivo de externalización cognitiva que amplifica capacidades humanas tales como la memoria, el análisis y la producción simbólica.

Este enfoque supone un desplazamiento desde modelos centrados en la internalización del conocimiento hacia esquemas de cognición distribuida, donde el aprendizaje emerge de la interacción entre sujetos, artefactos y contextos. En el ámbito latinoamericano, esta idea dialoga con los planteamientos de García Canclini (2004), quien destaca la hibridación cultural y tecnológica como rasgo constitutivo de las sociedades contemporáneas, así como con las reflexiones de Jesús Martín-Barbero (2003) sobre mediaciones tecnológicas y producción de sentido. En este marco, la IA no solo amplía la cognición, sino que reconfigura las condiciones de producción del conocimiento.

2.2 Aprendizaje mediado y profesional reflexivo

Desde la tradición sociocultural, el aprendizaje se concibe como un proceso mediado por herramientas y prácticas sociales (Vygotsky, 1978). La IA puede operar, en este sentido, como un mediador avanzado dentro de la Zona de Desarrollo Próximo, facilitando procesos de andamiaje cognitivo siempre que su uso esté pedagógicamente orientado. Sin dicha mediación docente, existe el riesgo de que la tecnología sustituya —en lugar de potenciar— los procesos de aprendizaje significativo.

Complementariamente, el concepto de profesional reflexivo desarrollado por Schön (1983) adquiere particular relevancia en entornos mediados por IA. La capacidad de “pensar en la acción” y “sobre la acción” se vuelve crítica cuando los sujetos interactúan con sistemas que generan respuestas aparentemente completas, pero que requieren interpretación, validación y contextualización. En América Latina, Freire (1970) aporta una dimensión clave al enfatizar la conciencia crítica como núcleo del proceso educativo, lo que permite problematizar el uso de la IA no como una solución técnica, sino como un objeto de reflexión pedagógica y política.

2.3 Competencias en la era de la automatización

La expansión de la automatización y la inteligencia artificial ha reconfigurado la estructura del valor en el trabajo y, por extensión, en la formación profesional. De acuerdo con Brynjolfsson y McAfee (2014), las tecnologías digitales tienden a sustituir tareas rutinarias, desplazando la demanda hacia habilidades de orden superior. En este contexto, la educación superior enfrenta el desafío de formar competencias que no sean fácilmente replicables por sistemas automatizados.

Entre estas competencias destacan el pensamiento crítico (Paul & Elder, 2006), la autorregulación del aprendizaje (Zimmerman, 2002), la alfabetización digital crítica (Selwyn, 2019) y la ética de la responsabilidad tecnológica (Jonas, 1984). En el contexto latinoamericano, Brunner (2011) ha señalado que la expansión de la educación superior requiere una actualización de los perfiles de egreso,

incorporando capacidades para desenvolverse en entornos complejos, inciertos y altamente tecnologizados.

2.4 Gobernanza de la IA en educación superior

La noción de gobernanza aplicada a la educación superior remite a la articulación de múltiples dimensiones —normativa, pedagógica, tecnológica y organizacional— que permiten orientar los procesos institucionales en contextos de cambio. En el caso de la IA, esta gobernanza debe integrar tanto criterios regulatorios como principios pedagógicos, evitando reduccionismos tecnocráticos.

En América Latina, Didriksson (2016) y Rama (2020) coinciden en la necesidad de modelos de gobernanza flexibles y adaptativos, capaces de responder a dinámicas de innovación sin comprometer la función crítica de la universidad. Asimismo, organismos internacionales como la UNESCO (2023) han enfatizado la importancia de desarrollar marcos éticos y pedagógicos para el uso de la IA, centrados en la equidad, la transparencia y la responsabilidad.

Desde esta perspectiva, la gobernanza pedagógica de la IA no se limita a regular su uso, sino que implica diseñar condiciones institucionales que permitan su integración significativa en los procesos de enseñanza y aprendizaje. Esto supone reconocer la IA como un elemento estructural del ecosistema educativo contemporáneo y, por tanto, como un objeto legítimo de intervención pedagógica y política.

3. Metodología

El estudio adopta un diseño cualitativo de tipo analítico-comparado, basado en:

1. Revisión documental
 - literatura internacional sobre IA y educación,
 - marcos normativos institucionales,

- informes de organismos internacionales.

2. Análisis comparado

- experiencias en México (enfoque normativo y ético),
- experiencias en España (enfoque sistémico e institucional).
- guía uso IA de Universidad de Valparaíso (modelo de gobernanza pedagógica incipiente con base normativa–ética y proyección sistémico-institucional).

3. Modelamiento conceptual

- construcción de un modelo de gobernanza pedagógica,
- adaptación al caso PIAESS como escenario de implementación.

El enfoque metodológico es de carácter interpretativo, orientado a generar un marco aplicable más que a validar hipótesis causales.

4. Análisis comparado: marcos institucionales y prácticas de gobernanza de la IA en México, España y la Universidad de Valparaíso

La respuesta institucional al uso de inteligencia artificial generativa en educación superior ha evolucionado desde enfoques centrados en la prohibición hacia modelos progresivamente más articulados de regulación. Esta evolución no ha sido homogénea y presenta diferencias significativas según contexto, tradición institucional y capacidad de innovación pedagógica. El análisis de experiencias en México y España, junto con el caso de la Universidad de Valparaíso, permite identificar niveles diferenciados de desarrollo en la gobernanza de la IA.

4.1 México: normatividad emergente y énfasis ético-formativo

En el contexto mexicano, la incorporación de la IA en educación superior ha sido impulsada principalmente por universidades de referencia, aunque sin una política

nacional homogénea. Instituciones como la Universidad Nacional Autónoma de México (UNAM) y el Tecnológico de Monterrey han desarrollado lineamientos orientados a promover el uso responsable de la IA, enfatizando principios de integridad académica, ética digital y alfabetización tecnológica.

La UNAM, a través de su Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC), ha difundido orientaciones sobre el uso de herramientas generativas, centradas en la protección de datos, la transparencia y la responsabilidad en la producción académica. Por su parte, el Tecnológico de Monterrey ha promovido activamente la integración de IA en el aprendizaje, desarrollando guías institucionales y programas de formación docente orientados a su uso pedagógico.

No obstante, estos avances se concentran principalmente en el plano normativo y formativo. Como advierte Didriksson, uno de los desafíos estructurales en América Latina radica en la dificultad de traducir las innovaciones discursivas en transformaciones pedagógicas concretas (Didriksson, 2016). En la misma línea, Ornelas sostiene que las políticas educativas tienden a perder efectividad cuando no se articulan con las prácticas de aula, especialmente en lo relativo a evaluación (Ornelas, 2018).

En consecuencia, la evidencia sugiere que, en el caso mexicano, la IA ha sido incorporada principalmente como un objeto de regulación ética, con menor desarrollo en la dimensión metodológica y evaluativa.

4.2 España: institucionalización progresiva e integración sistémica

El caso español presenta un nivel más avanzado de desarrollo, caracterizado por una progresiva institucionalización del uso de IA en el sistema universitario. Diversas universidades han desarrollado marcos integrales que no solo regulan el uso de herramientas generativas, sino que también promueven su integración en el diseño pedagógico y evaluativo.

La Universitat de Barcelona ha emitido orientaciones que recomiendan explícitamente adaptar los sistemas de evaluación para incorporar la trazabilidad del proceso, el análisis crítico y la validación de fuentes. De manera similar, la Universidad Autónoma de Madrid ha desarrollado guías docentes que integran el uso de IA en la planificación de actividades y evaluaciones.

Por su parte, la Universidad de Salamanca ha avanzado en la formalización de lineamientos institucionales que establecen la obligación de declarar el uso de IA, justificar su pertinencia pedagógica y evaluar su integración como parte del proceso formativo.

Estos desarrollos reflejan un tránsito desde la regulación hacia la integración, en línea con lo señalado por Selwyn (2019), quien advierte que el impacto de las tecnologías en educación depende de su articulación con las prácticas pedagógicas. Asimismo, UNESCO (2023) ha enfatizado la necesidad de avanzar hacia modelos de gobernanza que integren dimensiones éticas, pedagógicas y tecnológicas.

4.3 Universidad de Valparaíso: operacionalización metodológica de la gobernanza

En este escenario, el caso de la Universidad de Valparaíso adquiere particular relevancia, en tanto representa una estrategia de implementación situada en el nivel metodológico, que permite traducir principios normativos en prácticas pedagógicas concretas.

La ruta orientativa de uso de IA desarrollada por la institución establece un conjunto de dispositivos que operan directamente en el aula:

- declaración explícita del uso de IA en los programas de asignatura,
- definición de usos permitidos, restringidos y prohibidos,
- acuerdos pedagógicos con estudiantes,
- citación obligatoria de herramientas utilizadas,

- prohibición de delegar la autoría intelectual,
- resguardo de datos personales.

A diferencia de otros enfoques centrados exclusivamente en la normativa, esta propuesta introduce mecanismos que inciden directamente en el proceso de enseñanza-aprendizaje. Desde una perspectiva analítica, estos dispositivos cumplen funciones específicas dentro de la gobernanza pedagógica:

- transparencia, al visibilizar el uso de IA;
- trazabilidad, al permitir reconstruir el proceso de producción del conocimiento;
- formación, al transformar la regulación en un instrumento pedagógico;
- anticipación evaluativa, al generar condiciones para el rediseño de evaluaciones.

Un elemento clave es que esta implementación no requiere modificaciones curriculares estructurales, lo que permite su adopción inmediata y transversal. En este sentido, la Universidad de Valparaíso representa un caso de gobernanza metodológica, donde la regulación se materializa en prácticas docentes concretas.

4.4 Discusión comparada: niveles de gobernanza y trayectorias de desarrollo

El análisis de los tres contextos permite identificar distintos niveles de desarrollo en la gobernanza de la IA en educación superior:

Tabla 1

Niveles de gobernanza de la inteligencia artificial en educación superior: síntesis comparada

Nivel	Características principales	Ejemplo predominante
Normativo	Enfoque en ética, principios y uso responsable	México
Metodológico	Traducción de normas en prácticas pedagógicas	Universidad de Valparaíso
Sistémico	Integración institucional completa (políticas, evaluación, formación)	España

Nota. Elaboración propia a partir de Axel Didriksson (2016), Carlos Ornelas (2018), Claudio Rama (2020) y documentos institucionales de universidades europeas (2022–2023).

Este esquema permite situar la experiencia de la Universidad de Valparaíso en un nivel intermedio, con una ventaja estratégica significativa: su capacidad de implementación rápida mediante ajustes metodológicos, sin necesidad de rediseño curricular.

Al mismo tiempo, el análisis evidencia que la consolidación de un modelo robusto de gobernanza requiere avanzar hacia una mayor articulación institucional, particularmente en las dimensiones evaluativa y tecnológica.

4.5 Implicancias para el modelo de gobernanza propuesto

La evidencia comparada permite sostener que la efectividad de la regulación del uso de IA no depende exclusivamente de la existencia de normas, sino de su capacidad para integrarse en las prácticas pedagógicas.

En este sentido:

- el caso mexicano evidencia los límites de una regulación centrada en principios,
- el caso español muestra el potencial de una integración sistémica,

- el caso de la Universidad de Valparaíso demuestra la viabilidad de una implementación metodológica como punto de partida.

Esta triangulación refuerza la pertinencia de un modelo de gobernanza progresivo, en el que la regulación evoluciona desde el nivel normativo hacia el metodológico y, eventualmente, hacia el sistémico, proceso que puede ser catalizado mediante programas como PIAESS.

5. Propuesta: modelo de gobernanza pedagógica de la inteligencia artificial en educación superior

La evidencia comparada, junto con el análisis del caso de la Universidad de Valparaíso, permite sostener que la regulación efectiva del uso de inteligencia artificial en educación superior no se agota en la definición de normas, sino que requiere su institucionalización como un sistema de gobernanza pedagógica. Este sistema debe articular principios, dispositivos metodológicos, capacidades docentes e infraestructura tecnológica en coherencia con los fines formativos de la universidad.

En este marco, se propone un modelo de gobernanza pedagógica de la IA estructurado en cinco niveles interdependientes, acompañado de un proceso de implementación progresivo y de criterios de evaluación orientados a su validación empírica, con especial énfasis en su aplicación en el programa PIAESS.

5.1 Fundamentos conceptuales del modelo

El modelo se sustenta en cuatro tradiciones teóricas convergentes que permiten comprender la integración de la IA desde una perspectiva pedagógica y no meramente instrumental.

En primer lugar, la cognición distribuida y la teoría de los artefactos cognitivos permiten conceptualizar la IA como una extensión de los procesos cognitivos

humanos (Clark & Chalmers, 1998), desplazando el foco desde la producción de conocimiento hacia su validación, interpretación y uso.

En segundo lugar, el enfoque sociocultural del aprendizaje plantea que las herramientas tecnológicas amplían las capacidades del estudiante, pero requieren mediación pedagógica para evitar procesos de externalización pasiva del pensamiento (Vygotsky, 1978).

En tercer lugar, la noción de práctica reflexiva profesional enfatiza que el aprendizaje relevante ocurre en la reflexión sobre la acción, especialmente en contextos complejos donde intervienen tecnologías avanzadas (Schön, 1983).

Finalmente, desde la perspectiva de la gobernanza universitaria en contextos digitales, autores latinoamericanos han destacado la necesidad de modelos flexibles que articulen innovación, calidad y equidad (Didriksson, 2016; Rama, 2020).

En conjunto, estos enfoques convergen en una idea central: la inteligencia artificial no debe ser únicamente regulada, sino integrada pedagógicamente mediante dispositivos que hagan visible el proceso de aprendizaje.

5.2 Estructura del modelo de gobernanza

El modelo se organiza en cinco niveles articulados que permiten operacionalizar la gobernanza en distintos planos de la institución.

a) Nivel institucional: principios y marco normativo

Este nivel define los principios rectores que orientan el uso de IA, entre los que destacan la integridad académica, la transparencia, la responsabilidad, la equidad en el acceso y la protección de datos.

Siguiendo las orientaciones de UNESCO (2023), estos principios deben traducirse en políticas institucionales claras, pero suficientemente flexibles para adaptarse a distintas disciplinas y contextos formativos.

b) Nivel curricular: alineación sin rediseño estructural

A diferencia de enfoques que proponen reformas curriculares profundas, el modelo plantea una alineación funcional con los perfiles de egreso existentes.

En términos de coherencia constructiva (Biggs, 1996), se busca asegurar la articulación entre resultados de aprendizaje, actividades formativas y evaluación. En este contexto, la IA no introduce nuevas competencias formales, sino que intensifica el desarrollo de competencias previamente definidas, particularmente el pensamiento crítico y la autorregulación.

c) Nivel metodológico: traducción en prácticas docentes

Este nivel constituye el núcleo operativo del modelo, en tanto la gobernanza se materializa en prácticas pedagógicas concretas.

Entre los dispositivos principales se incluyen la declaración obligatoria del uso de IA, la definición de acuerdos pedagógicos explícitos, la integración de la IA como herramienta de apoyo en análisis y simulación, y el diseño de actividades que requieran interpretación y contextualización.

El caso de la Universidad de Valparaíso ilustra este nivel, al traducir principios normativos en prácticas de aula, generando condiciones para una evaluación más auténtica, centrada en el desempeño y no en la reproducción (Wiggins, 1998).

d) Nivel evaluativo: rediseño de la evidencia de aprendizaje

El nivel evaluativo resulta crítico para la efectividad del modelo, ya que la IA obliga a replantear qué se evalúa y cómo se evalúa.

Se proponen cuatro dispositivos principales:

- trazabilidad del proceso mediante bitácoras de uso de IA,
- evaluación comparativa entre producción asistida y no asistida,
- defensa oral para validar la comprensión del estudiante,

- evaluación del juicio crítico, especialmente en la detección de errores o sesgos.

Estos dispositivos se alinean con enfoques de evaluación auténtica y con la necesidad de evaluar competencias de orden superior (Paul & Elder, 2006).

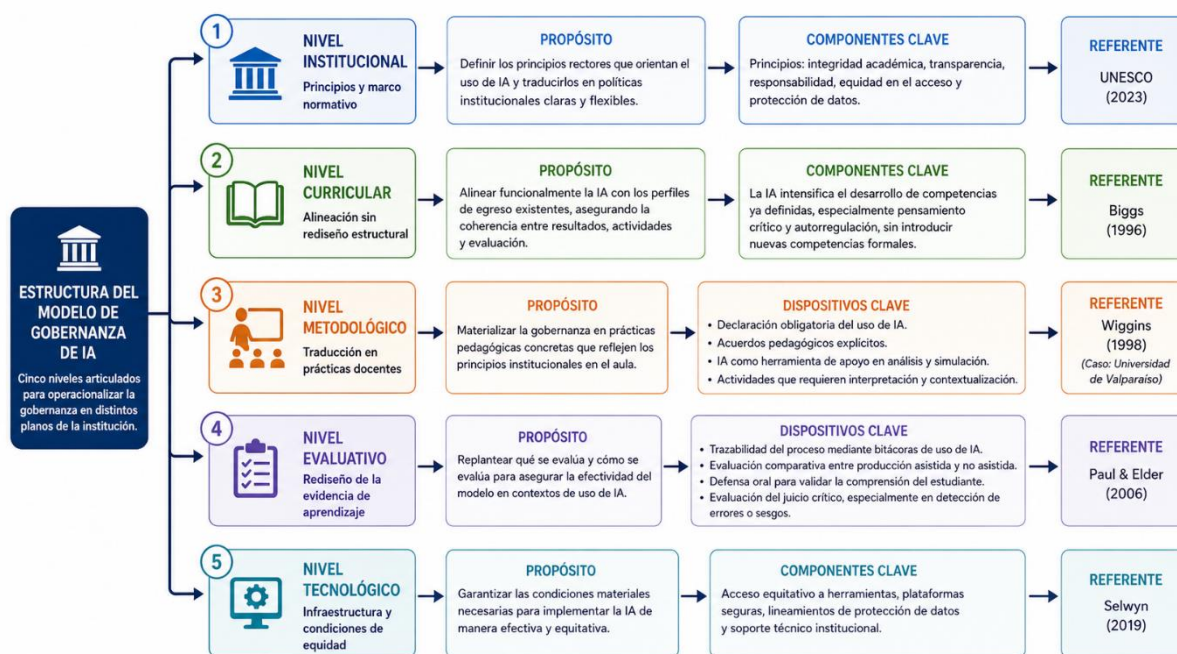
e) Nivel tecnológico: infraestructura y condiciones de equidad

La gobernanza de la IA requiere condiciones materiales que garanticen su implementación efectiva. Entre ellas destacan el acceso equitativo a herramientas, la disponibilidad de plataformas seguras, la existencia de lineamientos de protección de datos y el soporte técnico institucional.

Como advierte Selwyn (2019), la incorporación de tecnología sin políticas de acceso puede amplificar desigualdades, por lo que este nivel resulta fundamental para asegurar condiciones de equidad.

Figura 1

Modelo de gobernanza de IA en educación superior: estructura en cinco niveles



Nota. Elaboración propia a partir de UNESCO (2023), John Biggs (1996), Grant Wiggins (1998), Richard Paul & Linda Elder (2006), Neil Selwyn (2019), con apoyo de ChatGPT para la síntesis y visualización.

5.3 Proceso de implementación: un enfoque gradual y escalable

El modelo se implementa mediante un proceso gradual estructurado en cuatro fases.

La primera fase corresponde a la sensibilización y alineación institucional, incluyendo la difusión de principios, la capacitación inicial docente y la identificación de asignaturas piloto. La segunda fase implica la implementación metodológica, con la incorporación de dispositivos pedagógicos en el aula y ajustes en las evaluaciones existentes.

La tercera fase corresponde a la consolidación evaluativa, mediante el desarrollo de rúbricas específicas, la incorporación sistemática de evaluación de procesos y la generación de evidencia de aprendizaje.

Finalmente, la cuarta fase contempla el escalamiento institucional, integrando el modelo en políticas universitarias, articulándolo con programas estratégicos como PIAESS y promoviendo el desarrollo de investigación aplicada.

Este enfoque progresivo permite reducir resistencias y facilitar la adopción institucional.

5.4 Aplicación al programa PIAESS: laboratorio de validación

El programa PIAESS se configura como un espacio privilegiado para la implementación del modelo, actuando como un laboratorio de gobernanza pedagógica de la inteligencia artificial.

Su diseño interdisciplinario, internacional y basado en la resolución de problemas reales permite evaluar el uso de IA en contextos complejos, validar dispositivos metodológicos y generar evidencia empírica.

En este contexto, la gobernanza se operacionaliza mediante protocolos comunes de trabajo entre Chile y México, el uso sistemático de IA como herramienta de análisis, la evaluación centrada en procesos y la formación conjunta de docentes y estudiantes.

5.5 Indicadores y evaluación del modelo

Para la validación del modelo se proponen indicadores en tres dimensiones.

En la dimensión formativa, se considera el desarrollo del pensamiento crítico, la capacidad de validación de información y el nivel de autorregulación del aprendizaje.

En la dimensión evaluativa, se evalúa la coherencia entre proceso y producto, la calidad de la argumentación y la evidencia de aprendizaje auténtico.

En la dimensión institucional, se analizan el grado de adopción docente, la integración en programas formativos y la disponibilidad de infraestructura tecnológica.

En conjunto, estos indicadores permiten evaluar la efectividad del modelo y generar evidencia para su escalamiento.

Tabla 2

Indicadores para la validación del modelo de gobernanza pedagógica de la inteligencia artificial en educación superior

Dimensión	Indicador	Definición conceptual	Operacionalización	Instrumento / Fuente	Métrica / Escala
Formativa	Pensamiento crítico	Capacidad del estudiante para analizar, evaluar y emitir juicios fundamentados frente a información generada o mediada por IA	Identificación de sesgos, contrastación de fuentes, argumentación crítica sobre outputs de IA	Rúbrica de evaluación de trabajos escritos y orales	Escala Likert 1–5 (bajo–alto nivel crítico)
Formativa	Validación de información	Capacidad para verificar la veracidad, pertinencia y confiabilidad de la información generada por IA	Uso de fuentes externas, triangulación de información, justificación de decisiones	Bitácoras de uso de IA + análisis documental	% de fuentes verificadas / nivel de consistencia argumentativa
Formativa	Autorregulación del aprendizaje	Capacidad del estudiante para planificar, monitorear y evaluar su propio proceso de aprendizaje mediado por IA	Registro de decisiones, reflexión sobre uso de IA, ajustes en el proceso de trabajo	Bitácora reflexiva + autoevaluación	Escala Likert + análisis cualitativo
Evaluativa	Coherencia proceso–producto	Grado de correspondencia entre el proceso declarado y el resultado final del trabajo	Comparación entre bitácora de uso de IA y producto final	Rúbrica de coherencia metodológica	Alto / medio / bajo
Evaluativa	Calidad de la argumentación	Nivel de estructuración lógica, evidencia y profundidad del razonamiento	Evaluación de consistencia, uso de evidencia y articulación conceptual	Rúbrica analítica de evaluación	Puntaje (0–100) o escala Likert
Evaluativa	Aprendizaje auténtico	Evidencia de comprensión real y transferible del conocimiento por parte del estudiante	Defensa oral, resolución de problemas, aplicación contextualizada	Evaluación oral + estudio de caso	Nivel de desempeño (incipiente–avanzado)
Institucional	Adopción docente	Nivel de incorporación del modelo por parte del profesorado	Uso de IA en planificación, evaluación y actividades formativas	Encuestas docentes +	% de docentes que implementan el modelo

Dimensión	Indicador	Definición conceptual	Operacionalización	Instrumento / Fuente	Métrica / Escala
				revisión de programas	
Institucional	Integración curricular	Grado de incorporación del uso regulado de IA en asignaturas y programas	Presencia de lineamientos en syllabus y actividades evaluativas	Análisis documental de programas	% de asignaturas con integración IA
Institucional	Infraestructura tecnológica	Disponibilidad y acceso a herramientas, plataformas y soporte institucional	Acceso a software, conectividad, soporte técnico	Diagnóstico institucional + encuestas	Nivel de disponibilidad (alto/medio/bajo)

Nota. Elaboración propia a partir del modelo de gobernanza pedagógica de la inteligencia artificial y de la literatura sobre evaluación auténtica, aprendizaje autorregulado y competencias en entornos digitalmente mediados (Biggs, 1996; Wiggins, 1998; Zimmerman, 2002; Selwyn, 2019; UNESCO, 2023).

5.6 Síntesis del aporte del modelo

El modelo de gobernanza pedagógica propuesto representa un avance sustantivo frente a enfoques fragmentados, al desplazar el foco desde la mera regulación normativa hacia la articulación efectiva entre principios, prácticas pedagógicas y dispositivos evaluativos.

Su contribución principal radica en la configuración de un marco integrado y operativamente viable, que reconoce la complejidad del fenómeno y establece mecanismos concretos para su implementación en contextos reales de enseñanza.

En particular, el modelo permite integrar de manera coherente fundamentos teóricos y prácticas pedagógicas, articular distintos niveles institucionales, facilitar una implementación progresiva y reconfigurar la evaluación del aprendizaje hacia enfoques centrados en procesos, trazabilidad y juicio crítico.

Asimismo, el programa PIAESS refuerza este aporte al constituirse como un espacio de validación empírica, donde el modelo puede ser implementado en condiciones reales, evaluado sistemáticamente y proyectado hacia su escalamiento institucional.

6. Resultados esperados

La implementación del modelo de gobernanza pedagógica de la inteligencia artificial permite proyectar resultados en tres planos interrelacionados: formativo, evaluativo e institucional. Estos resultados no deben entenderse como efectos automáticos de la incorporación tecnológica, sino como consecuencia de su integración pedagógicamente orientada.

En el plano formativo, se espera un fortalecimiento del pensamiento crítico, entendido como la capacidad de analizar, contrastar y validar información en contextos de alta disponibilidad de contenidos generados por IA (Paul & Elder, 2006; Selwyn, 2019). En este sentido, el estudiante deja de centrarse en la producción de respuestas para focalizarse en la evaluación del conocimiento generado.

Asimismo, el modelo favorece una mayor trazabilidad del aprendizaje, al incorporar dispositivos que hacen visible el proceso cognitivo, tales como bitácoras, declaraciones de uso y justificación de decisiones. Este enfoque se alinea con la evaluación auténtica (Wiggins, 1998) y el aprendizaje autorregulado (Zimmerman, 2002).

En el plano evaluativo, se proyecta una mejora en la validez de las evaluaciones, al privilegiar el proceso de razonamiento por sobre el producto final. Experiencias europeas han demostrado que la incorporación de evaluaciones híbridas mejora la confiabilidad en contextos digitalizados (Comisión Europea, 2022).

En el plano institucional, el modelo permite generar evidencia sistemática para la toma de decisiones, contribuyendo a políticas universitarias basadas en datos. En este sentido, PIAESS se configura como un espacio estratégico para la producción de dicha evidencia.

7. Discusión

Los resultados proyectados permiten sostener que la gobernanza pedagógica de la IA constituye una alternativa superadora frente a enfoques dicotómicos — prohibición versus uso irrestricto—, posicionando la tecnología como parte de un proceso pedagógico deliberado.

Desde una perspectiva comparada, el modelo se alinea con tendencias observadas en sistemas universitarios europeos, donde la regulación se articula con prácticas pedagógicas (Comisión Europea, 2022). Al mismo tiempo, responde a desafíos latinoamericanos asociados a la brecha entre política educativa y práctica docente (Ornelas, 2018).

En este contexto, el modelo contribuye a:

- alinear prácticas docentes con la transformación digital,
- fortalecer la calidad del proceso formativo,
- mitigar riesgos asociados al uso no regulado de IA.

No obstante, su implementación requiere condiciones habilitantes claras:

- formación docente en integración pedagógica de IA (Mishra & Koehler, 2006),
- infraestructura tecnológica adecuada,
- articulación institucional sostenida.

En términos de contribución científica, el estudio propone una vía intermedia entre regulación normativa e integración sistémica, aportando evidencia situada en América Latina.

8. Conclusiones e implicancias

La gobernanza pedagógica de la inteligencia artificial se configura como un eje estratégico para la educación superior contemporánea, al permitir articular innovación tecnológica con calidad formativa e integridad académica.

El modelo propuesto demuestra que es posible integrar la IA mediante adaptaciones metodológicas progresivas, sin necesidad de rediseños curriculares estructurales. Esta característica le otorga viabilidad y potencial de replicabilidad en contextos latinoamericanos.

Uno de los principales hallazgos es que la efectividad de la regulación no depende de restringir el uso de IA, sino de integrarla como objeto pedagógico, haciendo visible el proceso de aprendizaje y fortaleciendo competencias de orden superior.

En este marco, el programa PIAESS se proyecta como un laboratorio de gobernanza pedagógica, con potencial para validar empíricamente el modelo y escalarlo institucionalmente.

Implicancias para investigación y política

Desde la investigación, el estudio abre líneas futuras:

- evaluación empírica del modelo,
- estudios comparados longitudinales,
- desarrollo de métricas de trazabilidad del aprendizaje.

Desde la política universitaria, aporta insumos para:

- diseño de lineamientos institucionales,
- integración en procesos de acreditación,
- fortalecimiento de la formación docente.

En síntesis, la incorporación de la IA representa una oportunidad para redefinir las bases pedagógicas de la educación superior, avanzando hacia modelos más reflexivos, transparentes y orientados al desarrollo integral de competencias.

Referencias

- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364.
- Brunner, J. J. (2011). Educación superior en Iberoamérica. Universia.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.
- Carr, N. (2010). *The shallows: What the Internet is doing to our brains*. W. W. Norton & Company.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Comisión Europea. (2022). Directrices éticas sobre el uso de la inteligencia artificial y los datos en la enseñanza y el aprendizaje. Oficina de Publicaciones de la Unión Europea.
- Didriksson, A. (2016). *La universidad en la sociedad del conocimiento*. Universidad Nacional Autónoma de México.
- Freire, P. (1970). *Pedagogía del oprimido*. Siglo XXI Editores.
- García Canclini, N. (2004). *Diferentes, desiguales y desconectados: Mapas de la interculturalidad*. Gedisa.
- Jonas, H. (1984). *The imperative of responsibility: In search of an ethics for the technological age*. University of Chicago Press.
- Martín-Barbero, J. (2003). *De los medios a las mediaciones: Comunicación, cultura y hegemonía*. Convenio Andrés Bello.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Ornelas, C. (2018). *Política, poder y pupitres: Crítica al nuevo federalismo educativo*. Siglo XXI Editores.
- Paul, R., & Elder, L. (2006). *Critical thinking: Tools for taking charge of your learning and your life*. Pearson.
- Rama, C. (2020). *La educación superior en la era digital*. Unión de Universidades de América Latina y el Caribe (UDUAL).

Redecker, C. (2017). European framework for the digital competence of educators (DigCompEdu). European Commission.

Schön, D. A. (1983). The reflective practitioner: How professionals think in action. Basic Books.

Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. Polity Press.

UNESCO. (2023). Orientaciones para el uso de la inteligencia artificial generativa en la educación y la investigación. UNESCO Publishing.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wiggins, G. (1998). Educative assessment: Designing assessments to inform and improve student performance. Jossey-Bass.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70.

APORTES PARA LA REGULACIÓN DE UNA IA SOSTENIBLE

Contributions to the regulation of sustainable AI

Karen Liliana Chip Domínguez

Universidad Nacional Autónoma de México, Laboratorio de Innovación y
Futuros, Instituto de Energías Renovables |
karenchip@comunidad.unam.mx
<https://orcid.org/0009-0006-9409-9583>

Jorge Iván González Gómez

Universidad Autónoma Metropolitana, Unidad Iztapalapa, Posgrado en
Estudios Organizacionales
jorgeivangonzalez@jigcom.com
<https://orcid.org/0009-0009-1567-153X>

Resumen

Este capítulo analiza la regulación de la inteligencia artificial desde una perspectiva de sostenibilidad. Parte de la premisa de que la IA no debe entenderse únicamente como una herramienta digital, algorítmica o funcional, sino como una infraestructura sociotécnica sostenida por recursos naturales, energía, agua, minerales, centros de datos, redes de comunicación, datos, trabajo humano e instituciones. A partir de esta mirada, se cuestiona la idea de una “IA para la sostenibilidad” cuando no se consideran las condiciones materiales, laborales, territoriales y geopolíticas que hacen posible su desarrollo. El texto propone distinguir entre los usos de la IA orientados a objetivos ambientales o sociales y la sostenibilidad de la propia IA como sistema productivo. Para ello, se revisan tensiones vinculadas con la cadena de valor, la división internacional del trabajo, la justicia espacial, la explotación del trabajo digital, la extracción de datos y los impactos de los centros de datos sobre los servicios de agua y energía. Se sostiene que una regulación guiada por la

sostenibilidad debe ser integral, dinámica, distributiva y territorialmente situada, capaz de evaluar no solo riesgos técnicos, sino también quién captura los beneficios, quién asume los costos y en qué condiciones públicas debe permitirse la expansión de la IA.

Palabras clave: regulación de la IA; sostenibilidad; inteligencia artificial; cadena de valor.

Abstract

This chapter analyzes the regulation of artificial intelligence from a sustainability perspective. It starts from the premise that AI should not be understood solely as a digital, algorithmic, or functional tool, but rather as a sociotechnical infrastructure sustained by natural resources, energy, water, minerals, data centers, communication networks, data, human labor, and institutions. From this standpoint, the chapter questions the notion of “AI for sustainability” when the material, labor, territorial, and geopolitical conditions that make its development possible are not taken into account. The text proposes a distinction between uses of AI aimed at environmental or social objectives and the sustainability of AI itself as a productive system. To this end, it examines tensions related to the value chain, the international division of labor, spatial justice, the exploitation of digital labor, data extraction, and the impacts of data centers on water and energy services. It argues that regulation guided by sustainability must be comprehensive, dynamic, distributive, and territorially situated, capable of assessing not only technical risks, but also who captures the benefits, who bears the costs, and under what public conditions the expansion of AI should be permitted.

Keywords: AI regulation; sustainability; artificial intelligence; value chain.

Introducción

La Inteligencia Artificial (IA) se ha convertido en una de las tecnologías más influyentes en la sociedad contemporánea, sus aplicaciones tienen alcance en sectores diversos y cruciales como la economía, la administración pública, la

educación, la salud, la investigación científica, la comunicación y la producción cultural. Por ejemplo, hoy en día la incorporación de la IA a la solución de tareas laborales puede significar el aumento en la eficiencia laboral, automatizar procesos, mejorar la toma de decisiones en ámbitos diversos y procesar grandes volúmenes de información (United Nations, s.f.; Villagómez, 2025). Empero la inmensa lista de cosas en las que la IA puede ayudarnos existe la necesidad de ampliar el horizonte de análisis de manera integral sobre esta herramienta ya que desafortunadamente en el capitalismo, la tecnología es utilizada para resolver los grandes problemas del mundo, sí y solo sí esto produce ganancias -y vaya que la IA las está produciendo¹¹- y que generalmente los efectos negativos socioambientales quedan en segundo término. Estos efectos negativos suelen ser de muy diversa índole, en el caso de la IA, muchos de ellos aún se están discutiendo, por lo que este trabajo busca abonar a la discusión sobre el impacto socioambiental de la IA.

La intención es exponer que la IA no es una tecnología etérea, neutral ni autónoma que nos resuelve la vida -solamente-, sino toda una infraestructura sociotécnica que depende de energía, agua, minerales, centros de datos, redes de comunicación, dispositivos, datos, trabajo humano e instituciones. Que mientras nos resuelve la vida. En otras palabras, mientras nos resuelve la vida, también es proclive de afectar su desarrollo por los impactos negativos que produce a muy distintos niveles. Si bien, ahora mismo los debates se centran en torno a su incipiente regulación sobre los usos éticos, es fundamental amplificar los debates más allá de las personas usuarias solamente. Si bien, la protección de datos personales, transparencia algorítmica, sesgos, seguridad, responsabilidad civil y gestión de riesgos asociados a sistemas automatizados son temas indispensables sobre los que existe un debate y marcos legales como la Ley de IA publicada por la Unión Europea (Comisión Europea, s.f.) y las Medidas para la Administración de Servicios de Inteligencia Artificial Generativa en China (Administración del Ciberespacio de

¹¹ “Se prevé que el mercado de la IA alcance los 4,8 billones de dólares en 2033 y se consolide como la principal tecnología de punta”, véase <https://unctad.org/es/news/se-preve-que-el-mercado-de-la-ia-alcance-los-48-billones-de-dolares-en-2033-y-se-consolide>

China, 2023), estos son apenas la punta del iceberg. Si la regulación se limita al funcionamiento del algoritmo, al diseño del producto o al comportamiento de las personas usuarias, deja fuera las condiciones materiales que hacen posible la IA, así como los impactos ambientales, laborales, sociales y territoriales que se producen a lo largo de su cadena de valor.

Dicho lo anterior, merece el esfuerzo preguntarse ¿cómo se relaciona la IA con la sostenibilidad? Es decir, ¿Puede la sostenibilidad guiar una regulación justa e integral de la inteligencia artificial? La definición clásica del Informe Brundtland plantea la sostenibilidad como la posibilidad de satisfacer las necesidades del presente sin comprometer la capacidad de las generaciones futuras para satisfacer sus propias necesidades (Brundtland, 1987). Bajo esta concepción, una IA alineada con la sostenibilidad debería ser evaluada por su compatibilidad con los límites ecológicos, la justicia social, la responsabilidad intergeneracional y la protección de derechos.

Actualmente, el vínculo entre inteligencia artificial y sostenibilidad se aborda como una IA *para* la sostenibilidad en sentido funcional. Esto es, hace alusión a aplicaciones como modelación climática, monitoreo ambiental, optimización energética, agricultura de precisión, gestión de redes eléctricas, planeación urbana o análisis de información para políticas públicas (Siqueira y otros., 2024; Raihan y otros., 2024). Todas ellas son posibilidades relevantes, empero no abonan a un desarrollo estructural sostenible de la inteligencia artificial. La contradicción es evidente: la misma herramienta con el potencial de contribuir con los objetivos de desarrollo sostenible, se mantiene a costa de procesos intensivos en consumo energético, uso de agua, extracción de minerales, generación de residuos electrónicos, distribución extremadamente desigual de ganancias y trabajo humano invisibilizado.

El problema regulatorio se vuelve más complejo si se considera que los beneficios y costos de la inteligencia artificial quedan en manos de unos cuantos capitalistas. Las empresas que desarrollan entrenan y comercializan sistemas de IA pueden concentrar ganancias económicas exorbitantes así como influencia política - ejemplos de ello son Elon Musk¹², Alex Karp, CEO de Palantir y Sam Altman, CEO de OpenAI-, mientras que los costos ambientales, sociales y territoriales se desplazan hacia comunidades desfavorecidas, trabajadores precarizados y ecosistemas legalmente desprotegidos (Carrera, 2026; UNAM Global, 2023). Asimismo, el acceso a sus beneficios depende de condiciones sociales estructurales, como conectividad, capacidades técnicas, infraestructura pública, recursos económicos, alfabetización digital y posición geopolítica.

También es necesario reconocer que la inteligencia artificial se sostiene en el trabajo humano. Aunque suele presentarse como automatización, muchos sistemas dependen de tareas de etiquetado, clasificación, moderación, limpieza de datos, validación, entrenamiento y producción intelectual previa (Miceli, 2024). Además, el uso de productos intelectuales para entrenar modelos -como obras, textos, imágenes y códigos- plantea debates sobre derechos de autor, remuneración y apropiación de valor (Ordelin, 2025). Así, las personas usuarias de la IA, recibimos respuestas que son el producto de entrenamientos con datos colectivos al mismo tiempo que alimentamos nuevas bases de datos sobre nuestras formas de buscar, construir información y hábitos de consumo de la inteligencia artificial. En cuanto a información, es un sistema con lazos de retroalimentación.

En este marco, regular únicamente la relación entre persona usuaria y sistema de IA resulta insuficiente. Una regulación de la tecnología en clave de sostenibilidad debe considerar para qué y para quién se diseña la tecnología, quién provee los

¹² Elon Musk, cofundador de siete empresas, entre ellas el fabricante de coches eléctricos Tesla, el productor de cohetes SpaceX y la empresa emergente de inteligencia artificial xAI concentra en su patrimonio 775.800 millones de dólares. Véase <https://www.forbes.com/profile/elon-musk/>

datos, quién realiza el trabajo necesario para entrenarla, qué recursos naturales consume, dónde se instala -y por qué-, quién se queda con las ganancias y qué poblaciones son vulneradas en cualquier parte de la cadena de valor, en el corto, mediano y largo plazo. Si además consideramos que el ritmo del cambio tecnológico avanza más rápido que los procesos de normatividad y regulación, la regulación de la IA en este caso, cualesquiera que sea su orientación, queda supeditada a la incertidumbre política y rigidez normativa, lo que puede inhibir el desarrollo e innovación de esta cada que exista incertidumbre ante cualquier novedad en su aplicación (Aithor, 2024). En otras palabras, hoy se puede llegar a regular un estado de la cuestión que mañana puede cambiar aceleradamente ya que la evolución de la IA es dinámica y por momentos va a saltos, entonces si regulamos rígidamente podríamos detener este flujo, este de hecho es el argumento de algunas empresas tecnológicas como Antropic u Open AI.

En las siguientes líneas se abordará el problema de la regulación de la inteligencia artificial como una tarea integral, dinámica y orientada por criterios de sostenibilidad. Integral, porque no puede limitarse al algoritmo ni al usuario final, sino abarcar toda la cadena de valor que hace posible la IA: recursos naturales, infraestructura, datos, trabajo humano, propiedad intelectual, impactos territoriales y concentración de beneficios. Dinámica, porque debe responder a una tecnología en constante transformación sin caer ni en la parálisis normativa ni en la fe ingenua en la autorregulación del mercado. Y orientada por sostenibilidad, porque su objetivo no debe ser únicamente hacer más eficiente el uso de la IA, sino garantizar que su desarrollo no comprometa derechos, territorios, ecosistemas, condiciones laborales ni la capacidad de las generaciones presentes y futuras para sostener una vida digna. En suma, la pregunta regulatoria no es solo cómo aprovechar la inteligencia artificial, sino bajo qué condiciones públicas, sociales y ambientales debe permitirse su expansión.

¿IA sustentable o para la sustentabilidad?

Si se pretende debatir sobre las relaciones entre IA y sostenibilidad, en primer lugar, conviene delinear brevemente algunas discusiones alrededor del término a fin de contextualizar al lector. La emblemática definición del Informe Brundtland¹³ colocó en el centro de la discusión la relación entre generaciones presentes y futuras, al plantear el desarrollo sostenible como aquel que permite satisfacer las necesidades del presente sin comprometer la capacidad de las generaciones futuras para satisfacer las propias (Brundtland, 1987). Esta formulación permitió articular dimensiones económicas, sociales y ambientales en el debate internacional; sin embargo, también ha sido cuestionada por su poca precisión conceptual y por la tensión que mantiene entre la idea de crecimiento económico continuo y los límites ecológicos del planeta (Redclift, 2006). Es decir, la contradicción que subyace en el centro del sistema capitalista.

Asimismo, existe un debate abierto entre los términos “sostenibilidad” y “sustentabilidad”. En algunos enfoques, lo sostenible se asocia con la permanencia de un sistema en el largo plazo mediante la articulación de dimensiones ambientales, sociales y económicas; mientras que lo sustentable se vincula con la capacidad de sostener o preservar determinados sistemas, en ocasiones con mayor énfasis en la dimensión ambiental (Rodríguez, 2018; UNAM, s.f.). No obstante, en este capítulo no se profundizará en esa discusión lingüística acerca del término. Para efectos del análisis, sostenibilidad y sustentabilidad se utilizarán como equivalentes operativos, entendidas como la capacidad de sostener,

¹³ El término "Desarrollo Sustentable" fue formalizado en el informe Nuestro Futuro Común (1987), elaborado por la Comisión Mundial sobre el Medio Ambiente y el Desarrollo (CMMAD) bajo la presidencia de Gro Harlem Brundtland. Conocido por la entonces primera ministra de Noruega, el "Informe de Brundtland" estableció que la satisfacción de las necesidades presentes no debe comprometer la capacidad de las futuras generaciones (equidad social y justicia intergeneracional) y sentó bases de la concepción interdependiente de las dimensiones social, ambiental y económica. Este informe sentó las bases de la Cumbre de Río 1992 y los actuales Objetivos de Desarrollo Sostenible (ODS).

preservar y reproducir en el largo plazo las condiciones materiales, sociales, ecológicas e institucionales que hacen posible la vida colectiva.

Esta decisión no implica asumir de manera acrítica el discurso del desarrollo sostenible. En cambio, permite distinguir, por un lado, entre una visión *greenwashing* que busca el menor impacto posible, como reducción de emisiones, cuotas de género, entre otras, pero sin cuestionar la estructura que sostiene al capitalismo. Por otro lado, una perspectiva crítica en torno a los modelos de producción, consumo y acumulación que generan deterioro ambiental y profundización de la desigualdad social. Desde la mirada crítica, la sostenibilidad no puede reducirse a compensar impactos o mejorar indicadores aislados; sino que implica consciencia profunda sobre la justicia social, el lugar de la sociedad en el ecosistema, la distribución territorial de costos y beneficios, y la responsabilidad frente a las generaciones presentes y futuras (Gudynas, 2010).

En relación con la regulación de la IA, la sustentabilidad se entenderá como una guía para garantizar que su desarrollo, despliegue y uso no comprometa los límites ecológicos, los derechos humanos, la justicia territorial, las condiciones laborales ni la capacidad de las generaciones presentes y futuras para satisfacer sus necesidades. Esta definición implica analizar la IA no solo por sus beneficios funcionales o económicos, sino por el conjunto de recursos, infraestructuras, datos, trabajo humano, territorios e impactos socioambientales que hacen posible su operación.

Al respecto, cabe recuperar la categorización propuesta por Eduardo Gudynas, construida según la profundidad de la crítica a la ideología del progreso y el nivel de incorporación de las dimensiones ambiental, ecológica y social. En un extremo se ubica la insustentabilidad, caracterizada por la permanencia de metas de crecimiento económico, lucro y competitividad sin una integración sustantiva de los límites ecológicos. Posteriormente aparece la sustentabilidad débil, que reconoce

el valor económico de la naturaleza y asume que los límites ecológicos pueden ser manejados sin cuestionar la ideología del progreso. La sustentabilidad fuerte introduce una crítica mayor al progresismo y reconoce la necesidad de preservar un stock natural crítico, aunque mantiene un enfoque técnico-político. Finalmente, la sustentabilidad superfuerte plantea una crítica sustantiva al desarrollo convencional, incorpora la ética de los valores propios de la naturaleza y propone la búsqueda de nuevos estilos de desarrollo desde un enfoque político más profundo (Gudynas, 2010). Esta clasificación resulta útil para superar la dicotomía entre “sostenible” e “insostenible”, pues permite reconocer distintos grados de sustentabilidad y nombrar con mayor precisión el alcance de las acciones, regulaciones y objetivos, sin caer en falsos optimismos ni en pesimismo fatalistas.

Ahora bien, ningún desarrollo es sustentable en algún grado per se. Antes de etiquetarlo, es necesario analizar cómo se produce, para qué y para quién se utiliza. De acuerdo con ello, conviene preguntarse si la IA puede contribuir con los objetivos de sustentabilidad de un sistema. La respuesta es sí, ya que su uso puede optimizar procesos, mejorar diagnósticos, reducir tiempos de trabajo, ahorrar recursos en actividades específicas e incluso identificar asimetrías en las relaciones a fin de no concentrar el poder en la toma de decisiones (Alcántara y otros., 2024).

Sin embargo, ¿Cómo deberían ser sus condiciones de producción y operación para ser un sistema sostenible en algún grado? Esta pregunta nos obliga a considerar su ciclo de vida completo: extracción de materias primas, fabricación de dispositivos, operación de centros de datos, consumo eléctrico e hídrico, almacenamiento de información, entrenamiento de modelos, generación de residuos electrónicos, extracción de datos y trabajo humano asociado (García y García, 2026).

Llegados a este punto, no basta con promover principios generales de innovación responsable y usos éticos; es necesaria una planeación tecnológica en función de

los recursos materiales donde no se comprometa el acceso a derechos humanos básicos (alimentación, agua y energía, por ejemplo) ni la protección ambiental. Una regulación orientada por sostenibilidad tendría que preguntar qué recursos consume un sistema de IA, qué emisiones genera, qué infraestructura requiere, qué territorios afecta, qué trabajo humano utiliza, qué datos extrae, quién obtiene beneficios económicos y quién asume los costos ambientales o sociales (Cadierno, 2025). Esto sería una caracterización básica desde un análisis sistémico de corte materialista histórico, heredado por la sostenibilidad -de acuerdo con las categorías de Gudynas *super-fuerte*-.

En este orden de ideas, regular la IA en sentido sostenible implica cuestionar las relaciones políticas y de mercado que la hacen posible. La literatura reciente muestra que sus efectos ambientales no son lineales ni pueden asumirse como positivos por definición. Wang, Li y Li (2024), al analizar datos de 67 países entre 1993 y 2019, encontraron que la IA puede contribuir a reducir la huella ecológica y las emisiones de carbono, así como a promover la transición energética; sin embargo, también advierten que estos efectos dependen de condiciones externas como la estructura industrial, la apertura comercial, el nivel de desarrollo de la propia IA y el avance previo de la transición energética. Esto refuerza la necesidad de una regulación que no parta de una confianza abstracta en el potencial de la tecnología y quienes dirigen estos desarrollos, sino de una evaluación situada de sus condiciones materiales, institucionales y territoriales.

Así, la IA puede ser una herramienta útil para objetivos ambientales, pero solo bajo ciertas condiciones de gobernanza. Wang, Li y Li (2024) también señala que una mayor participación del sector industrial puede debilitar el efecto de la IA en la reducción de huella ecológica y emisiones, aunque fortalezca su contribución a la transición energética; es decir, la tecnología puede funcionar como palanca ambiental o como acelerador de impactos, según el contexto productivo en el que se inserte. Por ello, la regulación sostenible no debe limitarse a promover “más IA”,

sino a definir dónde, cómo, para qué y bajo qué límites debe desplegarse. En términos regulatorios, esto supone pasar de una lógica de adopción tecnológica acelerada, basada en la ganancia privada, a una planeación pública capaz de medir impactos, anticipar riesgos, distribuir responsabilidades y evitar que la eficiencia tecnológica encubra nuevas formas de presión sobre recursos, territorios y derechos.

Ahora bien, en los últimos años, diversas empresas tecnológicas han adoptado discursos sobre inteligencia artificial responsable, ética algorítmica, reducción de emisiones o neutralidad de carbono (Pacto Mundial, 2026). Aunque estos instrumentos pueden aportar información, criterios éticos y buenas prácticas, resultan insuficientes si no existen mecanismos públicos de verificación, auditoría, sanción y control democrático. La sostenibilidad, sustantivo que por definición indica la capacidad de proteger con perspectiva intergeneracional los bienes comunes¹⁴, no puede depender únicamente de la voluntad de los actores que obtienen beneficios económicos de la expansión tecnológica.

Todo esto implica, como ya se dijo, reconocer que la tecnología no es neutral respecto del modelo de desarrollo de la época ni de sus condiciones materiales de reproducción. Como plantea Langdon Winner, los artefactos técnicos no solo deben evaluarse por su eficiencia, productividad o efectos ambientales, sino también por las formas de poder y autoridad que pueden encarnar (Winner, 1980). Esta idea resulta clave para la inteligencia artificial: sus modelos, bases de datos, infraestructuras de cómputo, centros de datos, arquitecturas propietarias y condiciones de acceso no son simples herramientas disponibles para cualquier uso, sino arreglos sociotécnicos que pueden favorecer determinadas formas de control, concentración económica, dependencia tecnológica o exclusión. En palabras de

¹⁴ El término “bienes comunes”, en inglés “*the commons*”, se refiere a los bienes que un grupo, comunidad o sociedad utiliza en común. Este concepto es retomado y adaptado al español por Elinor Ostrom (2009); a partir de este problematiza la categorización económica que refiere a “bienes públicos, privados y de club”, pues los bienes comunes, gobernados por reglas, usos y acuerdos colectivos, más que solo por mercado o Estado, representan una nueva categoría.

Winner, las tecnologías pueden funcionar como formas de ordenamiento social, pues sus diseños y arreglos influyen en cómo las personas trabajan, se comunican, se desplazan, consumen y participan en la vida pública.

Desde esta perspectiva, regular la IA no puede reducirse a sancionar usos indebidos después de que ocurren, sino que debe intervenir en las condiciones de diseño, despliegue e infraestructura que hacen posibles esos usos. Winner advierte que muchas decisiones técnicas, una vez materializadas en equipos, inversiones, hábitos e instituciones, pierden flexibilidad y se vuelven difíciles de revertir; incluso compara ciertas innovaciones tecnológicas con actos legislativos o fundacionales que establecen marcos duraderos de orden público. Por ello, una regulación sustentable debe actuar antes de que la arquitectura técnica y económica de la IA quede fijada bajo lógicas de concentración, vigilancia, opacidad o desplazamiento de costos socioecológicos. Dicho de otro modo: no basta con preguntar si la IA puede usarse bien o mal; hay que preguntar qué formas de poder deja instaladas, quién puede modificarlas y quién queda obligado a vivir bajo sus consecuencias.

Necesidades materiales y reconfiguraciones espaciales

Esto resulta especialmente importante para América Latina, donde persisten brechas en conectividad, infraestructura digital y capacidades tecnológicas (Banco Mundial, 2021), así como una dependencia creciente de servicios en la nube y centros de datos fuera de la región, que socavan la soberanía sobre datos e infraestructura computacional (Moreno, 2025). Desde una lectura *centro-periferia*, estas brechas no son simples rezagos técnicos, sino expresiones de una inserción desigual en la división internacional del trabajo. Kiljunen (1986) advierte que la posición de un país en dicha división depende del alcance y profundidad de su industrialización, de su autonomía relativa frente al exterior en insumos y mercados, de la diversidad de su estructura productiva, del uso de trabajo especializado y del valor agregado de su producción. Trasladado al campo de la inteligencia artificial, esto implica preguntar qué lugar ocupa la región en la cadena global de valor: si

participa como desarrolladora de modelos, productora de conocimiento, generadora de infraestructura y capacidades propias, o si queda reducida a mercado consumidor, fuente de datos, proveedor de recursos y territorio receptor de costos socioecológicos.

Así, el problema no es únicamente que los costos materiales del desarrollo tecnológico se desplacen hacia el sur global, sino que los beneficios de mayor valor también tienden a capturarse fuera de él. La periferia global puede aportar minerales, energía, agua, datos, trabajo digital o territorios para infraestructura, sin necesariamente apropiarse de las ganancias, las patentes, los modelos, las capacidades computacionales o las decisiones estratégicas. Esta asimetría reproduce lo que el enfoque estructuralista ha señalado sobre las relaciones *centro-periferia*: los beneficios del comercio y la especialización pueden distribuirse de forma desigual cuando los países parten de estructuras productivas y tecnológicas profundamente distintas. En el caso de la IA, una regulación sustentable debe evitar que América Latina sea integrada a la economía digital solo como soporte material, mercado cautivo o zona de externalización, mientras las capacidades de innovación, gobernanza y captura de valor permanecen concentradas en el centro tecnológico global.

En este punto, la noción de justicia espacial resulta útil como lente para observar la distribución territorial de beneficios y daños. Edward Soja plantea que las desigualdades no solo ocurren en el espacio, sino que también se producen espacialmente mediante decisiones políticas, económicas e institucionales (Soja, 2010). En el contexto de la IA, esto permite analizar cómo la infraestructura digital puede concentrar beneficios en ciertos actores y territorios, mientras desplaza impactos ambientales o sociales hacia otros.

La expansión de centros de datos ilustra con claridad esta dimensión. Estas infraestructuras requieren suelo, electricidad, agua, conectividad y condiciones

técnicas específicas para operar de forma continua. En territorios con estrés hídrico, redes eléctricas saturadas o desigualdad en el acceso a servicios básicos, su instalación puede generar tensiones adicionales sobre bienes públicos y recursos comunes (S&P Global Ratings, 2025; FUNIBER, 2025) además de conflictos sociales y políticos¹⁵. La cuestión regulatoria no es solo si estos proyectos son tecnológicamente factibles, sino si son ambientalmente viables y socialmente aceptados bajo condiciones territoriales concretas.

Esta mirada permite conectar la IA con debates más amplios sobre justicia ambiental. Cuando los costos ecológicos de una actividad económica se concentran en comunidades específicas, mientras los beneficios se captan en otros lugares, se producen conflictos ecológicos-distributivos (Martínez-Alier, 2007; Walter, 2009). En el caso de la IA, estos conflictos pueden aparecer en torno al uso intensivo de agua y energía, la presión sobre infraestructura pública, la extracción de minerales, la disposición de residuos electrónicos o la localización de centros de datos en territorios con infraestructura que no está preparada y problemas asociados a sus procesos de urbanización¹⁶.

Por ello, una regulación sostenible de la IA debe incluir criterios de evaluación territorial. No basta con medir el consumo agregado de energía o agua; es necesario identificar dónde ocurre ese consumo, qué condiciones socioambientales existen en el territorio, qué comunidades pueden verse afectadas, qué servicios públicos podrían ser presionados y qué mecanismos de participación, consulta,

¹⁵ Conflictos sociales y políticos alrededor de los centros de datos se extienden por Estados Unidos, ejemplos de ellos en las siguientes notas periodísticas: <https://www.aiconvolution.com/blog/2025/12/24/activistas-frenan-centros-datos-eeuu-2025/> <https://es.wired.com/articulos/en-ee-uu-protestan-contralos-centros-de-datos-pero-abrazan-la-infraestructura-que-los-sostiene>

¹⁶ En México, el caso Querétaro, véase <https://arteesetica.org/queretaro-sin-agua-y-sin-luz-por-la-ia/>

compensación o mitigación deben establecerse. Sin esta información, la promesa de innovación puede encubrir nuevas formas de desigualdad espacial.

La dimensión territorial también obliga a pensar en escalas. La IA opera en cadenas globales, pero sus impactos se viven localmente. Una empresa puede ofrecer servicios digitales a escala mundial, alojar datos en un país, consumir energía en otro, depender de minerales extraídos en un tercero y generar beneficios económicos concentrados en sedes corporativas o mercados financieros específicos. Esta dispersión dificulta la asignación de responsabilidades y muestra la necesidad de marcos regulatorios capaces de articular escalas locales, nacionales e internacionales.

Desde esta perspectiva, la pregunta por la justicia espacial complementa la discusión sobre sostenibilidad. Si la IA ocupa territorio, consume recursos y reorganiza infraestructuras, entonces su regulación debe evitar que su incidencia produzca zonas de sacrificio (Donoso, 2022). Esto implica impedir que la expansión tecnológica se apoye en territorios donde los costos ambientales, sociales o económicos sean menos visibles o políticamente más fáciles de desplazar.

En síntesis, la inteligencia artificial no solo transforma procesos digitales: también produce espacios materiales y sociales. Sus infraestructuras modifican relaciones entre energía, agua, suelo, trabajo, datos y poder económico. Incorporar una mirada territorial permite formular preguntas indispensables para una regulación sostenible: dónde se instala la infraestructura, qué recursos consume, qué comunidad afecta, quién decide sobre su localización, quién obtiene los beneficios y quién asume los costos. Sin esta dimensión, la regulación corre el riesgo de proteger al usuario digital mientras deja desprotegidos los territorios que sostienen materialmente a la IA. O lo que es lo mismo, podemos usar la IA como agente para facilitarnos la vida, mientras cada prompt produce impactos socioambientales en lugares que no vemos y no conocemos.

Desde esta perspectiva, el criterio de sostenibilidad desplaza la discusión desde la eficiencia hacia la justicia socioambiental. No se trata únicamente de saber si la IA permite hacer más cosas en menos tiempo, sino de preguntar si esa eficiencia reduce daños o simplemente traslada costos hacia otros cuerpos, territorios o generaciones. La eficiencia tecnológica puede ser compatible con la insostenibilidad si aumenta el consumo total de recursos, intensifica la demanda energética, acelera la extracción de datos o precariza el trabajo humano que sostiene los sistemas automatizados.

Esta tensión puede explicarse mediante la paradoja de Jevons que se refiere a la implementación de mecanismos para hacer más eficientes los sistemas, enmarcados en objetivos de crecimiento continuo. Mientras más eficiente se vuelve un sistema, más se produce, entonces lejos de ahorrar recursos (energéticos o de trabajo humano), rápido éstos se agotan. Foster, Clark y York (2010) recuperan esta idea para señalar que los aumentos en eficiencia energética o material no necesariamente reducen el consumo total de recursos. Por el contrario, bajo una lógica de expansión económica, pueden abaratar costos, estimular nuevas inversiones, ampliar la escala de producción y, finalmente, incrementar el uso absoluto de energía y materiales.

La paradoja, formulada originalmente por William Stanley Jevons respecto del carbón en el siglo XIX, muestra que la “economía” en el uso de un recurso puede conducir a su consumo más extensivo cuando esa eficiencia se integra a un sistema orientado a la acumulación y al crecimiento continuo. En el caso de la IA, esta advertencia resulta central: hacer modelos más eficientes, centros de datos más optimizados o algoritmos menos intensivos no garantiza por sí mismo una reducción de impactos socioecológicos si al mismo tiempo se multiplican los usos, se expande la demanda computacional y se acelera la incorporación de IA en cada esfera de la vida económica y social.

Para recapitular, incorporar la sostenibilidad como criterio regulatorio implica cuestionar paradigmas. En primer lugar, pasar de una visión de la IA como producto digital a una visión de la IA como infraestructura material. Segundo, trascender la concepción regulatoria centrada en el usuario final a una regulación de toda la cadena de valor. Tercero, establecer obligaciones verificables de transparencia, medición y responsabilidad en lugar de dejarlo como acto voluntario. Cuarto, hacer estudios prospectivos sociales que permitan anticipar riesgos sobre los derechos humanos, los territorios, la relación capital-trabajo y las modificaciones al ecosistema.

De esta manera, la sostenibilidad deja de ser una etiqueta o un objetivo externo al sistema. La pregunta regulatoria central no es solo cómo hacer que la IA funcione mejor para aportar así a soluciones pro-sostenibilidad, sino cómo asegurar que funcione sin comprometer las condiciones materiales y sociales que sostienen la vida.

Datos y trabajo invisible

Hasta este momento, ya se ha discutido sobre la necesidad de regular la IA desde una concepción material, pero ¿qué implica esto realmente? Pensar la IA desde su materialidad implica saber que su ciclo de vida no comienza en el algoritmo ni termina en la aplicación final; inicia con el acceso a materias primas estratégicas — como tierras raras, silicio de alta pureza y otros minerales críticos—, continúa con el diseño y fabricación de semiconductores, la disponibilidad de chips especializados, la instalación de centros de datos, la infraestructura energética, las redes de fibra óptica y cables submarinos, las plataformas de nube, los datos disponibles para entrenamiento, el talento especializado, la investigación algorítmica y, finalmente, la adopción de sistemas de IA por gobiernos, empresas y personas usuarias. En este sentido, la IA debe entenderse como una arquitectura

industrial, energética, digital y geopolítica, no solo como software o modelo computacional.

Esta lectura permite identificar que cada eslabón concentra vulnerabilidades distintas. Las materias primas pueden depender de cadenas extractivas altamente concentradas; los chips avanzados requieren capacidades industriales que pocos países poseen; los centros de datos demandan energía, agua, suelo y conectividad; las plataformas de nube pueden reforzar dependencias frente a grandes corporaciones tecnológicas; los datos plantean dilemas de privacidad, acceso, soberanía y apropiación de valor; y el talento especializado se concentra en regiones con mayores capacidades científicas, financieras e institucionales.

Por tanto, regular la IA desde la sostenibilidad implica regular el conjunto del sistema que la hace posible: los recursos que consume, las infraestructuras que instala, los datos que extrae, el trabajo que utiliza, las dependencias que genera y los beneficios que distribuye o concentra. He ahí que una regulación limitada al producto final resulta insuficiente. La pregunta no puede ser solo si un sistema de IA es transparente, explicable o no discriminatorio; también debe preguntarse si su cadena de valor es ambientalmente viable, socialmente justa, territorialmente responsable y políticamente gobernable. En otras palabras, la regulación en sentido sostenible de la IA requiere pasar del control del algoritmo al gobierno de la infraestructura completa que lo sostiene.

En el plano laboral, la IA puede automatizar tareas, modificar procesos productivos y transformar perfiles profesionales, pero no sustituye de manera homogénea todo tipo de trabajo. Algunas actividades pueden ser desplazadas, otras intensificadas y otras invisibilizadas. En particular, las tareas de cuidado, históricamente feminizadas y poco reconocidas, no desaparecen con la automatización digital, lo que obliga a cuestionar las promesas generales de liberación de tiempo o aumento universal de productividad (Collett, Neff, & Gouvea 2022).

Además, los beneficios productivos de la IA no necesariamente se traducen en mejoras distributivas. Una tecnología puede aumentar la eficiencia de empresas o instituciones sin mejorar las condiciones de vida de quienes trabajan, habitan o sostienen los territorios donde opera su infraestructura (Georgieva, 2024). Por ello, la pregunta regulatoria no debe limitarse a cuánto valor económico genera la IA, sino a dónde se queda ese valor, quién lo captura y qué mecanismos existen para redistribuir beneficios o compensar impactos

La falsa idea de inmaterialidad también oculta que la cadena de valor de la IA demanda trabajo humano. A esta cadena debe incorporarse, además, el trabajo humano que sostiene a la inteligencia artificial. Miceli y otros. (2025) advierten que la industria global de IA depende de trabajo oculto y precarizado: personas que recolectan y anotan datos, moderan contenido, verifican salidas algorítmicas, mantienen centros de datos, extraen minerales críticos y producen contenidos — textos, traducciones, imágenes, actuaciones o materiales culturales— que posteriormente alimentan sistemas de IA generativa. Desde esta perspectiva, el trabajo no es un elemento periférico ni un “impacto” posterior de la automatización, sino una condición constitutiva de la IA: sin trabajo humano, no hay datos limpios, modelos entrenados, infraestructura operativa ni sistemas funcionales.

Esto obliga a ampliar la pregunta regulatoria. Con frecuencia, el debate sobre IA y empleo se concentra en los puestos que podrían ser sustituidos por sistemas automatizados; sin embargo, esa mirada deja fuera el trabajo previo y permanente que hace posible la automatización. Las tareas de etiquetado, clasificación, moderación, validación, mantenimiento e incluso extracción de minerales suelen realizarse en condiciones de subcontratación, dispersión geográfica, baja remuneración, inestabilidad laboral, exposición psicológica a contenidos dañinos y escasa capacidad de organización colectiva. En consecuencia, una regulación sustentable de la IA debe incluir obligaciones laborales y de trazabilidad sobre toda la cadena de valor, no solo reglas sobre el funcionamiento final del sistema, como ya se dijo.

Además, Miceli también propone que las personas trabajadoras no sean tratadas únicamente como objetos de estudio o como población afectada, sino como actores con autoridad epistémica sobre las condiciones reales de producción de la IA. Esta idea es clave para la regulación: si quienes sostienen la infraestructura de datos, moderación, mantenimiento y extracción conocen de primera mano los daños, riesgos y formas de explotación de la cadena, por lo que sus experiencias deben formar parte del diseño regulatorio, de las auditorías, de las evaluaciones de impacto y de los mecanismos de rendición de cuentas. Regular la IA desde la sostenibilidad implica, entonces, reconocer que sus costos no son solo energéticos o ambientales, sino también laborales, corporales, psicosociales y políticos.

Desde la sostenibilidad, esta discusión es central porque permite identificar costos que no siempre aparecen en los balances de innovación. La IA puede mejorar procesos, reducir tiempos y apoyar decisiones complejas, pero esas ventajas no deben evaluarse separadas de su consumo energético, su demanda hídrica, su dependencia mineral, su huella territorial, su generación de residuos y las condiciones laborales que la sostienen. La pregunta no es solo qué tan eficiente es un sistema y qué tantos beneficios pueden aportar a otro, sino qué costos materiales y sociales requiere para ello.

Por ello, una regulación sostenible de la inteligencia artificial debe tratarla como infraestructura socioambiental y no solo como producto digital. Esto implica exigir información verificable sobre consumo energético, uso de agua, emisiones, procedencia de insumos, condiciones laborales, ubicación de infraestructura y manejo de residuos. También implica evitar que los costos de operación de la IA se trasladen indirectamente a la ciudadanía mediante presión sobre servicios públicos, tarifas, disponibilidad de agua o afectaciones territoriales asociadas a los centros de datos.

Estas tensiones ya no son hipotéticas. En España, el crecimiento acelerado de los centros de datos ha abierto preguntas sobre opacidad, capacidad eléctrica, consumo hídrico y competencia por recursos. Tuñas Matilla (2025) señala que en el país existen entre 100 y casi 200 centros de procesamiento de datos operativos, en construcción o planificados, pero no siempre hay información pública suficiente sobre su capacidad, consumos o impactos acumulados. Esta falta de transparencia dificulta planificar su despliegue sin comprometer el suministro de energía y agua para otros usos sociales, productivos y ambientales.

La tensión territorial se expresa, sobre todo, en la demanda de electricidad, agua y suelo. Los centros de datos requieren energía constante, redes robustas y, en muchos casos, agua para refrigeración. Aunque algunas empresas anuncian sistemas de circuito cerrado, eficiencia energética o estrategias de “agua positiva”, estas medidas no eliminan por sí mismas la presión sobre territorios con estrés hídrico, infraestructura limitada o competencia entre usos urbanos, agrícolas e industriales (Tuñas Matilla, 2025). Por ello, el problema no es solo cuánta energía o agua consume cada instalación, sino cómo se distribuyen esos consumos, quién los autoriza, quién se beneficia y quién asume los costos.

En América Latina, estas tensiones son mayores por la desigualdad en el acceso a infraestructura, agua, energía y capacidades regulatorias. Sierra Praeli (2025) documenta conflictos vinculados con centros de datos en países como Brasil, México y Chile, donde comunidades y especialistas han cuestionado la presión sobre recursos hídricos, redes eléctricas, territorio e información pública. De acuerdo con la Agencia Internacional de Energía, los centros de datos consumen entre el 1.5 % y 2 % de la electricidad a nivel mundial y se calcula que para 2030 esta demanda se duplicará. Por su parte, algunos centros de gran escala pueden utilizar entre uno y cinco millones de litros de agua al día, lo que equivale al consumo diario de miles de hogares.

Desde esta perspectiva, la infraestructura de IA produce tensiones territoriales al menos en cuatro planos: compite por electricidad en redes ya presionadas; demanda agua en contextos de estrés hídrico; transforma usos de suelo; y genera conflictos por falta de información, consulta y beneficios locales. Por ello, una regulación sostenible debe exigir evaluaciones ambientales, sociales y territoriales previas, así como información pública sobre consumo energético, uso de agua, emisiones, ubicación, capacidad instalada, beneficios locales y mecanismos de participación comunitaria para generar políticas button up que involucren a la población en la toma de decisiones con respecto a la infraestructura, desde su instalación, durante su tiempo de vida útil, el desmontaje y el tratamiento adecuado de residuos. Si la regulación omite estas tensiones, esta tecnología con tanto potencial de beneficio está condenada a generar más problemas que soluciones, o como dice el refrán mexicano *“salir más caro el remedio que la enfermedad”*.

A todo esto, uno de los elementos centrales de esta cadena es el dato. Pero ¿qué pasa con los datos? Un dato puede entenderse como la expresión codificada de una variable cualitativa o cuantitativa que describe *algo* y que, al organizarse y procesarse, se convierte en información útil para el análisis. En informática y sistemas, se entiende por dato al símbolo que representa propiedades de objetos y eventos con propiedades básicas sin refinar ni filtrar que deben procesarse. (Torregrosa, 2022). Muchos sistemas de IA se entrenan con grandes volúmenes de información generada por personas, instituciones, comunidades, plataformas digitales, archivos públicos, producción cultural, investigación científica y actividad cotidiana en línea. En ese sentido, la aparente gratuidad de ciertos servicios digitales debe examinarse críticamente: lo que no se paga con dinero puede pagarse con datos, atención, comportamiento, producción intelectual o exposición a nuevas formas de vigilancia y perfilamiento.

Sin embargo, también es necesario visibilizar una asimetría fundamental: quienes han incorporado el análisis de datos a su quehacer profesional representan apenas

una fracción mínima frente al universo de personas que los generan. La producción de datos es masiva, cotidiana y muchas veces involuntaria; pero su interpretación, procesamiento y aprovechamiento estratégico requiere formación especializada, alfabetización estadística y computacional, capacidades de programación, acceso a infraestructura digital y poder de cómputo. Esto vuelve problemática la promesa de que la IA “democratizará” automáticamente el acceso al conocimiento. En contextos de desigualdad, brecha digital y pobreza energética¹⁷, el simple hecho de generar datos no implica tener capacidad para analizarlos, gobernarlos o beneficiarse de ellos. Por el contrario, puede profundizar la distancia entre quienes producen datos y quienes cuentan con los medios técnicos, económicos e institucionales para convertirlos en valor.

Esta discusión dialoga con Saifer y Dacin (2022), quienes advierten sobre el “imperativo de los datos” y la “fetichización de los datos” en las organizaciones. Para estos autores, los datos no deben entenderse únicamente como números, estadísticas o evidencia objetiva; también operan estética, emocional y discursivamente en la vida social y organizacional. La sola invocación de “lo basado en datos” puede producir confianza, legitimidad y autoridad, incluso cuando los datos no son accesibles, comprensibles o verificables para la mayoría de las personas. De ahí que sea necesario evitar tanto el culto tecnocrático al dato como el rechazo anticientífico: el desafío consiste en construir una relación politizada, histórica y democrática con los datos, capaz de servir a la justicia social, económica y ecológica.

Por ello, la regulación de la IA debe preguntarse no solo qué datos se usan, sino cómo se producen, clasifican, limpian, interpretan y reutilizan. ¿En qué contextos

¹⁷ . Según Thomson y otros. (2022), la pobreza energética se define como “la incapacidad de alcanzar un nivel de servicios energéticos domésticos social y materialmente necesario. De acuerdo con el artículo *Reframing how we talk about ‘energy poverty’* se reconoce la pobreza energética como resultado de estructuras históricas, políticas y territoriales de injusticia (Bouzarovski y otros., 2025).

se levantan encuestas o se capturan registros? ¿Quién nombra y categoriza los datos? ¿Qué información queda fuera y por qué? ¿Qué sesgos se introducen durante la limpieza? ¿Qué marcos éticos orientan estas decisiones? Estas preguntas son centrales porque los datos no son neutrales ni “naturales”: son contruidos en contextos sociales, técnicos y políticos específicos. Como señalan Saifer y Dacin (2022), el discurso de los datos puede despolitizar problemas sociales al presentarlos como asuntos meramente técnicos, administrables por expertos y sistemas automatizados. A esto se suma que las personas usuarias no solo consumen productos generados por IA, sino que también producen nuevos datos que alimentan la mejora, expansión y rentabilidad de esos sistemas.

Sin mecanismos claros de trazabilidad, consentimiento, transparencia y límites de uso, resulta difícil saber cómo se recolectan los datos, con qué fines se reutilizan, durante cuánto tiempo se conservan, qué actores acceden a ellos y qué beneficios económicos generan. Frente a ello, una regulación de la IA debe contar con elementos para reconocer el trabajo involucrado (y por lo tanto, retribuirlo). Debe exigir apertura, transparencia y trazabilidad de los datos, así como los derechos laborales de quienes sostienen la infraestructura de datos como los derechos de autoría, reconocimiento y remuneración de quienes alimentan indirectamente estos sistemas. Asimismo, la discusión regulatoria debe promover el desarrollo científico crítico capaz de revisar los sesgos, doctrinas y relaciones de poder involucradas en la inteligencia artificial.

Conclusión

Hasta ahora hemos señalado algunos elementos que abonan a la caracterización de una regulación sostenible que de manera integral analice la cadena de valor de la IA para poder formular preguntas y objetivos de carácter sistémico, sean estos: quién produce los datos, quién los controla, quién trabaja para entrenar los sistemas, qué recursos naturales se consumen, qué territorios soportan la

infraestructura, quién captura las ganancias y quién asume los daños. O lo que es lo mismo, corregir la tendencia actual que busca desplazar la responsabilidad al usuario final omitiendo pasar factura a los capitalistas tecnológicos. Este desplazamiento es clave porque permite pasar de una regulación centrada en el “buen uso” de la tecnología a una regulación orientada a las condiciones de producción, despliegue y apropiación de valor.

Además, es importante decir que la regulación de la IA requiere de un enfoque global. Un marco normativo limitado a una sola región con fronteras rígidas resulta insuficiente y riesgoso, pues puede favorecer el desplazamiento de costos socioecológicos hacia países con regulaciones laxas, capacidades institucionales limitadas o vacíos normativos. Si solo se considera el consumo, siempre existirán vías legales o ilegales para utilizar la inteligencia artificial con fines destructivos, como la detección automatizada de objetivos en conflictos armados, la violencia digital y de género mediante *deepfake porn*, o la producción masiva de desinformación mediante *deepfakes*.

Analizar globalmente para actuar localmente, como ya apuntaba Patrick Geddes a principios del siglo XX. Además, se deben incorporar criterios de justicia distributiva, no solo criterios de seguridad técnica o eficiencia operativa. Gabriel (2022) sostiene que, cuando la IA se integra en instituciones y prácticas centrales de la sociedad, debe entenderse como parte de la estructura sociotécnica que distribuye oportunidades, cargas, derechos y beneficios. Una IA orientada al interés público tendría que ser públicamente justificable, proteger derechos y producir resultados sustantivamente justos, especialmente para las personas y comunidades desfavorecidas por el desarrollo desigual.

Por supuesto que la regulación de la inteligencia artificial no puede aspirar a la perfección, pero sí debe ser holística y dinámica. Entonces, cual sistema de control cuyo objetivo sea el bien común, debe regularse con preguntas iterativas sobre qué

puede hacer, para quién se hace, bajo qué condiciones, a costa de qué y de quién. Si sus beneficios se concentran mientras sus cargas se dispersan, la IA puede convertirse en una herramienta que está llamada a profundizar las desigualdades que ya existen.

Obligaciones de transparencia, redistribución de costos y beneficios, protección laboral, evaluación de viabilidad territorial y control sin prohibición sobre las condiciones de despliegue tecnológico son algunas posibles herramientas para que la tecnología no sea un arma de doble filo. Para volverlas realidad al menos se necesita en un primer momento, la discusión pública a todos los niveles que decante, en un segundo momento, en diagnósticos integrales de la cadena de valor que den luz a criterios concretos para orientar una regulación de la IA compatible con un mundo que apueste por la vida en todas sus formas.

La regulación de la IA debe ser también dinámica y apostarse a la creatividad. Pensar en un futuro lejano hoy día puede ser pensar en el día de mañana. La regulación debe descansar también en bases claras sobre los documentos que regulan los espacios fuera de la tierra, ya que habrá centros de datos que orbiten e impacten directamente sobre el planeta. Si la infraestructura de la inteligencia artificial es posible gracias al trabajo (reconocido y no reconocido) de millones de personas, así como a los recursos naturales de *nuestra casa común*, sería indispensable que también se lleve a la discusión pública quiénes podrán acceder a sus beneficios futuros.

En los últimos años, se ha demostrado la facilidad con la que grandes potencias violan el Derecho Internacional, terribles casos de esto son Gaza, Venezuela, Cuba y Líbano, lo que sienta un precedente para toda regulación y legislación en la escala global. En estos términos, habrá que pensar cómo legitimarla para que no sean esfuerzos en vano. Como primer ejercicio, sería importante depurar los instrumentos internacionales regulatorios que han demostrado ser obsoletos en el beligerante contexto actual y aprender de esas experiencias, para que usuarios y

trabajadores de la IA también podemos defender la sostenibilidad de esta herramienta tan útil y no vernos obligados a elegir entre vida digna o vida artificial. Si la IA es una tecnología que puede producir bienestar a la humanidad, entonces su regulación, en favor de todas las personas y seres vivos que habitan el planeta, en clave sostenible, está más que justificada.

Referencias

Administración del Ciberespacio de China. (2023). Generative Artificial Intelligence Service Management Measures. Recuperado de https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

Alcántara Gutiérrez, M. E., Sánchez Linares, É., Aguirre Hernández, I., Villafán Amezcua, L. A., Jarrín Rivadeneira, E. P., & León Torres, L. M. (2024). Inteligencia artificial y su impacto en la toma de decisiones. Estudios y Perspectivas. Revista Científica y Académica, 4(4), 1073–1089.

Aithor. (2024, 22 de junio). El futuro de las tecnologías emergentes y su impacto en la sociedad. Recuperado de <https://aithor.com/essay-examples/el-futuro-de-las-tecnologias-emergentes-y-su-impacto-en-la-sociedad#10-políticas-públicas-y-regulaciones-en-el-contexto-de-tecnologías-emergentes>

Banco Mundial. (2021). El escaso acceso digital frena a América Latina y el Caribe ¿Cómo solucionar este problema? Blogs del Banco Mundial. <https://blogs.worldbank.org/es/latinamerica/el-escaso-acceso-digital-frena-america-latina-y-el-caribe-como-solucionar-este>

Brundtland, G. (1987). Informe de la Comisión Mundial sobre Medio Ambiente y Desarrollo: Nuestro Futuro Común. Documento A/42/427 de la Asamblea General de las Naciones Unidas.

Bouzarovski, S., Cedano-Villavicencio, K.G., Delina, L.L. y otros. Reframing how we talk about ‘energy poverty’. Nat Energy 10, 806–810 (2025). <https://doi.org/10.1038/s41560-025-01794-w>

Cadierno Redondo, P. (2025). El impacto regulatorio de la inteligencia artificial en la sostenibilidad. Universidad Pontificia Comillas, Repositorio

Institucional

<https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/91623/TFG%20-%20Cadierno%20Redondo,%20Pablo.pdf?sequence=2>

Carrera, A. (2026). Mirar el bosque: IA, centros de datos e impactos ambientales en América Latina. SocialTIC. Recuperado de <https://socialtic.org/blog/ia-centros-de-datos-e-impactos-ambientales-en-latam/>

Comisión Europea. (s.f.). Marco regulatorio para la inteligencia artificial. Estrategia Digital de la Unión Europea. Recuperado de <https://digital-strategy.ec.europa.eu/es/policies/regulatory-framework-ai>

Foster, John & Clark, Brett & York, Richard. (2010). Capitalism and the Curse of Energy Efficiency: The Return of the Jevons Paradox. Monthly Review. 62. 1-12. 10.14452/MR-062-06-2010-10_1.

Collett, C., Neff, G., & Gouvea Gomes, L. (2022). Los efectos de la IA en la vida laboral de las mujeres. UNESCO; Banco Interamericano de Desarrollo; Organización para la Cooperación y el Desarrollo Económicos. ISBN 978-92-3-300179-4.

Donoso Cáceres, A. (2022). Extractivismo: Abusos e injusticias en las “zonas de sacrificio”. Instituto de Investigaciones Jurídicas, Universidad Nacional Autónoma de México. Recuperado de <https://archivos.juridicas.unam.mx/www/bjv/libros/15/7266/3.pdf>

FUNIBER. (2025). El impacto ambiental de los centros de datos de inteligencia artificial: desafíos de sostenibilidad. <https://www.funiber.blog/medio-ambiente/2025/12/11/el-impacto-ambiental-de-los-centros-de-datos-de-inteligencia-artificial-desafios-de-sostenibilidad>

García Retuerta, Ó., & García Retuerta, D. (2026, 12 de enero). La cadena de valor de la inteligencia artificial: Estrategias de autonomía para España. Ministerio de Defensa de España, CESEDEN. Recuperado de <https://www.defensa.gob.es/ceseden/-/ieeee/cadena-valor-ia-estrategias-espanna>

Gabriel, I. (2022). Toward a theory of justice for artificial intelligence. *Daedalus*, 151(2), 218–231. https://doi.org/10.1162/DAED_a_01911

Georgieva, K. (2024, 16 de enero). La economía mundial transformada por la inteligencia artificial ha de beneficiar a la humanidad. Fondo Monetario Internacional. Recuperado de <https://www.imf.org/es/blogs/articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity>

Gudynas, E. (2010). Desarrollo sostenible: Una guía básica de conceptos y tendencias hacia otra economía. *Otra Economía. Revista Latinoamericana de Economía Social y Solidaria*, 4(6), 43–59. Centro Latinoamericano de Ecología Social (CLAES).

Kiljunen, K. (1986). La división internacional del trabajo industrial y el concepto centro-periferia. *Revista de la CEPAL*, (30), 103–124. Comisión Económica para América Latina y el Caribe (CEPAL).

Martínez-Alier, J. (2007). Conflictos ecológicos y justicia ambiental. *Ecología política*, 33, 5–16. <https://dialnet.unirioja.es/servlet/articulo?codigo=2765210>

Miceli, M. (2024). How we made Data Workers' Inquiry. Data Workers' Inquiry. Recuperado de <https://data-workers.org/about/how-to-dwi/>

Miceli, M., Dinika, A.-A., Kauffman, K., Salim Wagner, C., Sachenbacher, L., Hanna, A., & Gebru, T. (2025). Methodological considerations for centering workers' epistemic authority in AI research. En Proceedings of the AAAI Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence. https://data-workers.org/wp-content/uploads/2025/08/DWI_Method_AIES-6.pdf

Moreno, A. (2025). Tecnología sin soberanía: el flujo de datos en América Latina. <https://soberaniadigital.org/tecnologia-sin-soberania-el-flujo-de-datos-en-america-latina/>

Naciones Unidas. (1987). Nuestro futuro común (Informe Brundtland). Informe de la Comisión Mundial sobre el Medio Ambiente y el Desarrollo. Nueva York: Naciones Unidas. https://www.ecominga.ugam.ca/PDF/BIBLIOGRAPHIE/GUIDE_LECTURE_1/CMMAD-Informe-Comision-Brundtland-sobre-Medio-Ambiente-Desarrollo.pdf

Ordellin Font, J. L. (2025). Derechos de autor y entrenamiento de sistemas de IA generativos: Las obligaciones de transparencia y la minería de textos y datos en la normativa europea. IDP. Revista de Internet, Derecho y Política, (42), 1–13. <https://doi.org/10.7238/idp.v0i42.431327>

Ostrom, E. (2009). El gobierno de los bienes comunes: La evolución de las instituciones de acción colectiva. Universidad Nacional Autónoma de México, Centro Regional de Investigaciones Multidisciplinarias; Fondo de Cultura Económica.

Pacto Mundial de la ONU España. (s.f.). Inteligencia artificial y empresas. Recuperado de <https://www.pactomundial.org/biblioteca/inteligencia-artificial-y-empresas/>

Raihan, A., Arindrajit Paul, Rahman, M. S., Islam, S., Paul, P., & Karmakar, S. (2024). Artificial Intelligence (AI) for Environmental Sustainability: A Concise Review of Technology Innovations in Energy, Transportation, Biodiversity, and Water Management. *Journal of Technology Innovations and Energy*, 3(2), 64–73. <https://doi.org/10.56556/jtie.v3i2.953>

Redclift, M. R. (2006). Desarrollo sostenible (1987–2005): Un oxímoron que alcanza la madurez. *Horizontes Antropológicos*, 12(25), 65–84.

Rodríguez Lacherre, M. (2018). SOSTENIBLE O SUSTENTABLE. *Sagasteguiana*, 1(2), 91-92. Recuperado a partir de <https://revistas.unitru.edu.pe/index.php/REVSAGAS/article/view/1798>

Saifer, A., & Dacin, M. T. (2022). Data and organization studies: Aesthetics, emotions, discourse and our everyday encounters with data. *Organization Studies*, 43(4), 623–636. <https://doi.org/10.1177/01708406211006250>

Sierra Praeli, Y. (2025, 11 de marzo). Inteligencia artificial: crecen los conflictos alrededor de los centros de datos por el agua, la energía y el territorio. *Mongabay Latam*. <https://socialtic.org/blog/ia-centros-de-datos-e-impactos-ambientales-en-latam/>

Siqueira, M. B., Santos, V. M. dos, Diniz, E. H., & Cruz, A. P. A. (2024). Artificial Intelligence for Sustainability: A Systematic Literature Review in Information

Soja, E. W. (2010). Seeking Spatial Justice. Minneapolis: University of Minnesota Press.
https://www.researchgate.net/publication/229770913_Seeking_Spatial_Justice
- By EDWARD SOJA

S&P Global Ratings. (2025). Rápido crecimiento de los centros de datos afronta desafíos de sostenibilidad: aumento de las emisiones y del estrés hídrico.
<https://www.spglobal.com/ratings/es/articulos/rapido-crecimiento-de-los-centros-de-datos-afronta-desafios-de-sostenibilidad-aumento-de-las-emisiones-y-del-estrés-hídrico>

Systems. Revista De Gestão Social E Ambiental - RGSA, 18(3), e07885.
<https://doi.org/10.24857/rgsa.v18n3-178>

Thomson, H., Day, R., Ricalde, K., Brand-Correa, L. I., Cedano, K., Martinez, M., Santillán, O., Delgado Triana, Y., Luis Cordova, J. G., Milian Gómez, J. F., Garcia Torres, D., Mercado, C., Castelao Caruana, M. E., & Giannini Pereira, M. (2022). Understanding, recognizing, and sharing energy poverty knowledge and gaps in Latin America and the Caribbean – because conocer es resolver. Energy Research & Social Science, 87, 102475.
<https://doi.org/10.1016/j.erss.2021.102475>

Torregrosa García, B. (2022). Introducción a los datos (PID_00272095). Universitat Oberta de Catalunya. Recuperado de
<https://openaccess.uoc.edu/server/api/core/bitstreams/fd2d7eba-e2a0-4ea1-b877-2cfba2791c1b/content>

Tuñas Matilla, A. (2025, 22 de diciembre). Centros de datos, ¿hay energía y agua para todos sin comprometer otros usos? EFE Verde.

<https://efeverde.com/centros-de-datos-hay-energia-y-agua-para-todos-sin-comprometer-otros-usos/>

UNAM Global. (2023). Inteligencia artificial y desigualdades: ¿cómo impacta la IA en México y el mundo? Universidad Nacional Autónoma de México. Recuperado de https://unamglobal.unam.mx/global_revista/ia-desigualdades-unam/

United Nations. (n.d.). Artificial intelligence. United Nations. <https://www.un.org/es/global-issues/artificial-intelligence>

Universidad Nacional Autónoma de México. (s.f.). Sostenibilidad y sustentabilidad. Biblioteca Digital de la UNAM (Bunam). Recuperado de https://uapa.bunam.unam.mx/u/ciencias/sostenibilidad_sustentabilidad/

Villagómez Palacios, Á. H. (2025). El impacto de la inteligencia artificial en la sociedad: Una revisión sistemática de su influencia en ámbitos sociales, económicos y tecnológicos. Ciencia Latina Revista Científica Multidisciplinar, 9(1). https://doi.org/10.37811/cl_rcm.v9i1.16468

Walter, M. (2009). Conflictos ambientales, socioambientales, ecológico-distributivos.... Centro de Estudios de Trabajo y Organización, Fuhem. https://www.fuhem.es/media/ecosocial/File/Boletin%20ECOS/Boletin%206/Conflictos%20ambientales_M.WALTER_mar09_final.pdf

Wang, Q., Li, Y., & Li, R. (2024). Ecological footprints, carbon emissions, and energy transitions: The impact of artificial intelligence (AI). Humanities & Social Sciences Communications, 11(1043). Springer Nature. <https://doi.org/10.1057/s41599-024-03520-5>

Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136.
The MIT Press on behalf of the American Academy of Arts & Sciences.
<http://www.jstor.org/stable/20024652>

CAPÍTULO XXIX

GOBERNANZA ALGORÍTMICA Y TRABAJO DIGNO: HACIA UN MARCO REGULATORIO DE LA INTELIGENCIA ARTIFICIAL EN LAS RELACIONES LABORALES CONTEMPORÁNEAS

Algorithmic governance and decent work: towards a regulatory framework
for artificial intelligence in contemporary labour relations

Aylin Torres Ramírez

aytora@azc.uam.mx

Universidad Autónoma Metropolitana

<https://orcid.org/0000-0001-6300-1666>

Resumen

Este capítulo analiza el impacto de la inteligencia artificial en las relaciones laborales contemporáneas desde una perspectiva organizacional, jurídica y crítica, centrándose en la transformación de los mecanismos tradicionales de supervisión hacia modelos de gestión algorítmica sustentados en automatización, monitoreo digital y toma automatizada de decisiones. A partir de enfoques provenientes de la teoría organizacional, los estudios críticos del trabajo y la regulación tecnológica, se examina cómo los sistemas de inteligencia artificial reconfiguran las dinámicas de poder, control y subordinación dentro de las organizaciones mediante procesos de vigilancia permanente, opacidad algorítmica y análisis masivo de datos. El capítulo sostiene que la expansión de la gestión algorítmica no representa únicamente una innovación tecnológica, sino una profundización digital de modelos históricos de control organizacional que generan tensiones entre eficiencia empresarial, dignidad humana y derechos fundamentales de los trabajadores. Asimismo, se argumenta que los marcos tradicionales del derecho laboral resultan insuficientes frente a estas nuevas formas de subordinación tecnológica, por lo que se propone la construcción de derechos laborales digitales orientados a garantizar

transparencia algorítmica, revisión humana significativa, límites a la vigilancia digital y protección frente a decisiones automatizadas arbitrarias. Finalmente, se concluye que la regulación de la inteligencia artificial en el trabajo constituye no solo un desafío técnico-jurídico, sino también una disputa ética y política sobre los límites democráticos del poder algorítmico dentro de las organizaciones contemporáneas.

Palabras clave: Inteligencia artificial, gestión algorítmica, derechos laborales digitales, vigilancia algorítmica, transparencia algorítmica, revisión humana, automatización laboral, organizaciones digitales, control organizacional, opacidad algorítmica, dignidad laboral, gobernanza algorítmica, trabajo digital, regulación tecnológica, relaciones laborales.

Abstract:

This chapter analyzes the impact of artificial intelligence on contemporary labor relations from an organizational, legal, and critical perspective, focusing on the transformation of traditional supervision mechanisms into models of algorithmic management based on automation, digital monitoring, and automated decision-making. Drawing on approaches from organizational theory, critical labor studies, and technology regulation, the chapter examines how artificial intelligence systems reconfigure dynamics of power, control, and subordination within organizations through processes of permanent surveillance, algorithmic opacity, and massive data analysis. The study argues that the expansion of algorithmic management does not merely represent a technological innovation, but rather a digital intensification of historical models of organizational control that generate tensions between business efficiency, human dignity, and workers' fundamental rights. Likewise, it contends that traditional labor law frameworks are insufficient to address these new forms of technological subordination, proposing instead the development of digital labor rights aimed at guaranteeing algorithmic transparency, meaningful human review, limits on digital surveillance, and protection against arbitrary automated decisions. Finally, the chapter concludes that regulating artificial intelligence in the workplace

constitutes not only a technical-legal challenge, but also an ethical and political dispute over the democratic limits of algorithmic power within contemporary organizations.

Keywords: artificial intelligence, algorithmic management, digital labor rights, algorithmic surveillance, algorithmic transparency, human review, labor automation, digital organizations, organizational control, algorithmic opacity, labor dignity, algorithmic governance, digital work, technology regulation, labor relations.

Introducción

La expansión acelerada de la inteligencia artificial (IA) en los entornos organizacionales está transformando profundamente las relaciones laborales, las dinámicas de supervisión y las formas de control dentro de las organizaciones contemporáneas. A diferencia de procesos previos de automatización industrial centrados principalmente en tareas mecánicas, los sistemas actuales de inteligencia artificial intervienen directamente en procesos de contratación, evaluación del desempeño, asignación de tareas y monitoreo laboral. Este fenómeno ha dado origen a nuevas formas de administración basadas en sistemas automatizados de decisión y vigilancia digital permanente.

En este contexto emerge la denominada gestión algorítmica (*algorithmic management*), entendida como el uso de algoritmos y sistemas automatizados para coordinar, supervisar y evaluar trabajadores. De acuerdo con Lee y otros. (2015), la gestión algorítmica implica que “las funciones administrativas tradicionalmente realizadas por supervisores humanos son delegadas a sistemas algorítmicos”, transformando significativamente las relaciones de autoridad dentro de las organizaciones. Estas transformaciones generan nuevas formas de control laboral sustentadas en monitoreo constante, evaluación automatizada y recopilación masiva de datos sobre el comportamiento de los trabajadores.

Desde la perspectiva de los estudios organizacionales, la incorporación de inteligencia artificial en los espacios laborales no representa únicamente un cambio tecnológico, sino una reconfiguración estructural del poder organizacional. Kellogg, Valentine y Christin (2020) sostienen que los sistemas algorítmicos están modificando los mecanismos tradicionales de supervisión laboral mediante formas automatizadas de control que reducen la autonomía de los trabajadores y fortalecen dinámicas de vigilancia permanente. En consecuencia, los trabajadores quedan subordinados a sistemas automatizados cuyos criterios de decisión frecuentemente resultan opacos e incomprensibles.

Paralelamente, diversos estudios han advertido que los sistemas de inteligencia artificial utilizados en procesos laborales pueden reproducir y amplificar desigualdades sociales existentes. O'Neil (2016) argumenta que muchos algoritmos empleados en procesos organizacionales operan como “armas de destrucción matemática”, reproduciendo sesgos discriminatorios relacionados con género, raza y condición socioeconómica. Esto resulta particularmente problemático en sistemas automatizados de reclutamiento y evaluación laboral, donde decisiones aparentemente objetivas pueden perpetuar exclusiones estructurales.

Asimismo, el crecimiento de mecanismos de vigilancia digital dentro de las organizaciones ha generado profundas preocupaciones éticas y jurídicas relacionadas con privacidad, autonomía y dignidad humana. Zuboff (2019) advierte que el capitalismo contemporáneo ha evolucionado hacia formas de “capitalismo de vigilancia”, donde la recopilación masiva de datos permite incrementar el control corporativo sobre las conductas humanas, incluyendo los espacios laborales.

A pesar de estos desafíos, los marcos tradicionales del derecho laboral continúan mostrando limitaciones significativas frente a la expansión de sistemas automatizados de gestión organizacional. Gran parte de las regulaciones laborales actuales fueron diseñadas para relaciones laborales mediadas por supervisión

humana directa y no para contextos caracterizados por toma automatizada de decisiones y vigilancia algorítmica permanente.

Frente a este escenario, resulta necesario avanzar hacia la construcción de derechos laborales digitales orientados a proteger a los trabajadores frente al creciente poder algorítmico organizacional. Este capítulo sostiene que la regulación de la inteligencia artificial en el trabajo debe centrarse no solo en el problema del desplazamiento laboral, sino también en la defensa de la dignidad humana, la transparencia y la justicia dentro de entornos laborales automatizados. Para ello, se analizan cinco dimensiones fundamentales: el derecho a explicación algorítmica, el derecho a revisión humana, el derecho a desconexión de vigilancia algorítmica, las auditorías laborales algorítmicas y la transparencia en sistemas automatizados de recursos humanos.

La transformación algorítmica de las relaciones laborales

La incorporación de sistemas de inteligencia artificial en los entornos organizacionales contemporáneos ha impulsado una transformación profunda en las formas de administrar, supervisar y controlar el trabajo. Actualmente, numerosas organizaciones utilizan sistemas algorítmicos para gestionar procesos de reclutamiento, evaluación del desempeño, asignación de tareas, monitoreo de productividad y predicción de comportamientos laborales. Esta transición ha dado paso a modelos de administración sustentados en análisis masivo de datos, automatización predictiva y vigilancia digital continua, configurando nuevas dinámicas de poder y subordinación dentro de las organizaciones.

La denominada gestión algorítmica (*algorithmic management*) implica la transferencia de funciones tradicionalmente desempeñadas por supervisores humanos hacia sistemas automatizados capaces de tomar decisiones mediante modelos predictivos y procesamiento de datos (Lee y otros., 2015). Este fenómeno

representa no solo una innovación tecnológica, sino una transformación estructural en las formas de organización del trabajo y en los mecanismos mediante los cuales se ejerce autoridad dentro de las organizaciones.

Desde la perspectiva clásica de la Teoría de la Organización, esta transformación puede entenderse como una nueva etapa en la evolución de los mecanismos de control organizacional. Desde principios del siglo XX, autores como Frederick Winslow Taylor ya planteaban la necesidad de racionalizar el trabajo mediante métodos científicos orientados a maximizar eficiencia y productividad. En *The Principles of Scientific Management* (1911), Taylor defendía la fragmentación de tareas, la estandarización de procesos y el control sistemático del desempeño laboral como mecanismos para incrementar la eficiencia organizacional.

En muchos sentidos, la gestión algorítmica contemporánea representa una profundización digital del modelo taylorista. Mientras el taylorismo clásico buscaba controlar tiempos y movimientos del trabajador mediante supervisión directa y medición científica del trabajo, los sistemas algorítmicos actuales amplían esta lógica mediante monitoreo permanente, análisis predictivo y automatización de decisiones laborales. Como señalan Kellogg, Valentine y Christin (2020), los algoritmos permiten ejercer formas de control organizacional mucho más extensas y sofisticadas que los mecanismos tradicionales de supervisión jerárquica.

Sin embargo, la transformación algorítmica del trabajo también entra en tensión con los principios desarrollados por la escuela de las Relaciones Humanas. Autores como Elton Mayo demostraron, a partir de los estudios de Hawthorne realizados en la década de 1920, que las relaciones sociales, el reconocimiento y los factores emocionales desempeñan un papel fundamental en la productividad y bienestar de los trabajadores (Mayo, 1933). Desde esta perspectiva, las organizaciones no pueden comprenderse únicamente como estructuras técnicas orientadas a maximizar eficiencia, sino también como sistemas sociales donde las dinámicas humanas influyen profundamente en el comportamiento organizacional.

La expansión de sistemas automatizados de supervisión puede debilitar precisamente estos componentes humanos de las organizaciones. La sustitución de interacción humana por evaluación algorítmica reduce espacios de comunicación interpersonal, reconocimiento emocional y negociación social dentro del trabajo. En lugar de relaciones construidas mediante interacción humana, los trabajadores quedan sujetos a procesos automatizados impersonales cuyos criterios frecuentemente resultan desconocidos.

En este sentido, la crítica desarrollada por Douglas McGregor en *The Human Side of Enterprise* (1960) adquiere especial relevancia. McGregor diferenciaba entre la Teoría X, basada en supuestos de control rígido y desconfianza hacia los trabajadores, y la Teoría Y, centrada en autonomía, participación y desarrollo humano. La gestión algorítmica contemporánea parece acercarse nuevamente a modelos de control intensivo asociados con la Teoría X, donde los trabajadores son monitoreados constantemente bajo lógicas automatizadas de vigilancia y productividad.

Desde los estudios organizacionales contemporáneos, diversos autores advierten que la inteligencia artificial está reconfigurando las relaciones de poder dentro de las organizaciones. La autoridad ya no se ejerce exclusivamente mediante estructuras jerárquicas visibles, sino también a través de mecanismos automatizados de vigilancia permanente y toma algorítmica de decisiones. Esta transformación genera nuevas formas de subordinación tecnológica donde los trabajadores quedan sujetos a sistemas automatizados difíciles de comprender, cuestionar o impugnar.

Uno de los principales problemas de esta transformación radica en la opacidad algorítmica. En numerosos casos, los trabajadores desconocen cómo son evaluados, qué variables son utilizadas para medir desempeño o bajo qué criterios se toman decisiones que afectan contratación, promoción o permanencia laboral. Esta opacidad genera profundas asimetrías de información entre organizaciones y trabajadores, fortaleciendo formas invisibles de control organizacional.

Asimismo, el monitoreo digital continuo ha intensificado dinámicas de vigilancia laboral que afectan privacidad, autonomía y bienestar psicológico. Desde una perspectiva foucaultiana, algunos autores sostienen que la gestión algorítmica configura nuevas formas de disciplinamiento organizacional sustentadas en vigilancia permanente y autorregulación conductual inducida por sistemas tecnológicos (Ball, 2010).

En consecuencia, la transformación algorítmica de las relaciones laborales no debe analizarse únicamente como un proceso de innovación tecnológica, sino como una reconfiguración profunda de las estructuras organizacionales, las relaciones de poder y los mecanismos de control sobre el trabajo humano. Esto plantea importantes desafíos éticos, organizacionales y jurídicos relacionados con dignidad humana, autonomía laboral y protección de derechos en contextos crecientemente automatizados.

Derechos laborales digitales frente a la inteligencia artificial

La transformación algorítmica de las relaciones laborales ha evidenciado que los marcos tradicionales de protección laboral resultan insuficientes frente a las nuevas formas de control organizacional mediadas por inteligencia artificial. Como se ha señalado anteriormente, la incorporación de sistemas automatizados en los espacios de trabajo no solo modifica procesos operativos, sino que reconfigura profundamente las dinámicas de autoridad, supervisión y subordinación dentro de las organizaciones. En este escenario, los trabajadores enfrentan crecientes asimetrías de poder derivadas de sistemas automatizados capaces de tomar decisiones sobre contratación, evaluación, productividad y permanencia laboral mediante mecanismos frecuentemente opacos e inaccesibles.

Ante esta realidad, surge la necesidad de construir un marco de derechos laborales digitales orientado a responder a los desafíos derivados de la gestión algorítmica

del trabajo. Estos derechos representan una extensión contemporánea de las garantías históricas construidas por el derecho laboral frente a formas tradicionales de subordinación, pero adaptadas a contextos caracterizados por automatización organizacional, vigilancia digital y toma algorítmica de decisiones.

Desde una perspectiva histórica, el derecho laboral surgió como un mecanismo de protección frente a las desigualdades estructurales existentes entre empleadores y trabajadores durante el desarrollo industrial. Sin embargo, las transformaciones digitales contemporáneas han generado nuevas formas de subordinación tecnológica que exceden muchos de los supuestos sobre los cuales fueron construidas las regulaciones laborales tradicionales. Como señala Antonio Aloisi, las tecnologías digitales están redefiniendo las relaciones de trabajo mediante sistemas automatizados capaces de intensificar mecanismos de control y vigilancia sobre los trabajadores (Aloisi & De Stefano, 2022).

En este contexto, los derechos laborales digitales buscan limitar los riesgos asociados al ejercicio del poder algorítmico dentro de las organizaciones. Su objetivo no consiste únicamente en regular tecnologías emergentes, sino en garantizar que los procesos de automatización respeten principios fundamentales vinculados con dignidad humana, autonomía laboral, privacidad, igualdad y transparencia organizacional.

Uno de los principales desafíos derivados de la inteligencia artificial en el trabajo radica en la consolidación de formas invisibles de control organizacional. A diferencia de los modelos tradicionales de supervisión jerárquica, donde las relaciones de autoridad eran identificables y podían ser cuestionadas directamente, los sistemas algorítmicos operan mediante procesos automatizados cuya lógica interna frecuentemente permanece inaccesible para los trabajadores. Esta situación genera lo que algunos autores han denominado una “caja negra algorítmica”, donde decisiones fundamentales sobre la vida laboral son tomadas mediante sistemas difíciles de interpretar o impugnar (Pasquale, 2015).

La opacidad algorítmica no solo limita la capacidad de los trabajadores para comprender cómo son evaluados, sino que también dificulta la identificación de prácticas discriminatorias o arbitrarias. Diversas investigaciones han demostrado que los algoritmos utilizados en procesos organizacionales pueden reproducir sesgos estructurales presentes en los datos históricos utilizados para entrenar sistemas automatizados. Como advierte Cathy O'Neil (2016), muchos modelos algorítmicos amplifican desigualdades sociales bajo una apariencia de objetividad técnica, legitimando prácticas discriminatorias mediante procesos automatizados aparentemente neutrales.

Asimismo, la expansión de sistemas digitales de monitoreo ha intensificado prácticas de vigilancia laboral capaces de afectar derechos fundamentales relacionados con privacidad y autonomía personal. El monitoreo constante del desempeño, comportamiento y actividad digital de los trabajadores ha permitido a las organizaciones desarrollar mecanismos de supervisión mucho más invasivos que los modelos tradicionales de control laboral. Desde una perspectiva crítica, estas dinámicas reflejan la consolidación de nuevas formas de disciplinamiento organizacional sustentadas en extracción masiva de datos y vigilancia permanente (Zuboff, 2019).

Frente a estos riesgos, diversos organismos internacionales y especialistas en derecho laboral han comenzado a plantear la necesidad de reconocer nuevos derechos específicos frente a la automatización organizacional. Entre estos destacan el derecho a explicación algorítmica, el derecho a revisión humana en decisiones automatizadas, la protección frente a vigilancia digital excesiva y la obligación de transparencia en sistemas automatizados de recursos humanos.

Estos derechos buscan garantizar que la incorporación de inteligencia artificial en las organizaciones no implique la erosión de principios fundamentales históricamente vinculados con el trabajo digno. En este sentido, la discusión sobre derechos laborales digitales no debe interpretarse como un rechazo a la innovación

tecnológica, sino como un esfuerzo por establecer límites éticos y jurídicos frente al creciente poder algorítmico en los espacios laborales.

La construcción de estos derechos adquiere particular relevancia en un contexto donde las decisiones automatizadas tienen impactos directos sobre estabilidad laboral, bienestar psicológico, oportunidades profesionales y condiciones de trabajo. Por ello, el desafío contemporáneo no consiste únicamente en promover eficiencia tecnológica dentro de las organizaciones, sino en garantizar que los procesos de automatización permanezcan subordinados a principios democráticos, derechos fundamentales y mecanismos efectivos de protección laboral.

El derecho a explicación algorítmica como garantía de transparencia y justicia laboral

La consolidación de derechos laborales digitales frente a la inteligencia artificial encuentra uno de sus pilares fundamentales en el derecho a explicación algorítmica. Como se ha señalado en el apartado anterior, uno de los principales riesgos derivados de la gestión algorítmica del trabajo radica en la creciente opacidad de los sistemas automatizados utilizados dentro de las organizaciones. A medida que las empresas incorporan herramientas de inteligencia artificial en procesos relacionados con contratación, evaluación del desempeño, promociones, asignación de tareas y despidos, los trabajadores enfrentan dificultades cada vez mayores para comprender cómo y bajo qué criterios se toman decisiones que afectan directamente sus condiciones laborales.

Este fenómeno ha sido ampliamente discutido en la literatura contemporánea bajo el concepto de “opacidad algorítmica” (*algorithmic opacity*), entendido como la dificultad —e incluso imposibilidad— de conocer los procesos internos mediante los cuales un sistema automatizado produce decisiones o recomendaciones. Según Pasquale (2015), muchos sistemas algorítmicos operan como verdaderas “cajas

negras”, donde los individuos afectados desconocen los criterios utilizados para evaluar su comportamiento o determinar decisiones con consecuencias significativas sobre sus vidas.

En los entornos laborales, esta opacidad adquiere implicaciones particularmente sensibles debido a las profundas asimetrías de poder existentes entre organizaciones y trabajadores. Mientras las empresas poseen acceso y control sobre sistemas automatizados, los trabajadores frecuentemente desconocen qué datos son recopilados, cómo son procesados o qué variables determinan evaluaciones relacionadas con productividad, desempeño o permanencia laboral.

La opacidad algorítmica transforma así las relaciones laborales en dinámicas crecientemente inaccesibles para quienes se encuentran sujetos a ellas. Como advierte Burrell (2016), la complejidad técnica de muchos modelos algorítmicos dificulta que las personas afectadas comprendan no solo las decisiones finales, sino también la lógica subyacente que las produce. Esto genera escenarios donde los trabajadores pueden ser evaluados, clasificados o incluso despedidos mediante sistemas cuyo funcionamiento permanece prácticamente invisible.

En este contexto, el derecho a explicación algorítmica surge como una respuesta orientada a reducir las asimetrías de información producidas por la automatización organizacional. Este derecho implica que los trabajadores tengan acceso comprensible a los criterios generales utilizados por sistemas automatizados para tomar decisiones que afectan sus derechos, oportunidades y condiciones laborales.

Más allá de una simple exigencia técnica de transparencia, el derecho a explicación algorítmica constituye un mecanismo fundamental de protección frente al ejercicio arbitrario del poder algorítmico dentro de las organizaciones. Diversos especialistas en ética digital y regulación tecnológica sostienen que sin mecanismos adecuados de explicabilidad resulta imposible garantizar principios básicos de justicia, rendición de cuentas y debido proceso en contextos de toma automatizada de decisiones (Wachter, Mittelstadt & Floridi, 2017).

La relevancia de este derecho resulta particularmente evidente en procesos laborales sensibles como contratación, promoción, evaluación del desempeño, asignación salarial o despido. En ausencia de transparencia algorítmica, los trabajadores carecen de herramientas efectivas para identificar posibles errores, sesgos discriminatorios o decisiones arbitrarias derivadas de sistemas automatizados.

Numerosas investigaciones han demostrado que los algoritmos utilizados en sistemas de recursos humanos pueden reproducir y amplificar desigualdades estructurales presentes en los datos históricos utilizados para entrenar modelos predictivos. Un caso ampliamente citado es el sistema experimental de reclutamiento desarrollado por Amazon, el cual mostró sesgos sistemáticos contra mujeres debido a patrones históricos presentes en los datos utilizados para entrenar el algoritmo (Dastin, 2018). Este tipo de situaciones evidencia que la aparente neutralidad tecnológica puede ocultar mecanismos automatizados de discriminación difíciles de detectar sin acceso a información sobre funcionamiento algorítmico.

Desde el ámbito jurídico, el derecho a explicación algorítmica ha comenzado a adquirir relevancia dentro de debates internacionales sobre regulación de inteligencia artificial y protección de derechos fundamentales. Instrumentos regulatorios recientes, como el Reglamento General de Protección de Datos de la European Union (GDPR), han impulsado discusiones sobre el derecho de las personas a obtener información significativa respecto a decisiones automatizadas que produzcan efectos relevantes sobre sus vidas.

Aplicado al ámbito laboral, este derecho podría traducirse en obligaciones concretas para las organizaciones, entre ellas:

- informar explícitamente cuando se utilicen sistemas de inteligencia artificial en procesos laborales;

- proporcionar explicaciones comprensibles sobre criterios generales utilizados por algoritmos;
- notificar a los trabajadores cuando una decisión haya sido tomada mediante sistemas automatizados;
- establecer mecanismos de revisión e impugnación frente a decisiones algorítmicas;
- garantizar acceso a información sobre datos utilizados en evaluaciones automatizadas.

Estas medidas no buscan revelar necesariamente secretos industriales o códigos fuente completos de los sistemas utilizados por las organizaciones, sino garantizar niveles mínimos de transparencia compatibles con principios de justicia laboral y protección de derechos fundamentales.

Desde una perspectiva organizacional, la transparencia algorítmica también representa un elemento central para preservar confianza dentro de las relaciones laborales. Cuando los trabajadores desconocen cómo son evaluados o perciben que decisiones importantes son tomadas mediante mecanismos opacos e inaccesibles, pueden generarse dinámicas de incertidumbre, desconfianza y despersonalización organizacional.

En consecuencia, el derecho a explicación algorítmica constituye una condición indispensable para limitar formas invisibles de control organizacional y garantizar relaciones laborales compatibles con principios democráticos y trabajo digno. En contextos crecientemente mediados por inteligencia artificial, la posibilidad de comprender y cuestionar decisiones automatizadas representa no solo un problema técnico o jurídico, sino una cuestión fundamental relacionada con dignidad humana, autonomía laboral y equilibrio de poder dentro de las organizaciones.

El derecho a revisión humana frente a la automatización de decisiones laborales

La exigencia de transparencia y explicabilidad en los sistemas automatizados conduce necesariamente a una segunda dimensión fundamental de los derechos laborales digitales: el derecho a revisión humana. Si el derecho a explicación algorítmica busca reducir la opacidad de los sistemas automatizados utilizados en las organizaciones, el derecho a revisión humana pretende establecer límites al grado de autonomía decisional que puede delegarse a la inteligencia artificial en asuntos que afectan de manera significativa la vida laboral de las personas.

Uno de los riesgos más preocupantes derivados de la automatización organizacional radica precisamente en la posibilidad de transferir decisiones laborales críticas hacia sistemas algorítmicos capaces de operar sin intervención humana significativa. En numerosos entornos laborales contemporáneos, herramientas de inteligencia artificial ya participan activamente en procesos de selección de personal, evaluación del desempeño, asignación de productividad, establecimiento de métricas laborales e incluso recomendaciones vinculadas con despidos o sanciones disciplinarias.

Sin embargo, la creciente automatización de estos procesos plantea profundas interrogantes éticas y jurídicas sobre los límites aceptables de la toma automatizada de decisiones en contextos laborales. Como advierte Virginia Eubanks, la delegación de decisiones sociales a sistemas algorítmicos puede producir formas de exclusión y deshumanización donde las personas son reducidas a conjuntos de datos procesables (Eubanks, 2018).

En el ámbito laboral, esta problemática resulta especialmente delicada debido a que decisiones relacionadas con contratación, promoción, despido o reducción salarial afectan directamente estabilidad económica, bienestar psicológico y desarrollo profesional de los trabajadores. Permitir que tales decisiones dependan

exclusivamente de modelos automatizados implica trasladar procesos profundamente humanos hacia sistemas incapaces de comprender plenamente dimensiones contextuales, emocionales y sociales presentes en las relaciones laborales.

El derecho a revisión humana sostiene, en consecuencia, que ninguna decisión laboral con efectos significativos sobre un trabajador debería depender exclusivamente de sistemas automatizados sin intervención humana sustancial. Este principio parte del reconocimiento de que los algoritmos, aunque capaces de procesar grandes volúmenes de información, carecen de capacidades interpretativas necesarias para valorar adecuadamente circunstancias humanas complejas presentes en los entornos organizacionales.

Desde una perspectiva crítica, diversos autores han señalado que la automatización excesiva de decisiones laborales puede fortalecer procesos de deshumanización administrativa. Hannah Arendt (1958), aunque en un contexto distinto, advertía sobre los riesgos de estructuras burocráticas donde las decisiones se distancian progresivamente de consideraciones humanas y éticas concretas. En la actualidad, la automatización algorítmica puede profundizar esta lógica mediante sistemas capaces de ejecutar decisiones laborales sin espacios reales para deliberación humana o consideración contextual.

La necesidad de mantener intervención humana significativa también responde a los límites inherentes de los sistemas algorítmicos. Como señalan Selbst y Barocas (2018), los algoritmos operan mediante simplificaciones técnicas de realidades sociales complejas, transformando fenómenos humanos en variables cuantificables que frecuentemente omiten dimensiones fundamentales del comportamiento social y organizacional. En consecuencia, decisiones automatizadas pueden generar resultados injustos cuando ignoran circunstancias específicas imposibles de capturar completamente mediante datos estructurados.

Asimismo, múltiples investigaciones han demostrado que los sistemas automatizados pueden reproducir errores y sesgos discriminatorios derivados tanto

de datos históricos como de decisiones de diseño tecnológico. En este sentido, la revisión humana funciona también como un mecanismo de protección frente a errores algorítmicos capaces de afectar injustamente a trabajadores.

Desde el ámbito regulatorio internacional, el principio de intervención humana significativa ha comenzado a consolidarse como uno de los ejes centrales en debates sobre inteligencia artificial y derechos fundamentales. El Reglamento General de Protección de Datos (GDPR) de la European Union establece protecciones frente a decisiones basadas exclusivamente en procesamiento automatizado cuando estas producen efectos significativos sobre las personas. Del mismo modo, recientes propuestas regulatorias sobre inteligencia artificial en Europa han enfatizado la necesidad de mantener supervisión humana efectiva en sistemas automatizados considerados de alto riesgo.

Aplicado al ámbito laboral, el derecho a revisión humana implicaría que decisiones relacionadas con despidos, sanciones laborales, promociones, contratación o modificaciones sustanciales de condiciones de trabajo no puedan ejecutarse exclusivamente mediante sistemas automatizados sin participación humana real y sustantiva.

Esto supone que las organizaciones deberían garantizar:

- supervisión humana efectiva sobre decisiones automatizadas;
- mecanismos de apelación frente a decisiones algorítmicas;
- posibilidad de revisión individualizada de casos;
- intervención humana en decisiones laborales críticas;
- responsabilidad organizacional sobre decisiones tomadas mediante sistemas automatizados.

No obstante, algunos autores advierten que la simple presencia formal de intervención humana no garantiza necesariamente protección efectiva de derechos. Binns (2018) señala que en muchos casos la supervisión humana puede

convertirse en una validación automática de recomendaciones algorítmicas, fenómeno conocido como *automation bias*, donde los individuos tienden a confiar excesivamente en sistemas automatizados incluso cuando estos presentan errores.

Por ello, el derecho a revisión humana no debe entenderse únicamente como participación simbólica de supervisores humanos, sino como capacidad real de cuestionar, modificar o revertir decisiones automatizadas mediante procesos deliberativos efectivos.

Desde una perspectiva ética y organizacional, preservar espacios de intervención humana representa también una defensa de la dignidad laboral frente a dinámicas crecientes de automatización despersonalizada. Las relaciones laborales no pueden reducirse exclusivamente a cálculos de eficiencia algorítmica, ya que involucran dimensiones humanas vinculadas con reconocimiento, empatía, contexto social y justicia organizacional.

En consecuencia, el derecho a revisión humana constituye un elemento fundamental dentro de los derechos laborales digitales contemporáneos. Más allá de limitar técnicamente la automatización, este derecho busca preservar principios fundamentales de responsabilidad, justicia y dignidad humana frente al avance de sistemas algorítmicos capaces de transformar profundamente las relaciones de poder dentro de las organizaciones.

El derecho a desconexión de vigilancia algorítmica y los límites al control digital del trabajo

La necesidad de garantizar transparencia algorítmica y preservar espacios de revisión humana adquiere una dimensión aún más compleja cuando se analiza el crecimiento de los sistemas de vigilancia digital en los entornos laborales contemporáneos. La expansión de la inteligencia artificial dentro de las organizaciones no solo ha transformado los procesos de toma de decisiones, sino

también las capacidades de monitoreo y supervisión sobre los trabajadores. En este contexto, el desarrollo de mecanismos automatizados de vigilancia permanente ha consolidado nuevas formas de control organizacional sustentadas en recopilación masiva de datos, seguimiento conductual y monitoreo continuo del desempeño laboral.

Actualmente, numerosas organizaciones utilizan tecnologías capaces de rastrear actividad digital, medir productividad en tiempo real, registrar ubicación geográfica, analizar patrones de comportamiento e incluso monitorear indicadores biométricos y emocionales de los trabajadores. Estas herramientas permiten construir sistemas de supervisión mucho más invasivos que los modelos tradicionales de control laboral, ampliando significativamente la capacidad de intervención organizacional sobre la vida cotidiana de los individuos.

Desde una perspectiva histórica, el control sobre el trabajo ha sido una dimensión central dentro de la teoría organizacional. Sin embargo, las tecnologías contemporáneas han transformado radicalmente la intensidad y alcance de los mecanismos de supervisión. Mientras los modelos clásicos de control organizacional dependían de supervisión jerárquica visible y limitada espacialmente, los sistemas digitales actuales permiten desarrollar formas de vigilancia continua capaces de extenderse incluso fuera de los espacios físicos de trabajo.

Diversos autores han advertido que esta expansión del monitoreo digital configura nuevas formas de disciplinamiento organizacional. Inspirándose en las reflexiones de Michel Foucault sobre vigilancia y control social, investigaciones contemporáneas sostienen que los sistemas digitales de monitoreo producen dinámicas de autocontrol permanente donde los trabajadores modifican su comportamiento ante la posibilidad constante de ser observados (Ball, 2010). La vigilancia deja de operar únicamente como supervisión externa para convertirse en un mecanismo internalizado de regulación conductual.

En el contexto de la inteligencia artificial, esta lógica alcanza niveles sin precedentes. La capacidad de recopilar y analizar datos masivos permite a las organizaciones monitorear no solo resultados laborales, sino también ritmos de trabajo, interacciones digitales, pausas, movimientos físicos e incluso estados emocionales inferidos mediante sistemas automatizados. Como señala Shoshana Zuboff, estas dinámicas forman parte de una lógica más amplia de “capitalismo de vigilancia”, caracterizada por extracción permanente de datos conductuales orientados al control y predicción del comportamiento humano (Zuboff, 2019).

La incorporación de inteligencia artificial en estos procesos amplifica aún más las capacidades de vigilancia organizacional. Los sistemas automatizados no solo registran información, sino que también producen perfiles predictivos capaces de anticipar comportamientos, evaluar riesgos laborales o identificar patrones considerados problemáticos por las organizaciones. Esto genera escenarios donde los trabajadores permanecen sometidos a procesos constantes de evaluación automatizada difíciles de percibir o cuestionar.

Las consecuencias de estas prácticas trascienden cuestiones meramente técnicas o administrativas. Diversas investigaciones han demostrado que la vigilancia digital intensiva puede producir efectos negativos sobre bienestar psicológico, autonomía laboral y salud mental de los trabajadores. El monitoreo constante genera ambientes laborales caracterizados por presión psicológica permanente, ansiedad y sensación de pérdida de control sobre la propia actividad laboral (Moore, Upchurch & Whittaker, 2018).

Además, la expansión de mecanismos de vigilancia digital plantea serios riesgos para derechos fundamentales relacionados con privacidad y autonomía individual. En muchos casos, las fronteras entre vida laboral y vida privada se vuelven cada vez más difusas debido a tecnologías capaces de extender monitoreo fuera de horarios y espacios tradicionales de trabajo. El crecimiento del trabajo remoto y las plataformas digitales ha intensificado particularmente esta problemática,

permitiendo formas de supervisión continua que invaden espacios domésticos y tiempos personales.

Frente a este escenario, emerge la necesidad de reconocer un derecho a desconexión de vigilancia algorítmica orientado a establecer límites éticos y jurídicos frente al monitoreo digital excesivo. Este derecho parte del reconocimiento de que la capacidad tecnológica de vigilar permanentemente a los trabajadores no implica necesariamente legitimidad para hacerlo.

El derecho a desconexión de vigilancia algorítmica busca garantizar que los trabajadores conserven espacios mínimos de privacidad, autonomía y desconexión frente a sistemas automatizados de monitoreo organizacional. En este sentido, constituye una extensión contemporánea de debates previos sobre derecho a desconexión digital surgidos en respuesta a la expansión de tecnologías laborales permanentes.

Aplicado al ámbito organizacional, este derecho podría traducirse en diversas obligaciones regulatorias para las empresas, entre ellas:

- establecer límites temporales al monitoreo digital;
- prohibir vigilancia fuera de la jornada laboral;
- restringir uso desproporcionado de tecnologías biométricas;
- garantizar consentimiento informado para sistemas de monitoreo;
- proteger datos personales recopilados mediante vigilancia automatizada;
- limitar recopilación de información no vinculada directamente con funciones laborales.

Particularmente preocupante resulta el crecimiento del uso de tecnologías biométricas en los espacios laborales. Sistemas de reconocimiento facial, monitoreo ocular, análisis de voz o seguimiento corporal introducen formas altamente invasivas de recopilación de datos sensibles que pueden afectar

profundamente privacidad y dignidad humana. Diversos especialistas han advertido que estas tecnologías incrementan riesgos de discriminación, manipulación conductual y vulneración de derechos fundamentales (Ajunwa, Crawford & Schultz, 2017).

Desde una perspectiva ética, el reconocimiento de límites frente a vigilancia algorítmica constituye una defensa fundamental de la dignidad humana en el trabajo. Las organizaciones no pueden transformarse en espacios de supervisión total donde cada conducta, interacción o emoción del trabajador sea susceptible de monitoreo permanente mediante sistemas automatizados.

En consecuencia, el derecho a desconexión de vigilancia algorítmica representa un componente esencial dentro de los derechos laborales digitales contemporáneos. Más allá de regular herramientas tecnológicas específicas, este derecho busca establecer límites democráticos frente al creciente poder de vigilancia organizacional posibilitado por la inteligencia artificial y la automatización digital del trabajo.

Hacia un marco regulatorio para la gobernanza algorítmica laboral

Los procesos de automatización analizados a lo largo de este capítulo evidencian que la incorporación de inteligencia artificial en las organizaciones no puede entenderse únicamente como una transformación tecnológica orientada a incrementar eficiencia y productividad. La expansión de sistemas algorítmicos en los espacios laborales implica una reconfiguración profunda de las relaciones de poder, los mecanismos de control organizacional y las condiciones bajo las cuales se ejerce autoridad sobre el trabajo humano. En consecuencia, los desafíos derivados de la gestión algorítmica requieren algo más que soluciones técnicas aisladas: demandan la construcción de marcos regulatorios capaces de proteger

derechos fundamentales frente al creciente poder automatizado de las organizaciones.

Como se ha desarrollado anteriormente, la opacidad algorítmica, la automatización de decisiones laborales y la expansión de sistemas de vigilancia digital generan riesgos significativos para principios fundamentales vinculados con dignidad humana, autonomía laboral, privacidad y justicia organizacional. Frente a este escenario, resulta necesario avanzar hacia modelos regulatorios centrados en derechos humanos y trabajo digno, capaces de establecer límites democráticos frente a la automatización del control laboral.

Desde esta perspectiva, la regulación de la inteligencia artificial en el trabajo no debe reducirse a discusiones sobre innovación tecnológica o competitividad empresarial. Tal como advierte Shoshana Zuboff, las tecnologías digitales contemporáneas forman parte de estructuras económicas y organizacionales orientadas a expandir capacidades de vigilancia, predicción y control sobre el comportamiento humano (Zuboff, 2019). Por ello, regular la inteligencia artificial implica también intervenir en disputas relacionadas con distribución del poder dentro de las organizaciones y protección de derechos frente a nuevas formas de subordinación tecnológica.

En este sentido, diversos especialistas han señalado que los marcos regulatorios futuros deben construirse desde enfoques interdisciplinarios capaces de integrar dimensiones jurídicas, organizacionales, éticas y sociales. Como sostienen De Stefano y Aloisi (2022), la automatización del trabajo plantea desafíos estructurales para el derecho laboral contemporáneo debido a que los sistemas algorítmicos modifican profundamente las formas tradicionales de supervisión, responsabilidad y subordinación laboral.

Un marco regulatorio integral para la gobernanza algorítmica laboral debería incorporar, al menos, seis dimensiones fundamentales (ver Figura 1).



Figura 1. Dimensiones fundamentales del marco regulatorio integral para la gobernanza algorítmica laboral.

Fuente: elaboración propia.

Discusión

La expansión de la inteligencia artificial en los entornos laborales contemporáneos ha reconfigurado profundamente las relaciones entre tecnología, organización y trabajo. A lo largo de este capítulo se ha argumentado que la incorporación de sistemas algorítmicos en procesos de supervisión, evaluación y toma de decisiones laborales no constituye únicamente una innovación técnica orientada a incrementar eficiencia organizacional, sino una transformación estructural en las formas de ejercer poder y control dentro de las organizaciones. En este sentido, los debates sobre regulación de inteligencia artificial en el trabajo deben analizarse no solo desde perspectivas tecnológicas o jurídicas, sino también desde una mirada organizacional capaz de comprender cómo los sistemas automatizados modifican dinámicas de autoridad, subordinación y legitimidad dentro de las estructuras laborales contemporáneas.

Uno de los principales hallazgos que emerge del análisis desarrollado es que la gestión algorítmica representa una continuidad histórica —aunque profundamente sofisticada— de modelos organizacionales orientados al control racional del trabajo. Desde la administración científica de Frederick Winslow Taylor hasta las formas contemporáneas de vigilancia algorítmica, persiste una lógica organizacional centrada en maximizar eficiencia mediante medición, estandarización y supervisión constante del desempeño laboral. Sin embargo, mientras el taylorismo clásico dependía de supervisión humana directa y fragmentación física del trabajo, los sistemas algorítmicos actuales permiten desarrollar mecanismos automatizados de control permanente capaces de operar en tiempo real mediante extracción masiva de datos y monitoreo digital continuo.

Esta transformación plantea una tensión fundamental entre eficiencia organizacional y dignidad humana. Desde perspectivas gerenciales tradicionales, la inteligencia artificial aparece como una herramienta capaz de optimizar procesos, reducir costos y mejorar capacidad predictiva en la gestión de recursos humanos.

No obstante, desde enfoques críticos de los estudios organizacionales, esta automatización intensifica riesgos relacionados con deshumanización del trabajo, debilitamiento de autonomía laboral y consolidación de formas invisibles de control organizacional.

En este punto resulta particularmente relevante contrastar las lógicas contemporáneas de gestión algorítmica con los aportes históricos de la escuela de las Relaciones Humanas. Mientras autores como Elton Mayo sostenían que la productividad organizacional dependía también de factores emocionales, reconocimiento social y relaciones humanas dentro del trabajo, la automatización contemporánea parece desplazar progresivamente estos componentes hacia sistemas impersonales basados en datos cuantificables. La sustitución de interacción humana por evaluación algorítmica transforma las relaciones laborales en dinámicas crecientemente mediadas por procesos técnicos donde dimensiones subjetivas y contextuales del trabajo pierden relevancia frente a métricas automatizadas de rendimiento.

Esta situación revela una paradoja organizacional contemporánea: mientras las organizaciones modernas reconocen discursivamente la importancia del bienestar laboral, liderazgo humanista y cultura organizacional, simultáneamente implementan tecnologías capaces de intensificar vigilancia, control y presión sobre los trabajadores. En consecuencia, emerge una contradicción entre discursos organizacionales centrados en capital humano y prácticas tecnológicas orientadas hacia monitoreo permanente y automatización disciplinaria.

Asimismo, el análisis realizado permite observar que la opacidad algorítmica constituye uno de los mecanismos centrales mediante los cuales se reproduce el poder organizacional contemporáneo. A diferencia de las estructuras jerárquicas tradicionales, donde la autoridad podía identificarse claramente en figuras directivas o cadenas formales de mando, los sistemas automatizados producen formas difusas de autoridad difíciles de cuestionar o responsabilizar. La “caja negra algorítmica” descrita por Pasquale (2015) no solo representa un problema técnico

relacionado con comprensión de sistemas automatizados, sino también un problema político vinculado con distribución desigual del conocimiento y poder dentro de las organizaciones.

Desde esta perspectiva, el derecho a explicación algorítmica adquiere una relevancia que trasciende la simple transparencia tecnológica. Permitir que los trabajadores comprendan cómo son evaluados implica también democratizar parcialmente estructuras organizacionales crecientemente opacas. La transparencia algorítmica representa, en este sentido, un mecanismo de redistribución del poder informacional dentro de las relaciones laborales.

No obstante, la discusión desarrollada evidencia también límites importantes respecto a la idea de transparencia como solución suficiente frente a los riesgos derivados de la inteligencia artificial. Diversos autores han advertido que incluso sistemas técnicamente explicables pueden continuar reproduciendo relaciones estructurales de desigualdad y subordinación. Esto resulta especialmente evidente en contextos donde los trabajadores, aun comprendiendo parcialmente funcionamiento algorítmico, carecen de capacidad real para impugnar decisiones organizacionales o negociar condiciones de implementación tecnológica.

Por ello, el derecho a revisión humana adquiere especial importancia como mecanismo orientado a preservar espacios de deliberación ética y responsabilidad organizacional frente a decisiones automatizadas. Sin embargo, esta dimensión también enfrenta tensiones relevantes. En numerosos casos, la supervisión humana puede convertirse en un proceso meramente formal donde operadores humanos validan automáticamente recomendaciones algorítmicas sin cuestionarlas críticamente. Esto sugiere que la presencia humana por sí sola no garantiza necesariamente protección efectiva frente a injusticias automatizadas.

Otro elemento central que emerge de esta discusión es la expansión de nuevas formas de vigilancia organizacional posibilitadas por inteligencia artificial. Los sistemas contemporáneos de monitoreo digital transforman profundamente fronteras históricas entre espacio laboral y vida privada. Tecnologías capaces de

rastrear actividad digital, ubicación geográfica o comportamiento emocional permiten extender supervisión organizacional más allá de límites tradicionales del trabajo.

Desde una mirada organizacional crítica, esta expansión de vigilancia refleja procesos más amplios de intensificación del control laboral en contextos de capitalismo digital. Como sostiene Shoshana Zuboff, la extracción masiva de datos se ha convertido en un mecanismo central para producción de valor económico y control conductual en las sociedades contemporáneas. En consecuencia, las organizaciones ya no solo administran fuerza laboral, sino también información detallada sobre comportamientos, emociones e interacciones de los trabajadores.

Esto plantea interrogantes fundamentales sobre los límites éticos del control organizacional. ¿Hasta qué punto resulta legítimo monitorear permanentemente a los trabajadores mediante sistemas automatizados? ¿Puede existir trabajo digno en contextos caracterizados por vigilancia algorítmica continua? ¿Qué ocurre con autonomía y privacidad cuando cada acción laboral se convierte en dato susceptible de análisis predictivo?

Frente a estas problemáticas, la propuesta de construir derechos laborales digitales representa un intento por reequilibrar relaciones de poder profundamente transformadas por automatización organizacional. Sin embargo, la discusión desarrollada sugiere que estos derechos enfrentarán importantes desafíos prácticos y políticos. Las organizaciones contemporáneas operan en contextos económicos altamente competitivos donde tecnologías de automatización y monitoreo son frecuentemente justificadas mediante discursos de eficiencia, productividad y optimización empresarial.

Por ello, cualquier intento regulatorio deberá enfrentar tensiones estructurales entre protección laboral y racionalidades económicas orientadas hacia maximización de rendimiento organizacional. Esto implica reconocer que la regulación de inteligencia artificial en el trabajo no constituye únicamente un problema técnico-jurídico, sino

también una disputa política sobre el futuro del trabajo y los límites del poder organizacional en sociedades crecientemente automatizadas.

En última instancia, la discusión presentada en este capítulo permite sostener que los desafíos derivados de la inteligencia artificial en las relaciones laborales obligan a replantear preguntas fundamentales sobre naturaleza del trabajo, autoridad organizacional y dignidad humana en contextos digitales. La cuestión central ya no radica únicamente en determinar qué tareas pueden automatizarse, sino en definir qué límites democráticos deben establecerse frente a tecnologías capaces de transformar profundamente las relaciones humanas dentro de las organizaciones.

En este sentido, la construcción de marcos regulatorios orientados hacia transparencia, supervisión humana, protección de privacidad y participación laboral representa no solo una respuesta jurídica frente a tecnologías emergentes, sino también una defensa ética y política de la dignidad humana frente al avance del poder algorítmico en el mundo del trabajo contemporáneo.

Conclusión

La incorporación de inteligencia artificial en las organizaciones contemporáneas ha transformado profundamente las relaciones laborales, desplazando los mecanismos tradicionales de supervisión hacia formas de gestión algorítmica sustentadas en automatización, monitoreo digital y análisis masivo de datos. Este proceso no solo modifica dinámicas operativas dentro de las organizaciones, sino que reconfigura estructuras de poder, mecanismos de control y formas de subordinación laboral en contextos crecientemente mediados por tecnologías automatizadas.

A lo largo de este capítulo se ha sostenido que los principales desafíos derivados de la inteligencia artificial en el trabajo no pueden reducirse únicamente al problema del desplazamiento laboral o la sustitución de empleos. El problema central radica

también en la consolidación de formas automatizadas de control organizacional capaces de afectar derechos fundamentales vinculados con dignidad humana, autonomía, privacidad, transparencia y justicia laboral. La opacidad algorítmica, la vigilancia digital permanente y la automatización de decisiones críticas evidencian las limitaciones de los marcos tradicionales del derecho laboral frente a nuevas formas de subordinación tecnológica.

En este escenario, la construcción de derechos laborales digitales emerge como una necesidad indispensable para responder a los desafíos derivados de la gestión algorítmica del trabajo. Derechos como la explicación algorítmica, la revisión humana significativa, la desconexión frente a vigilancia digital y la transparencia en sistemas automatizados representan mecanismos fundamentales para limitar el ejercicio arbitrario del poder algorítmico dentro de las organizaciones y garantizar condiciones laborales compatibles con principios democráticos y trabajo digno.

Estas preocupaciones han comenzado a ocupar un lugar central en el discurso de diversos organismos internacionales. La International Labour Organization ha advertido que la transformación digital del trabajo exige fortalecer mecanismos de protección laboral frente a nuevas formas de precarización, vigilancia y desigualdad derivadas de la automatización. Desde esta perspectiva, el desarrollo tecnológico debe mantenerse subordinado al principio de trabajo decente y a la protección de la dignidad humana en el empleo.

Del mismo modo, la UNESCO ha sostenido que la inteligencia artificial debe desarrollarse bajo principios éticos orientados hacia derechos humanos, transparencia, supervisión humana y justicia social. Asimismo, la Organisation for Economic Co-operation and Development ha enfatizado la necesidad de garantizar sistemas de inteligencia artificial centrados en las personas, responsables y sujetos a mecanismos de rendición de cuentas.

Sin embargo, la regulación de la inteligencia artificial en el trabajo no debe entenderse únicamente como un desafío técnico o jurídico. Se trata, en realidad, de una disputa profundamente organizacional y política relacionada con los límites

del control tecnológico sobre el trabajo humano. El avance de sistemas automatizados obliga a replantear preguntas fundamentales sobre autoridad organizacional, autonomía laboral y los límites éticos de la vigilancia digital dentro de las organizaciones contemporáneas.

En última instancia, el desafío central consiste en construir modelos de innovación tecnológica que no sacrifiquen derechos fundamentales en nombre de la eficiencia organizacional. El futuro del trabajo dependerá de la capacidad de las sociedades, instituciones y organizaciones para establecer marcos regulatorios capaces de subordinar el desarrollo tecnológico a principios de justicia laboral, transparencia, participación y respeto por la dignidad humana. En ello no solo se define el futuro de las relaciones laborales, sino también el tipo de organizaciones y sociedades que emergerán en la era de la inteligencia artificial.

Referencias

- Ajunwa, I., Crawford, K., & Schultz, J. (2017). *Limitless worker surveillance*. California Law Review, 105(3), 735–776.
- Aloisi, A., & De Stefano, V. (2022). *Your boss is an algorithm: Artificial intelligence, platform work and labour*. Hart Publishing.
- Arendt, H. (1958). *The human condition*. University of Chicago Press.
- Ball, K. (2010). *Workplace surveillance: An overview*. Labor History, 51(1), 87–106. <https://doi.org/10.1080/00236561003654776>
- Binns, R. (2018). *Fairness in machine learning: Lessons from political philosophy*. Proceedings of Machine Learning Research, 81, 149–159.
- Burrell, J. (2016). *How the machine ‘thinks’: Understanding opacity in machine learning algorithms*. Big Data & Society, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>
- Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters.
- De Stefano, V., & Aloisi, A. (2022). *Your boss is an algorithm: Artificial intelligence, platform work and labour*. Hart Publishing.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin’s Press.
- Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Pantheon Books.
- Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). *Algorithms at work: The new contested terrain of control*. Academy of Management Annals, 14(1), 366–410. <https://doi.org/10.5465/annals.2018.0174>
- Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). *Working with machines: The impact of algorithmic and data-driven management on human workers*. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 1603–1612. <https://doi.org/10.1145/2702123.2702548>
- Mayo, E. (1933). *The human problems of an industrial civilization*. Macmillan.
- McGregor, D. (1960). *The human side of enterprise*. McGraw-Hill.
- Moore, P. V., Upchurch, M., & Whittaker, X. (2018). *Humans and machines at work: Monitoring, surveillance and automation in contemporary capitalism*. Palgrave Macmillan.

- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Selbst, A. D., & Barocas, S. (2018). *The intuitive appeal of explainable machines*. Fordham Law Review, 87(3), 1085–1139.
- Taylor, F. W. (1911). *The principles of scientific management*. Harper & Brothers.
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*. International Data Privacy Law, 7(2), 76–99. <https://doi.org/10.1093/idpl/ix005>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

CAPÍTULO XXX

**CENTRO DE INTELIGENCIA DE MERCADOS Y
ANALÍTICA DEL CONSUMIDOR (CIMEC-IUE):
CONCEPTO, FUNCIONES ESTRATÉGICAS Y LÍNEAS
DE ACCIÓN EN EL MARCO DE LA REGULACIÓN DE
LA INTELIGENCIA ARTIFICIAL**

Center for Market Intelligence and Consumer Analytics (CIMEC-IUE):
concept, strategic functions and lines of action within the framework of
artificial intelligence regulation

Francisco Javier Arias Vargas

fjarias@correo.iue.edu.co

Facultad de Ciencias Empresariales y Comunicaciones, Institución
Universitaria de Envigado, Colombia.

<https://orcid.org/0000-0002-4483-1741> .

Oscar Lozano Carrillo

exato@azc.uam.mx

Universidad Autónoma Metropolitana, unidad Azcapotzalco, México.

<https://orcid.org/0000-0001-6166-3033>

Rafael Alejandro Betancourt Durango

Institución Universitaria de Envigado, Colombia.

<https://orcid.org/0000-0002-6423-4864>

Valery Gabriela Galvis Monterroza

vgalvis@correo.iue.edu.co

Facultad de Ciencias Empresariales y Comunicaciones, Institución
Universitaria de Envigado, Colombia.

<https://orcid.org/0009-0002-6421-7367>

Valentina Chavarria Calle

vchavarria@correo.iue.edu.co

Facultad de Ciencias Empresariales y Comunicaciones, Institución
Universitaria de Envigado, Colombia.

<https://orcid.org/0009-0007-5587-5989>.

Laura Echeverry Sierra

Lecheverry@correo.iue.edu.co

Institución Universitaria de Envigado, Colombia.

<https://orcid.org/0009-0003-2624-2200>

Este trabajo se enmarca en el proyecto de investigación titulado “Propuesta para la estructuración de un Centro de Inteligencia de Mercados y Analítica del Consumidor (CIMEC-IUE)” de la Institución Universitaria de Envigado, el cual sumo esfuerzos desde el Programa de Inteligencia Artificial Estratégica Social y Solidaria (PIAESS) en colaboración entre la Institución Universitaria de Envigado (Colombia) y la Universidad Autónoma Metropolitana unidad Azcapotzalco (México).

Resumen

La acelerada incorporación de la inteligencia artificial (IA) en los procesos de mercadeo, gestión y toma de decisiones empresariales ha redefinido el modo en que las organizaciones interactúan con los mercados, los consumidores y los entornos competitivos; este capítulo aborda la propuesta conceptual y operativa del Centro de Inteligencia de Mercados y Analítica del Consumidor de la Institución Universitaria de Envigado (CIMEC-IUE), entendido como una unidad académico-investigativa orientada a articular las funciones sustantivas de docencia, investigación y proyección social en torno al uso responsable de datos e IA, el análisis se enmarca en el debate contemporáneo sobre la regulación de la IA y sus implicaciones para el trabajo, la economía y la innovación, considerando referentes normativos como el Reglamento (UE) 2024/1689 y la Circular Externa 002 de 2024 de la Superintendencia de Industria y Comercio de Colombia; a partir de una revisión documental de literatura indexada con identificadores DOI verificables del periodo 2021-2026, se delimita la definición del CIMEC-IUE, se sistematizan sus funciones estratégicas (investigación aplicada, formación, proyección social y articulación regulatoria) y se proponen cinco líneas de acción; investigación aplicada en inteligencia de mercados con IA, formación de capacidades analíticas; asesoría y consultoría especializada; observatorio de tendencias en IA, trabajo y

consumo y ética, gobernanza algorítmica y cumplimiento normativo. La propuesta sostiene que un centro de esta naturaleza, anclado en una institución universitaria territorial, puede contribuir a democratizar el acceso a herramientas de inteligencia, fortalecer la competitividad del Aburrá Sur y aportar al desarrollo de marcos institucionales que armonicen innovación tecnológica con protección de derechos.

Palabras clave: *inteligencia de mercados; analítica del consumidor; inteligencia artificial; regulación; gobernanza algorítmica; innovación.*

Abstract

The accelerated incorporation of artificial intelligence (AI) into marketing, management, and business decision-making processes has redefined the way organizations interact with markets, consumers, and competitive environments. This chapter addresses the conceptual and operational proposal of the Market Intelligence and Consumer Analytics Center of the Institución Universitaria de Envigado (CIMEC-IUE), understood as an academic-research unit aimed at articulating the substantive functions of teaching, research, and social outreach around the responsible use of data and AI. The analysis is framed within the contemporary debate on AI regulation and its implications for labor, the economy, and innovation, considering regulatory references such as Regulation (EU) 2024/1689 and External Circular 002 of 2024. Based on a documentary review of indexed literature with verifiable DOI identifiers from the 2021–2026 period, the chapter defines the CIMEC-IUE, systematizes its strategic functions (applied research, training, social outreach, and regulatory articulation), and proposes five lines of action: applied research in AI-driven market intelligence; development of analytical capabilities; specialized advisory and consulting services; an observatory of trends in AI, labor, and consumption; and ethics, algorithmic governance, and regulatory compliance. The proposal argues that a center of this nature, anchored in a territorial higher education institution, can contribute to democratizing access to intelligence tools, strengthening the competitiveness of the Aburrá Sur region, and

supporting the development of institutional frameworks capable of harmonizing technological innovation with the protection of rights.

Keywords: *market intelligence; consumer analytics; artificial intelligence; regulation; algorithmic governance; innovation.*

1. Introducción

La inteligencia artificial ha pasado, en menos de una década, de ser una tecnología emergente confinada a laboratorios especializados a convertirse en una infraestructura económica que atraviesa prácticamente todos los procesos empresariales contemporáneos; la adopción de modelos de aprendizaje automático, procesamiento de lenguaje natural y sistemas de decisión automatizada (Chowdhury y otros., 2022); ha modificado profundamente las prácticas de mercadeo, la analítica del consumidor, la gestión de la cadena de valor y los modelos de innovación organizacional; Acatrinei y otros. (2025) muestran, con base en un estudio empírico aplicado a 501 usuarios de redes sociales que la integración de la IA en marketing digital opera no solo sobre la eficiencia operativa, sino también sobre dimensiones centrales de la experiencia del consumidor, como la confianza, la transparencia percibida y la disposición a adoptar comportamientos sostenibles, esta transformación, sin embargo, ocurre en paralelo a una creciente preocupación social y regulatoria por los riesgos derivados del uso intensivo de datos personales, los sesgos algorítmicos, la erosión de la privacidad y la concentración de poder informacional en manos de unos pocos actores tecnológicos (Chulvi, 2023).

En este escenario, las organizaciones empresariales y públicas enfrentan un doble desafío; por una parte deben desarrollar capacidades analíticas avanzadas que les permitan transformar grandes volúmenes de datos en decisiones estratégicas

sustentadas en evidencia (Cristofaro y otros., 2024; Theodorakopoulos & Theodoropoulou, 2024) y por otro lado, deben navegar un ecosistema normativo cada vez más complejo que incluye desde el Reglamento (UE) 2024/1689 sobre inteligencia artificial (Minssen y otros., 2025), hasta directrices nacionales como la Circular Externa 002 de 2024 emitida por la Superintendencia de Industria y Comercio de Colombia; la tensión entre el imperativo competitivo de innovar con IA y el imperativo ético-jurídico de hacerlo de manera responsable define gran parte del debate actual sobre el lugar de la analítica avanzada en la economía contemporánea (Madanchian & Taherdoost, 2025; Mahmutovic, 2025).

Este reto adquiere connotaciones particulares en contextos de economías emergentes y específicamente, en territorios subnacionales como el municipio de Envigado y la subregión del Aburrá Sur, en Antioquia, Colombia; las organizaciones del territorio enfrentan procesos acelerados de digitalización, mayor disponibilidad de datos provenientes de canales digitales y de Internet de las Cosas y crecientes exigencias competitivas derivadas de la globalización de los mercados (Dávalos, 2022); no obstante, persisten brechas significativas en términos de capacidades técnicas, metodológicas, formativas y normativas para aprovechar de manera efectiva y responsable estas nuevas posibilidades; frente a ello, la Institución Universitaria de Envigado (IUE) en el marco de su misión institucional y de las funciones sustantivas de docencia, investigación y proyección social, ha formulado una propuesta para la estructuración de un Centro de Inteligencia de Mercados y Analítica del Consumidor (CIMEC-IUE); concebido como una unidad académica especializada que articule capacidades analíticas, formativas e investigativas para apoyar la toma de decisiones de organizaciones públicas y privadas del territorio.

El presente capítulo se inscribe en la línea temática de inteligencia artificial, trabajo, economía e innovación dentro del debate sobre la regulación de la IA, su propósito es definir el concepto, las funciones estratégicas y las líneas de acción del CIMEC-IUE, con el fin de ofrecer una propuesta conceptual y operativa coherente con las

exigencias del entorno regulatorio contemporáneo; para ello, el documento se estructura en seis secciones, además de esta introducción, la segunda sección desarrolla la aproximación conceptual al CIMEC-IUE, partiendo de la evolución de la inteligencia de mercados hacia esquemas potenciados por IA; la tercera sección examina el marco regulatorio internacional y colombiano de la IA y sus implicaciones para los centros académicos de inteligencia; la cuarta sección sistematiza las funciones estratégicas del CIMEC-IUE, la quinta propone sus cinco líneas de acción y la sexta sitúa al CIMEC-IUE en el ecosistema de IA, trabajo, economía e innovación; finalmente, se ofrecen consideraciones finales y prospectiva.

Este capítulo adopta un enfoque cualitativo de revisión documental, en el cual se priorizan fuentes académicas indexadas en los últimos cinco años (2021-2026); esta delimitación responde a la necesidad de fundamentar la propuesta del CIMEC-IUE en evidencia académica reciente y verificable; dada la rapidez con que evoluciona el campo de la inteligencia artificial y su regulación, lo cual se complementa esta revisión con el análisis de instrumentos normativos vigentes y con la consulta de informes técnicos de organismos multilaterales citados aquí únicamente como contexto normativo y no como evidencia empírica primaria.

2. Aproximación conceptual al CIMEC-IUE

2.1 De la inteligencia de mercados clásica a la inteligencia analítica con IA

La inteligencia de mercados, en su acepción tradicional, ha sido entendida como el proceso sistemático de recolección, análisis y difusión de información sobre clientes, competidores, proveedores y entornos competitivos, con el fin de apoyar la toma de decisiones estratégicas (Sarango y otros., 2023); esta concepción, anclada en métodos clásicos de investigación de mercados (encuestas, grupos focales, análisis de ventas), ha sido objeto de profundas transformaciones en las

dos últimas décadas. Vishnoi y otros. (2025), tras mapear sistemáticamente la literatura producida entre 1962 y 2022 e identificar 127 artículos académicos con quince definiciones formales de marketing intelligence, proponen una definición integradora que reconoce cinco temas constitutivos como insumos informacionales, capacidades analíticas, propósitos estratégicos, aplicaciones tácticas y resultados organizacionales; esta sistematización permite superar la fragmentación conceptual prevalente y entender la inteligencia de mercados como un marco comprehensivo que articula datos, métodos, decisiones y resultados (Moraes y otros., 2015).

La incorporación masiva de big data y de algoritmos de aprendizaje automático ha desplazado la frontera de lo posible; Theodorakopoulos & Theodoropoulou (2024), en una revisión sistemática siguiendo el protocolo PRISMA sobre la analítica de big data aplicada al comportamiento del consumidor en mercadeo digital, documentan cómo la integración de fuentes heterogéneas de datos (transaccionales, de navegación, de redes sociales, de dispositivos conectados); permiten niveles de comprensión del consumidor que no eran alcanzables mediante métodos tradicionales, las técnicas de minería de texto, análisis de sentimientos, clustering y redes neuronales abren la posibilidad de identificar patrones latentes, anticipar tendencias y personalizar experiencias a escalas previamente inéditas; Cristofaro y otros. (2024) profundizan esta perspectiva al proponer al growth hacking como un enfoque científico de toma de decisiones basada en datos, articulado en siete pasos secuenciales que van desde el análisis de escenarios hasta la revisión de resultados, y que combina experimentación, validación empírica y aprendizaje organizacional iterativo.

En este contexto, Huang & Rust (2021) plantean un marco estratégico de tres etapas para la incorporación de IA en el ámbito del mercadeo, distinguiendo entre IA mecánica (orientada a la automatización de tareas rutinarias), IA pensante (centrada en el procesamiento de datos para la generación de decisiones) e IA

emocional (focalizada en el análisis de interacciones y emociones humanas); cada una de estas modalidades genera beneficios diferenciados (Hermann, 2021), la IA mecánica posibilita la estandarización; la IA pensante habilita la personalización y la IA emocional facilita lo que los autores denominan la relacionalización (Nidhishree, 2024); esto es, la construcción de vínculos personalizados con cada consumidor; esta tipología constituye una herramienta heurística útil para clasificar los desarrollos recientes en analítica del consumidor y para situar las capacidades que un centro de inteligencia debería desarrollar.

La transición de la inteligencia de mercados clásica a la inteligencia analítica potenciada por IA no es un mero cambio tecnológico, implica transformaciones organizacionales sustantivas, particularmente en la cultura analítica, los modelos de gobernanza de datos y los procesos de toma de decisiones; (Szukits & Móricz, 2023); con base en una encuesta a 305 empresas medianas y grandes, demuestran que el énfasis estratégico sobre la analítica de negocio favorece el desarrollo de una cultura analítica organizacional que, a su vez, facilita un uso más intensivo de datos en las distintas fases del proceso decisorio; no obstante, los autores identifican una paradoja y es que aunque la cultura analítica incrementa la centralización del uso de datos, esta centralización no se traduce automáticamente en una toma de decisiones efectivamente basada en datos, lo cual sugiere que la efectividad analítica depende de configuraciones organizacionales y socio-técnicas más complejas que la mera disponibilidad tecnológica.

Xu & Lu (2026), aplicando modelos de lenguaje basados en BERT y ChatGLM2-6B para cuantificar la toma de decisiones basada en datos a nivel micro, demuestran empíricamente que esta práctica impacta positivamente el desempeño internacional de las firmas, los autores identifican cuatro mecanismos mediante los cuales la decisión basada en datos crea valor sostenible; prevención de la contaminación, innovación verde, divulgación de información de sostenibilidad y desempeño ambiental, social y de gobernanza donde esta evidencia respalda la

idea de que un centro académico orientado a la inteligencia de mercados y la analítica del consumidor no opera únicamente como un proveedor de información para decisiones tácticas, sino que cumple una función estructural en el desarrollo de capacidades organizacionales con efectos económicos, sociales y ambientales mensurables.

2.2 Definición conceptual del CIMEC-IUE

A partir del marco anterior, el Centro de Inteligencia de Mercados y Analítica del Consumidor de la Institución Universitaria de Envigado (CIMEC-IUE) se define como una unidad académica y de investigación aplicada, adscrita a la Facultad de Ciencias Empresariales y Comunicaciones de la IUE, orientada al análisis sistemático de la información de mercados del comportamiento del consumidor y de las tendencias competitivas, mediante la integración de métodos cualitativos, cuantitativos, computacionales y de inteligencia artificial, con el propósito de fortalecer la toma de decisiones estratégicas en organizaciones públicas y privadas del municipio de Envigado, la subregión del Aburrá Sur y, prospectivamente, otras geografías.

Esta definición contempla cinco componentes esenciales, su carácter académico-investigativo, que lo distingue de las consultoras comerciales y lo articula con la misión universitaria de generación, apropiación y transferencia de conocimiento y su orientación aplicada que privilegia la pertinencia y el impacto sobre el desarrollo teórico abstracto; luego su anclaje territorial, que reconoce las particularidades del entorno productivo y social del Aburrá Sur sin renunciar a una proyección internacional; sumado a su naturaleza metodológicamente híbrida que combina aproximaciones tradicionales (entrevistas, etnografía, análisis estadístico); con técnicas avanzadas de IA (aprendizaje automático, procesamiento de lenguaje natural, análisis de sentimientos) y por último, su vocación articuladora, que busca

integrar las funciones sustantivas de docencia, investigación y proyección social en una unidad operativa coherente.

El CIMEC-IUE asume, además, una perspectiva crítica respecto a las potencialidades y limitaciones de la IA en el ámbito del mercadeo y la analítica del consumidor; la sofisticación tecnológica no garantiza, por sí sola, la generación de valor estratégico, como muestran (Madanchian & Taherdoost, 2025), la adopción ética y efectiva de la IA en organizaciones requiere de marcos integrados que articulen teorías éticas, modelos de gobernanza, evaluación de riesgos, transparencia algorítmica y cultura organizacional (Hernández, 2024); un centro académico de inteligencia de mercados que aspire a generar valor sostenible debe, por tanto, incorporar desde su diseño estos componentes éticos y de gobernanza, no como elementos accesorios, sino como dimensiones constitutivas de su propuesta de valor.

2.3 Fundamentos teóricos: enfoque sociotécnico y orientación al mercado

La conceptualización del CIMEC-IUE se sustenta en dos cuerpos teóricos complementarios, el primero es el enfoque de sistemas sociotécnicos, según el cual los centros de inteligencia organizacional no son meros conjuntos de herramientas tecnológicas ni equipos aislados de analistas, sino configuraciones complejas en las que elementos técnicos y sociales coevolucina (Capolupo y otros., 2025; Metcalf & Daniel, 2025); el desempeño del sistema emerge de la calidad de la integración entre componentes tecnológicos, procesos de trabajo, competencias humanas y estructuras organizacionales, esta perspectiva ayuda a explicar por qué implementaciones tecnológicamente similares pueden generar resultados radicalmente distintos en organizaciones con culturas, gobernanzas o modelos de talento diferentes (Szukits & Móricz, 2023).

El segundo cuerpo teórico es el de la orientación al mercado, entendida como una capacidad organizacional que combina la generación sistemática de inteligencia sobre clientes y competidores, su diseminación interna y la respuesta coordinada a sus implicaciones; la literatura reciente complementa esta perspectiva con el enfoque de capacidades dinámicas, que enfatiza la habilidad de las organizaciones para integrar, construir y reconfigurar competencias internas y externas frente a entornos cambiantes; Cristofaro y otros. (2024) muestran cómo el growth hacking, en tanto enfoque científico de decisión basada en datos, opera como un mecanismo concreto de capacidad dinámica al permitir a las organizaciones probar hipótesis, aprender rápidamente y ajustar estrategias mediante ciclos iterativos de experimentación.

Estos enfoques convergen en una premisa central para el diseño del CIMEC-IUE donde la efectividad de un centro de inteligencia no depende exclusivamente de la sofisticación de sus herramientas, sino de la calidad de la articulación entre tecnología, talento, procesos y cultura organizacional.

Un centro académico-investigativo añade a esta articulación una dimensión adicional, la integración con las funciones de docencia y proyección social, lo cual lo distingue cualitativamente de centros corporativos puramente operativos, esta integración multifuncional es, simultáneamente, una fortaleza distintiva y una fuente de complejidad de gestión, pues requiere balancear los tiempos, lógicas e incentivos propios de cada función sustantiva universitaria.

3. Marco regulatorio de la IA y sus implicaciones para el CIMEC-IUE

3.1 Panorama internacional: el AI Act europeo y otros marcos regulatorios

El Reglamento (UE) 2024/1689, conocido como AI Act, constituye la primera regulación integral sobre inteligencia artificial en el mundo y representa un referente

obligado para cualquier reflexión académica o institucional sobre el tema; Mahmutovic (2025) analiza este reglamento como un marco proactivo que busca abordar simultáneamente los desafíos de innovación, gobernanza ética, derechos fundamentales y seguridad jurídica que plantea la IA, el reglamento entró en vigor el 1 de agosto de 2024 y será plenamente aplicable el 2 de agosto de 2026, con disposiciones transitorias específicas; su arquitectura se basa en una clasificación de los sistemas de IA en cuatro categorías de riesgo: inaceptable, alto, limitado y mínimo; los sistemas de riesgo inaceptable, como aquellos que manipulan el comportamiento humano causando daño físico o psicológico, están directamente prohibidos.

Los sistemas de alto riesgo, particularmente relevantes para la inteligencia de mercados y la analítica del consumidor deben cumplir requisitos estrictos en materia de gestión del riesgo, calidad de los datos, documentación técnica, trazabilidad, transparencia, supervisión humana, robustez, exactitud y ciberseguridad; al respecto Bignami y otros. (2025) destacan que el AI Act introduce mecanismos proactivos novedosos como los “*regulatory sandboxes*” y los marcos de gobernanza de datos, los cuales buscan armonizar la innovación con la supervisión ética; sin embargo los autores también identifican tensiones inherentes al equilibrio entre el impulso a la innovación y el control ético, particularmente en contextos de incertidumbre geopolítica y competencia tecnológica global.

Alvarado (2025) sintetiza las principales lecciones del AI Act para la comunidad científica y empresarial, destacando que el reglamento establece un marco para el desarrollo, comercialización y aplicación de IA en la Unión Europea bajo la lógica de promover una IA centrada en el ser humano y respetuosa de los derechos fundamentales. Importante para los centros de inteligencia de mercados es la exigencia de transparencia respecto al contenido generado por IA, mediante mecanismos como metadatos y marcas de agua que permitan identificar el origen sintético del contenido; por su parte, Guadamuz (2024) examina las implicaciones del AI Act para los derechos de autor, particularmente en lo relativo a los modelos

de IA de propósito general entrenados con grandes volúmenes de datos, las disposiciones sobre transparencia de los datos de entrenamiento y respeto a la legislación europea sobre derechos de autor tienen implicaciones directas para cualquier centro académico que incorpore técnicas de procesamiento de lenguaje natural en sus análisis de mercados.

Teniendo en cuenta lo anterior, el panorama regulatorio internacional muestra una creciente diversidad de aproximaciones; mientras la Unión Europea opta por una regulación horizontal basada en riesgos, otros marcos privilegian aproximaciones sectoriales o autorregulatorias; esta diversidad genera retos para las organizaciones que operan en múltiples jurisdicciones (Minssen y otros., 2025), pero también plantea oportunidades para que centros académicos como el CIMEC-IUE contribuyan al análisis comparado y a la formulación de directrices contextualmente apropiadas para realidades como la colombiana.

3.2 Marco normativo colombiano: protección de datos y guías sobre IA

Colombia cuenta con un marco normativo en construcción para la regulación de la inteligencia artificial, que se articula con la legislación preexistente sobre protección de datos personales; la Ley 1581 de 2012 establece el régimen general de protección de datos personales y define principios fundamentales como el de legalidad, finalidad, libertad, veracidad, transparencia, acceso y circulación restringida, seguridad y confidencialidad; los cuales resultan directamente aplicables al tratamiento de datos en sistemas de IA, constituyendo un piso normativo que cualquier centro de inteligencia de mercados debe cumplir.

En agosto de 2024, la Superintendencia de Industria y Comercio (SIC) de Colombia emitió la Circular Externa 002 de 2024, que establece directrices específicas sobre el tratamiento de datos personales en sistemas de inteligencia artificial, la circular subraya la importancia de garantizar entornos seguros que cumplan con la regulación vigente antes de iniciar cualquier tratamiento de datos personales, e

incluso antes de la recolección de información y establece como principios rectores la necesidad, idoneidad, razonabilidad y proporcionalidad en el uso de sistemas de IA para el tratamiento de datos personales y enfatiza la realización de estudios de impacto sobre la privacidad; estas exigencias trascienden el cumplimiento formal y sugieren la necesidad de incorporar la evaluación de impacto desde las fases de diseño de los proyectos analíticos, en una lógica de privacidad por diseño.

La hoja de ruta colombiana para la IA, impulsada desde el Ministerio de Ciencia, Tecnología e Innovación (MinCiencias) y otros actores institucionales, reconocen que la adopción de IA en américa latina, se convierte una oportunidad significativa pero también un desafío en términos de gobernanza, teniendo en cuenta principios como privacidad, transparencia, responsabilidad, accesibilidad y justicia (Vanegas y otros., 2024); lo cual plantea la necesidad de fortalecer marcos legales y regulatorios que aborden cuestiones éticas, responsabilidades y derechos relacionados con la implementación y uso de la IA; la Red Iberoamericana de Protección de Datos, también ha emitido directrices que reflejan buenas prácticas en el procesamiento de datos personales mediante IA, lo cual ofrece referentes adicionales para el CIMEC-IUE.

El marco regulatorio colombiano se encuentra en una fase dinámica de actualización con propuestas legislativas tendientes a alinear la normativa nacional con estándares internacionales como el Reglamento General de Protección de Datos europeo; para el CIMEC-IUE, esta dinámica regulatoria implica la necesidad de mantener una vigilancia normativa permanente y de incorporar el análisis regulatorio como un componente sustantivo de su trabajo no solo para asegurar el cumplimiento propio, sino también para apoyar a las organizaciones del territorio en la navegación de un entorno normativo cambiante.

3.3 Implicaciones éticas, gobernanza algorítmica y sesgos

Más allá del cumplimiento estricto de la normativa, la incorporación de IA en procesos de inteligencia de mercados plantea desafíos éticos sustantivos que cualquier centro académico responsable debe abordar de manera explícita; (Madanchian & Taherdoost, 2025) examinan las teorías éticas, modelos de gobernanza y marcos estratégicos que habilitan una adopción responsable de IA en contextos organizacionales, los autores identifican una brecha persistente entre los principios éticos articulados en la literatura académica y su aplicación efectiva en contextos industriales, lo cual exige una colaboración interdisciplinaria continua y mecanismos de comunicación con múltiples grupos de interés.

El primer eje crítico es el de los sesgos algorítmicos, los sistemas de aprendizaje automático aprenden de datos que reflejan, en gran medida, las desigualdades, estereotipos y patrones discriminatorios preexistentes en la sociedad; cuando estos sistemas se aplican a decisiones de mercadeo (segmentación de clientes, fijación de precios, asignación de crédito, focalización publicitaria); pueden replicar e incluso amplificar esas desigualdades, Acatrinei y otros. (2025) destacan que la transparencia percibida y las normas sociales son determinantes significativos de la confianza en herramientas de IA aplicadas a marketing, mientras que los riesgos percibidos y las emociones negativas erosionan esa confianza; lo cual refuerzan la necesidad de abordar los sesgos no solo como un problema técnico, sino como un desafío socio ético con consecuencias directas sobre la legitimidad de las prácticas comerciales.

El segundo es el de la gobernanza de datos; la calidad, integridad, representatividad y trazabilidad de los datos que alimentan los sistemas analíticos constituyendo condiciones necesarias para la generación de inteligencia confiable; Madanchian & Taherdoost (2025), enfatizan que las organizaciones se enfrentan al reto de balancear la transparencia con la complejidad inherente a los algoritmos modernos, frecuentemente opacos; esta tensión exige el desarrollo de herramientas de IA explicable que permitan a usuarios y grupos de interés

comprender la lógica subyacente a las decisiones automatizadas, así como la incorporación de procesos de auditoría algorítmica.

Un tercer eje es el de la autonomía del consumidor; la capacidad de los sistemas de IA para personalizar experiencias y mensajes a niveles micro plantea preguntas profundas sobre la libertad de elección y la posibilidad de manipulación, la regulación europea responde a estas preocupaciones prohibiendo expresamente sistemas de IA que manipulen el comportamiento humano causando daño físico o psicológico (Davola y otros., 2023); para un centro académico de inteligencia de mercados, este horizonte normativo implica el deber de promover prácticas analíticas que respeten la dignidad y autonomía de los consumidores como sujetos de derechos, y no únicamente como fuentes de datos o destinatarios de mensajes optimizados.

El cuarto es el de la rendición de cuentas, donde la trazabilidad de las decisiones automatizadas y la posibilidad de identificar responsabilidades cuando estas decisiones generan daños constituyen requisitos para que la IA opere dentro de un Estado de Derecho donde los marcos regulatorios emergentes exigen documentación técnica detallada, registros de operación y mecanismos de revisión humana significativa (Pérez, 2024); particularmente en sistemas de alto riesgo, estos requisitos representan no solo cargas de cumplimiento, sino también oportunidades para que centros académicos contribuyan al desarrollo de metodologías rigurosas de evaluación y monitoreo de sistemas analíticos.

4. Funciones estratégicas del CIMEC-IUE

A partir del marco conceptual y regulatorio expuesto, el CIMEC-IUE se concibe como una unidad que articula cuatro funciones estratégicas interdependientes, cada una de las cuales se vincula con las funciones sustantivas de la educación superior y con los desafíos contemporáneos de la regulación de la IA.

4.1 Función investigativa: generación de conocimiento aplicado

La primera función estratégica del CIMEC-IUE es la generación de conocimiento aplicado en inteligencia de mercados, analítica del consumidor y aplicaciones responsables de IA; esta función responde directamente a la misión investigativa de la IUE y se articula con la línea temática de IA, trabajo, economía e innovación que enmarca este capítulo, la investigación aplicada se distingue de la investigación básica por su orientación hacia problemas concretos del entorno productivo y social y por su compromiso con la generación de resultados utilizables por organizaciones públicas y privadas.

Esta función comprende cinco subfunciones; la primera es la producción académica indexada, expresada en artículos científicos, capítulos de libro y ponencias internacionales que aporten al estado del arte y posicionen a la IUE como un actor relevante en el campo, la segunda es la formulación de proyectos de investigación con financiación interna y externa, que canalicen recursos hacia el estudio sistemático de problemáticas específicas del territorio; la tercera es el desarrollo metodológico particularmente en la integración de métodos clásicos con técnicas avanzadas de IA, incluyendo análisis de sentimientos, minería de texto, modelos predictivos y experimentación digital; la cuarta es la validación empírica de marcos teóricos en contextos colombianos y latinoamericanos, lo cual contribuye a contextualizar y enriquecer literaturas frecuentemente dominadas por evidencias provenientes de economías desarrolladas y la quinta es la generación de productos de apropiación social del conocimiento, como notas técnicas, boletines divulgativos y materiales formativos accesibles a públicos no especializados.

La función investigativa se articula con las exigencias regulatorias contemporáneas mediante el desarrollo de investigación sobre los efectos de la IA en mercados y consumidores, lo cual aporta evidencia empírica para el debate público y la formulación de políticas; también se propone la investigación sobre metodologías

éticas de analítica avanzada, que contribuya al desarrollo de prácticas de gobernanza algorítmica; y por último mediante el análisis comparado de marcos regulatorios, que oriente a las organizaciones del territorio en su navegación de entornos normativos en transformación; al respecto Ju (2024) ofrece un ejemplo paradigmático de este tipo de investigación al desarrollar un marco de inteligencia competitiva en redes sociales que combina detección de tópicos de marca y predicción de engagement con clientes, mostrando cómo la rigurosidad metodológica puede coexistir con la pertinencia aplicada.

4.2 Función formativa: desarrollo de capacidades en analítica e IA

La segunda función estratégica del CIMEC-IUE es la formación de capacidades analíticas y en aplicaciones responsables de IA; esta se articula con la misión de la IUE y responde a una necesidad estructural del territorio; la formación de profesionales que combinen competencias técnicas avanzadas con comprensión profunda del comportamiento del consumidor, sensibilidad ética y capacidad de comunicación, la literatura es enfática al señalar que la efectividad de los centros de inteligencia depende críticamente del capital humano y particularmente de equipos multidisciplinarios capaces de articular ciencia de datos, mercadeo, comportamiento del consumidor, ética e ingeniería de datos (Cristofaro y otros., 2024).

La función formativa del CIMEC-IUE comprende cuatro subfunciones, siendo la primera es la incorporación curricular de contenidos sobre inteligencia de mercados, analítica del consumidor y aplicaciones de IA en los programas de pregrado y posgrado de la Facultad de Ciencias Empresariales y Comunicaciones; otra es el desarrollo de programas de educación continua, diplomados y cursos cortos dirigidos a profesionales en ejercicio que requieren actualización en estos campos; adicionalmente la formación de auxiliares de investigación y semilleros, en línea con la apuesta institucional por el desarrollo de jóvenes investigadores y por

último la generación de materiales formativos abiertos que democratizen el acceso a herramientas y conocimientos de inteligencia de mercados.

Un componente especialmente relevante es la formación en alfabetización en IA, en línea con las exigencias del AI Act europeo que desde febrero de 2025 requiere a proveedores y usuarios de sistemas de IA implementar acciones para desarrollar conocimientos suficientes sobre estas tecnologías en su personal; aunque esta exigencia regulatoria es directamente vinculante en el ámbito europeo (Minssen y otros., 2025), su lógica resulta extrapolable a contextos como el colombiano, donde el desarrollo de competencias críticas sobre IA constituye un imperativo formativo más allá del cumplimiento normativo, allí la alfabetización en IA debe abarcar tanto las capacidades técnicas básicas como la comprensión de los fundamentos éticos, sociales y regulatorios que enmarcan el uso de estas tecnologías.

4.3 Función de proyección social: vinculación con el entorno productivo

La tercera función estratégica del CIMEC-IUE es la proyección social, entendida como la articulación con el entorno productivo y social del Aburrá Sur para canalizar el conocimiento académico hacia la solución de problemas reales y para retroalimentar la academia con las preocupaciones, retos y aprendizajes del tejido empresarial, esta función adquiere relevancia particular en un contexto territorial donde coexisten organizaciones de distinto tamaño, sector y nivel de madurez digital, muchas de las cuales enfrentan brechas significativas para incorporar herramientas avanzadas de inteligencia de mercados.

La función de proyección social comprende tres subfunciones principales. la primera es la prestación de servicios de asesoría y consultoría a organizaciones del territorio, en proyectos de inteligencia de mercados, analítica del consumidor y diagnóstico de capacidades, la otra es el desarrollo de alianzas estratégicas con cámaras de comercio, gremios, agencias de desarrollo, gobiernos locales y otras

instituciones que permitan ampliar el alcance e impacto del centro; además de la organización de eventos, foros y espacios de diálogo entre academia, empresa y sector público, que faciliten la circulación de conocimiento y el aprendizaje colectivo.

Un aspecto particularmente importante en esta función es la atención a las pequeñas y medianas empresas (PYMES), que constituyen el grueso del tejido productivo del Aburrá Sur y que típicamente enfrentan barreras para adoptar herramientas avanzadas de inteligencia; la literatura reciente muestra que las PYMES pueden desarrollar capacidades dinámicas mediante la adopción incremental de analítica de datos, particularmente cuando esta adopción se acompaña de procesos de aprendizaje organizacional y de redes de cooperación con universidades y otras instituciones de conocimiento (Cristofaro y otros., 2024); teniendo en cuenta lo anterior, el CIMEC-IUE puede operar como un nodo de estas redes, ofreciendo a las PYMES acceso a capacidades analíticas que individualmente serían inaccesibles.

4.4 Función articuladora con políticas públicas y regulación

La cuarta función estratégica del CIMEC-IUE es la articulación con políticas públicas y marcos regulatorios; esta función, frecuentemente subestimada en la literatura sobre centros de inteligencia, adquiere centralidad en un contexto en el que la regulación de la IA se transforma rápidamente y en el que la calidad de las políticas públicas depende en buena medida de la disponibilidad de evidencia académica rigurosa, la función articuladora se ejerce en al menos cuatro dimensiones.

La primera dimensión es la vigilancia regulatoria, que implica el seguimiento sistemático de los desarrollos normativos en materia de IA, protección de datos, propiedad intelectual y derechos del consumidor, tanto en el ámbito nacional como

internacional, esta vigilancia debe traducirse en informes periódicos accesibles a las organizaciones del territorio, que les permitan anticipar exigencias y oportunidades; la segunda es la generación de evidencia para la formulación de políticas, mediante estudios que documenten los efectos de las prácticas analíticas sobre consumidores, trabajadores y ecosistemas competitivos; la tercera es la participación en espacios de consulta pública, mesas de trabajo regulatorio y foros multiactor donde se discuten marcos normativos en construcción y la cuarta es la promoción del cumplimiento responsable entre las organizaciones del territorio, mediante guías, protocolos y herramientas que faciliten la adopción de prácticas regulatoriamente adecuadas.

Esta función articuladora cobra especial sentido en el contexto de la línea temática IA, trabajo, economía e innovación; los marcos regulatorios emergentes están redefiniendo las condiciones bajo las cuales la IA puede ser desplegada en distintos sectores, con implicaciones directas sobre los modelos de negocio, las relaciones laborales y los procesos de innovación. Centros académicos como el CIMEC-IUE pueden contribuir a que estas redefiniciones ocurran de manera informada por evidencia, sensible a contextos territoriales y orientada al bienestar agregado.

5. Líneas de acción del CIMEC-IUE

Las cuatro funciones estratégicas anteriores se operacionalizan a través de cinco líneas de acción que son diferenciables analíticamente, se entrelazan en la práctica cotidiana del centro; estas líneas constituyen los ejes programáticos sobre los cuales el CIMEC-IUE estructura su agenda de trabajo, asigna recursos, evalúa resultados y se articula con actores externos.

5.1 Línea 1: Investigación aplicada en inteligencia de mercados con IA

La primera línea de acción es la investigación aplicada en inteligencia de mercados con incorporación de técnicas de IA. Esta línea constituye el eje sustantivo del centro y articula la función investigativa con las demás, sus actividades incluyen el desarrollo de proyectos de investigación con financiación interna y externa, la producción de artículos indexados, capítulos de libro y ponencias internacionales, la formulación de tesis de pregrado y posgrado vinculadas a las problemáticas del centro, y la sistematización metodológica de estudios de caso aplicables a la realidad del Aburrá Sur.

Las temáticas prioritarias de esta línea incluyen, entre otras, la analítica del comportamiento del consumidor mediante minería de texto y análisis de sentimientos en redes sociales, siguiendo aproximaciones como las desarrolladas por (Ju, 2024); Igualmente relevante es la investigación sobre segmentación inteligente de clientes mediante algoritmos de aprendizaje automático, en línea con desarrollos como los documentados por (Kasem y otros., 2023), los cuales muestran cómo la combinación de análisis RFM (recencia, frecuencia, valor monetario) con técnicas de *clustering K-means* permite identificar segmentos de clientes con características diferenciadas y orientar estrategias de mercadeo personalizadas; la investigación aplicada del CIMEC-IUE puede aportar significativamente a la adaptación contextual de estas técnicas a sectores económicos y poblaciones específicas del territorio.

Otras temáticas prioritarias incluyen la analítica predictiva del consumo en sectores estratégicos del Aburrá Sur (alimentos, comercio, servicios financieros, salud, educación); el estudio de la confianza, transparencia percibida y aceptación de herramientas de IA en mercadeo, siguiendo aproximaciones como las de Acatrinei y otros. (2025) y la investigación sobre sostenibilidad y consumo responsable, áreas en las que la IA ofrece oportunidades para alinear personalización con orientación ética; al respecto Theodorakopoulos & Theodoropoulou (2024) destacan vacíos

importantes en la literatura, particularmente sobre los efectos longitudinales de la personalización basada en big data sobre la confianza y lealtad del consumidor, vacíos que la investigación del CIMEC-IUE puede contribuir a llenar mediante estudios contextualizados al territorio.

Esta línea contempla el desarrollo metodológico mediante la integración de técnicas cualitativas tales como entrevistas semiestructuradas, etnografía digital, talleres participativos con técnicas cuantitativas y computacionales; la hibridez metodológica constituye una fortaleza distintiva del centro frente a aproximaciones puramente algorítmicas, pues permite capturar dimensiones simbólicas, culturales y emocionales del comportamiento del consumidor que resultan difíciles de aprehender mediante datos estructurados.

5.2 Línea 2: Formación y capacidades analíticas

La segunda línea de acción es el desarrollo de capacidades analíticas a través de la formación, articulada con la función docente de la IUE; esta línea contempla la incorporación curricular de contenidos sobre analítica de datos, IA aplicada al mercadeo y ética digital en los programas de pregrado de Administración de Empresas, Mercadeo, Comunicación Digital y afines, igualmente incluye la formulación de programas de posgrado en analítica del consumidor, ciencia de datos aplicada al mercadeo o áreas afines, dependiendo de la viabilidad institucional y la demanda territorial.

Más allá de la formación formal, la línea contempla el desarrollo de programas de educación continúa dirigidos a profesionales en ejercicio, estos programas en formatos de diplomados, cursos cortos, talleres y bootcamps; deben combinar rigor académico con orientación práctica, y deben aprovechar modalidades virtuales y mixtas para ampliar su cobertura territorial. Madanchian & Taherdoost (2025) subrayan la importancia de la formación interdisciplinaria que integre teorías éticas

con conocimiento técnico, perspectiva que orienta el diseño de los programas formativos del CIMEC-IUE.

La línea formativa contempla también el fortalecimiento de los semilleros de investigación y de programas estructurados de auxiliares de investigación, que permitan formar investigadores con competencias avanzadas en analítica del consumidor y aplicaciones de IA, así mismo, la articulación con redes nacionales e internacionales de formación y movilidad académica complementa esta apuesta, al ampliar las oportunidades de exposición de los estudiantes a frontera del conocimiento global (Hernández y otros., 2023).

Un componente transversal de esta línea es el desarrollo de competencias éticas y críticas, la formación en analítica avanzada no debe limitarse a la transferencia de habilidades técnicas; sino que debe incorporar la reflexión sobre las implicaciones sociales, laborales y políticas del despliegue de estas tecnologías, la línea formativa del CIMEC-IUE asume, en este sentido, una responsabilidad pedagógica explícita; la cual se centra en formar profesionales capaces no solo de implementar soluciones analíticas eficientes, sino también de cuestionar críticamente sus supuestos, reconocer sus limitaciones y contribuir a su orientación ética.

5.3 Línea 3: Asesoría y consultoría especializada

La tercera línea de acción es la asesoría y consultoría especializada, entendida como una modalidad concreta de proyección social que canaliza las capacidades analíticas del centro hacia la solución de problemas específicos de organizaciones del territorio; esta línea se diferencia de la consultoría comercial tradicional en al menos tres aspectos como su anclaje académico, que asegura rigor metodológico y vigilancia ética, la orientación formativa, que busca dejar capacidades instaladas en las organizaciones atendidas y la finalidad de retroalimentación investigativa, que convierte cada proyecto en una oportunidad de aprendizaje sistematizable.

Las modalidades específicas de asesoría que esta línea contempla incluyen, los diagnósticos de capacidades analíticas, que permiten a las organizaciones identificar fortalezas, brechas y oportunidades en el uso de datos para la toma de decisiones; los estudios de mercado y de comportamiento del consumidor adaptados a sectores y problemáticas específicas y el acompañamiento en la implementación de proyectos de analítica avanzada, particularmente en organizaciones con recursos limitados que requieren orientación experta para hacer un uso eficiente de tecnologías como aprendizaje automático y procesamiento de lenguaje natural; y la asesoría en cumplimiento normativo y gobernanza de datos, particularmente relevante en el contexto de las exigencias regulatorias emergentes sobre IA.

Cristofaro y otros. (2024) destaca que la efectividad de las metodologías de toma de decisiones basadas en datos depende de su adaptación contextual y de la integración entre rigor metodológico y aprendizaje organizacional; esta perspectiva orienta la práctica consultiva del CIMEC-IUE donde cada proyecto se concibe no como una transacción aislada, sino como un proceso de co-construcción de capacidades entre el centro y la organización atendida, esta lógica permite que el impacto de las asesorías trascienda los productos entregables y se materialice en transformaciones de mediano plazo en la cultura analítica de las organizaciones.

5.4 Línea 4: Observatorio de tendencias en IA, trabajo y consumo

La cuarta línea de acción es el desarrollo de un Observatorio de Tendencias en IA, Trabajo y Consumo, concebido como un dispositivo permanente de vigilancia, análisis y comunicación sobre las transformaciones que la IA introduce en estos tres ámbitos interrelacionados, esta línea articula las funciones investigativas, de proyección social y articuladora con políticas públicas, y constituye un mecanismo concreto para que el CIMEC-IUE genere valor sistemático para múltiples grupos de interés del territorio.

El Observatorio comprende tres componentes principales. El primero es un sistema de vigilancia tecnológica y regulatoria sobre desarrollos en IA aplicada al mercadeo, la analítica del consumidor y los mercados de trabajo; este sistema combina monitoreo automatizado de fuentes académicas, técnicas y regulatorias con interpretación experta por parte del equipo del centro; otro componente es la producción periódica de informes, boletines y notas técnicas que sistematicen los hallazgos del observatorio y los pongan a disposición de empresas, instituciones públicas y comunidad académica y por último se plantea el desarrollo de paneles temáticos, foros y eventos que permitan la discusión cualificada de las tendencias identificadas.

Un eje temático central del Observatorio son los efectos de la IA sobre los mercados de trabajo; la evidencia disponible sugiere que estos efectos son significativos pero heterogéneos, en América Latina y el Caribe, organismos multilaterales han estimado que gran parte de los empleos se verán influenciados por la IA generativa, con potencial de mejorar la productividad de millones de trabajadores pero también con riesgos de desplazamiento laboral, particularmente en ocupaciones rutinarias (Banh & Strobel, 2023); la gran informalidad laboral de la región matiza estos efectos respecto a las economías avanzadas; al respecto para el Aburrá Sur donde coexisten sectores manufactureros tradicionales con servicios avanzados y emprendimientos digitales, la comprensión local de estas dinámicas es crítica y constituye un objeto privilegiado del Observatorio.

Otro eje temático es la transformación de los hábitos de consumo bajo el influjo de la IA. La personalización a escala, los sistemas de recomendación, los asistentes conversacionales y la convergencia entre comercio electrónico y plataformas sociales están redefiniendo lo que significa ser consumidor en el siglo XXI (Bhuiyan, 2024, p. 166); Theodorakopoulos & Theodoropoulou (2024) muestran cómo la analítica de big data ha transformado fundamentalmente la comprensión y la influencia sobre el comportamiento del consumidor, abriendo posibilidades inéditas

pero también planteando interrogantes sobre confianza, lealtad y autonomía; por lo tanto el Observatorio del CIMEC-IUE puede aportar evidencia contextualizada sobre estas dinámicas en el territorio colombiano.

Un tercer eje son las dinámicas de innovación empresarial vinculadas a la adopción de IA; la incorporación de técnicas analíticas avanzadas modifica los procesos de descubrimiento de oportunidades, prototipado, validación y escalamiento de productos y servicios; Cristofaro y otros. (2024) sistematizan, mediante el concepto de growth hacking, una aproximación científica a la decisión basada en datos que se ha vuelto particularmente relevante en startups y empresas digitalmente nativas pero cuya lógica es extrapolable a organizaciones tradicionales que buscan acelerar sus procesos de innovación; el Observatorio puede documentar la difusión y adaptación de estas prácticas en el tejido productivo del Aburrá Sur.

5.5 Línea 5: Ética, gobernanza algorítmica y cumplimiento normativo

La quinta línea de acción es la consolidación de un eje transversal de ética, gobernanza algorítmica y cumplimiento normativo; esta línea atraviesa las cuatro anteriores y se erige como un componente sustantivo (no accesorio) del valor que el CIMEC-IUE ofrece a sus grupos de interés, su justificación es doble; por una parte, las exigencias regulatorias emergentes en materia de IA, datos personales y derechos del consumidor configuran un nuevo piso de obligaciones para cualquier organización que opere con sistemas analíticos avanzados y por otra la legitimidad social del despliegue de IA en mercadeo y analítica depende crecientemente de la capacidad de los actores para demostrar prácticas éticamente sostenibles.

Esta línea comprende cuatro componentes operativos; el primero es el desarrollo de marcos éticos y protocolos de gobernanza algorítmica adaptados a la realidad de las organizaciones colombianas; Madanchian & Taherdoost (2025) plantean que la transición desde principios éticos abstractos hacia prácticas operativas constituye

una de las brechas más significativas en el campo, brecha que centros académicos como el CIMEC-IUE pueden contribuir a cerrar mediante el desarrollo de herramientas concretas, listas de verificación, evaluaciones de impacto algorítmico y procedimientos de revisión humana significativa.

El segundo es el acompañamiento a organizaciones en la implementación de programas de cumplimiento normativo, particularmente en lo relativo a la Ley 1581 de 2012, la Circular Externa 002 de 2024 de la SIC y otras disposiciones aplicables; este acompañamiento incluye la formulación de políticas internas de protección de datos, la realización de evaluaciones de impacto sobre la privacidad y el diseño de procedimientos para el ejercicio de derechos por parte de los titulares de datos y aquí la asesoría en cumplimiento normativo se ofrece como un servicio especializado, articulado con la línea de asesoría y consultoría.

El tercer componente es la promoción de la auditoría algorítmica como práctica estándar; las auditorías algorítmicas permiten verificar el comportamiento de los sistemas de IA frente a criterios de equidad, transparencia, exactitud y cumplimiento de objetivos declarados; Bignami y otros. (2025) destacan que el AI Act europeo introduce mecanismos de evaluación de conformidad obligatorios para sistemas de alto riesgo y aunque estos requisitos no son directamente vinculantes en el contexto colombiano, su lógica resulta extrapolable y puede orientar el desarrollo de prácticas voluntarias de auditoría que fortalezcan la confianza en los sistemas analíticos; el CIMEC-IUE puede operar como una entidad académica imparcial capaz de realizar auditorías independientes para organizaciones del territorio.

El cuarto componente es el desarrollo de capacidades formativas específicas en ética digital, gobernanza algorítmica y cumplimiento normativo, dirigidas a estudiantes de la IUE, profesionales en ejercicio y actores institucionales, las cuales son capacidades cada vez más demandadas en el mercado laboral, dado que las organizaciones requieren talento capaz de navegar la complejidad técnica, ética y

regulatoria de los sistemas analíticos contemporáneos; la oferta formativa de esta línea complementa la línea 2 (Formación) con un énfasis específico en las dimensiones normativas y deontológicas.

6. CIMEC-IUE en el ecosistema de IA, trabajo, economía e innovación

La inscripción del CIMEC-IUE en la línea temática de IA, trabajo, economía e innovación del libro sobre regulación de la inteligencia artificial requiere explicitar las articulaciones entre la propuesta del centro y los grandes debates regulatorios contemporáneos. Esta sección desarrolla cuatro articulaciones principales.

En primera instancia, el CIMEC-IUE se sitúa en la intersección entre innovación empresarial y protección del consumidor. Las aplicaciones contemporáneas de IA en mercadeo y analítica del consumidor abren posibilidades inéditas para la personalización, la eficiencia y la generación de valor pero también plantean riesgos en términos de manipulación, sesgos, privacidad y autonomía; los marcos regulatorios emergentes buscan precisamente equilibrar estos dos imperativos. Acatrinei y otros. (2025) muestran empíricamente que la confianza del consumidor en herramientas de IA depende de factores como la transparencia percibida y las normas sociales, lo cual sugiere que la sostenibilidad económica de la innovación con IA está estrechamente vinculada a la calidad de las prácticas de protección del consumidor; el CIMEC-IUE asume esta articulación como principio orientador, evitando aproximaciones que privilegien la eficiencia técnica sobre la legitimidad social; adicionalmente el CIMEC-IUE aporta a la comprensión de las transformaciones del trabajo bajo el influjo de la IA.

Aunque la analítica del consumidor no es directamente una temática laboral, sus desarrollos modifican significativamente las competencias requeridas en los mercados de trabajo, particularmente en sectores como mercadeo, ventas, servicio al cliente, investigación de mercados y gestión de marca (Estupiñán & Mesa, 2023); la automatización de tareas analíticas rutinarias coexiste con una creciente demanda de profesionales capaces de combinar competencias técnicas avanzadas

con sensibilidad ética y comunicación estratégica; el CIMEC-IUE, mediante su función formativa y su Observatorio, contribuye a anticipar y dar respuesta a estas transformaciones, promoviendo trayectorias laborales sostenibles para los profesionales del territorio.

Otro articulador se revisa en el hecho de que el CIMEC-IUE contribuye a la dinamización económica del Aburrá Sur; la capacidad de las organizaciones del territorio para incorporar inteligencia analítica avanzada en sus procesos de decisión constituye un determinante crítico de su competitividad; Xu & Lu (2026) demuestran empíricamente que la toma de decisiones basada en datos impacta positivamente el desempeño internacional de las firmas, mediante mecanismos como la prevención de la contaminación, la innovación verde y la divulgación de información de sostenibilidad, La transferencia de capacidades analíticas desde el ámbito académico hacia las organizaciones del territorio, por la vía de las líneas de acción del CIMEC-IUE, contribuye a fortalecer un ecosistema económico más productivo, innovador y sostenible.

El CIMEC-IUE se concibe como un actor del ecosistema de innovación regional; Valencia-Arías y otros. (2025); en una revisión sistemática sobre la transferencia de tecnología en Colombia, evidencian las dinámicas y desafíos que enfrenta este proceso en el país, particularmente en lo relativo a la articulación universidad-empresa, las políticas públicas de innovación y los sistemas regionales de conocimiento, los autores identifican brechas importantes en la conexión entre investigación universitaria y aplicación productiva, brechas que centros como el CIMEC-IUE pueden contribuir a reducir mediante mecanismos formales de transferencia de conocimiento, asesoría especializada y formación contextualizada.

La regulación de la IA introduce en este ecosistema, una dimensión adicional; las exigencias regulatorias contemporáneas no son meras restricciones, sino que pueden operar como catalizadores de innovación responsable; los marcos como el

Al Act europeo prevén explícitamente mecanismos como los “*regulatory sandboxes*” que permiten experimentar con sistemas innovadores en entornos supervisados; Mahmutovic (2025) destaca que estos mecanismos buscan armonizar la innovación con la supervisión ética; el CIMEC-IUE puede contribuir a explorar la viabilidad y adaptación de instrumentos análogos en el contexto colombiano, aportando al desarrollo de un ecosistema de innovación que reconcilie dinamismo tecnológico con derechos y bienestar.

7. Consideraciones finales y prospectiva

La propuesta del Centro de Inteligencia de Mercados y Analítica del Consumidor de la Institución Universitaria de Envigado (CIMEC-IUE) responde a un momento histórico particular, en el que la inteligencia artificial reconfigura aceleradamente las prácticas de mercadeo y la analítica del consumidor, mientras los marcos regulatorios buscan adaptarse a esta transformación; este capítulo ha definido el concepto del CIMEC-IUE como una unidad académica y de investigación aplicada que articula las funciones sustantivas de docencia, investigación y proyección social en torno al uso responsable de datos e IA y ha sistematizado sus cuatro funciones estratégicas: investigativa, formativa, de proyección social y articuladora con políticas públicas; además ha propuesto cinco líneas de acción operativas: investigación aplicada, formación, asesoría especializada, observatorio temático y eje transversal de ética y gobernanza.

Varios elementos de la propuesta merecen ser resaltados a modo de síntesis; primero, el centro se concibe explícitamente como una entidad académica con orientación aplicada y anclaje territorial, lo cual lo distingue de modelos puramente comerciales o puramente teóricos; segundo, se asume desde su diseño la dimensión ética y regulatoria como componente sustantivo, no como elemento accesorio o restricción, en tercer lugar se privilegia la hibridez metodológica, articulando técnicas tradicionales con desarrollos recientes en IA; como cuarto

punto se busca explícitamente la democratización del acceso a herramientas analíticas, particularmente para las pequeñas y medianas empresas del territorio y por último se reconoce la necesidad de articular el centro con ecosistemas de innovación, redes académicas y marcos regulatorios en transformación.

La materialización de esta propuesta enfrenta, sin embargo, varios desafíos como los recursos porque la estructuración efectiva de un centro con las características descritas requiere inversión sostenida en talento, infraestructura tecnológica, alianzas y comunicaciones (Cristofaro y otros., 2024); el financiamiento institucional inicial debe complementarse, a mediano plazo, con recursos provenientes de proyectos externos, servicios de asesoría y alianzas estratégicas.

Otro desafío es el del talento, puesto que la conformación de un equipo multidisciplinario con competencias en analítica de datos, IA, mercadeo, comportamiento del consumidor, ética y comunicación constituye una de las tareas más complejas, particularmente en un mercado laboral en el que estas competencias son altamente demandadas; la articulación con redes académicas y la formación de talento joven mediante semilleros y auxiliares de investigación son estrategias críticas para abordar este desafío.

Otro aspecto a considerar es la legitimidad; el posicionamiento del CIMEC-IUE como un actor relevante en el ecosistema de inteligencia de mercados del Aburrá Sur requiere la construcción gradual de credibilidad mediante producción académica visible, proyectos de impacto demostrable y alianzas con actores reputados; la literatura sobre centros de inteligencia coincide en que esta legitimidad se construye en plazos medianos a largos y exige inversión sostenida en comunicación estratégica, eventos y alianzas; otro desafío es el de la articulación interna, puesto que la integración efectiva de las funciones sustantivas (docencia, investigación, proyección social) en una unidad operativa coherente exige mecanismos de gobernanza interna que reconozcan las lógicas y tiempos diferenciados de cada función.

Mirando hacia adelante, es posible trazar distintos horizontes para el CIMEC-IUE según el momento del proceso; en el corto plazo, la tarea más urgente será afianzar una estructura interna funcional, establecer focos temáticos claros y generar productos concretos que acrediten su pertinencia. Más adelante, cuando esa base esté consolidada, surgirán condiciones para buscar nuevas fuentes de recursos, sumar talento al equipo y tejer vínculos con redes de investigación dentro y fuera del país. A más largo plazo, el centro tiene condiciones reales para posicionarse como punto de referencia en la región en temas como la analítica del consumidor con inteligencia artificial y el uso ético de sistemas algorítmicos, aportando al mismo tiempo al crecimiento académico de la IUE, a la dinámica productiva del Aburrá Sur y a la construcción de marcos normativos pensados desde y para el contexto local.

La regulación de la inteligencia artificial será, en los años venideros, uno de los asuntos más debatidos en el plano jurídico y ético a nivel mundial. Instrumentos como el Reglamento (UE) 2024/1689 marcarán un camino que otras jurisdicciones seguirán, ajustarán o pondrán en cuestión según sus propias circunstancias.

Colombia, como otras economías en desarrollo, tiene ante sí el reto de diseñar su propio andamiaje regulatorio: uno que retome las lecciones del debate internacional, pero que al mismo tiempo reconozca las características de su aparato productivo, los alcances de sus instituciones y las prioridades que definen su modelo de desarrollo. En ese recorrido, los centros académicos especializados pueden cumplir una función que difícilmente puede ser reemplazada: producir evidencia sólida, formar capacidades técnicas y críticas, propiciar conversaciones entre actores diversos y contribuir a consolidar una cultura analítica que sea, a la vez, democrática, responsable y viable en el tiempo.

El CIMEC-IUE aspira a contribuir, modestamente, pero con consistencia, a este propósito; su éxito no se medirá únicamente por sus indicadores de productividad académica o por el número de organizaciones atendidas, sino por su capacidad de

fortalecer las condiciones bajo las cuales la inteligencia artificial despliega sus potenciales sin sacrificar derechos, diversidades y posibilidades de bienestar humano.

En la convergencia entre IA, trabajo, economía e innovación, hay un campo amplio para que la academia haga aportes sustantivos, la propuesta aquí presentada constituye un primer paso en esa dirección.

Referencias

- Acatrinei, C., Apostol, I. G., Barbu, L. N., Chivu, R.-G., & Orzan, M. (2025). Artificial Intelligence in Digital Marketing: Enhancing Consumer Engagement and Supporting Sustainable Behavior Through Social and Mobile Networks. *Sustainability*, 17(14), 6638–6638. <https://doi.org/10.3390/su17146638>
- Alvarado, A. (2025). Lessons from the EU AI Act. *Patterns*, 6(2), 101183–101183. <https://doi.org/10.1016/j.patter.2025.101183>
- Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33(1). <https://doi.org/10.1007/s12525-023-00680-1>
- Bhuiyan, M. S. (2024). The Role of AI-Enhanced Personalization in Customer Experiences. *Journal of Computer Science and Technology Studies*, 6(1), 162–169. <https://doi.org/10.32996/jcsts.2024.6.1.17>
- Bignami, E., Russo, M., Semeraro, F., & Bellini, V. (2025). Balancing Innovation and Control: The European Union AI Act in an Era of Global Uncertainty. *JMIR AI*, 4. <https://doi.org/10.2196/75527>
- Capolupo, N., Romeo, E., Giampaola, V., & Adinolfi, P. (2025). To AI or not to AI? Unpacking the Hamletian dilemma in public organizations' work processes through a socio-technical perspective. *Business Process Management Journal*, 1–29. <https://doi.org/10.1108/bpmj-07-2025-1188>
- Chowdhury, S., Dey, P. K., Joel-Edgar, S., Bhattacharya, S., Rodríguez-Espíndola, O., Abadie, A., & Truong, L. (2022). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review*, 33(1), 100899–100899. <https://doi.org/10.1016/j.hrmr.2022.100899>
- Chulvi, C. P. (2023). Transparencia algorítmica en los medios de comunicación y las plataformas digitales. *Revista Española de La Transparencia*, 107–136. <https://doi.org/10.51915/ret.308>
- Cristofaro, M., Giardino, P. L., & Barboni, L. (2024). Growth hacking: A scientific approach for data-driven decision making. *Journal of Business Research*, 186, 115030–115030. <https://doi.org/10.1016/j.jbusres.2024.115030>
- Dávalos, R. M. F. (2022). Editorial Vol. 2, Núm. 4 - Transformación Digital y Sostenibilidad Empresarial. *Revista de Análisis y Difusión de Perspectivas Educativas y Empresariales*, 2(4), 6–8. <https://doi.org/10.56216/radee022022dic.e00>
- Davola, A., Querci, I., & Romani, S. (2023). No Consumer Is an Island—Relational Disclosure as a Regulatory Strategy to Advance Consumer Protection Against Microtargeting. *Journal of Consumer Policy*, 46(1), 1–25. <https://doi.org/10.1007/s10603-022-09530-7>
- Estupiñán, A. M. L., & Mesa, L. P. (2023). Inteligencia Artificial: el futuro del empleo. *Revista Lecciones Vitales*. <https://doi.org/10.18046/rlv.2023.6118>
- Guadamuz, A. (2024). The EU's Artificial Intelligence Act and copyright. *The Journal of World Intellectual Property*, 28(1), 213–219. <https://doi.org/10.1111/jwip.12330>
- Hermann, E. (2021). Leveraging Artificial Intelligence in Marketing for Social Good—An Ethical Perspective. *Journal of Business Ethics*, 179(1), 43–61. <https://doi.org/10.1007/s10551-021-04843-y>
- Hernández, E. (2024). Hacia una implementación ética e inclusiva de la Inteligencia Artificial en las organizaciones: un marco multidimensional. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2405.00225>
- Hernández, R. J. Á., Islas, A. C., & Barrios, A. V. (2023). Redes estratégicas e investigación colaborativa en el marketing. *Vinculatégica EFAN*, 9(6), 193–207. <https://doi.org/10.29105/vtga9.6-698>

- Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50. <https://doi.org/10.1007/s11747-020-00749-9>
- Ju, X. (2024). A social media competitive intelligence framework for brand topic identification and customer engagement prediction. *PLoS ONE*, 19(11). <https://doi.org/10.1371/journal.pone.0313191>
- Kasem, M. S., Hamada, M. A., & Taj-Eddin, I. (2023). Customer profiling, segmentation, and sales prediction using AI in direct marketing. *Neural Computing and Applications*, 36(9), 4995–5005. <https://doi.org/10.1007/s00521-023-09339-6>
- Madanchian, M., & Taherdoost, H. (2025). Ethical theories, governance models, and strategic frameworks for responsible AI adoption and organizational success. *Frontiers in Artificial Intelligence*, 8, 1619029–1619029. <https://doi.org/10.3389/frai.2025.1619029>
- Mahmutovic, A. (2025). The EU AI Act: a proactive framework for comprehensive AI regulation. *International Journal of Law and Information Technology*, 33. <https://doi.org/10.1093/ijlit/eaaf028>
- Metcalfe, G. S., & Daniel, T. A. (2025). Working with organizational systems. In *American Psychological Association eBooks* (pp. 603–623). American Psychological Association. <https://doi.org/10.1037/0000432-030>
- Minssen, T., Smith, H. J., Mann, S. P., García, A. M., & Negrete, J. (2025). THE EUROPEAN AI ACT: PROMISES, PERILS & TRADE-OFFS IN A COMPLEX REGULATORY ECOSYSTEM. *Research Portal Denmark*, 1, 330–372. <https://local.forskningsportal.dk/local/dki-cgi/ws/cris-link?src=ku&id=ku-2fbec26-4adb-487a-92d7-469ae5a7b774&ti=THE%20EUROPEAN%20AI%20ACT%20%3A%20PROMISES%2C%20PERILS%20%26amp%3B%20TRADE-OFFS%20IN%20A%20COMPLEX%20REGULATORY%20ECOSYSTEM>
- Moraes, C. A. de, Toledo, L. A., & Garber, M. F. (2015). Market Intelligence: An Essay on the Competitiveness. *Future Studies Research Journal Trends and Strategies*, 7(2), 3–20. <https://doi.org/10.7444/future.v7i2.195>
- Nidhishree, M. S. (2024). Emotional Artificial Intelligence: Methodologies, Benefits, and Drawbacks. *International Journal for Research in Applied Science and Engineering Technology*, 12(4), 4004–4009. <https://doi.org/10.22214/ijraset.2024.60851>
- Pérez, J. A. B. (2024). La dignidad humana frente a la inteligencia artificial: un análisis ético y normativo en América Latina. *Trayectorias Humanas Trascontinentales*, 18. <https://doi.org/10.25965/trahs.6367>
- Sarango, A. F. H., Yacelga, A. P. M., Sevilla, R. M. N., Sailema, M. E. C., & Lescano, J. C. P. (2023). Inteligencia de negocios en la gestión empresarial: un análisis a las investigaciones científicas mundiales. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4(1). <https://doi.org/10.56712/latam.v4i1.493>
- Szukits, Á., & Móricz, P. (2023). Towards data-driven decision making: the role of analytical culture and centralization efforts. *Review of Managerial Science*, 18(10), 2849–2887. <https://doi.org/10.1007/s11846-023-00694-1>
- Theodorakopoulos, L., & Theodoropoulou, A. (2024). Leveraging Big Data Analytics for Understanding Consumer Behavior in Digital Marketing: A Systematic Review. *Human Behavior and Emerging Technologies*, 2024(1). <https://doi.org/10.1155/2024/3641502>
- Valencia-Arías, A., Restrepo, K. C. B., Enciso, E. M. V., Valencia, J., Hernandez, J. C. R., & Ramírez-Ramírez, D. M. (2025). Dynamics and challenges of technology transfer in Colombia: a systematic literature review. *Frontiers in Research Metrics and Analytics*, 10, 1628141–1628141. <https://doi.org/10.3389/firma.2025.1628141>

- Vanegas, W. J., Santamaria, M., & Molina, M. K. R. (2024). Políticas públicas ante la revolución de la inteligencia artificial en Colombia. *Revista Venezolana de Gerencia*, 29(106), 865–883. <https://doi.org/10.52080/rvgluz.29.106.26>
- Vishnoi, S. K., Mathur, S., & Sharma, S. (2025). Marketing intelligence for intelligent marketing: a comprehensive definition and framework. *Journal of Advances in Management Research*, 23(1), 121–147. <https://doi.org/10.1108/jamr-02-2024-0040>
- Xu, M., & Lu, B. (2026). The data-driven decision-making, sustainable value creation, and international firm performance: Micro-level evidence based on AI language models. *PLoS ONE*, 21(2). <https://doi.org/10.1371/journal.pone.0340731>

ENTRE ALGORITMOS Y EMOCIONES: LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN EL DEPORTE Y SUS LÍMITES HUMANOS

Between algorithms and emotions: the regulation of artificial intelligence in
sport and its human limits

Marlene Atenea Grimaldo Ramírez

ramireznea51@gmail.com

Universidad Autónoma Metropolitana unidad Azcapotzalco

<https://orcid.org/0009-0009-7784-941X>

Resumen

El avance de la inteligencia artificial (IA) ha transformado el ecosistema deportivo, impactando el rendimiento, la toma de decisiones y la interacción con los aficionados. Sin embargo, su implementación plantea desafíos éticos, jurídicos y sociales relacionados con la equidad, la transparencia, la privacidad de los datos y la autonomía de los deportistas. Este estudio analiza la necesidad de una regulación específica de la IA en el deporte mediante un enfoque que integra el análisis teórico, el estado del arte y una propuesta normativa. Asimismo, examina los efectos de la automatización y los posibles escenarios futuros que podrían modificar la naturaleza del deporte. A partir de ello, se propone un modelo regulatorio basado en principios como la autonomía humana, la equidad, la transparencia y la protección de datos, así como en un enfoque de gobernanza multinivel. Se concluye que la regulación de la IA debe ir más allá de la gestión de riesgos técnicos y orientarse a preservar los valores que sustentan el deporte como actividad humana.

Palabras clave: Inteligencia artificial; deporte; regulación; ética; gobernanza; automatización; privacidad de datos; equidad deportiva

Abstract

The advancement of artificial intelligence (AI) has transformed the sports ecosystem, impacting performance, decision-making, and interaction with fans. However, its implementation raises ethical, legal, and social challenges related to fairness, transparency, data privacy, and athletes' autonomy. This study analyzes the need for a specific regulation of AI in sports through an approach that integrates theoretical analysis, the state of the art, and a regulatory proposal. It also examines the effects of automation and possible future scenarios that could modify the nature of sports. Based on this, a regulatory model is proposed grounded on principles such as human autonomy, fairness, transparency, and data protection, as well as on a multi-level governance approach. The study concludes that AI regulation must go beyond the management of technical risks and focus on preserving the values that underpin sports as a human activity.

Keywords: artificial intelligence; sports; regulation; ethics; governance; automation; data privacy; sports fairness.

1. Introducción

En las últimas décadas, la inteligencia artificial (IA) ha emergido como una de las tecnologías más influyentes en la transformación de los sistemas sociales, económicos y culturales a nivel global. Su capacidad para procesar grandes volúmenes de datos, aprender de ellos y generar predicciones o decisiones automatizadas ha permitido optimizar procesos en múltiples sectores, desde la medicina hasta la seguridad, pasando por las finanzas, la educación y la industria. En este contexto, el deporte no ha permanecido ajeno a esta revolución tecnológica, sino que, por el contrario, se ha convertido en un espacio particularmente fértil para la adopción y experimentación de sistemas basados en IA.

La incorporación de la inteligencia artificial en el ámbito deportivo ha modificado de manera sustancial tanto la práctica como la gestión del deporte. Actualmente, la IA se utiliza para analizar el rendimiento de los atletas, diseñar estrategias de competencia, prevenir lesiones, optimizar entrenamientos, mejorar el arbitraje y fortalecer la interacción con los aficionados. Este fenómeno responde, en gran medida, a la creciente disponibilidad de datos y al desarrollo de algoritmos capaces de identificar patrones complejos en contextos altamente dinámicos. Como se ha señalado en la literatura, “la IA promete mejorar la toma de decisiones, haciéndolas más rápidas, eficientes, seguras y fiables” (Pérez Triviño, 2023), lo que explica su rápida adopción en el ecosistema deportivo.

Sin embargo, el uso de la IA en el deporte no puede entenderse únicamente desde una perspectiva instrumental o técnica. A diferencia de otros sectores, el deporte no solo constituye una actividad económica o profesional, sino también un fenómeno social y cultural profundamente arraigado en valores como la equidad, el juego limpio, la meritocracia, la competencia justa y la emoción colectiva. En este sentido, la introducción de tecnologías que alteran la forma en que se entrena, se compete y se evalúa el desempeño plantea interrogantes fundamentales sobre la preservación de dichos valores.

Uno de los aspectos más relevantes en este debate es el impacto de la IA sobre los derechos de los deportistas. La literatura especializada advierte que el uso de sistemas automatizados puede afectar dimensiones esenciales como la autonomía, la privacidad, la seguridad y la igualdad. En particular, como señala Pérez Triviño (2023) “esta puede impactar en varios principios y derechos de los deportistas: a) la pérdida de autonomía; b) la transparencia; c) la discriminación; d) la seguridad” . Estas preocupaciones adquieren una relevancia especial en el deporte, donde los atletas se encuentran en una posición de vulnerabilidad frente a organizaciones, clubes y federaciones que controlan el acceso a recursos, información y oportunidades.

A ello se suma el papel central que desempeñan los datos en el funcionamiento de los sistemas de IA. En el deporte, estos datos no solo incluyen estadísticas de rendimiento, sino también información altamente sensible, como datos biométricos, fisiológicos y de salud. La recopilación y análisis de esta información ha permitido avances significativos en la optimización del rendimiento y la prevención de lesiones. No obstante, también ha incrementado los riesgos asociados a la privacidad y al uso indebido de los datos. En este sentido, se advierte que “a medida que los sistemas de IA recopilan y analizan los datos de los jugadores, aumenta el riesgo de fuga o mal uso de los datos” (Kim y otros., 2025), lo que podría tener consecuencias negativas tanto para la carrera profesional de los atletas como para sus derechos fundamentales.

Desde una perspectiva ética, los desafíos asociados al uso de la IA en el deporte han sido ampliamente discutidos en la literatura reciente. Diversos estudios coinciden en que las principales preocupaciones se concentran en cuatro dimensiones clave: la equidad y el sesgo, la transparencia y explicabilidad, la privacidad de los datos y la responsabilidad en la toma de decisiones. En este sentido, una revisión sistemática destaca que “las preocupaciones éticas se agrupan en equidad y sesgo, transparencia y explicabilidad, privacidad de los datos y responsabilidad en la aplicación de la IA” (Kim y otros., 2025). Estas dimensiones no solo reflejan riesgos técnicos, sino también tensiones normativas y sociales que requieren ser abordadas mediante marcos regulatorios adecuados.

Particularmente preocupante es el problema del sesgo algorítmico. Dado que los sistemas de IA se entrenan a partir de datos históricos, existe el riesgo de que reproduzcan y amplifiquen desigualdades preexistentes. En el contexto deportivo, esto puede traducirse en decisiones injustas en procesos como la selección de jugadores, la evaluación del rendimiento o la detección de talento. Como se ha señalado, “si los datos de entrenamiento contienen sesgos, pueden generar resultados injustos” (Kim y otros., 2025), lo que pone en entredicho el principio de equidad que sustenta la competencia deportiva.

Asimismo, la falta de transparencia en los sistemas de IA representa un desafío significativo. Muchos de estos sistemas operan como “cajas negras”, lo que dificulta comprender cómo se generan sus resultados. Esta opacidad no solo limita la capacidad de supervisión y control, sino que también puede afectar la confianza de los atletas, entrenadores y aficionados. En ámbitos como el arbitraje o la detección de dopaje, donde las decisiones pueden tener consecuencias determinantes, la falta de explicabilidad resulta especialmente problemática.

Más allá de los dilemas éticos y jurídicos, el avance de la IA en el deporte plantea interrogantes sobre el futuro mismo de esta actividad. En particular, surge la cuestión de hasta qué punto la automatización puede transformar la naturaleza del deporte sin comprometer su esencia. La posibilidad de que sistemas de IA asuman un papel cada vez más central en la competencia —e incluso la eventual sustitución de los atletas humanos por entidades tecnológicas— ha dejado de ser un escenario puramente especulativo para convertirse en un tema de discusión relevante.

En este sentido, algunos estudios advierten que el uso excesivo de tecnologías automatizadas podría afectar la experiencia de los aficionados. Como Kim y otros., (2025) han señalado, “depender exclusivamente de las técnicas de IA puede disminuir el entusiasmo y la pasión de los aficionados” . Esta afirmación pone de relieve una dimensión frecuentemente ignorada en el debate sobre la IA: su impacto en la dimensión emocional y simbólica del deporte. La emoción, la incertidumbre y la identificación con los atletas son elementos fundamentales que podrían verse alterados en un entorno altamente automatizado.

Adicionalmente, el desarrollo de la IA en el deporte plantea desafíos en términos de gobernanza. A diferencia de otros sectores, el deporte se caracteriza por una compleja interacción entre actores públicos y privados, lo que dificulta la implementación de marcos regulatorios homogéneos. La coexistencia de regulaciones nacionales, normas internacionales y políticas internas de organizaciones deportivas genera un entorno fragmentado que complica la supervisión del uso de la IA.

En este contexto, la necesidad de una regulación específica para la IA en el deporte se vuelve evidente. No se trata únicamente de establecer límites al uso de la tecnología, sino de diseñar un marco que permita aprovechar sus beneficios sin comprometer los valores fundamentales del deporte. Esto implica abordar cuestiones como la protección de los datos, la transparencia de los algoritmos, la equidad en la competencia y la responsabilidad en la toma de decisiones.

A partir de estas consideraciones, el presente artículo tiene como objetivo analizar de manera integral los desafíos que plantea la regulación de la inteligencia artificial en el deporte. En particular, se busca: (i) examinar los principales usos de la IA en el sector deportivo; (ii) identificar los riesgos éticos y jurídicos asociados a su implementación; (iii) proponer un marco de gobernanza que garantice un uso responsable de la tecnología; y (iv) explorar escenarios futuros sobre el impacto de la IA en la práctica deportiva y en la sociedad.

De esta manera, el estudio no solo pretende contribuir al debate académico sobre la regulación de la IA, sino también ofrecer herramientas conceptuales y normativas que permitan orientar la toma de decisiones en el ámbito deportivo. En última instancia, se busca responder a una pregunta central: ¿es posible integrar la inteligencia artificial en el deporte sin perder aquello que lo hace profundamente humano?

2. Inteligencia artificial en el deporte: marco teórico, estado del arte y bases para su regulación

2.1 Conceptualización de la inteligencia artificial y su inserción en el deporte

La inteligencia artificial puede entenderse como un conjunto de sistemas capaces de analizar su entorno, aprender de los datos y tomar decisiones con cierto grado de autonomía. En términos generales, su funcionamiento se basa en la combinación de grandes volúmenes de datos, algoritmos y capacidad computacional, lo que le permite replicar —y en algunos casos superar— determinadas capacidades humanas.

En el ámbito deportivo, la IA ha evolucionado desde aplicaciones básicas de análisis estadístico hacia sistemas complejos de aprendizaje automático que permiten optimizar el rendimiento, predecir resultados y automatizar procesos de toma de decisiones. Como se ha señalado, Kim y otros. (2025), “cuando la IA se introdujo por primera vez en el deporte, se utilizó para analizar datos estadísticos de los partidos y predecir sus resultados”. Sin embargo, con el avance del machine learning, su alcance se ha ampliado significativamente.

Actualmente, la IA interviene en múltiples dimensiones del deporte, incluyendo el entrenamiento, la medicina deportiva, el arbitraje, la gestión organizacional y la experiencia de los aficionados. Este proceso de expansión responde a la creciente digitalización del deporte y a la disponibilidad de datos cada vez más detallados sobre el desempeño de los atletas.

No obstante, esta evolución también ha generado una dependencia estructural de los datos, lo que convierte a la información en un recurso estratégico dentro del ecosistema deportivo. En este sentido, la IA no solo transforma la práctica deportiva, sino también las relaciones de poder entre los actores involucrados, particularmente entre atletas, organizaciones y empresas tecnológicas.

2.2 Estado del arte: aplicaciones actuales de la IA en el deporte

El desarrollo de la inteligencia artificial en el deporte ha transitado de herramientas básicas de análisis estadístico hacia sistemas complejos capaces de incidir directamente en la toma de decisiones, el rendimiento competitivo y la experiencia del espectador. Este proceso no solo refleja un avance tecnológico, sino también una transformación estructural del ecosistema deportivo, en el que los datos se han convertido en un recurso central para la generación de valor y la toma de decisiones estratégicas.

En sus primeras etapas, la IA fue utilizada principalmente para analizar datos históricos y predecir resultados. Como se ha señalado, “cuando la IA se introdujo

por primera vez en el deporte, se utilizó para analizar datos estadísticos de los partidos y predecir sus resultados” Kim y otros. (2025). Sin embargo, con el desarrollo del aprendizaje automático, estas herramientas han evolucionado hacia sistemas capaces de aprender de manera autónoma, identificar patrones complejos y ofrecer recomendaciones en tiempo real.

Actualmente, uno de los ámbitos donde la IA tiene mayor impacto es el rendimiento deportivo. A través del uso de sensores, dispositivos portátiles y sistemas de monitoreo continuo, es posible recopilar información detallada sobre variables fisiológicas, técnicas y tácticas de los atletas. Estos datos permiten diseñar entrenamientos personalizados, optimizar el rendimiento y reducir el riesgo de lesiones. En este sentido, la literatura destaca que “la IA se utiliza actualmente para analizar los datos físicos de los atletas para diseñar programas de entrenamiento personalizados, predecir el riesgo de lesiones, rehabilitar lesiones y detectar el dopaje” Kim y otros. (2025) . Este tipo de aplicaciones ha transformado la preparación deportiva, desplazando progresivamente la intuición hacia modelos basados en datos.

De manera complementaria, la inteligencia artificial ha redefinido el análisis táctico en la competencia. Los modelos predictivos permiten anticipar patrones de juego, evaluar estrategias y simular escenarios, lo que otorga a entrenadores y atletas una ventaja competitiva significativa. No obstante, esta ventaja no se distribuye de manera equitativa, ya que el acceso a estas tecnologías depende en gran medida de los recursos económicos de los equipos y organizaciones, lo que introduce nuevas formas de desigualdad en la competencia.

El arbitraje constituye otro de los ámbitos en los que la IA ha tenido una influencia notable. Sistemas como el seguimiento de trayectoria del balón o el análisis de video han sido incorporados para asistir en la toma de decisiones, reduciendo el margen de error humano. Tecnologías como el VAR o sistemas similares en otros deportes han contribuido a aumentar la precisión de las decisiones arbitrales; sin embargo, también han generado cuestionamientos sobre la transparencia de los

procesos y la legitimidad de las decisiones automatizadas. La dificultad para comprender el funcionamiento interno de estos sistemas —muchas veces opacos— plantea problemas de confianza y aceptación por parte de jugadores y aficionados.

En el ámbito de la medicina deportiva y la detección del dopaje, la IA ha abierto nuevas posibilidades para el análisis de datos biológicos y la identificación de patrones anómalos. Estos avances permiten mejorar la prevención de lesiones y fortalecer los mecanismos de control antidopaje. No obstante, también introducen riesgos significativos. El uso de datos biométricos y de salud implica el manejo de información altamente sensible, cuyo uso indebido puede tener consecuencias graves. Como se advierte, “los datos relacionados con la salud de los atletas [...] pueden poner en peligro sus carreras y los derechos humanos fundamentales” (Kim y otros., 2025) . Además, la posibilidad de errores en los sistemas —como falsos positivos— plantea interrogantes sobre la fiabilidad de estas tecnologías y sus implicaciones éticas.

Más allá de la práctica deportiva, la IA también ha transformado la gestión organizacional y la industria del deporte. Los modelos predictivos se utilizan para la selección de jugadores, la planificación de fichajes y la negociación de contratos, permitiendo evaluar el rendimiento y el valor potencial de los atletas con mayor precisión. Estas herramientas han modificado la forma en que las organizaciones toman decisiones estratégicas, aumentando la eficiencia, pero también reforzando asimetrías de información entre los distintos actores, particularmente entre clubes y deportistas.

Asimismo, la inteligencia artificial ha redefinido la relación entre el deporte y sus aficionados. A través del análisis de datos y el uso de sistemas automatizados, las organizaciones deportivas pueden ofrecer experiencias personalizadas, mejorar la interacción con los usuarios y optimizar sus estrategias de marketing. Los chatbots, los sistemas de recomendación y el análisis del comportamiento del consumidor permiten crear entornos más dinámicos e inmersivos. Sin embargo, estas prácticas

también plantean preocupaciones en términos de privacidad y vigilancia, ya que implican la recopilación y explotación de datos personales con fines comerciales.

En conjunto, el estado del arte evidencia que la inteligencia artificial se ha consolidado como una herramienta transversal en el deporte, capaz de influir en prácticamente todos sus ámbitos. No obstante, este desarrollo tecnológico no ha estado acompañado de una regulación proporcional a su impacto. La rápida adopción de la IA ha generado un desfase entre la innovación y la capacidad de los marcos normativos para gestionar sus riesgos.

Además, las implicaciones de la IA en el deporte no se limitan a cuestiones técnicas, sino que abarcan dimensiones éticas, sociales y culturales. La creciente dependencia de sistemas automatizados plantea desafíos relacionados con la equidad, la transparencia, la privacidad y la autonomía de los actores involucrados. En este sentido, el análisis del estado del arte no solo permite identificar los beneficios de la IA, sino también evidenciar la urgencia de desarrollar mecanismos de gobernanza que orienten su uso de manera responsable.

Finalmente, resulta pertinente destacar que el avance de la IA no solo transforma la forma en que se practica y se gestiona el deporte, sino que también plantea interrogantes sobre su futuro. Como se ha señalado, “depender exclusivamente de las técnicas de IA puede disminuir el entusiasmo y la pasión de los aficionados” (Kim y otros., 2025). Esta advertencia pone de relieve la necesidad de encontrar un equilibrio entre innovación tecnológica y preservación de la dimensión humana del deporte, cuestión que será fundamental en el desarrollo de propuestas regulatorias.

2.3 Dilemas éticos y riesgos emergentes en el uso de la IA

El estado del arte muestra que el uso de la IA en el deporte no solo implica beneficios, sino también una serie de riesgos que deben ser abordados desde una perspectiva ética y regulatoria. En particular, la literatura identifica cuatro grandes ejes de preocupación: el sesgo algorítmico, la falta de transparencia, la privacidad de los datos y la responsabilidad en la toma de decisiones.

En primer lugar, el sesgo algorítmico constituye uno de los principales desafíos. Dado que los sistemas de IA se entrenan con datos históricos, existe el riesgo de que reproduzcan desigualdades existentes. En el contexto deportivo, esto puede afectar procesos como la selección de talento, la evaluación del rendimiento y la toma de decisiones estratégicas. Como se ha señalado, “los algoritmos pueden perpetuar desigualdades existentes si los datos de entrenamiento contienen sesgos” (Kim y otros., 2025).

En segundo lugar, la falta de transparencia en los sistemas de IA limita la comprensión de sus decisiones y dificulta la rendición de cuentas. Muchos sistemas funcionan como “cajas negras”, lo que impide a los usuarios conocer los criterios utilizados para generar resultados. Esta opacidad resulta especialmente problemática en contextos donde las decisiones tienen un alto impacto, como el arbitraje o la detección de dopaje.

En tercer lugar, la privacidad de los datos representa un riesgo crítico. La recopilación de información antropométrica, biométrica y fisiológica de los atletas implica el manejo de datos altamente sensibles, cuyo uso indebido puede tener consecuencias graves. En este sentido, se advierte que “los datos relacionados con la salud de los atletas [...] pueden poner en peligro sus carreras y los derechos humanos fundamentales” (Kim y otros., 2025) .

Finalmente, el problema de la responsabilidad plantea interrogantes sobre quién debe asumir las consecuencias de las decisiones tomadas por sistemas automatizados. En un entorno donde las decisiones pueden ser parcialmente autónomas, resulta complejo determinar si la responsabilidad recae en los desarrolladores, las organizaciones deportivas o los propios usuarios.

Estos dilemas reflejan la necesidad de desarrollar marcos normativos que permitan gestionar los riesgos asociados al uso de la IA en el deporte, sin limitar su potencial innovador.

2.4 La gobernanza de la IA en el deporte: hacia un enfoque regulatorio

La gobernanza de la inteligencia artificial en el deporte implica establecer principios, normas y mecanismos que regulen su uso de manera ética y responsable. A diferencia de otros sectores, el deporte presenta características particulares que complican la implementación de marcos regulatorios homogéneos.

En primer lugar, existe una diversidad de actores involucrados, incluyendo organismos internacionales, federaciones, clubes, empresas tecnológicas y atletas. Esta multiplicidad de actores genera un entorno fragmentado en el que las decisiones no dependen de una única autoridad.

En segundo lugar, la coexistencia de regulaciones públicas y privadas dificulta la armonización normativa. Como se ha señalado, el deporte se caracteriza por “la variedad de regulaciones (públicas, privadas)” , lo que complica la implementación de estándares comunes.

En tercer lugar, la lógica competitiva del deporte puede incentivar el uso intensivo de tecnologías para obtener ventajas, lo que aumenta el riesgo de desigualdad y de prácticas poco éticas.

En este contexto, la regulación de la IA en el deporte debe orientarse a garantizar principios fundamentales como la equidad, la transparencia, la protección de datos y la responsabilidad. Asimismo, debe considerar la necesidad de preservar la esencia del deporte como una actividad humana basada en la competencia justa y la emoción.

2.5 La inteligencia artificial y el futuro del deporte: automatización, límites y transformaciones sociales

El avance de la inteligencia artificial en el deporte no solo está transformando sus dinámicas actuales, sino que plantea una reconfiguración profunda de su futuro. Más allá de su papel como herramienta de apoyo, la IA introduce la posibilidad de

una progresiva automatización de la práctica deportiva, lo que obliga a reflexionar sobre sus implicaciones en términos éticos, sociales y culturales. En este sentido, la discusión ya no se limita a cómo mejorar el rendimiento o la eficiencia, sino que se desplaza hacia una cuestión más compleja: hasta qué punto la tecnología puede intervenir en el deporte sin alterar su esencia.

En su estado actual, la inteligencia artificial funciona principalmente como un instrumento complementario que amplifica las capacidades humanas. Su aplicación en el análisis de datos, la prevención de lesiones y la toma de decisiones ha permitido optimizar el desempeño de los atletas y mejorar la precisión en distintos ámbitos del deporte. No obstante, incluso en este nivel de integración, comienzan a emerger tensiones relacionadas con la equidad, el acceso a la tecnología y la dependencia de sistemas automatizados.

A medida que la IA adquiere mayor sofisticación, es posible imaginar un escenario en el que su papel deje de ser meramente auxiliar y se convierta en un elemento dominante en la toma de decisiones. En este contexto, los algoritmos no solo procesan información, sino que orientan de manera decisiva las estrategias, evaluaciones y resoluciones dentro del deporte. Esta transición implica un desplazamiento progresivo del juicio humano, lo que puede afectar directamente la autonomía de los actores involucrados. Como advierte la literatura, la IA puede generar “pérdida de autonomía” en los deportistas (Pérez Triviño, 2023), situación que plantea cuestionamientos sobre el papel del criterio humano en un entorno cada vez más mediado por sistemas algorítmicos.

En un nivel más avanzado, la relación entre humanos y tecnología podría evolucionar hacia formas híbridas de competencia. En este tipo de escenario, la inteligencia artificial no solo influye en la toma de decisiones, sino que participa directamente en la práctica deportiva mediante sistemas tecnológicos avanzados o entornos virtuales. Esta posibilidad ya comienza a vislumbrarse en el desarrollo de deportes digitales y simulaciones, donde la frontera entre lo humano y lo tecnológico se vuelve cada vez más difusa. Sin embargo, esta integración también

introduce tensiones en torno a la identidad del deporte, ya que desplaza el énfasis del esfuerzo humano hacia la capacidad tecnológica.

El escenario más radical plantea la automatización total del deporte, en el que la participación humana podría ser sustituida por sistemas autónomos, como robots o entidades controladas por inteligencia artificial. Aunque este panorama puede parecer lejano, el desarrollo tecnológico sugiere que no debe ser descartado. De hecho, se ha señalado que en el futuro podrían existir sistemas capaces de desplegar “capacidades físicas y cognitivas similares o superiores a las de los mejores deportistas actuales” (Pérez Triviño, 2023) . Este escenario implica una transformación profunda del deporte, que dejaría de ser una actividad basada en la competencia humana para convertirse en un espectáculo tecnológico.

Las implicaciones de estos posibles escenarios trascienden el ámbito deportivo y se proyectan sobre la sociedad en su conjunto. Desde una perspectiva de salud pública, el deporte cumple una función fundamental como promotor de la actividad física y el bienestar. Si la práctica deportiva pierde su dimensión humana, podría disminuir su capacidad de incentivar la participación de la población, lo que tendría efectos negativos en términos de sedentarismo y calidad de vida.

En el plano cultural, el deporte constituye un espacio de construcción de identidad, pertenencia y emoción colectiva. La relación entre los aficionados y los atletas se basa en la admiración y la identificación, elementos que podrían verse debilitados en un contexto de automatización. En este sentido, se ha advertido que la tecnología, lejos de fortalecer el deporte, podría erosionar algunos de sus componentes más esenciales.

Desde una perspectiva económica, la automatización también podría transformar la estructura del sector deportivo. La posible reducción del papel de los atletas humanos implicaría cambios en el mercado laboral, mientras que la industria podría orientarse hacia nuevas formas de entretenimiento basadas en la tecnología. Esto plantea interrogantes sobre la sostenibilidad del modelo actual del deporte y sobre el papel que desempeñarán los distintos actores en este nuevo contexto.

En conjunto, estos escenarios evidencian que el desarrollo de la inteligencia artificial en el deporte no es un proceso neutral, sino una transformación que afecta múltiples dimensiones de la vida social. La cuestión central no radica en si la IA debe formar parte del deporte, sino en definir los límites de su intervención. En este sentido, el desafío consiste en encontrar un equilibrio entre innovación tecnológica y preservación de los valores que dan sentido al deporte.

En última instancia, la regulación de la inteligencia artificial deberá orientarse a garantizar que la tecnología actúe como un medio para potenciar la experiencia deportiva, y no como un sustituto de la misma. Esto implica reconocer que el deporte no solo es rendimiento y eficiencia, sino también emoción, incertidumbre y humanidad, elementos que difícilmente pueden ser replicados por sistemas automatizados.

3. Hacia una regulación de la inteligencia artificial en el deporte: principios, riesgos y gobernanza

El desarrollo acelerado de la inteligencia artificial en el deporte ha generado un desfase significativo entre la innovación tecnológica y la capacidad de los marcos normativos para regular sus implicaciones. Si bien diversas regiones del mundo han comenzado a establecer lineamientos generales para el uso de la IA, estos no contemplan de manera específica las particularidades del ecosistema deportivo, caracterizado por la interacción entre múltiples actores, la centralidad del rendimiento y la relevancia de valores como la equidad, la transparencia y el juego limpio.

En este contexto, la necesidad de una regulación específica para la inteligencia artificial en el deporte se vuelve evidente. La gobernanza deportiva no puede limitarse únicamente a aspectos administrativos o económicos, sino que debe incorporar el impacto de las tecnologías emergentes en la práctica deportiva y en los derechos de los actores involucrados. En este sentido, se ha definido la gobernanza como “el conjunto de responsabilidades y medidas [...] que una entidad

deportiva asume [...] al objeto de imponer [...] comportamientos éticos que regulen la gestión” (Pérez Triviño, 2023). Esta definición permite entender que la regulación de la IA no debe concebirse únicamente desde el derecho positivo, sino también desde un enfoque ético y organizacional.

A partir de esta perspectiva, es posible establecer una serie de principios rectores que orienten el uso de la inteligencia artificial en el deporte. En primer lugar, el principio de autonomía humana resulta fundamental, en la medida en que la IA no debe sustituir completamente la toma de decisiones de los actores, sino complementarla. La literatura advierte que el uso de estas tecnologías puede implicar una “pérdida de autonomía” (Pérez Triviño, 2023) , lo que obliga a establecer límites claros sobre su grado de intervención, especialmente en decisiones que afectan la carrera y la salud de los deportistas.

En segundo lugar, el principio de equidad y no discriminación adquiere especial relevancia en un entorno donde los sistemas de IA se alimentan de datos históricos. Como se ha señalado, “si los datos de entrenamiento contienen sesgos, pueden generar resultados injustos” (Kim y otros., 2025) . Esto implica que la regulación debe contemplar mecanismos para prevenir la reproducción de desigualdades, tales como auditorías algorítmicas, revisión de datos y evaluación constante de los sistemas.

Asimismo, la transparencia y la explicabilidad constituyen elementos clave para garantizar la legitimidad de las decisiones basadas en IA. La opacidad de los sistemas algorítmicos puede dificultar la rendición de cuentas y generar desconfianza entre los actores. Por ello, resulta necesario que los sistemas utilizados en el deporte permitan comprender los criterios que sustentan sus decisiones, especialmente en ámbitos sensibles como el arbitraje o la selección de jugadores.

Otro aspecto central es la protección de datos y la privacidad. El deporte implica el uso de información altamente sensible, particularmente en lo que respecta a datos biométricos y de salud. En este sentido, se exige la implementación de medidas

estrictas de protección, así como el establecimiento de protocolos de consentimiento informado y control de acceso a la información.

De igual forma, el principio de responsabilidad y rendición de cuentas resulta indispensable en un entorno donde las decisiones pueden estar mediadas por sistemas automatizados. La complejidad de la IA dificulta la identificación de responsabilidades, lo que puede generar vacíos normativos. En consecuencia, la regulación debe establecer con claridad quién responde por las decisiones tomadas con apoyo de estas tecnologías, ya sean desarrolladores, organizaciones deportivas o usuarios finales.

Más allá de estos principios, resulta necesario estructurar la regulación a partir de un enfoque basado en niveles de riesgo, que permita diferenciar las aplicaciones de la IA según su impacto potencial. En este sentido, las aplicaciones de bajo riesgo —como los sistemas de atención al aficionado o recomendación de contenido— pueden operar con requisitos mínimos de supervisión. Por otro lado, las aplicaciones de riesgo medio —como el análisis de rendimiento— requieren mecanismos de transparencia y supervisión humana. En contraste, aquellas aplicaciones de alto riesgo —como el uso de datos biométricos, el arbitraje automatizado o la detección de dopaje— deben estar sujetas a controles estrictos, incluyendo evaluaciones de impacto, auditorías y supervisión constante. Finalmente, aquellas tecnologías que vulneren la autonomía de los deportistas o comprometan su integridad deben considerarse inaceptables y, por tanto, prohibidas.

La implementación de este modelo regulatorio no puede depender exclusivamente de normas jurídicas formales. Dado el carácter global del deporte, resulta necesario articular mecanismos de gobernanza que combinen instrumentos de hard law con estrategias de soft law. Mientras que los Estados pueden establecer marcos legales obligatorios en materia de protección de datos y derechos fundamentales, las organizaciones deportivas tienen un papel clave en el desarrollo de códigos de conducta, estándares éticos y buenas prácticas.

En este sentido, se vuelve fundamental la creación de estructuras institucionales que permitan supervisar el uso de la IA en el deporte. Entre ellas, destacan las auditorías algorítmicas independientes, los comités de ética tecnológica, las evaluaciones de impacto en derechos humanos y los registros de sistemas de IA utilizados por organizaciones deportivas. Estos mecanismos permiten traducir los principios normativos en prácticas concretas y garantizar su cumplimiento.

Asimismo, la regulación de la inteligencia artificial en el deporte requiere un enfoque de gobernanza multinivel. A nivel internacional, organismos como el Comité Olímpico Internacional pueden establecer principios globales que orienten el uso de la IA. A nivel regional, marcos regulatorios como el europeo pueden adaptarse al contexto deportivo. A nivel nacional, los Estados pueden desarrollar normativas específicas, mientras que, a nivel organizacional, clubes y federaciones deben implementar políticas internas que regulen el uso de estas tecnologías.

Finalmente, es necesario destacar que la regulación de la IA en el deporte no debe limitarse a mitigar riesgos, sino que debe orientarse a preservar la esencia misma de esta actividad. Esta advertencia pone de relieve que el desafío no consiste únicamente en regular la tecnología, sino en garantizar que su uso sea compatible con los valores que definen al deporte.

En última instancia, la regulación de la inteligencia artificial en el deporte debe construirse a partir de un equilibrio entre innovación y ética, eficiencia y equidad, tecnología y humanidad. Solo a través de este equilibrio será posible aprovechar el potencial de la IA sin comprometer aquello que hace del deporte una de las expresiones más significativas de la experiencia humana.

4. Escenarios futuros e impacto social de la inteligencia artificial en el deporte

El avance de la inteligencia artificial (IA) en el deporte no solo transforma sus dinámicas actuales, sino que proyecta una serie de escenarios futuros que cuestionan la naturaleza misma de esta actividad. Este proceso debe entenderse en el marco más amplio de la digitalización de la sociedad, donde los sistemas

algorítmicos tienden a desplazar progresivamente la toma de decisiones humanas (Coeckelbergh, 2020). En el ámbito deportivo, esta transformación resulta particularmente relevante, ya que el deporte no es únicamente un espacio de optimización del rendimiento, sino también un fenómeno cultural basado en la incertidumbre, la emoción y la identificación colectiva.

En el escenario más inmediato, la inteligencia artificial continúa operando como una herramienta de apoyo que amplifica las capacidades humanas. Su integración en el análisis de rendimiento, la prevención de lesiones y la toma de decisiones tácticas ha demostrado beneficios claros en términos de eficiencia y precisión. Sin embargo, incluso en este nivel emergen tensiones relacionadas con la equidad, ya que el acceso desigual a estas tecnologías puede generar ventajas competitivas significativas. En este sentido, la literatura advierte que “las preocupaciones éticas se agrupan en equidad y sesgo, transparencia y explicabilidad, privacidad de los datos y responsabilidad en la aplicación de la IA” (Kim y otros., 2025), lo que refleja la complejidad de su implementación en contextos deportivos.

A medida que la IA evoluciona, es posible proyectar un escenario en el que su papel deje de ser auxiliar y se convierta en dominante en la toma de decisiones. En este contexto, los algoritmos no solo procesan información, sino que orientan estrategias, evaluaciones y resoluciones con un alto grado de autonomía. Este desplazamiento del juicio humano plantea un problema central: la erosión de la autonomía de los actores deportivos. Como advierte la literatura, la IA puede implicar una “pérdida de autonomía” en los deportistas (Pérez Triviño, 2023) , especialmente cuando decisiones clave se fundamentan en sistemas cuya lógica no es completamente transparente.

Un escenario más avanzado sugiere la consolidación de formas híbridas de práctica deportiva, en las que humanos y sistemas tecnológicos interactúan de manera directa. Este modelo ya se vislumbra en disciplinas como los deportes electrónicos y los entornos virtuales, donde la inteligencia artificial no solo asiste, sino que participa activamente en la dinámica competitiva. Sin embargo, esta transformación

introduce tensiones en torno al concepto de mérito, ya que desplaza el énfasis del esfuerzo humano hacia la capacidad tecnológica, cuestionando los principios tradicionales del deporte.

El escenario más radical plantea la automatización total del deporte, en el que los atletas humanos podrían ser sustituidos por sistemas autónomos, como robots o entidades controladas por inteligencia artificial. Aunque este panorama puede parecer lejano, no es completamente descartable. De hecho, se ha señalado que podrían desarrollarse sistemas con “capacidades físicas y cognitivas similares o superiores a las de los mejores deportistas actuales” (Pérez Triviño, 2023) . Este escenario implicaría una transformación profunda del deporte, que dejaría de ser una actividad centrada en el desempeño humano para convertirse en un espectáculo tecnológico.

Las implicaciones de estos escenarios trascienden el ámbito deportivo y se proyectan sobre la sociedad en su conjunto. En primer lugar, el impacto en la salud pública resulta particularmente relevante. El deporte no solo cumple una función competitiva, sino también social, al fomentar la actividad física y el bienestar. Una posible pérdida de centralidad del atleta humano podría disminuir la motivación social hacia la práctica deportiva, lo que tendría efectos negativos en indicadores de salud (World Health Organization, 2022).

En segundo lugar, el deporte constituye un espacio de construcción cultural y simbólica. La relación entre los aficionados y los atletas se basa en la identificación, la emoción y la narrativa del esfuerzo humano; lo que pone en evidencia el riesgo de deshumanización del deporte en contextos altamente automatizados.

Desde una perspectiva económica, la integración de la IA también podría reconfigurar la estructura del sector deportivo. La posible reducción del papel de los atletas humanos implicaría cambios en el mercado laboral, mientras que la industria podría orientarse hacia nuevas formas de entretenimiento basadas en la tecnología. Este proceso se inserta en una tendencia más amplia en la que la automatización transforma los modelos productivos y laborales (Brynjolfsson & McAfee, 2014).

En conjunto, estos escenarios evidencian que el desarrollo de la inteligencia artificial en el deporte no es un proceso neutral, sino una transformación que afecta múltiples dimensiones de la vida social. La cuestión central no radica en si la IA debe formar parte del deporte, sino en definir los límites de su intervención. En este sentido, la regulación debe garantizar que la tecnología actúe como un medio para potenciar la experiencia deportiva, y no como un sustituto de la misma.

En última instancia, el desafío consiste en preservar aquello que hace del deporte una actividad profundamente humana: la emoción, la incertidumbre y la conexión entre las personas. Sin estos elementos, el deporte corre el riesgo de perder no solo su esencia, sino también su relevancia social.

El análisis desarrollado a lo largo de este trabajo permite identificar una tensión estructural entre el avance tecnológico de la inteligencia artificial y la capacidad del deporte para preservar sus valores fundamentales. Si bien la literatura coincide en destacar los beneficios de la IA en términos de eficiencia, precisión y optimización del rendimiento, también reconoce una serie de riesgos éticos y sociales que no han sido suficientemente abordados por los marcos regulatorios actuales.

Uno de los principales hallazgos es que la regulación de la IA en el deporte se ha centrado predominantemente en aspectos técnicos, como la gestión de riesgos o la protección de datos, dejando en un segundo plano las implicaciones culturales y sociales de su implementación. En este sentido, aunque se han identificado preocupaciones relevantes —como “equidad y sesgo, transparencia y explicabilidad, privacidad de los datos y responsabilidad” (Kim y otros., 2025) —, estas han sido tratadas de manera fragmentada, sin integrarse en un enfoque holístico que contemple la complejidad del ecosistema deportivo.

Asimismo, el análisis evidencia que la introducción de la IA en el deporte no es un proceso neutral, sino que redefine las relaciones de poder entre los actores involucrados. La creciente dependencia de sistemas algorítmicos puede reforzar asimetrías existentes, particularmente entre organizaciones deportivas y atletas, quienes en muchos casos no tienen acceso ni control sobre los sistemas utilizados

para evaluarlos. Esta situación resulta especialmente problemática si se considera que dichos sistemas pueden influir directamente en decisiones clave, como la selección, el entrenamiento o la continuidad profesional.

En este contexto, el principio de autonomía humana adquiere una relevancia central. Como se ha señalado, la IA puede generar una “pérdida de autonomía” en los deportistas (Pérez Triviño, 2023) , lo que plantea interrogantes sobre el papel del juicio humano en un entorno cada vez más automatizado. La delegación excesiva de decisiones en sistemas algorítmicos no solo afecta la capacidad de acción de los individuos, sino que también puede diluir la responsabilidad, generando vacíos normativos difíciles de abordar.

Otro aspecto crítico identificado es la tensión entre transparencia y privacidad. Por un lado, la necesidad de explicar las decisiones algorítmicas exige mayor apertura en el funcionamiento de los sistemas. Por otro, la protección de datos — especialmente aquellos de carácter biométrico o de salud— requiere limitar el acceso y la difusión de información sensible. Esta tensión evidencia la dificultad de equilibrar principios que, en la práctica, pueden entrar en conflicto.

Adicionalmente, el análisis de los escenarios futuros sugiere que la regulación actual no está preparada para enfrentar las transformaciones más profundas que la IA podría generar en el deporte. En particular, la posibilidad de una automatización progresiva plantea desafíos que trascienden lo jurídico y se sitúan en el plano cultural. La advertencia de que “depender exclusivamente de las técnicas de IA puede disminuir el entusiasmo y la pasión de los aficionados” (Kim y otros., 2025) pone de relieve que el impacto de la IA no se limita a la eficiencia, sino que afecta la dimensión emocional del deporte.

En este sentido, la principal limitación de los enfoques actuales radica en su incapacidad para incorporar una visión centrada en el ser humano. La mayoría de los marcos regulatorios buscan mitigar riesgos, pero no cuestionan los límites de la intervención tecnológica ni su impacto en la esencia del deporte. Esto sugiere la necesidad de un cambio de enfoque, que no solo regule el uso de la IA, sino que

también defina hasta qué punto su implementación es compatible con los valores que sustentan la práctica deportiva.

A partir de estos elementos, se puede afirmar que el desafío de la regulación de la inteligencia artificial en el deporte no consiste únicamente en establecer normas, sino en construir un modelo de gobernanza que articule innovación tecnológica y principios éticos. Este modelo debe ser capaz de anticipar los efectos a largo plazo de la IA y de integrar dimensiones que van más allá de lo técnico, incluyendo el impacto social, cultural y humano.

6. Conclusión

El análisis desarrollado a lo largo de este trabajo permite afirmar que la inteligencia artificial se ha consolidado como un elemento transformador del deporte contemporáneo, con un impacto que trasciende lo técnico para incidir en dimensiones éticas, sociales y culturales. Su capacidad para optimizar el rendimiento, mejorar la toma de decisiones y ampliar las posibilidades de análisis ha generado beneficios innegables. Sin embargo, estos avances no están exentos de riesgos, particularmente en lo que respecta a la equidad, la privacidad, la transparencia y la autonomía de los actores involucrados.

Uno de los principales aportes de este estudio radica en evidenciar que la discusión sobre la regulación de la IA en el deporte no puede limitarse a un enfoque técnico o instrumental. Si bien la literatura ha identificado preocupaciones relevantes — como “equidad y sesgo, transparencia y explicabilidad, privacidad de los datos y responsabilidad” (Kim y otros., 2025) —, estas no han sido integradas en un marco que contemple la complejidad del deporte como fenómeno humano. En este sentido, la regulación no debe centrarse únicamente en mitigar riesgos, sino en definir los límites de la intervención tecnológica.

Asimismo, se ha demostrado que la introducción de la inteligencia artificial en el deporte redefine las relaciones de poder dentro del ecosistema deportivo. La creciente dependencia de sistemas algorítmicos puede generar asimetrías entre

organizaciones y atletas, afectando directamente la capacidad de estos últimos para tomar decisiones informadas. Esta situación se agrava si se considera que la IA puede implicar una “pérdida de autonomía” (Pérez Triviño, 2023) , lo que plantea un desafío central para cualquier marco regulatorio.

Por otro lado, los escenarios futuros analizados permiten identificar que el riesgo más significativo no radica únicamente en el uso indebido de la tecnología, sino en su posible expansión sin límites claros. La progresiva automatización del deporte plantea interrogantes sobre su naturaleza misma, especialmente en relación con el papel del ser humano. En este contexto, la advertencia de que “depender exclusivamente de las técnicas de IA puede disminuir el entusiasmo y la pasión de los aficionados” (Kim y otros., 2025) adquiere una relevancia particular, al evidenciar que la IA puede afectar no solo la práctica deportiva, sino también su significado social.

A partir de estos elementos, este trabajo sostiene que la regulación de la inteligencia artificial en el deporte debe construirse sobre un enfoque centrado en el ser humano. Esto implica reconocer que el deporte no es únicamente un espacio de eficiencia y rendimiento, sino una práctica social que depende de la emoción, la incertidumbre y la interacción humana. En consecuencia, la IA debe ser concebida como una herramienta que potencia estas dimensiones, y no como un sustituto de las mismas.

En última instancia, el desafío no consiste en frenar el avance de la inteligencia artificial, sino en orientarlo. La pregunta central no es si la IA debe formar parte del deporte, sino hasta qué punto su intervención es compatible con los valores que lo definen. Responder a esta pregunta requiere no solo marcos normativos sólidos, sino también una reflexión crítica sobre el futuro del deporte en una sociedad cada vez más automatizada.

En este sentido, puede afirmarse que el verdadero riesgo no es la inteligencia artificial en sí misma, sino el momento en que se permite que sustituya aquello que hace del deporte una expresión profundamente humana

Referencias

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Carrio, A. (2020). Ethical principles for artificial intelligence in sport. En *Artificial intelligence and ethics in sport* (pp. 25–35). Springer.

Coeckelbergh, M. (2020). *AI ethics*. MIT Press.

Kim, J.-H., Kim, J., Kang, H., & Youn, B.-Y. (2025). Ethical implications of artificial intelligence in sport: A systematic scoping review. *Journal of Sport and Health Science*, 14, 101047. <https://doi.org/10.1016/j.jshs.2025.101047>

Organización Mundial de la Salud. (2022). *Global status report on physical activity 2022*. World Health Organization. <https://www.who.int/publications/i/item/9789240059153>

Pérez Triviño, J. L. (2023). Los problemas de la gobernanza de la inteligencia artificial en el deporte. *Cuadernos Económicos de ICE*, 106, 129–152. <https://doi.org/10.32796/cice.2023.106.7700>

CAPÍTULO XXXII

POLÍTICAS CULTURALES EN LA ERA DE LA INTELIGENCIA ARTIFICIAL: REGULACIÓN, DIVERSIDAD Y TRABAJO CREATIVO.

Cultural Policies in the Age of Artificial Intelligence: Regulation, Diversity, and Creative Work.

Guadalupe Zaragoza Rodríguez

gZR@azc.uam.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0002-5863-1687>

José Antonio Martínez Barajas

josebarajas@azc.uam.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0002-4816-4691>

Resumen

Este capítulo analiza los desafíos que la inteligencia artificial (IA) plantea para las políticas culturales, con especial atención a sus implicaciones regulatorias en materia de diversidad cultural y trabajo creativo. A partir de un enfoque crítico, se argumenta que la IA no constituye únicamente una innovación tecnológica, sino una transformación estructural en los procesos de producción, circulación y acceso a los bienes culturales. En este contexto, los sistemas algorítmicos y las plataformas digitales emergen como nuevos intermediarios culturales, capaces de influir en la visibilidad de contenidos y en la configuración de los mercados creativos.

El texto examina, por un lado, los efectos de la IA en la diversidad cultural, destacando los riesgos de homogeneización y sesgo algorítmico; y, por otro, sus impactos en el trabajo creativo, particularmente en relación con el uso de obras

para el entrenamiento de modelos y las tensiones con los marcos tradicionales de los derechos de autor. Asimismo, se identifican las limitaciones de los enfoques regulatorios actuales, centrados principalmente en aspectos técnicos o éticos.

Finalmente, se proponen lineamientos para incorporar una perspectiva de política cultural en la regulación de la IA, incluyendo instrumentos orientados a la transparencia algorítmica, la protección de los creadores y la gobernanza de los datos culturales, con el fin de garantizar ecosistemas culturales más diversos y equitativos.

Palabras clave: inteligencia artificial, políticas culturales, diversidad cultural, trabajo creativo, regulación algorítmica, derechos de autor.

Abstract

This chapter analyzes the challenges that artificial intelligence (AI) poses for cultural policies, with special attention to its regulatory implications regarding cultural diversity and creative labor. From a critical perspective, it argues that AI does not merely constitute a technological innovation, but rather a structural transformation in the processes of production, circulation, and access to cultural goods. In this context, algorithmic systems and digital platforms emerge as new cultural intermediaries capable of influencing content visibility and shaping creative markets.

The text examines, on the one hand, the effects of AI on cultural diversity, highlighting the risks of homogenization and algorithmic bias; and, on the other hand, its impacts on creative labor, particularly concerning the use of works for model training and the tensions with traditional copyright frameworks. Likewise, it identifies the limitations of current regulatory approaches, which are mainly focused on technical or ethical aspects.

Finally, the chapter proposes guidelines for incorporating a cultural policy perspective into AI regulation, including instruments aimed at algorithmic

transparency, the protection of creators, and the governance of cultural data, in order to guarantee more diverse and equitable cultural ecosystems.

Keywords:

artificial intelligence, cultural policies, cultural diversity, creative labor, algorithmic regulation, copyright.

Introducción

En los últimos años, la inteligencia artificial (IA) se ha consolidado como una de las tecnologías más influyentes en la reconfiguración de los ecosistemas culturales contemporáneos. Más allá de su dimensión técnica, la IA interviene de manera directa en los procesos de producción, circulación y consumo de bienes culturales, transformando tanto las dinámicas de creación como las condiciones de visibilidad y acceso a contenidos. En este sentido, la IA no puede ser entendida únicamente como una herramienta, sino como un conjunto de infraestructuras sociotécnicas que reconfiguran el campo cultural en su conjunto (Mejías & Couldry, 2019).

Uno de los aspectos más relevantes de esta transformación es el papel que desempeñan los sistemas algorítmicos y las plataformas digitales como nuevos intermediarios culturales. A través de mecanismos de recomendación, curaduría automatizada y generación de contenidos, estas tecnologías influyen en la manera en que los públicos acceden a la cultura, condicionando la visibilidad de determinadas expresiones culturales frente a otras. Este fenómeno plantea interrogantes fundamentales en torno a la diversidad cultural, entendida como un principio central de las políticas culturales contemporáneas (UNESCO, 2005). Diversos estudios han advertido sobre los riesgos de homogeneización cultural y sesgo algorítmico, derivados de modelos entrenados sobre grandes volúmenes de datos que tienden a privilegiar contenidos dominantes o hegemónicos (Noble, 2018).

De manera paralela, el desarrollo de la IA generativa ha intensificado las tensiones en torno al trabajo creativo. Herramientas capaces de producir textos, imágenes, música y otros contenidos culturales a partir de datos preexistentes han abierto nuevas posibilidades expresivas, pero también han generado controversias en relación con el uso de obras protegidas para el entrenamiento de modelos, la atribución de autoría y la sostenibilidad económica de los creadores (Ginsburg & Budiardjo, 2020).

Estas dinámicas evidencian las limitaciones de los marcos tradicionales de derechos de autor para responder a un entorno en el que la creatividad se encuentra crecientemente mediada por procesos automatizados.

A pesar de la magnitud de estos cambios, las respuestas regulatorias han tendido a centrarse en aspectos técnicos o éticos, dejando en un segundo plano las implicaciones específicamente culturales de la IA (European Parliament, 2025).

Esta omisión resulta problemática si se considera que la cultura constituye un ámbito clave para la construcción de identidades, la diversidad simbólica y la vida democrática. En este contexto, se plantea la necesidad de incorporar una dimensión cultural en la regulación de la inteligencia artificial, por lo tanto, se sostiene que las políticas culturales ofrecen un marco pertinente para abordar los desafíos que la IA plantea en términos de diversidad cultural, condiciones del trabajo creativo y acceso a los bienes culturales.

La IA y transformaciones del campo cultural

La expansión de la inteligencia artificial en los ecosistemas digitales contemporáneos ha reconfigurado de manera profunda el campo cultural, entendido como el conjunto de relaciones, instituciones y prácticas que estructuran la producción, circulación y consumo de bienes simbólicos (Chihu, 1998). Estas transformaciones no se limitan a la incorporación de nuevas herramientas tecnológicas, sino que implican cambios en las condiciones de visibilidad, en las

jerarquías culturales y en la organización de los mercados creativos. En particular, tres dimensiones resultan clave para comprender este proceso: la diversidad cultural, el papel de las plataformas como intermediarios y la centralidad de los datos como recurso estratégico.

La diversidad cultural ha sido un principio normativo central en el desarrollo de las políticas culturales contemporáneas, especialmente a partir de su reconocimiento en instrumentos internacionales como la Convención de la UNESCO (UNESCO, 2005). Sin embargo, la creciente mediación algorítmica en la circulación de contenidos culturales plantea nuevos desafíos para su preservación y promoción. Los sistemas de recomendación, ampliamente utilizados por plataformas digitales, operan mediante el análisis de grandes volúmenes de datos para predecir y orientar las preferencias de los usuarios. Si bien estos sistemas pueden facilitar el acceso a una amplia variedad de contenidos, también tienden a reforzar patrones de consumo preexistentes, favoreciendo la visibilidad de contenidos populares o dominantes en detrimento de expresiones culturales minoritarias o locales (Mejías & Couldry, 2019). Este fenómeno ha sido descrito por Eli Pariser (2011) como una forma de encierro algorítmico o *filter bubble*, en la medida en que limita no solo el acceso la información sino la exposición a la diversidad cultural.

Asimismo, diversos estudios han señalado que los algoritmos no son neutrales, sino que incorporan sesgos derivados tanto de los utilizados para su entrenamiento como de las decisiones de diseño que los sustentan (Noble, 2018). En el ámbito cultural, esto puede traducirse en la reproducción de jerarquías lingüísticas, geográficas y/o estéticas, consolidando la hegemonía de ciertos repertorios culturales globales. En consecuencia, la diversidad cultural enfrenta no solo desafíos de mercado, sino también de infraestructura tecnológica, lo que plantea la necesidad de repensar los instrumentos de política cultural en entornos digitalizados.

Plataformas digitales como intermediarios culturales

El auge de las plataformas digitales ha redefinido las formas de intermediación en el campo cultural. Empresas como los servicios de streaming, redes sociales o plataformas de contenido generado por usuarios no solo distribuyen bienes culturales, sino que desempeñan funciones de curaduría, clasificación y jerarquización de contenidos mediante sistemas automatizados. En este sentido, actúan como “intermediarios culturales algorítmicos” con capacidad para influir significativamente en la visibilidad y el éxito de obras y creadores.

A diferencia de los intermediarios tradicionales, como editoriales, programadores culturales o medios de comunicación, las plataformas operan a escala global y bajo lógicas de extracción y monetización de datos, lo que introduce nuevas dinámicas de concentración económica y poder cultural. Como señalan Couldry y Mejias (2019), este modelo se inscribe en un proceso más amplio de “colonialismo de datos”, en el que la vida social es sistemáticamente capturada y convertida en insumo para la generación de valor.

En este contexto, la mediación algorítmica no solo organiza la oferta cultural, sino que también moldea las condiciones de competencia en los mercados creativos. La visibilidad, y, por ende, la sostenibilidad económica, de muchas y muchos creadores depende de su posicionamiento dentro de estos sistemas, lo que genera relaciones asimétricas entre plataformas y productores culturales. Esta reconfiguración plantea interrogantes fundamentales para las políticas culturales, particularmente en lo que respecta a la regulación de los intermediarios y la garantía de condiciones equitativas de acceso y participación.

De igual manera, la creciente centralidad de los datos constituye otro eje fundamental en la transformación del campo cultural en la era de la IA. Los sistemas de inteligencia artificial, especialmente aquellos basados en aprendizaje automático, dependen de grandes volúmenes de datos para su entrenamiento y funcionamiento. En el ámbito cultural, estos datos incluyen textos, imágenes,

música, audiovisuales y otras formas de expresión que, en muchos casos, forman parte del acervo cultural colectivo.

Sin embargo, la captura, almacenamiento y procesamiento de estos datos se encuentra mayoritariamente en manos de actores privados, lo que plantea problemas en términos de acceso, control y distribución de beneficios. Como advierten Couldry y Mejias (2019), este proceso implica una apropiación sistemática de recursos culturales que no siempre se traduce en reconocimiento o compensación para lxs creadores o comunidades de origen.

Además, la utilización de datos culturales para el entrenamiento de modelos de IA plantea desafíos significativos para los marcos regulatorios existentes, particularmente en materia de derechos de autor. La distinción entre uso legítimo, reproducción, transformación e inspiración se vuelve cada vez más difusa en entornos donde los contenidos son utilizados para generar nuevos productos culturales mediante procesos automatizados (Ginsburg & Budiardjo, 2020).

En este sentido, los datos culturales pueden ser entendidos no solo como insumos económicos, sino también como bienes de interés público, cuya gestión tiene implicaciones directas para la diversidad cultural, la equidad en el acceso y la sostenibilidad de los ecosistemas creativos. Esta perspectiva resulta clave para repensar el papel del Estado y de las políticas culturales en la regulación de la IA, particularmente en lo que respecta a la gobernanza de los datos y la protección del patrimonio cultural en entornos digitales.

IA y trabajo creativo: transformaciones y desafíos regulatorios

La expansión de la inteligencia artificial, en particular de los sistemas generativos, ha introducido cambios significativos en las condiciones de producción del trabajo creativo. A diferencia de otras innovaciones tecnológicas, la IA no solo actúa como herramienta de apoyo, sino que puede participar activamente en la generación de textos, imágenes, videos, música y otros contenidos culturales, reconfigurando las

nociones de autoría, originalidad y valor en el campo cultural. En este contexto, emergen tensiones relevantes en torno a los derechos de los creadores, el uso de sus obras como insumo para el entrenamiento de modelos y la sostenibilidad económica del trabajo creativo.

Los sistemas de IA generativa se basan en el entrenamiento con grandes volúmenes de datos culturales preexistentes, a partir de los cuales identifican patrones para producir nuevos contenidos. Este proceso desafía las concepciones tradicionales de autoría, centradas en la figura individual del creador y en la originalidad como criterio fundamental de protección jurídica (Bracha, 2016). En la medida en que los *outputs* generados por IA derivan de la recombinación de obras previas, se difuminan los límites entre creación, reproducción y derivación.

Diversxs autores han señalado que estos cambios ponen en crisis los fundamentos del derecho de autor, que históricamente ha operado bajo supuestos de creatividad humana individual y control sobre la obra puesto que la dificultad para atribuir autoría en obras generadas por IA, ya sea a quien programa, al usuario/a o al propio sistema, evidencia la necesidad de repensar las categorías jurídicas existentes. Además, la posibilidad de que sistemas automatizados reproduzcan estilos o características de artistas específicos plantea interrogantes sobre la protección de la identidad creativa y los derechos de lxs autores.

Uno de los debates más relevantes en la intersección entre IA y trabajo creativo se refiere al uso de obras protegidas por derechos de autor en el entrenamiento de modelos generativos. En muchos casos, estos sistemas se entrenan con grandes conjuntos de datos que incluyen libros, imágenes, música y otros contenidos, sin que exista necesariamente consentimiento explícito de lxs titulares de derechos

Este fenómeno ha generado controversias tanto en el ámbito artístico, como el académico y el jurídico, particularmente en relación con la aplicabilidad de excepciones como el *fair use* o los límites al derecho de autor en distintos contextos normativos. Mientras algunxs argumentan que el entrenamiento dichos modelos

generativos constituye un uso transformativo, otros sostienen que implica una forma de explotación no autorizada que debería ser objeto de compensación

En este escenario, se evidencian las limitaciones de los marcos regulatorios actuales para abordar prácticas que no encajan fácilmente en las categorías tradicionales de reproducción o comunicación pública. La escala y opacidad de los procesos de entrenamiento dificultan, además, la identificación de las obras utilizadas y la trazabilidad de sus contribuciones, lo que complica la implementación de mecanismos de compensación o reconocimiento.

Más allá de las cuestiones jurídicas, la IA plantea desafíos estructurales para la sostenibilidad del trabajo creativo. La capacidad de generar contenidos a gran escala y a bajo costo puede intensificar dinámicas de precarización en sectores culturales ya caracterizados por la inestabilidad laboral y la desigual distribución de ingresos. En particular, la competencia entre contenidos generados por humanos y por sistemas automatizados puede afectar la demanda de trabajo creativo, especialmente en áreas como la ilustración, la redacción o la producción audiovisual. Al mismo tiempo, la integración de IA en procesos productivos puede desplazar ciertas tareas, redefiniendo las habilidades requeridas y las condiciones de empleo en el sector cultural (Frey & Osborne, 2017)

No obstante, algunos autores también señalan que la IA puede abrir nuevas oportunidades para la creación y la experimentación artística, ampliando las herramientas disponibles para los creadores; esta ambivalencia sugiere que el impacto de la IA en el trabajo creativo no es unívoco, sino que depende en gran medida de los marcos institucionales y regulatorios que configuren su desarrollo y uso.

Frente a estos desafíos, resulta necesario avanzar hacia marcos regulatorios que reconozcan las particularidades del trabajo creativo en entornos mediados por IA. En primer lugar, se plantea la necesidad de establecer mecanismos de consentimiento y transparencia en el uso de obras para el entrenamiento de

modelos generativos, así como de explorar esquemas de compensación que reconozcan la contribución de los creadores.

En segundo lugar, se requiere actualizar los marcos de derechos de autor para abordar cuestiones como la autoría de contenidos generados por IA, la protección de estilos e identidades creativas y los límites al uso automatizado de obras. Esto podría implicar la creación de nuevas categorías jurídicas o la adaptación de las existentes a las condiciones del entorno digital.

Finalmente, desde una perspectiva de política cultural, resulta fundamental considerar medidas orientadas a garantizar la sostenibilidad del trabajo creativo, incluyendo apoyos públicos, regulación de plataformas y mecanismos que promuevan condiciones equitativas de participación en los mercados culturales digitales. En este sentido, la regulación de la IA no debe limitarse a la gestión de riesgos tecnológicos, sino que debe incorporar una dimensión cultural y laboral que atienda a las transformaciones del trabajo creativo en el contexto contemporáneo.

Límites de la regulación actual de la inteligencia artificial

A pesar de la creciente centralidad de la inteligencia artificial en los ecosistemas culturales, los marcos regulatorios vigentes presentan limitaciones significativas para abordar sus implicaciones en términos de diversidad cultural, trabajo creativo y gobernanza de datos. En términos generales, la regulación de la IA se ha desarrollado bajo enfoques predominantemente tecnocráticos, centrados en la gestión de riesgos, la seguridad y la protección de derechos fundamentales, sin incorporar de manera sistemática una perspectiva. Esta orientación ha dejado vacíos importantes en relación con las transformaciones que la IA introduce en el campo cultural.

Uno de los rasgos más visibles de las iniciativas regulatorias recientes es la adopción de marcos basados en el riesgo, ejemplo de ello es la propuesta de Reglamento Europeo de Inteligencia Artificial (REDEGAL, 2025), que clasifica los

sistemas de IA en función de su nivel de riesgo y establece obligaciones diferenciadas para su desarrollo y uso. Si bien este enfoque representa un avance en términos de gobernanza tecnológica, su énfasis se centra principalmente en aspectos como la seguridad, la transparencia y la no discriminación, dejando en segundo plano las implicaciones culturales de estos sistemas, es así como la diversidad cultural rara vez aparece como un objetivo explícito de la regulación de la IA.

Las preocupaciones sobre sesgos algorítmicos se abordan, en su mayoría, desde una perspectiva de igualdad y no discriminación, sin considerar cómo estos sesgos pueden afectar la visibilidad y circulación de expresiones culturales diversas y como resultado, la regulación tiende a tratar la cultura de manera tangencial, sin reconocer su especificidad como ámbito de intervención pública.

Otra limitación relevante es la fragmentación de los marcos normativos aplicables a la IA en el ámbito cultural. Las cuestiones relacionadas con plataformas digitales, derechos de autor, protección de datos y competencia económica suelen abordarse en regímenes jurídicos separados, lo que dificulta una respuesta integral a los desafíos que plantea la IA (Senftleben, y otros, 2018), por ejemplo, los marcos de derechos de autor no fueron diseñados para contextos en los que las obras se utilizan masivamente como datos para el entrenamiento de modelos, mientras que las regulaciones de plataformas no siempre contemplan su papel como intermediarios culturales con capacidad de influir en la diversidad de contenidos. Esta desconexión entre regímenes normativos genera vacíos regulatorios y zonas grises que son aprovechadas por actores tecnológicos para operar con amplios márgenes de discrecionalidad.

Asimismo, la velocidad del desarrollo tecnológico supera con frecuencia la capacidad de adaptación de las instituciones regulatorias, lo que acentúa el desajuste entre normas existentes y prácticas emergentes. En el ámbito cultural, esto se traduce en una falta de herramientas específicas para abordar fenómenos como la generación automatizada de contenidos o la circulación algorítmica.

Como se señaló en el apartado anterior, el derecho de autor constituye uno de los principales instrumentos para la regulación del trabajo creativo. Sin embargo, su aplicación en el contexto de la IA presenta limitaciones importantes. En particular, la dificultad para encuadrar el entrenamiento de modelos dentro de las categorías tradicionales de uso como reproducción, transformación o uso legítimo evidencia la necesidad de una actualización normativa.

Además, los sistemas de IA generativa plantean desafíos en relación con la autoría y la titularidad de derechos sobre los contenidos producidos. La ausencia de criterios claros en muchas jurisdicciones genera incertidumbre jurídica tanto para los creadores como para los desarrolladores de tecnología. A esto se suma la dificultad de garantizar mecanismos efectivos de compensación en contextos donde la contribución individual de cada obra al entrenamiento de un modelo es difícil de rastrear. En este sentido, el derecho de autor, en su configuración actual, resulta insuficiente para abordar las dinámicas de producción cultural mediadas por IA, lo que refuerza la necesidad de enfoques regulatorios complementarios desde la política cultural.

Por otro lado, a pesar del papel central que tienen las plataformas digitales en la organización de la circulación cultural, las plataformas digitales han sido reguladas principalmente desde perspectivas vinculadas a la competencia económica, la protección de datos o la moderación de contenidos, sin considerar plenamente su función como intermediarios culturales.

Las obligaciones de transparencia algorítmica, cuando existen, suelen ser limitadas y no permiten comprender de manera efectiva cómo se priorizan ciertos contenidos sobre otros. Asimismo, las políticas de recomendación y curaduría permanecen en gran medida opacas, lo que dificulta evaluar su impacto en la diversidad cultural. Esta falta de transparencia se traduce en una concentración de poder en manos de plataformas que operan como “guardianes” del acceso a la cultura (Gillespie, 2018)

Desde una perspectiva de política cultural, esta situación plantea la necesidad de avanzar hacia marcos regulatorios que reconozcan explícitamente el rol cultural de las plataformas y establezcan obligaciones orientadas a garantizar la pluralidad y diversidad de contenidos.

Finalmente, la gobernanza de datos, que es un componente central en la regulación de la IA, ha sido abordada principalmente desde enfoques vinculados a la privacidad y la protección de datos personales. Si bien estos aspectos son fundamentales, resultan insuficientes para captar las implicaciones culturales del uso masivo de datos en sistemas de IA (Mejías & Couldry, 2019)

En consecuencia, la ausencia de una perspectiva cultural en la gobernanza de datos limita la capacidad de los Estados para intervenir en la configuración de infraestructuras digitales que tienen un impacto directo en la producción y circulación de la cultura.

Hacia una política cultural de la IA: lineamientos regulatorios

Las limitaciones de los marcos regulatorios actuales evidencian la necesidad de incorporar de manera explícita una perspectiva de política cultural en la gobernanza de la inteligencia artificial (IA). Lejos de constituir un ámbito exclusivamente técnico, la regulación de la IA implica decisiones normativas sobre la producción, circulación y acceso a la cultura. En este sentido, las políticas culturales pueden aportar un conjunto de principios e instrumentos orientados a garantizar la diversidad cultural, proteger el trabajo creativo y promover formas más equitativas de organización de los ecosistemas digitales (UNESCO, 2005).

A continuación, se proponen cuatro líneas de intervención que, sin pretender ser exhaustivas, delinean un marco posible para una política cultural de la IA.

1 Regulación de plataformas como intermediarios culturales

Dado el papel central de las plataformas digitales en la organización de la circulación cultural, resulta necesario avanzar hacia su reconocimiento explícito como intermediarios culturales sujetos a obligaciones específicas. En este sentido, una política cultural de la IA debería contemplar mecanismos de transparencia algorítmica, que permitan comprender los criterios mediante los cuales se priorizan ciertos contenidos (Gillespie, 2018).

Asimismo, podrían establecerse medidas orientadas a promover la diversidad cultural en los sistemas de recomendación, tales como cuotas de visibilidad para contenidos locales o independientes, o la incorporación de parámetros de diversidad en el diseño de algoritmos. Este tipo de instrumentos retoma experiencias previas en sectores como el audiovisual, donde las políticas culturales han buscado equilibrar la oferta cultural frente a dinámicas de concentración de mercado (Napoli, 2014)

Adicionalmente, se podrían impulsar esquemas de auditoría independiente de algoritmos, con el fin de evaluar su impacto en la pluralidad cultural y detectar posibles sesgos sistémicos. Estas medidas contribuirían a reducir la opacidad que caracteriza actualmente a los sistemas de recomendación y a fortalecer la rendición de cuentas de las plataformas.

2 Protección del trabajo creativo en entornos de IA

En relación con el trabajo creativo, resulta fundamental desarrollar instrumentos regulatorios que garanticen condiciones justas para los creadores en contextos mediados por IA. En primer lugar, se plantea la necesidad de establecer mecanismos de consentimiento informado para el uso de obras en el entrenamiento de modelos, así como de avanzar hacia esquemas de remuneración o compensación que reconozcan la contribución de lxs creadores (Ginsburg & Budiardjo, 2020)

En segundo lugar, es necesario actualizar los marcos de derechos de autor para abordar cuestiones emergentes como la autoría de contenidos generados por IA, la protección de estilos e identidades creativas y los límites al uso automatizado de obras. Esto podría implicar la introducción de nuevas categorías jurídicas o la adaptación de las existentes a las condiciones del entorno digital.

Asimismo, se propone la implementación de sistemas de etiquetado de contenido generado por IA, que permitan a los usuarios distinguir entre producciones humanas y automatizadas. Esta medida no solo contribuye a la transparencia, sino que también puede tener implicaciones en la valoración cultural y económica de las obras.

Finalmente, desde una perspectiva de política cultural, se podrían fortalecer los mecanismos de apoyo público al sector creativo, incluyendo programas de financiamiento, formación y adaptación tecnológica, con el fin de mitigar los efectos potencialmente precarizadores de la automatización

3 Gobernanza de datos culturales

La regulación de la IA requiere también abordar la cuestión de los datos culturales como recurso estratégico. En este sentido, se propone avanzar hacia modelos de gobernanza de datos que reconozcan su carácter de bien de interés público, especialmente cuando se trata de contenidos que formen parte del patrimonio cultural o que hayan sido producidos con financiamiento público (Mejías & Couldry, 2019).

Una posible línea de acción es la creación de repositorios públicos de datos culturales, gestionados por instituciones públicas o en esquemas de gobernanza participativa, que permitan el acceso a conjuntos de datos diversos y representativos. Estos repositorios podrían contribuir a reducir la dependencia de *datasets* privados y a promover el desarrollo de sistemas de IA más inclusivos.

Asimismo, se podrían desarrollar modelos de licenciamiento colectivo para el uso de datos culturales en el entrenamiento de IA, que faciliten la compensación a los creadores y simplifiquen los procesos de autorización. Este tipo de mecanismos ha sido explorado en otros ámbitos de la gestión colectiva de derechos y podría adaptarse al contexto de la IA.

Finalmente, resulta clave establecer principios de transparencia y trazabilidad en el uso de datos, que permitan identificar las fuentes utilizadas en el entrenamiento de modelos y evaluar sus implicaciones en términos de representación cultural.

4 El rol del Estado en la política cultural digital

La implementación de estos lineamientos requiere un rol activo del Estado como regulador, facilitador y garante de derechos culturales en entornos digitales. En particular, es necesario fortalecer las capacidades institucionales para diseñar, implementar y supervisar políticas públicas en contextos tecnológicos complejos.

En este contexto, el rol del Estado no se limita a la formulación de principios generales, sino que implica el diseño de marcos regulatorios específicos que orienten el desarrollo y uso de la inteligencia artificial en el ámbito cultural. Esto supone reconocer que la regulación de la IA no es únicamente una cuestión técnica, sino una intervención directa en las condiciones de producción, circulación y acceso a la cultura.

Entre los instrumentos posibles destacan la incorporación de criterios de diversidad cultural en la evaluación de sistemas de IA, la creación de instancias regulatorias con competencias en cultura digital, así como el desarrollo de mecanismos de evaluación de impacto cultural que permitan anticipar los efectos de estas tecnologías en los ecosistemas creativos. Asimismo, resulta clave promover la coordinación entre distintos ámbitos regulatorios, incluyendo cultura, tecnología, competencia y derechos de autor, con el fin de construir enfoques integrales frente a la complejidad de la IA (Napoli, 2014).

El Estado no solo debe establecer marcos normativos, sino también participar en la construcción de infraestructuras públicas digitales, tales como repositorios de datos culturales, plataformas de acceso y herramientas de IA orientadas al interés público. Esta dimensión resulta fundamental para evitar una dependencia exclusiva de actores privados en la organización de los ecosistemas culturales (Mejías & Couldry, 2019)

De este modo, el Estado se configura como un actor clave no solo en la gestión de riesgos tecnológicos, sino en la definición de las condiciones bajo las cuales se produce, circula y accede a la cultura en la era de la inteligencia artificial. Desde esta perspectiva, las políticas culturales adquieren un papel estratégico en la gobernanza de la IA, al articular objetivos de diversidad, equidad y acceso en un entorno crecientemente mediado por tecnologías digitales.

Conclusión

La expansión de la inteligencia artificial (IA) en los ecosistemas culturales contemporáneos no constituye únicamente un cambio tecnológico, sino una transformación estructural que reconfigura las condiciones de producción, circulación y acceso a la cultura. A lo largo de este capítulo se ha argumentado que la mediación algorítmica, el papel central de las plataformas digitales y la creciente dependencia de datos culturales como insumo estratégico introducen tensiones significativas en torno a la diversidad cultural, el trabajo creativo y la gobernanza de los bienes culturales. En este sentido, la IA no es neutral: incorpora y amplifica relaciones de poder preexistentes, al tiempo que genera nuevas formas de concentración y desigualdad en el campo cultural (Mejías & Couldry, 2019).

Uno de los principales hallazgos del análisis es que los marcos regulatorios actuales resultan insuficientes para abordar estas transformaciones. El predominio de enfoques tecnocráticos y basados en el riesgo ha permitido avances importantes en materia de seguridad y derechos fundamentales, pero ha dejado en un segundo plano la dimensión cultural de la IA. Esta omisión es particularmente problemática

si se considera que la cultura no solo constituye un sector económico, sino un ámbito fundamental para la diversidad simbólica, la construcción de identidades y el ejercicio de la ciudadanía (UNESCO, 2005).

En respuesta a este vacío, el capítulo ha propuesto la necesidad de avanzar hacia una política cultural de la IA, entendida como un conjunto de lineamientos regulatorios orientados a integrar la dimensión cultural en la gobernanza de estas tecnologías. En particular, se han delineado cuatro ejes de intervención: la regulación de plataformas como intermediarios culturales, la protección del trabajo creativo, la gobernanza de los datos culturales y el fortalecimiento del rol del Estado en la política cultural digital. Estos ejes no deben ser entendidos como ámbitos aislados, sino como componentes interrelacionados de una estrategia integral.

En primer lugar, reconocer a las plataformas como intermediarios culturales implica avanzar hacia mayores niveles de transparencia y rendición de cuentas en los sistemas de recomendación, así como incorporar criterios de diversidad cultural en su funcionamiento. En segundo lugar, la protección del trabajo creativo requiere actualizar los marcos de derechos de autor y desarrollar mecanismos que garanticen el consentimiento y la compensación de los creadores en el uso de sus obras para el entrenamiento de modelos de IA. En tercer lugar, la gobernanza de los datos culturales plantea la necesidad de concebir estos recursos no solo como insumos económicos, sino como bienes de interés público, cuya gestión tiene implicaciones directas en la equidad y la representación cultural.

Finalmente, el fortalecimiento del rol del Estado emerge como una condición necesaria para articular estos distintos ámbitos de intervención. Lejos de limitarse a la regulación de riesgos, el Estado está llamado a desempeñar un papel activo en la definición de las reglas que estructuran los ecosistemas culturales digitales, así como en la construcción de infraestructuras públicas que garanticen condiciones más equitativas de acceso y participación. En este contexto, las políticas culturales adquieren una relevancia estratégica como espacio de articulación entre objetivos tecnológicos, económicos y sociales.

En suma, regular la inteligencia artificial desde una perspectiva de política cultural implica reconocer que lo que está en juego no es únicamente el desarrollo de una tecnología, sino la configuración de los marcos que determinan qué culturas circulan, quiénes pueden producirlas y en qué condiciones. Frente a escenarios de creciente concentración y homogeneización, incorporar la dimensión cultural en la regulación de la IA constituye no solo una necesidad analítica, sino un imperativo político orientado a la construcción de ecosistemas culturales más diversos, justos y democráticos.

Bibliografía

- Chihu, A. (1998). La teoría de los campos en Pierre Bourdieu. *POLIS 98 Anuario de Sociología.*, 179-198.
- Bracha, O. (2016). *Owning Ideas : The Intellectual Origins of American Intellectual Property, 1790–1909*. Cambridge: Cambridge University Press.
- European Parliament. (19 de 02 de 2025). *European Parliament*. Recuperado el abril de 2026, de EU AI Act: first regulation on artificial intelligence: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 254-280.
- Gillespie, T. (2018). *Researchgate*. Obtenido de Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media: https://www.researchgate.net/publication/327186182_Custodians_of_the_internet_Platforms_content_moderation_and_the_hidden_decisions_that_shape_social_media#fullTextFileContent
- Ginsburg, J. C., & Budiardjo, L. A. (2020). AUTHORS AND MACHINES. *BERKELEY TECHNOLOGY LAW JOURNAL*, 34, 343-448.
- Mejías, U. A., & Couldry, N. (2019). Colonialismo de datos: repensando la relación de los datos masivos con el sujeto contemporáneo. *Virtualis*, 10(18), 78-97.
- Napoli, P. M. (17 de agosto de 2014). Obtenido de Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers: <http://dx.doi.org/10.2139/ssrn.2481886>
- Noble, S. U. (2018). Conclusion. En S. U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (págs. 171-182). New York: NYU Press.

REDEGAL. (30 de diciembre de 2025). *Redegal*. Obtenido de Reglamento Europeo de Inteligencia Artificial: <https://www.redegala.com/es/blog/reglamento-ue-inteligencia-artificial/>

Senftleben, M., Angelopoulos, C., Frosio, G., Moscon, V., Peguera, M., & Rognstadf, O.-A. (2018). The Recommendation on Measures to Safeguard Fundamental Rights and the Open Internet in the Framework of the EU Copyright Reform. *European Intellectual Property Review*, 40(3), 149-163.

UNESCO. (3 AL 21 de OCT de 2005). *Diversidad de las Expresiones Culturales*. Obtenido de La Convención sobre la Diversidad de las Expresiones Culturales de 2005: <https://www.unesco.org/creativity/es/2005-convention>

UNESCO. (20 de Oct de 2005). *UNESCO*. Obtenido de Convention on the Protection and Promotion of the Diversity of Cultural Expressions: <https://www.unesco.org/en/legal-affairs/convention-protection-and-promotion-diversity-cultural-expressions>

CAPÍTULO XXXIII

DIGNIDAD, BASE DE TODA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL: HACIA UN CONSTITUCIONALISMO TECNOLÓGICO DE LA VIDA COMPARTIDA

Dignity, the basis of all regulation of artificial intelligence: towards a technological constitutionalism of shared life

Filiberto Eduardo R. Manrique Molina**

filiberto.manrique@uabc.edu.mx
Universidad Autónoma de Baja California
<https://orcid.org/0000-0001-9407-6548>.

Mauricio Iván Vargas Mendoza***

vargasm38@uabc.edu.mx
Universidad Autónoma de Baja California
<https://orcid.org/0000-0003-0290-8250>

Resumen

Este capítulo analiza la inteligencia artificial (IA) desde una perspectiva jurídico-filosófica centrada en la dignidad como fundamento de toda regulación tecnológica. A partir de un enfoque crítico, se sostiene que la IA no constituye únicamente una herramienta técnica, sino una tecnología con profundas implicaciones sociales, éticas, jurídicas y ecológicas. El estudio examina cómo los sistemas algorítmicos

** Doctor en Derecho y Globalización; Maestro en Derecho e investigación, abogado, realizó estudios de posdoctorado en la Facultad de Derecho, Ciencias Políticas y Sociales de la Universidad Nacional de Colombia- sede Bogotá; Investigador Nacional SNII nivel I- SECIHTI, Investigador Asociado. Minciencias, es profesor investigador de tiempo completo en FCIAS de la Universidad Autónoma de Baja California. ID ORCID: <https://orcid.org/0000-0001-9407-6548>, mail: filiberto.manrique@uabc.edu.mx

*** Posdoctorado en Derechos Humanos, Doctor en Derecho, Maestro en Derechos Humanos y Maestro en Derecho Civil, Director Facultad de Ciencias de la Ingeniería, Administrativas y Sociales de la Universidad Autónoma de Baja California. ORCID: <https://orcid.org/0000-0003-0290-8250>, Correo: vargasm38@uabc.edu.mx

pueden afectar la autonomía, la igualdad, las condiciones materiales de existencia y la protección frente a formas de humillación o exclusión, así como su impacto sobre otros seres vivos, los ecosistemas y la biodiversidad.

Asimismo, se propone ampliar la comprensión tradicional de la dignidad humana hacia una noción de “vida compartida”, orientada a reconocer la interdependencia entre humanidad, naturaleza y tecnología. Desde esta perspectiva, se plantea la necesidad de un constitucionalismo tecnológico capaz de someter el poder algorítmico a principios de derechos humanos, sostenibilidad, justicia social y protección ambiental. El capítulo también analiza marcos regulatorios internacionales, como las recomendaciones de la UNESCO y la Ley Europea de Inteligencia Artificial, destacando sus avances y limitaciones. Finalmente, se concluye que la regulación legítima de la IA debe orientarse no solo a controlar riesgos técnicos, sino a garantizar una convivencia democrática, digna y sostenible entre seres humanos, otros seres vivos y el mundo común.

Palabras clave: inteligencia artificial; dignidad humana; constitucionalismo tecnológico; vida compartida; regulación; derechos humanos.

Abstract

This chapter analyzes artificial intelligence (AI) from a legal-philosophical perspective centered on dignity as the foundation of all technological regulation. From a critical approach, it argues that AI is not merely a technical tool, but rather a technology with profound social, ethical, legal, and ecological implications. The study examines how algorithmic systems may affect autonomy, equality, material living conditions, and protection against forms of humiliation or exclusion, as well as their impact on other living beings, ecosystems, and biodiversity.

Likewise, the chapter proposes expanding the traditional understanding of human dignity toward a notion of “shared life,” aimed at recognizing the interdependence between humanity, nature, and technology. From this perspective, it advocates for a form of technological constitutionalism capable of subjecting algorithmic power to

principles of human rights, sustainability, social justice, and environmental protection. The chapter also examines international regulatory frameworks, such as UNESCO recommendations and the European Union Artificial Intelligence Act, highlighting both their progress and limitations. Finally, it concludes that the legitimate regulation of AI should not only focus on controlling technical risks, but also on ensuring a democratic, dignified, and sustainable coexistence among human beings, other living beings, and the shared world.

Keywords: artificial intelligence; human dignity; technological constitutionalism; shared life; regulation; human rights.

Introducción.

La inteligencia artificial se ha convertido en una de las fuerzas más influyentes en la transformación de la vida contemporánea. Su presencia atraviesa ámbitos tan diversos que no solo impactan la vida de los seres humanos, también de otras especies y seres con los que compartimos. Junto con sus oportunidades que son en su mayoría muy conocidas por la amplia divulgación en internet, también emergen riesgos profundos, el más extremo es el debilitamiento de la responsabilidad humana en la toma de decisiones que no solo afectan al ser humano, también a otros sujetos. Frente a este escenario, la pregunta central que guía el presente capítulo ya no es únicamente qué puede hacer la inteligencia artificial, sino bajo qué principios debe ser regulada.

En este contexto, la dignidad humana aparece como el fundamento esencial de toda regulación de la inteligencia artificial. No se trata de un concepto abstracto ni meramente declarativo, sino de un principio jurídico-filosófico que exige colocar a los seres vivos en el centro del desarrollo tecnológico. Regular la inteligencia artificial desde la dignidad implica reconocer que ninguna innovación puede justificar la reducción, instrumentalización, experimentación o aniquilación de cualquier forma de vida, además de la humana. La tecnología debe estar al servicio

de la vida humana y no la vida humana subordinada a la lógica de la eficiencia técnica de la Inteligencia Artificial

La metodología adoptada para este capítulo se fundamenta en un enfoque cualitativo, analítico y crítico, orientado a examinar la dignidad humana como principio rector de la regulación de la inteligencia artificial. Para ello, se recurre a la revisión de literatura jurídica, filosófica y tecnológica, así como al análisis de documentos normativos, declaraciones éticas e informes de organismos internacionales que abordan los riesgos y desafíos de la IA desde la perspectiva de los derechos fundamentales.

A partir de esta revisión, el capítulo desarrolla una reflexión teórico-normativa sobre la necesidad de avanzar hacia un constitucionalismo tecnológico, entendido como un marco de garantías que proyecta los valores constitucionales sobre los entornos digitales. De este modo, la metodología no se limita a describir los efectos de la inteligencia artificial, sino que busca valorar críticamente sus implicaciones para la vida compartida, proponiendo la dignidad humana como base indispensable para una regulación legítima, democrática y orientada al bien común.

II. Aproximación conceptual a la inteligencia artificial desde la dignidad humana

La inteligencia artificial puede comprenderse como un conjunto de sistemas, métodos y tecnologías diseñados para realizar tareas que, tradicionalmente, han sido asociadas con la imitación de las capacidades cognitivas humanas. Entre estas funciones se encuentran el aprendizaje, la percepción, el razonamiento, la resolución de problemas, el reconocimiento de patrones y la toma de decisiones, etc. En este sentido, la IA no debe entenderse únicamente como una máquina que ejecuta instrucciones, sino como una tecnología capaz de procesar información, identificar relaciones entre datos y generar respuestas o predicciones a partir de modelos previamente entrenados (Russell & Norvig, 2020). No estamos equiparando la capacidad de la IA con otras capacidades que la naturaleza le ha

dotado al ser humano y otros seres de la familia *animalia*, tales como lo son regulación emocional, la conciencia propia, metacognición, asociación de estímulos del entorno, etc.

La imitación de las capacidades del razonamiento, no son nada nuevas, se han popularizado actualmente por el boom tecnológico del siglo XXI, pero el origen conceptual de la inteligencia artificial se vincula con las reflexiones de Alan Turing, quien planteó la posibilidad de que una máquina pudiera imitar comportamientos inteligentes y participar en procesos de comunicación semejantes a los humanos. Su propuesta abrió el debate sobre la capacidad de las máquinas para “pensar” o, al menos, para producir resultados funcionalmente equivalentes a ciertas formas de inteligencia humana (Turing, 1950, p. 433). Desde entonces, la IA ha evolucionado desde sistemas basados en reglas lógicas y programación simbólica hacia modelos más complejos, capaces de aprender de grandes volúmenes de datos.

En sus primeras etapas, la inteligencia artificial se apoyó principalmente en algoritmos simbólicos, reglas predefinidas y estructuras lógicas. Estos sistemas permitieron resolver problemas específicos mediante instrucciones diseñadas por seres humanos, pero presentaban limitaciones importantes frente a contextos cambiantes o situaciones no previstas. Su funcionamiento dependía, en gran medida, de la calidad y exhaustividad de las reglas incorporadas, por lo que su capacidad de adaptación era reducida.

Con el desarrollo del aprendizaje automático y, especialmente, del aprendizaje profundo, la inteligencia artificial experimentó una transformación significativa. La disponibilidad masiva de datos, el incremento de la capacidad computacional y la expansión de internet facilitaron la creación de modelos capaces de identificar patrones complejos y mejorar su desempeño mediante el entrenamiento continuo. Las redes neuronales artificiales, inspiradas parcialmente en el funcionamiento del cerebro humano, han permitido avances relevantes en áreas como el

reconocimiento de imágenes, el procesamiento del lenguaje natural, la traducción automática y el reconocimiento de voz (LeCun, Bengio, & Hinton, 2015, p. 437), entre otras funciones que van aparaciendo con los nuevos modelos.

No obstante, definir la inteligencia artificial exige reconocer que no se trata de una tecnología neutral en sus efectos sociales. Su capacidad para intervenir en procesos de decisión, clasificar personas, objetos, predecir conductas o automatizar servicios públicos y privados la convierte en una herramienta de alto impacto que nos invita a abrir un nuevo momento dignatario por sus efectos directos en la vida de los seres humanos e indirectos a otros seres como los animales, plantas, biodiversidad y entornos naturales. Por ello, la IA puede ser considerada una tecnología de multidimensional, pues no solo transforma las relaciones humanas, sino también la relación de la humanidad con otros seres vivos y con los ecosistemas que sostienen la vida. Por ello, una regulación fundada en la dignidad humana debe proyectarse hacia una ética de la vida compartida, capaz de integrar la protección del ambiente, el bienestar animal y la preservación de la biodiversidad como condiciones indispensables para el desarrollo tecnológico legítimo.

Desde esta perspectiva, la inteligencia artificial no solo debe definirse por sus capacidades cuyos efectos y consecuencias son directos sobre la vida humana, también guarda una relación de vida compartida, donde la regulación de la IA no solo proteja a las personas, sino también las condiciones que hacen posible la existencia común con otros seres vivos. En consecuencia, la IA debe ser entendida como una tecnología capaz de transformar no solo las estructuras sociales e institucionales, sino también las formas en que los seres humanos se relacionan con otros seres vivos y con los ecosistemas que sostienen la existencia común. Su desarrollo y aplicación requieren marcos normativos sólidos que garanticen su orientación hacia el bien común, evitando que la eficiencia tecnológica se imponga sobre la dignidad humana, el bienestar animal, la protección ambiental y la preservación de la biodiversidad (UNESCO, 2020, p. 8).

Desde esta perspectiva, la inteligencia artificial no puede concebirse como una herramienta neutral ni aislada de sus consecuencias éticas, sociales y ecológicas. Sus usos en la agricultura, la ganadería, la vigilancia ambiental, la biotecnología o la gestión de recursos naturales pueden generar beneficios importantes, pero también riesgos cuando reducen la vida a datos, rendimiento, productividad o control. Por ello, una regulación adecuada debe avanzar hacia una idea de vida compartida, en la que la innovación tecnológica se oriente a proteger las condiciones que hacen posible la convivencia entre personas, comunidades, especies y naturaleza.

III. Dignidad base de toda regulación de la inteligencia artificial

La dignidad, entendida en términos amplios y no exclusivamente humanos, constituye un precepto esencial para orientar la regulación de la inteligencia artificial, porque permite reconocer el valor de la vida y de las condiciones que la hacen posible. No se limita únicamente a la protección individual de la persona humana, sino que también proyecta una exigencia ética, jurídica y social hacia la convivencia, el respeto por otros seres vivos, la protección de los ecosistemas y la preservación de la vida compartida. Desde esta perspectiva, la regulación de la IA no debe construirse únicamente desde un enfoque antropocéntrico, centrado de manera exclusiva en los intereses humanos, sino desde una comprensión más amplia de sus efectos sobre la existencia común. Esto implica valorar cómo los sistemas inteligentes inciden no solo en las personas, sino también en otros seres vivos, en los ecosistemas y en las condiciones materiales que hacen posible la vida compartida.

La dignidad opera, entonces, como un límite frente a los usos de la IA que puedan afectar la autonomía humana, profundizar desigualdades, deteriorar el ambiente, instrumentalizar a los animales o debilitar las condiciones materiales y simbólicas de la convivencia. Regular la IA desde una comprensión amplia de la dignidad implica asumir que el desarrollo tecnológico no ocurre en el vacío, sino dentro de

una comunidad de vida. Las decisiones algorítmicas pueden impactar derechos fundamentales, estructuras sociales, modelos económicos, prácticas ambientales y formas de relación con la biodiversidad. Por esta razón, la dignidad exige que la inteligencia artificial sea diseñada, implementada y supervisada conforme a criterios de responsabilidad, justicia, sostenibilidad, transparencia y respeto por el bien común.

En este sentido, la regulación de la IA debe estar guiada por un principio rector, que toda innovación tecnológica en la materia debe respetar, proteger y garantizar la dignidad en sus distintas dimensiones. La ausencia de regulación puede generar vacíos normativos que permitan el desarrollo y uso de sistemas de IA sin límites claros, sin responsabilidades definidas y sin mecanismos suficientes de control frente a sus posibles riesgos. Por ello, la legislación sobre inteligencia artificial no debe surgir únicamente como una respuesta técnica o económica, sino como una construcción jurídica orientada por valores superiores.

La dignidad debe ocupar el lugar de fundamento permanente en todo proceso legislativo sobre IA, desde el diseño de las normas hasta su aplicación e interpretación. Esto supone que cualquier marco regulatorio debe preguntarse, antes que nada, si el desarrollo tecnológico contribuye a proteger la vida, la autonomía, la igualdad, la justicia, la sostenibilidad y la convivencia democrática. Así, la inteligencia artificial no debe avanzar como un poder autónomo ni quedar sometida exclusivamente a las lógicas del mercado o de la eficiencia, sino integrarse en un orden jurídico que la encauce hacia el bien común y hacia el respeto de la vida compartida.

Derivado de las reflexiones anteriores, los siguientes apartados analizarán la relación entre dignidad e inteligencia artificial desde una doble perspectiva. En primer lugar, se abordará la dignidad humana e inteligencia artificial, con el propósito de examinar cómo los sistemas algorítmicos pueden incidir en la autonomía, la libertad, las condiciones materiales de existencia, la protección frente

a toda forma de humillación o degradación de las personas y el contar con tiempo de recreación. En segundo lugar, se estudiará la regulación de la inteligencia artificial desde la dignidad del mundo compartido, entendida como una mirada más amplia que no se agota en el ser humano, sino que también considera los efectos de la IA sobre otros seres vivos, los ecosistemas, la biodiversidad y las condiciones naturales que sostienen la vida. Así, el análisis propone comprender la dignidad como fundamento regulatorio tanto para proteger a las personas frente al poder tecnológico, como para orientar la innovación hacia el cuidado responsable del mundo y de la vida que nos rodea.

III. I Dignidad humana e inteligencia artificial

La dignidad humana ha sido reconocida ampliamente como el fundamento central sobre el cual se construyen los derechos humanos y el sistema jurídico de las sociedades democráticas. En la interpretación de algunos tribunales, la dignidad se ha consolidado como un principio y derecho fundamental que comprende, al menos, tres dimensiones esenciales: vivir como se quiera, es decir, la autonomía para diseñar un proyecto vital propio; vivir bien, entendido como el acceso a condiciones materiales mínimas de existencia; y vivir sin humillaciones, esto es, la protección frente a tratos degradantes, físicos o morales (Principio de dignidad humana, naturaleza, 2002). A estas dimensiones puede añadirse una cuarta exigencia contemporánea: disponer de tiempo libre real para descansar, convivir, participar en la vida cultural y recrearse, pues una vida digna no se agota en sobrevivir ni en producir, sino que requiere espacios efectivos para habitar la existencia con libertad, vínculos y sentido.

Estas dimensiones adquieren una relevancia especial frente al desarrollo de la inteligencia artificial. La IA no es únicamente una herramienta técnica destinada a procesar datos, automatizar o facilitar tareas; es una tecnología capaz de incidir en la forma en que las personas deciden, trabajan, acceden a servicios, son clasificadas por las instituciones y participan en la vida social. Por ello, cuando los

sistemas algorítmicos intervienen en decisiones relacionadas con la salud, la educación, el empleo, la seguridad, la justicia, la asistencia social o el acceso al crédito, pueden afectar directamente las condiciones que hacen posible una vida digna. En consecuencia, la regulación de la IA debe partir de la dignidad humana como criterio estructurante, no como una referencia decorativa, sino como un límite material frente a toda forma de instrumentalización tecnológica.

Desde esta perspectiva, la primera dimensión de la dignidad, vivir como se quiera, exige proteger la autonomía personal frente a sistemas de IA que puedan manipular, condicionar o sustituir indebidamente la voluntad humana. La autonomía no se reduce a la ausencia de interferencias, sino que implica la posibilidad real de decidir de manera informada, consciente y responsable sobre el propio proyecto de vida. En el entorno digital, esta exigencia se traduce en la necesidad de garantizar transparencia algorítmica, consentimiento informado, explicabilidad, control humano y protección frente a prácticas de vigilancia, perfilamiento o persuasión automatizada que limiten la libertad de elección. Cuando una persona es reducida a un perfil predictivo, a una puntuación de riesgo o a una categoría estadística que determina sus oportunidades, su autonomía puede verse gravemente comprometida.

Ahora bien, la autonomía digna no significa una libertad absoluta ni aislada. Como advierte Waldron, no es correcto contraponer derechos y responsabilidades como si fueran términos incompatibles, pues el ejercicio de las libertades implica deberes correlativos frente a los demás (Waldron, 2019, p. 109). Esta idea resulta particularmente importante para la inteligencia artificial, porque quienes diseñan, entrenan, comercializan o implementan sistemas algorítmicos no solo ejercen libertad de innovación, sino también responsabilidades frente a las personas afectadas por sus decisiones técnicas. La dignidad exige que la IA no sea desarrollada desde una lógica de poder unilateral, sino desde una ética de responsabilidad, prudencia y respeto por quienes pueden sufrir sus impactos.

En ese sentido, la dignidad humana no debe entenderse como una garantía ilimitada de cualquier pretensión individual, sino como un criterio estructurante del orden constitucional. Siguiendo la reflexión de Nogueira de Brito, la dignidad funciona mejor como un umbral infranqueable frente a tratos objetivamente incompatibles con la condición humana, sin absorber por completo el campo de los demás derechos fundamentales (Nogueira de Brito, 2019, pp. 131–132). Aplicado a la IA, esto significa que no toda innovación debe prohibirse por el solo hecho de generar riesgos, pero sí deben excluirse aquellas aplicaciones que cosifican, discriminan, degradan o privan a las personas de control significativo sobre decisiones que afectan su vida.

La segunda dimensión de la dignidad, vivir bien, exige analizar la inteligencia artificial desde sus efectos sobre las condiciones materiales de existencia. La dignidad no se agota en la libertad formal, sino que requiere alimentación, vivienda, salud, educación, trabajo digno, seguridad social y acceso a bienes básicos. En este punto, la IA puede convertirse en una herramienta para ampliar capacidades y mejorar servicios públicos, por ejemplo, mediante diagnósticos médicos más oportunos, sistemas de alerta temprana, gestión eficiente de recursos o apoyos personalizados en educación. Sin embargo, también puede profundizar desigualdades si reproduce sesgos, excluye a poblaciones vulnerables, automatiza decisiones sin revisión humana o concentra beneficios económicos en pocos actores tecnológicos.

Por ello, la regulación de la IA debe tener una orientación materialmente igualitaria. El contenido mínimo existencial, como señala Bilchitz, no debe entenderse solo como un medio instrumental para alcanzar otros fines, sino como una exigencia intrínseca de la vida digna, especialmente frente a amenazas graves a la supervivencia, el dolor extremo o la privación severa (Bilchitz, 2017, pp. 261–262). En el campo de la IA, esta idea obliga a preguntarse si los sistemas automatizados contribuyen efectivamente a garantizar mínimos vitales o si, por el contrario, se convierten en mecanismos de exclusión. Un algoritmo utilizado para asignar ayudas

sociales, evaluar riesgos médicos, priorizar servicios públicos o seleccionar trabajadores no puede operar como una barrera invisible que niega derechos bajo apariencia de neutralidad técnica.

Esta preocupación se intensifica respecto de las personas con discapacidad y otros grupos históricamente excluidos. Barclay advierte que una agenda de justicia digna exige terminar con la discriminación, hacer el entorno plenamente accesible, financiar servicios de cuidado personal, asegurar apoyos de ingresos adecuados y combatir el estigma (Barclay, 2019, p. 4). En el ámbito de la IA, ello implica diseñar tecnologías accesibles, inclusivas y no discriminatorias desde su origen. Un sistema algorítmico que no reconoce diversidad corporal, cognitiva, lingüística o sensorial puede reforzar barreras ya existentes. Por tanto, la dignidad exige que la IA no solo evite dañar, sino que contribuya a remover obstáculos reales para la participación igualitaria.

La tercera dimensión, vivir sin humillaciones, resulta igualmente decisiva. La inteligencia artificial puede producir formas nuevas de degradación cuando clasifica, expone, vigila o juzga a las personas mediante criterios opacos. Margalit sostiene que la humillación comprende comportamientos o condiciones que dan a una persona razones fundadas para considerar lesionado su autorrespeto (Margalit, 1996, p. 9). En el mundo algorítmico, la humillación puede manifestarse cuando una persona es rechazada automáticamente por un sistema sin explicación, cuando es tratada como sospechosa por patrones estadísticos, cuando su rostro o sus datos son utilizados sin consentimiento, o cuando una decisión automatizada reproduce prejuicios raciales, económicos, de género, edad, discapacidad u origen social.

Desde un enfoque negativo de la dignidad, las vulneraciones no solo se producen mediante violencia física, sino también mediante prácticas simbólicas de desprecio, estigmatización, silenciamiento o exclusión. Esta lectura es especialmente útil para la IA, porque muchos daños algorítmicos no aparecen como agresiones directas, sino como decisiones impersonales, automatizadas y aparentemente objetivas. Sin

embargo, sus efectos pueden ser profundamente degradantes: negar oportunidades, reforzar estereotipos, impedir la participación social o comunicar a ciertos grupos que valen menos que otros. Por ello, una regulación dignitaria de la IA debe prohibir sistemas que produzcan discriminación, vigilancia desproporcionada, manipulación abusiva o tratos incompatibles con el respeto debido a toda persona.

La cuarta dimensión de la dignidad, relativa al tiempo libre real, también debe incorporarse al debate sobre IA. La automatización puede liberar tiempo humano, reducir cargas repetitivas y mejorar la calidad de vida; pero también puede intensificar la productividad, ampliar la vigilancia laboral y someter a los trabajadores a ritmos definidos por sistemas algorítmicos. En este sentido, la IA no debe utilizarse para convertir a la persona en una extensión de la máquina ni para medir cada gesto, pausa o rendimiento como si el trabajador fuera únicamente una unidad productiva. La dignidad exige reconocer que el descanso, el ocio, la vida familiar, la participación cultural y la recreación son condiciones necesarias para una existencia humana plena.

Cicerón permite comprender que el ocio no es simple inactividad, sino un tiempo valioso para la reflexión, la recuperación y la vida interior, siempre que no signifique evasión de los deberes sociales (Cicerón, 1962, pp. 9, 100). Trasladado al campo de la inteligencia artificial, esto significa que la tecnología debe contribuir a humanizar el trabajo y no a intensificar la explotación. Una regulación adecuada debe impedir que la IA se utilice para extender jornadas, precarizar empleos, vigilar permanentemente a los trabajadores o eliminar márgenes de descanso. El progreso tecnológico solo puede considerarse compatible con la dignidad si amplía las posibilidades de vivir, convivir y participar en la cultura, no si reduce la existencia a rendimiento permanente.

En suma, la dignidad humana ofrece el criterio más sólido para orientar la regulación de la inteligencia artificial porque permite evaluar sus impactos sobre la

autonomía, las condiciones materiales, la integridad moral y física, y el tiempo vital de las personas. Regular la IA desde la dignidad significa afirmar que ningún sistema tecnológico puede situarse por encima del valor de la vida humana ni convertir a las personas en objetos de predicción, vigilancia, explotación o exclusión. La IA será legítima solo si contribuye a ampliar capacidades, reducir desigualdades, proteger derechos, evitar humillaciones y fortalecer una vida compartida más humana.

III.II La regulación de la inteligencia artificial desde la dignidad del mundo compartido

Pensar la dignidad del mundo compartido en el contexto de la inteligencia artificial exige partir de una premisa fundamental, que el ser humano no existe fuera de la naturaleza ni por encima de ella, sino dentro de una red de vida, límites y dependencias que hacen posible su propia existencia. En clave espinosista, el ser humano no es una realidad separada del orden natural, sino una expresión finita de la Naturaleza; por ello, la vida humana debe comprenderse como parte de una totalidad más amplia y no como un poder autorizado para dominarla sin restricciones (Spinoza, 2000, pp. 125, 189). Desde esta perspectiva, la IA no puede ser regulada únicamente como una herramienta técnica al servicio de intereses humanos inmediatos, sino como una tecnología cuyos efectos se proyectan sobre la vida común, los ecosistemas, los animales, las plantas, los recursos naturales y las condiciones materiales que sostienen la existencia.

La dignidad del mundo compartido supone superar una visión estrictamente antropocéntrica de la regulación tecnológica. Esto no significa negar la centralidad de la dignidad humana, sino reconocer que dicha dignidad se encuentra vinculada a un entorno vivo del cual depende. Una humanidad digna no puede construirse sobre la destrucción de aquello que hace posible su vida. Por ello, cuando la inteligencia artificial se aplica en ámbitos como la agricultura, la ganadería, la gestión ambiental, la vigilancia de ecosistemas, la biotecnología, la explotación de recursos o la administración de infraestructuras energéticas, sus impactos no

recaen solamente sobre individuos humanos, sino también sobre otros seres vivos y sobre el equilibrio del mundo natural.

En este sentido, la dignidad debe entenderse como una forma de ubicación ética en el mundo. No se trata de afirmar que el ser humano posee un valor que lo autoriza a disponer ilimitadamente de la naturaleza, sino de reconocer que su capacidad racional y técnica le impone mayores responsabilidades. Como advierte Kateb, una comprensión irreflexiva de la dignidad humana puede derivar en un orgullo de especie que justifique la explotación de la naturaleza y el sometimiento de otros seres vivos (Kateb, 2011, pp. 3-4). Trasladado al campo de la inteligencia artificial, este riesgo aparece cuando la innovación tecnológica se desarrolla bajo una lógica de dominio, productividad y extracción, sin atender a los límites ecológicos ni al valor propio de las formas de vida afectadas.

La inteligencia artificial, por tanto, debe ser sometida a una regulación que incorpore la conciencia de interdependencia. Las crisis ecológicas actuales no son simples efectos secundarios del progreso, sino manifestaciones de una forma de relación con la Tierra marcada por la apropiación acelerada y el desequilibrio. La intensidad y velocidad de la intervención humana han alcanzado niveles que impiden que los ecosistemas se reacomoden de manera natural, comprometiendo su capacidad regenerativa (Ávila Foucat & Perevochtchikova, 2019, p. 33). En este escenario, la IA puede contribuir a prevenir daños, monitorear riesgos y optimizar recursos, pero también puede intensificar modelos extractivos, productivistas o de vigilancia ambiental si no se encuentra jurídicamente encauzada por principios de dignidad, sostenibilidad y responsabilidad.

De ahí que la regulación de la IA deba evitar los vacíos normativos que permitirían su expansión sin límites claros. Una tecnología capaz de procesar grandes volúmenes de datos, automatizar decisiones y dirigir procesos productivos o ambientales no puede quedar sometida exclusivamente a criterios de eficiencia, rentabilidad o competitividad. Su desarrollo legislativo debe partir de la dignidad

como fundamento, entendida no solo como protección de la persona humana, sino también como respeto por la vida compartida. Esto exige que toda norma sobre inteligencia artificial considere sus consecuencias sobre la comunidad humana, los otros seres vivos y los ecosistemas que sostienen la existencia común.

La advertencia de Ferrajoli resulta especialmente pertinente en este punto: por primera vez, la humanidad enfrenta peligros que pueden comprometer su propia continuidad como consecuencia de su actividad irresponsable (Ferrajoli, 2022, p. 6). La IA forma parte de ese nuevo poder técnico que puede ampliar las capacidades humanas, pero también aumentar los riesgos si se emplea sin autocontención normativa. Por ello, no basta con preguntarse qué puede hacer la inteligencia artificial; es indispensable preguntarse qué debe hacer, bajo qué límites, con qué finalidad y a favor de qué modelo de vida. La regulación dignitaria de la IA debe impedir que la tecnología se convierta en un instrumento de aceleración destructiva del mundo común.

Esta perspectiva se relaciona con el giro biocéntrico y ecocéntrico desarrollado en el derecho contemporáneo. Durante mucho tiempo, la naturaleza fue concebida como objeto de uso o recurso disponible para la acción humana. Sin embargo, nuevas corrientes constitucionales y legales han comenzado a reconocer que la naturaleza, los animales y los ecosistemas poseen un valor que no depende exclusivamente de su utilidad para las personas. En ese sentido, el tránsito hacia modelos que reconocen derechos de la naturaleza, protección reforzada de los animales o deberes de cuidado ecológico revela una transformación profunda de la conciencia jurídica contemporánea (R. Manrique y otros., 2020, p. 127).

La inteligencia artificial debe insertarse dentro de ese cambio de paradigma. Si el derecho ha comenzado a reconocer que la naturaleza no es un simple objeto de apropiación, entonces la regulación de la IA tampoco puede tratar los ecosistemas y los seres vivos como meras fuentes de datos, recursos productivos o variables de optimización. La automatización aplicada al mundo natural debe estar guiada por

criterios de respeto, precaución y responsabilidad. Esto implica reconocer que no todo lo técnicamente posible es jurídicamente legítimo, especialmente cuando la intervención tecnológica puede alterar formas de vida, modificar ecosistemas o profundizar relaciones de explotación.

Un referente especialmente significativo para esta discusión se encuentra en la Constitución Suiza, cuyo artículo 120 incorpora una referencia expresa a la dignidad de los seres vivos en el ámbito de la tecnología genética no humana (Swiss Confederation, 1999, art. 120). Aunque esta disposición se refiere directamente a la tecnología genética, su lógica puede iluminar la regulación de la inteligencia artificial: las tecnologías emergentes que intervienen sobre la vida deben ser evaluadas no solo por su utilidad científica o económica, sino también por su compatibilidad con el valor propio de los seres vivos. En consecuencia, una regulación constitucional de la IA debe asumir que la intervención tecnológica sobre animales, plantas, organismos y ecosistemas exige límites éticos y jurídicos.

Esta idea resulta particularmente importante en relación con los animales. El derecho contemporáneo ha avanzado en reconocerlos como seres sintientes, capaces de experimentar dolor, sufrimiento y formas de bienestar. Este cambio normativo permite cuestionar los usos de IA que intensifican prácticas de explotación animal o que reducen la vida animal a rendimiento, productividad o control automatizado. Si la legislación comienza a exigir trato digno, bienestar y protección frente al sufrimiento, entonces los sistemas de IA aplicados a la ganadería, experimentación, vigilancia, reproducción o manejo de animales deben ser diseñados y regulados conforme a esos parámetros.

En América Latina y Europa pueden observarse avances en esta dirección. La Constitución de Ecuador reconoce a la naturaleza como sujeto de derechos, mientras que otros ordenamientos constitucionales y legales han incorporado deberes de protección ambiental, prohibición de crueldad animal, reconocimiento de seres sintientes o categorías vinculadas al trato digno (Asamblea Nacional del

Ecuador, 2008, arts. 10 y 71; Presidência da República do Brasil, 1988, art. 225, § 1, VII; Constitución Política de la Ciudad de México, 2025, art. 13, apartado B). Estos desarrollos son relevantes para la inteligencia artificial porque muestran que el derecho ya no puede limitarse a proteger únicamente intereses humanos directos, sino que debe atender a la vida en sus relaciones y dependencias.

La dignidad del mundo compartido exige, entonces, que la IA sea regulada desde una ética de la corresponsabilidad. Esto significa que su desarrollo debe orientarse hacia la protección de la vida y no hacia su instrumentalización ilimitada. La IA puede ser una herramienta valiosa para conservar biodiversidad, anticipar riesgos climáticos, mejorar el uso de recursos naturales, proteger especies o reducir daños ambientales. Pero esos beneficios solo serán legítimos si se insertan dentro de marcos normativos que impidan que la tecnología refuerce dinámicas de dominación, extracción o exclusión. La innovación no puede presentarse como progreso cuando compromete las condiciones ecológicas, sociales y simbólicas que hacen posible una vida digna.

En suma, la dignidad del mundo compartido permite ampliar el horizonte de la regulación de la inteligencia artificial. No se trata de abandonar la dignidad humana, sino de comprenderla dentro de una comunidad más amplia de vida. Una IA digna no es solo aquella que respeta derechos individuales, sino también aquella que reconoce límites, protege vínculos, evita daños al entorno y contribuye a una convivencia más justa entre seres humanos, otros seres vivos y naturaleza. Por ello, el constitucionalismo tecnológico debe asumir que la vida compartida constituye el marco último de legitimidad de toda innovación: ninguna inteligencia artificial puede considerarse verdaderamente avanzada si su desarrollo debilita las condiciones que permiten vivir, coexistir y cuidar el mundo común.

IV. Constitucionalismo en la era de la IA con protección de la vida compartida.

La expansión de la inteligencia artificial exige repensar el papel de las constituciones frente al poder tecnológico. Si las cartas nacionales han sido

históricamente el espacio normativo destinado a limitar el poder político y garantizar la dignidad, la libertad y los derechos fundamentales, hoy también deben proyectar esa función sobre los nuevos poderes de la IA. Esta, no solo transforma actividades económicas o procesos educativos o técnicos; también incide en la forma en que las personas acceden a derechos, son evaluadas por instituciones, participan en la vida pública, trabajan, se relacionan y habitan el mundo. Por ello, su regulación no puede quedar reducida a normas administrativas, políticas sectoriales o códigos éticos voluntarios, sino que debe encontrar fundamento en el orden constitucional.

Un constitucionalismo contemporáneo debe regular la IA reconociendo desde las cartas constitucionales que la dignidad debe operar como principio rector de toda innovación. Esto significa que ningún sistema de inteligencia artificial puede ser legítimo si reduce a las personas a datos, perfiles, objetos de predicción o simples medios para fines económicos, estatales o corporativos hablando en términos Kantianos, empleándolos como un medio (2012, pág. 148). Las constituciones deben establecer límites sustantivos frente a usos tecnológicos que afecten los derechos civiles, políticos, económicos, sociales, culturales y ambientales. De esta manera, la IA debe desarrollarse dentro de un marco de derechos, y no como una fuerza autónoma guiada únicamente por las decisiones de las grandes corporaciones y el mercado.

La regulación constitucional de la IA también resulta necesaria porque los riesgos tecnológicos suelen avanzar más rápido que las respuestas legislativas ordinarias. La ausencia de principios constitucionales claros puede generar vacíos regulatorios, zonas de impunidad y formas de afectación difícilmente reparables. Por ello, las cartas nacionales deben ofrecer criterios superiores que orienten al legislador, a las autoridades administrativas, a los jueces, a las empresas y a los desarrolladores. Tales criterios deben asegurar que el diseño, entrenamiento, implementación y supervisión de sistemas de IA respeten la dignidad y los derechos humanos

Sin embargo, este constitucionalismo del siglo XXI no debe limitarse a una mirada exclusivamente antropocéntrica, como lo hemos afirmado en el presente capítulo, pues nos encontramos ante un nuevo paradigma (Hernández Ramos, 2025, p. 22). La inteligencia artificial también puede afectar a otros seres vivos, a los ecosistemas, a la biodiversidad y a las condiciones naturales que sostienen la vida. Su uso en agricultura, ganadería, biotecnología, vigilancia ambiental, gestión de recursos naturales, modelos climáticos, explotación energética o administración territorial puede producir impactos profundos sobre el mundo compartido. Por ello, una regulación constitucional adecuada debe integrar una comprensión amplia de la dignidad, capaz de proteger no solo al ser humano, sino también las relaciones de interdependencia que hacen posible la existencia común.

En este sentido, las constituciones deben avanzar hacia la protección de la vida compartida como horizonte normativo. Esto implica reconocer que la naturaleza no es un simple recurso disponible para la explotación tecnológica, ni los animales meros objetos de administración productiva, ni la biodiversidad una variable secundaria del desarrollo económico. El constitucionalismo digital debe afirmar que toda innovación, incluida la IA, debe respetar los límites ecológicos, prevenir daños graves al ambiente, evitar la instrumentalización cruel de otros seres vivos y contribuir a la preservación de los ecosistemas. La protección de la vida humana depende, en última instancia, de la protección de las condiciones naturales que la sostienen.

De esta forma, regular la IA desde las constituciones significa colocar la tecnología bajo un marco superior de legitimidad democrática, jurídica y ética. No se trata de frenar la innovación, sino de orientarla hacia fines compatibles con la dignidad, los derechos humanos, la justicia social, la sostenibilidad y el bien común. La IA puede ser una herramienta poderosa para mejorar la vida, ampliar capacidades, proteger el ambiente y fortalecer instituciones; pero solo será constitucionalmente legítima si se somete a límites que impidan su uso como mecanismo de dominación, exclusión, degradación o destrucción de la vida compartida.

Así, el constitucionalismo tecnológico de la IA debe concebirse como una nueva etapa del Estado constitucional. Su tarea principal consiste en asegurar que el poder algorítmico no quede fuera del derecho, sino sometido a los principios superiores que organizan la convivencia democrática. En esta perspectiva, las cartas nacionales deben convertirse en el fundamento de una regulación capaz de proteger simultáneamente la dignidad humana, los derechos fundamentales, los derechos de otros seres vivos, los ecosistemas, la biodiversidad y las condiciones naturales de existencia. Solo así la inteligencia artificial podrá integrarse en un proyecto de sociedad orientado no al dominio técnico del mundo, sino al cuidado responsable de la vida que compartimos.

Así, la dignidad se convierte en el fundamento de un constitucionalismo orientado no solo a proteger a la persona frente al poder algorítmico, sino también a garantizar que la innovación contribuya a una vida compartida más justa, solidaria y sostenible. En este sentido, regular la inteligencia artificial desde la dignidad significa preguntarse no solo qué puede hacer la tecnología, sino qué tipo de sociedad, de convivencia y de relación con la vida queremos construir mediante ella.

Uno de los ejemplos constitucionales más avanzados para pensar una regulación amplia de la inteligencia artificial se encuentra en la Constitución Suiza, la cual incorpora, en su artículo 120, una referencia expresa a la dignidad de otros seres vivos no humanos en el ámbito de la tecnología (Swiss Confederation, 1999, art. 120). Aunque esta disposición fue formulada originalmente en relación con la ingeniería genética, su valor resulta especialmente relevante para el debate contemporáneo sobre la IA, pues muestra que la regulación de tecnologías emergentes no debe limitarse a criterios de seguridad, eficiencia o utilidad científica, sino que debe considerar también el valor propio de la vida y los efectos que la innovación puede producir sobre seres humanos, animales, plantas, organismos y ecosistemas.

Esta perspectiva permite trasladar al campo de la inteligencia artificial una comprensión más amplia de la dignidad. Los sistemas de IA no solo intervienen en decisiones humanas, sino que también pueden incidir en la gestión de recursos naturales, la agricultura de precisión, la ganadería automatizada, la vigilancia ambiental, la conservación de especies y la experimentación biotecnológica. Por ello, una regulación constitucionalmente orientada de la IA debe reconocer que la tecnología puede afectar la vida compartida en múltiples dimensiones, y que su desarrollo debe sujetarse a límites éticos y jurídicos capaces de proteger tanto la dignidad humana como las condiciones de existencia de otros seres vivos.

V. Normativas internacionales y propuestas legislativas sobre IA

Ante la rápida expansión de la inteligencia artificial y sus implicaciones sobre los derechos humanos, la dignidad, la vida social y las condiciones de existencia compartida, diversas organizaciones internacionales han comenzado a elaborar marcos regulatorios y éticos destinados a orientar su desarrollo, implementación y control. Es indudable que la IA tiene una capacidad para intervenir en decisiones públicas y privadas, por ello, exige respuestas jurídicas que eviten vacíos regulatorios y garanticen que la innovación tecnológica se mantenga subordinada a reglas jurídicas claras.

En este contexto, la UNESCO ha propuesto recomendaciones dirigidas a armonizar el desarrollo tecnológico con el respeto a la dignidad humana, la justicia social, la protección de los derechos humanos y el bien común (UNESCO, 2020). Esta orientación resulta especialmente relevante porque sitúa la regulación de la IA dentro de un marco ético y jurídico que no se limita a la eficiencia o al rendimiento tecnológico, sino que exige evaluar sus efectos sobre las personas, las comunidades y las condiciones que hacen posible una convivencia justa.

Asimismo, la Unión Europea ha impulsado iniciativas como la “Ley de Inteligencia Artificial”, orientada a establecer estándares normativos para asegurar que el desarrollo y uso de estas tecnologías se realice bajo criterios de transparencia,

seguridad, responsabilidad y no discriminación (European Commission, 2021). Este tipo de propuestas evidencia la necesidad de construir marcos jurídicos capaces de anticipar riesgos, definir responsabilidades y someter los sistemas de IA a controles adecuados, especialmente cuando sus aplicaciones pueden incidir en derechos fundamentales o en ámbitos sensibles de la vida social.

La crítica central a esta Ley es que la Ley Europea mantiene una matriz predominantemente antropocéntrica. Aunque incluye referencias a la protección ambiental, no desarrolla una comprensión robusta de la IA como tecnología capaz de afectar la vida no humana y el equilibrio ecológico. La naturaleza aparece más como un bien protegido o un ámbito de impacto que como parte de una comunidad de vida con valor propio. En consecuencia, la regulación europea avanza en la protección de la persona frente al poder algorítmico, pero no logra formular con igual fuerza una protección constitucional o normativa de la vida compartida.

En ese sentido, la Ley Europea representa un primer paso, pero no agota el horizonte regulatorio necesario. Su mayor fortaleza es haber colocado la IA dentro de un marco jurídico vinculante basado en riesgos y derechos fundamentales, ejemplo de ello son los recitales 27, 28, 31, 48, 58 y 80; su principal insuficiencia es no haber desarrollado una noción amplia de dignidad que incluya de manera expresa la vida no humana, la biodiversidad y los ecosistemas. Desde la perspectiva sostenida en este capítulo, estas propuestas legislativas deben comprenderse como parte de un proceso de regulación de la inteligencia artificial, aún y con sus imperfecciones.

Regular la IA no significa únicamente ordenar su funcionamiento técnico, sino establecer límites y finalidades compatibles con la dignidad. Por ello, toda normativa debe partir de la idea de que la tecnología no puede avanzar como un poder autónomo guiado exclusivamente por el mercado, la productividad o la innovación, sino como una herramienta jurídicamente orientada hacia la protección de la persona, la justicia social y la vida compartida.

Además, la regulación de la IA debe incorporar una mirada que vaya más allá del antropocentrismo estricto. Sus efectos pueden proyectarse sobre otros seres vivos, los ecosistemas, la biodiversidad y las condiciones naturales que sostienen la existencia. En consecuencia, los marcos internacionales y nacionales deben evaluar la inteligencia artificial no solo por su utilidad a la persona humana, sino también por su impacto sobre el mundo común. Así, la dignidad se convierte en el fundamento que permite conectar la protección de los derechos humanos con el cuidado responsable de la naturaleza y de las formas de vida que rodean al ser humano.

No obstante, la implementación efectiva de estas normativas enfrenta desafíos importantes, entre ellos la rápida evolución de las tecnologías y la dificultad para prever todas sus posibles aplicaciones y consecuencias. Precisamente por ello, la regulación debe contar con principios sólidos y flexibles, capaces de adaptarse al cambio tecnológico sin renunciar a su fundamento esencial: el respeto y garantía de la dignidad. Solo así la inteligencia artificial podrá convertirse en un verdadero instrumento de progreso, bienestar social, sostenibilidad y protección de la vida compartida.

VI. Tesis conclusivas

La tesis central de este capítulo es que la inteligencia artificial no puede ser comprendida únicamente como una herramienta técnica, sino como una tecnología de alto impacto jurídico, social, ético y ecológico. Su capacidad para procesar datos, automatizar decisiones, optimizar procesos y transformar relaciones obliga a que su regulación no se limite a criterios de mercado, eficiencia, innovación o productividad. Por el contrario, debe partir de un fundamento superior: la dignidad. Esta dignidad debe operar como principio rector para determinar qué usos de la IA son legítimos y cuáles resultan incompatibles con la protección de la vida humana y de la vida compartida. En ese sentido, una IA jurídicamente aceptable no es aquella que simplemente funciona con eficacia, sino aquella que respeta a las

personas, protege derechos, evita formas de dominación tecnológica y contribuye al bien común.

Desde la perspectiva de la dignidad humana, la regulación de la IA debe impedir que las personas sean reducidas a datos, perfiles, puntuaciones, predicciones o simples objetos de decisión automatizada. La dignidad exige proteger la autonomía, la libertad, la igualdad, las condiciones materiales de existencia, la integridad física y moral, así como el derecho a vivir sin humillaciones. Por ello, los sistemas algorítmicos deben ser evaluados no solo por su funcionamiento técnico, sino también por sus efectos sobre la vida concreta de las personas. Una IA que reproduce desigualdades, excluye a grupos vulnerables, dificulta el acceso a servicios esenciales o somete a los trabajadores a formas intensificadas de vigilancia y productividad contradice el ideal de una vida digna. La tecnología solo puede considerarse progreso cuando amplía las posibilidades de vivir mejor, decidir con libertad, acceder a derechos, descansar, convivir y participar plenamente en la vida social.

Ahora bien, la regulación de la inteligencia artificial no debe permanecer encerrada en un antropocentrismo estricto. Aunque la dignidad humana conserva un lugar central, la IA también puede afectar a otros seres vivos, los ecosistemas, la biodiversidad y las condiciones naturales que sostienen la existencia. Por ello, este capítulo sostiene la necesidad de avanzar hacia una dignidad del mundo compartido, entendida como una categoría ética y jurídica que reconoce la interdependencia entre humanidad, naturaleza y demás formas de vida. Regular la IA desde esta perspectiva implica impedir que la tecnología reduzca la vida humana a recurso, dato, rendimiento o variable de optimización. Su uso en agricultura, ganadería, biotecnología, vigilancia ambiental, conservación de especies o gestión de recursos naturales debe someterse a criterios de responsabilidad, precaución, sostenibilidad y respeto por la vida.

De este modo, el constitucionalismo aparece como una respuesta necesaria frente al poder algorítmico. Las constituciones no pueden permanecer ajenas a la transformación digital, pues la IA incide en derechos fundamentales, estructuras democráticas, relaciones económicas, condiciones ambientales y formas de convivencia. Por ello, las cartas nacionales deben ofrecer principios superiores que orienten al legislador, a los jueces, a las autoridades administrativas, a las empresas y a los desarrolladores. Las propuestas internacionales, como las recomendaciones de la UNESCO y la Ley Europea de Inteligencia Artificial, representan avances relevantes al reconocer la necesidad de regular la IA conforme a la dignidad humana, los derechos fundamentales, la transparencia, la seguridad y la no discriminación; sin embargo, siguen siendo insuficientes cuando no desarrollan una noción amplia de dignidad que incluya expresamente a otros seres vivos, los ecosistemas, la biodiversidad y las condiciones naturales de existencia. En síntesis, la inteligencia artificial solo será legítima si se integra en un proyecto jurídico, constitucional y ético orientado al cuidado de la vida: no basta preguntar qué puede hacer la IA, sino qué debe permitírsele hacer, bajo qué límites, con qué finalidad y en beneficio de qué modelo de sociedad y de mundo compartido.

Fuentes de información

- Ávila Foucat, V. S., & Perevochtchikova, M. (2019). *Sistemas socio-ecológicos*. Instituto de Investigaciones Económicas, Universidad Nacional Autónoma de México.
- Barclay, L. (2019). *Disability with dignity: Justice, human rights and equal status*. Routledge.
- Bilchitz, D. (2017). *Pobreza y derechos fundamentales: La justificación y efectivización de los derechos socioeconómicos* (J. A. Portocarrero Quispe, Trad.). Marcial Pons.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 1–15.
- Cicerón, M. T. (1962). *De los deberes* (B. Estrada Morán, Trad.). Universidad Nacional Autónoma de México.
- Corte Constitucional de Colombia. (2002, 17 de octubre). *Sentencia T-881/02. Principio de dignidad humana, naturaleza*.
- European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts COM(2021) 206 final*.
- Ferrajoli, L. (2022). *Per una Costituzione della Terra: L'umanità al bivio*. Feltrinelli.
- Hernández Ramos, M. (2025). *El impacto de la inteligencia artificial en el derecho constitucional*. Centro de Estudios Políticos y Constitucionales.
- Kant, I. (2012). *Fundamentación de la metafísica de las costumbres*. Alianza Editorial.
- Kateb, G. (2011). *Human dignity*. The Belknap Press of Harvard University Press.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.
- Margalit, A. (1996). *The decent society*. Harvard University Press.
- Nogueira de Brito, M. (2019). Is there any absolute concept of dignity? En M. Nogueira de Brito, R. Herdy, G. Damele, P. Moniz Lopes, & J. Silva Sampaio (Eds.), *The role of legal argumentation and human dignity in constitutional courts*:

Proceedings of the special workshops held at the 28th World Congress of the International Association for Philosophy of Law and Social Philosophy in Lisbon, 2017. Franz Steiner Verlag; Nomos.

Presidência da República do Brasil. (1988). *Constituição da República Federativa do Brasil de 1988.*

R. Manrique Molina, F., Huertas Díaz, O., & Carmona Arias, E. A. (2020). *Hacia una Constitución ecológica en México frente a la crisis ambiental causada por la disrupción corrupta del antropocentrismo. Utopía y Praxis Latinoamericana.*

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Spinoza, B. (2000). *Ética demostrada según el orden geométrico* (A. Domínguez, Ed. y trad.). Editorial Trotta.

Swiss Confederation. (1999). *Federal Constitution of the Swiss Confederation of 18 April 1999.*

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.

UNESCO. (2020). *Recommendation on the ethics of artificial intelligence.*
UNESCO. <https://unesdoc.unesco.org/>

Waldron, J. (2019). *Democratizar la dignidad: Estudios sobre dignidad humana y derechos* (V. F. Benítez R., J. Gallego Saade & L. García Jaramillo, Trans. y eds.). Universidad Externado de Colombia.

CAPÍTULO XXXIV

REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN EMPRESAS: GOBERNANZA, MARCOS INTERNACIONALES Y VACÍOS NORMATIVOS EN MÉXICO.

Regulation of artificial intelligence in companies: Governance, international frameworks and regulatory gaps in Mexico.

Wendy Sebastiana Hernández del Puerto

whernandez@ucaribe.edu.mx

Universidad del Caribe Cancún, México

Maestra en Innovación y Gestión del Aprendizaje

0009-0006-5097-4761

Resumen

El capítulo analiza la regulación de la inteligencia artificial (IA) desde un enfoque de gobernanza sociotécnica, enfatizando la necesidad de articular capacidades organizacionales, estándares internacionales, protección de derechos y mecanismos prácticos de cumplimiento, especialmente en las medianas empresas mexicanas. La IA no se limita a ser una herramienta tecnológica, sino que transforma procesos de decisión, redistribuye responsabilidades y genera nuevas exigencias de transparencia, privacidad, seguridad y rendición de cuentas.

La investigación destaca que la gobernanza de la IA implica supervisar todo el ciclo de vida de los sistemas inteligentes, incluyendo gestión de riesgos, explicabilidad, mitigación de sesgos, protección de datos, supervisión humana y responsabilidad organizacional. Asimismo, se subraya que las empresas medianas enfrentan desafíos particulares debido a limitaciones de infraestructura, financiamiento, capacidades técnicas y cumplimiento regulatorio, lo que hace necesaria una regulación proporcional y adaptable a su nivel de madurez digital.

El texto revisa las principales iniciativas internacionales de regulación y gobernanza de IA, como las recomendaciones de la OCDE y UNESCO, el NIST AI Risk Management Framework, el AI Act de la Unión Europea, las normas ISO/IEC 42001 y 23894, así como marcos de privacidad como GDPR, CCPA y POPIA. Estas iniciativas muestran una evolución desde principios éticos generales hacia modelos operativos de gestión de riesgos, auditoría y supervisión.

En el caso mexicano, se identifica un vacío normativo debido a la inexistencia de una ley integral de IA. Aunque existen disposiciones relacionadas con protección de datos, derechos del consumidor, propiedad intelectual y responsabilidad civil, aún no se regulan de manera específica aspectos como clasificación de riesgos, auditoría algorítmica, explicabilidad y supervisión humana. Finalmente, el capítulo propone avanzar hacia una agenda de gobernanza regulatoria basada en riesgos, gradualidad, fortalecimiento institucional y apoyo a las empresas, con el objetivo de impulsar una adopción responsable y confiable de la IA en México

Palabras clave: inteligencia artificial, gobernanza de IA, regulación tecnológica, gobernanza sociotécnica, medianas empresas, gestión de riesgos, transparencia algorítmica, supervisión humana, protección de datos, IA responsable, ética de la inteligencia artificial, transformación digital y auditoría algorítmica.

Abstract

This chapter analyzes the regulation of artificial intelligence (AI) from a socio-technical governance perspective, emphasizing the need to articulate organizational capabilities, international standards, rights protection, and practical compliance mechanisms, particularly in Mexican medium-sized enterprises. AI is not limited to being a technological tool; rather, it transforms decision-making processes, redistributes responsibilities, and creates new demands for transparency, privacy, security, and accountability.

The study highlights that AI governance involves supervising the entire lifecycle of intelligent systems, including risk management, explainability, bias mitigation, data protection, human oversight, and organizational accountability. It also emphasizes that medium-sized enterprises face particular challenges due to limitations in infrastructure, financing, technical capabilities, and regulatory compliance, making it necessary to establish proportional regulations adaptable to their level of digital maturity.

The text reviews the main international AI governance and regulatory initiatives, such as the OECD and UNESCO recommendations, the NIST AI Risk Management Framework, the European Union AI Act, ISO/IEC 42001 and 23894 standards, as well as privacy frameworks such as GDPR, CCPA, and POPIA. These initiatives demonstrate an evolution from general ethical principles toward operational models of risk management, auditing, and supervision.

In the Mexican context, the chapter identifies a regulatory gap resulting from the absence of a comprehensive AI law. Although there are legal provisions related to data protection, consumer rights, intellectual property, and civil liability, issues such as risk classification, algorithmic auditing, explainability, and human oversight are not yet specifically regulated. Finally, the chapter proposes advancing toward a regulatory governance agenda based on risk assessment, gradual implementation, institutional strengthening, and support for businesses, with the aim of promoting the responsible and trustworthy adoption of AI in Mexico.

Keywords: artificial intelligence, AI governance, technology regulation, socio-technical governance, medium-sized enterprises, risk management, algorithmic transparency, human oversight, data protection, responsible AI, artificial intelligence ethics, digital transformation, and algorithmic auditing.

Introducción

El presente capítulo analiza la regulación de la IA desde el enfoque de gobernanza, analizando los requerimientos de gobernanza, las regulaciones de iniciativas internacionales y los vacíos normativos en México, con énfasis en las medianas empresas. El argumento central sostiene que la adopción responsable de IA requiere articular capacidades organizacionales, estándares internacionales, protección de derechos y mecanismos prácticos de cumplimiento.

La inteligencia artificial (IA) se ha convertido en una infraestructura tecnológica transversal para la transformación digital en las organizaciones públicas y privadas. En las empresas, su incorporación se observa en la atención al cliente, análisis de datos, automatización administrativa, finanzas, marketing, compras, logística, gestión de talento y servicios digitales. Sin embargo, su adopción no puede analizarse únicamente como una decisión tecnológica, porque los sistemas inteligentes modifican procesos de decisión, redistribuyen responsabilidades y generan nuevas exigencias de transparencia, privacidad, seguridad y rendición de cuentas (Leon, 2026; Stein, 2026).

La literatura reciente coincide en que la IA introduce una diferencia sustantiva frente al Software tradicional: aprende a partir de datos, opera bajo incertidumbre probabilística, puede degradarse con el tiempo y requiere supervisión durante todo su ciclo de vida. Esta condición explica por qué la discusión regulatoria internacional ha transitado desde principios generales de ética hacia marcos de gestión, evaluación de riesgos, auditoría, documentación y supervisión humana. En este sentido, la IA no es solo una herramienta de eficiencia, sino un sistema sociotécnico que debe de ser gobernado (Leon, 2026; Kim y otros., 2025).

El enfoque en medianas empresas es relevante porque este tipo de organizaciones enfrentan una atención particular: necesita innovar para sostener competitividad, pero no siempre cuenta con áreas especializadas en cumplimiento normativo, ciencia de datos, auditoría de algoritmos o gestión de riesgos tecnológicos. Sagala y Ori (2024) muestran que las pymes deben considerar su

punto de partida, sus limitaciones y sus características específicas antes de diseñar estrategias de transformación digital. En la misma línea, Omrani y otros. (2022) identifican que la infraestructura tecnológica, las capacidades disponibles, la regulación corporativa, los recursos financieros y las habilidades internas influyen en la adopción digital.

De acuerdo con la OCDE, en el 2024 el uso de IA en empresas grandes alcanzó 40%, mientras que en empresas medianas fue del 20.4% y en pequeñas apenas el 11.1%, lo que evidencia una brecha persistente asociada con las capacidades digitales, acceso a datos, infraestructura, financiamiento y habilidades especializadas. Además, el organismo identifica cuatro habilidades claves para impulsar la adopción de IA en pymes: conectividad, insumos habilitadores de IA como datos, algoritmos y capacidad de cómputo, desarrollo de habilidades y acceso al financiamiento. Esta evidencia refuerza la necesidad de que la regulación y la gobernanza de la IA consideren las condiciones reales de las empresas, ya que una normativa excesivamente compleja podría limitar la innovación mientras que la ausencia de elementos puede profundizar riesgos de uso opaco, inseguridad o desigualdad de estas tecnologías (OCDE, 2025).

En contextos emergentes, la noción de IA se ve atravesada por condiciones institucionales heterogéneas. Boikanyo (2025) advierte que las pymes de mercados emergentes enfrentan dilemas éticos relacionados con privacidad, sesgos, transferencia, empleo, inclusión digital y marcos regulatorios incipientes. Bawa y Asongu (2026), agregan que, en economías emergentes, el impacto de la IA depende de la fortaleza institucional y de la capacidad de las empresas para articular métricas ESG -ambientales, sociales y de gobernanza-, gobernanza de responsabilidad social corporativa y desempeño ambiental. Esto muestra que la tecnología no produce beneficios responsables de manera automática.

Por ello, la regulación de IA debe comprenderse como una arquitectura de gobernanza multinivel: incluye leyes, estándares, guías voluntarias, códigos corporativos, comités internos, procesos de documentación, mecanismos de

auditoría, criterios de explicabilidad y prácticas de participación de actores interesados. La cuestión central no es si las empresas deben adoptar IA, sino bajo qué condiciones organizacionales, regulatorias y éticas pueden hacerlo sin comprometer derechos, confianza y responsabilidad (Leon, 2026).

1. Gobernanza de la IA como enfoque sociotécnico

La gobernanza de la IA se refiere al conjunto de reglas, prácticas, procesos, capacidades y herramientas que permiten dirigir, supervisar y evaluar el diseño, implementación y uso de sistemas inteligentes. Kim y otros. (2025) definen como un sistema de reglas, prácticas, procesos y herramientas tecnológicas orientadas a asegurar el desarrollo, despliegue y uso responsable de la IA. Esta definición es útil porque desplaza el énfasis desde un algoritmo aislado hacia la organización que lo implementa, las personas que interactúan con él y los efectos que produce.

En el foco sociotécnico resulta indispensable porque los riesgos de la IA no emergen únicamente del modelo matemático, sino de la interacción entre datos, infraestructura, decisiones humanas, cultura organizacional, proveedores externos, legislación y alternativas sociales. Leon (2026) sostiene que la confianza en sistemas de IA es una prioridad emergente asociada con robustez, confiabilidad, aplicabilidad, equidad, privacidad, transferencia y responsabilidad. Estas dimensiones deben analizarse en conjunto porque una sola puede ser técnicamente precisa y otra al mismo tiempo, socialmente problemáticas y reproduce sesgos o no permite explicar sus resultados.

En organizaciones medianas, la gobernanza sociotécnica adquiere relevancia de escenario porque hay muchas soluciones de IA son adquiridas con los servicios externos. La empresa usuaria puede no controlar el entrenamiento del modelo, pero si decide su finalidad, su integridad en procesos de negocios y la formación en que sus resultados afectan a clientes, colaboradores o proveedores. Stein (2026) destaca que la adopción de IA generativa en contextos empresariales requiere clasificar los casos de uso por complejidad, autonomía técnica, alineación

estratégica y gobernanza de datos. Esto permite distinguir usos de bajo riesgo, como asistentes internos, de uso más sensibles, como agentes conectados a sistemas empresariales.

Kim y otros. (2025) muestran que la diferencia entre adopción tecnológica y gobernanza es fundamental. El marco TOE permite comprender factores de adopción, pero no establece por sí mismo políticas, estructuras de supervisión y mecanismos de rendición de cuentas para gestionar la tecnología en el tiempo. Esta distinción es crítica para las empresas mexicanas: implementar una herramienta de IA en ventas, recursos humanos o atención a clientes no equivale a contar con controles sobre datos, explica habilidad, supervisión humana o responsabilidad ante errores.

La gobernanza también exige que las organizaciones definan roles. En la experiencia documentada por Kim y otros. (2025), el marco People, procesos, Technology and Operation (PPTO) permitió identificar personas requerido, comités de gobernanza, procesos de revisión, infraestructura técnica, patrocinio ejecutivo, presupuesto y métricas de efectividad. Aunque el estudio se ubica en salud, sus lecciones son trasladables a empresas medianas, sin personas responsables, procesos documentados, tecnologías adecuadas y capacidades operativas, la gobernanza queda reducida a una declaración ética sin capacidad de implementación.

Desde una perspectiva corporativa, Coovadia y otros. (2025) sostienen que la gobernanza ética de IA debe integrarse a los procesos de negocios y alinearse con la gobernanza corporativa. Su estudio identifica cinco elementos esenciales: transparencia, sesgo de máquina, privacidad y IA beneficiosa e IA responsable. Además, subraya la necesidad de vigilancia continua, participación de Stakeholders y cumplimiento con marcos legales como el King IV Code y la protección Off Personal Information Act (POPIA) en Sudáfrica. Este enfoque confirma que la IA debe formar parte de la agenda directiva y no sólo del área técnica.

2. Requerimientos de gobernanza de la IA

El primer requerimiento de gobernanza es la gestión del ciclo de vida. Los sistemas de guía deben supervisarse desde la definición del problema hasta su retiro pasando por adquisición de datos, desarrollo, entrenamiento, validación, despliegue, monitoreo y actualización. León (2026) plantea una gobernanza basada en ciclo de vida porque los sistemas inteligentes cambian los con los datos, el contexto y la retroalimentación. Esta perspectiva evita que la organización trate la IA como una compra puntual y obliga a establecer controles previos, concurrentes y posteriores al despliegue.

En segundo requerimiento es la transparencia explicativa. Low y Zhang (2026) muestran que la adopción de IA personalizada en sistemas TVET depende de factores como confianza, transparencia, riesgo percibido de privacidad, apoyo del instructor y condiciones facilitadoras. Aunque el estudio pertenece al ámbito educativo, su lógica es aplicable a empresas: cuando un sistema de IA recomienda, clasificar o automatizar decisiones los usuarios necesitan comprender cómo se usan los datos, qué variables son relevantes, y qué límites tiene la recomendación. La transparencia no es sólo publicación de información técnica, sino comunicación comprensible para las personas afectadas.

El tercer requerimiento es la responsabilidad y rendición de cuentas. Rahmatullah y Nozlan (2026) muestran que, en instituciones financieras, el aseguramiento de calidad desempeña un papel central para garantizar la implementación responsable, seguridad de cumplimiento regulatorio, detección de sesgos y explica habilidad. Esta evidencia indica que la gobernanza no puede limitarse al cumplimiento formal ya que requiere pruebas, validación, monitoreo, revisión de desempeño y mecanismos para corregir desviaciones muestran que, en instituciones financieras, el aseguramiento de calidad desempeña un papel central para garantizar la implementación responsable, seguridad de cumplimiento regulatorio, detección de sesgos y explicabilidad. En aplicaciones de alto impacto,

la pregunta relevante no es sólo si el sistema funciona, si no quien responde por sus resultados.

El cuarto requerimiento es la protección de datos y privacidad. La IA se alimenta de datos personales, transacciones, biométricos financieros o de complemento. Amarasiriya y Abston (2026) señalan que la adopción ética de IA en pequeñas empresas debe considerar prácticas transparentes y justas en redes y proveedores, así como aducción regulatoria dinámica. Además, mencionan la relevancia de marcos como el GDPR -General Data Protection Regulation- y la CCPA -California Consumer Privacy Act - para orientar privacidad, minimización de datos y protección de información sensible. En medianas empresas, estas exigencias deben traducirse en controles de acceso, anonimización, consentimiento y retención limitada de datos.

El quinto requerimiento es la equidad y mitigación de sesgos. Jaturat, Nan y Hu (2026) validan un Marco multidimensional de responsabilidad de IA compuesto por privacidad y seguridad, transparencia y rendición de cuentas, impacto en el empleo, sostenibilidad, diseño centrado en el usuario, impacto social e innovación y adaptación. Esa visión amplía la gobernanza más allá de la técnica, al incluir efectos laborales, sociales y ambientales. En medianas empresas la equidad es relevante en reclutamiento, evaluación de desempeño, seguimiento de clientes, crédito, precios publicidad y atención automatizada.

El sexto requerimiento es la supervisión humana. La literatura revisada insiste que la IA debe aumentar capacidades humanas, no desplazar de forma opaca el juicio profesional. En compras, Amarasuriya y Abston (2026) fórmula la AI de automatización inteligente y aumento de la decisión para resaltar que la IA debe apoyar, no reemplazar sin control, la decisión humana. En salud, Kim y otros. (2025) pues tranquila gobernanza requiere participación de actores clínicos, técnicos, operativos y experiencia del paciente. En el marketing, Ozturkcan y Bozdağ (2025) advierten que la pérdida de agencia humana puede alimentar desconfianza y rechazo social hacia la IA.

3. IA responsable, confianza y adopción organizacional

La IA responsable se ha convertido en un constructor central para explicar la aceptación de tecnologías inteligentes. Chadrawi y otros. (2025) integran la IA responsable al modelo UTAUT muestra que prácticas como apertura, justicia, privacidad y responsabilidad influyen positivamente en las intenciones conductuales de los usuarios. Este hallazgo es importante para las empresas porque muestra que la adopción no depende únicamente del rendimiento percibido o de la facilidad de uso; también depende de si las personas consideran que la tecnología respeta valores sociales y normas éticas.

Anwer (2026) aporta evidencia desde el uso de ChatGPT en escritura académica al mostrar que la confianza en esta herramienta se relaciona con la percepción de mejora y con la intención futura de Uso. Aunque en el contexto educativo, el resultado ofrece una lección organizacional. La confianza en IA no surge automáticamente de la disponibilidad de la herramienta, sino de la percepción de confiabilidad, utilidad y apoyo institucional. En empresas esto implica que la capacidad, las reglas de uso y la supervisión son condiciones de adopción responsable.

Kumar, Dwivedi y Anand (2023) profundizan esta discusión en el sector salud al conceptualizar la IA y responsable como un factor asociado con formación de valor, compromiso cognitivo del paciente y desempeño del mercado. Su estudio muestra que la IA responsable no sólo cumple con la función ética, sino también estratégica, pues fortalece la percepción de valor y la relación con usuarios. Esta evidencia sugiere que la gobernanza puede entenderse como una inversión de confianza, no únicamente como costo del cumplimiento.

El caso del marketing permite observar como la falta de transparencia erosionada la confianza. Ozturkacan y Bozdağ (2025) introduce los conceptos de IA Washing y IA Booing para explicar el ciclo de desconfianza en IA: las empresas exageran capacidades, los usuarios experimentan resultados opacos o decepcionantes y emerge una relación pública negativa.

Este ciclo es relevante para la regulación porque muestra que la desinformación empresarial sobre capacidades de IA también es un riesgo. No basta con regular sistemas técnicos, también debe regularse las afirmaciones comerciales engañosas sobre IA.

La confianza también se vincula con el contexto institucional. Yigitcanlar y otros. (2026) muestran que la adopción de IA en gobiernos locales y ciudades inteligentes depende de configuraciones entre capacidad tecnológica, preparación regulatoria, orientación comunitaria y reconocimiento de restricciones de implementación. Su análisis evidencia que no existe una única ruta de opción: algunas ciudades avanzan por infraestructura tecnológica, otros por mandatos institucionales y otras para equilibrio entre tecnología política y comunidad. Para México esta idea resulta útil porque empresas y gobiernos locales tendrían capacidades desiguales.

En síntesis, la IA responsable se sostiene por tres capas: confianza del usuario, gobernanza organizacional y legitimidad institucional (Anwer, 2026; Chedrawi y otros., 2025; Kumar y otros., 2023; Ozturkcan y Bozdağ, 2025) coinciden en que la aceptación de IA depende de la percepción de beneficio, pero también de transparencia, privacidad, responsabilidad y alineación con valores humanos. Por ello la regulación eficaz debe incentivar prácticas que fortalezcan confianza y no únicamente imponer sanciones posteriores a fallas.

4. Regulaciones de iniciativas internacionales identificadas en la literatura

A continuación, se mencionan un conjunto amplio de regulaciones, estándares e iniciativas que han configurado en la gobernanza internacional de la IA. Leon (2026) describe una evolución desde principios éticos generales así instrumentos más operativos: la recomendación de la OCDE sobre IA, la recomendación de UNESCO sobre la ética de la IA, el NIST AI Risk Management Framework 1.0, la Executive Order 114110 de Estados Unidos, IA Act europeo, el NIST Generative AI Profile, ISO/IEC 42001, ISO/IEC 23894 y el IEEE Ethicaly

Aligned Desing. En conjunto, estos instrumentos muestran que la gobernanza moderna combina principios, gestión de riesgos, documentación y responsabilidades.

La recomendación de la OCDE sobre IA, adaptada en 2019 se orienta a promover sistemas que benefician a las personas y el planeta, respeten derechos humanos democracia Estado de derecho transparencia robustez y rendición de cuentas. La UNESCO, por su parte aprobó en el 2021 un instrumento normativo global sobre ética de prácticas como sociales scoring y vigilancia masiva, así como gobernanza de datos y equidad (Leon, 2026).

El National Institute of Standards and Technology -NIST- presenta el AI Risk Management Framework 1.0 -RMF- representa una transición hacia herramientas operativas de gestión. Leon (2026) explica que este Marco propone actividades de gobernar, mapear, medir y gestionar riesgos, facilitando que las organizaciones traduzcan principios en prácticas. Kim y otros (2025) también señala que el NIST AI RMF ha ganado reconocimiento internacional como un Marco para evaluar riesgos de IA, aunque advierten que los marcos generales deben adaptarse a contextos organizacionales completos para convertirse en políticas aplicables.

La ley de inteligencia Artificial de la Unión Europea (AI Act), Regulation (EU) 2024/1689, constituye uno de los referentes más importantes por su enfoque basado en riesgos siendo la primera normativa integral del mundo sobre IA (Future of Life, s.f). Leon (2026) señala que este instrumento impone obligaciones diferenciadas para sistemas de alto riesgo, incluyendo gestión de riesgos, gobernanza de datos, documentación, supervisión humana, robustez y transparencia, así como un código de buenas prácticas y espacios aislados de regulación de IA. Su valor para México radica en mostrar que la regulación no debe tratar sólo de usos veía por igual ya que es una herramienta de apoyo interno no debe recibir el mismo tratamiento que un sistema de decisión sobre empleo, salud, crédito o derechos.

Las normas ISO/IEC 42001 e ISO/IEC 23894 representan el avance hacia sistemas de gestión. ISO/IEC 42001 define requisitos para establecer, implementar,

mantener y mejorar un sistema de gestión de IA; ISO/IEC 23894 orienta la gestión de riesgos específicos de IA. Estas normas, son relevantes ya que pueden convertirse en guías de madurez: liderazgo, objetivos, competencias, documentos, evaluación de desempeño, tratamiento de riesgos y mejora continua.

El NIST Generative AI Profile, aborda riesgos propios de la IA generativa como alucinaciones, filtración de privacidad, contenido engañoso, procedencia de información, impacto ambiental y protección de datos. Stein (2026) complementa esta discusión al clasificar los usos empresariales de IA generativa por niveles de complejidad: asistentes de trabajo, generación automatizada de código, generación integrada en sistemas y uso de herramientas o agentes. A mayor autonomía y conexión con sistemas empresariales, mayor debe ser la exigencia de gobernanza.

El enfoque IEEE Ethically Aligned Design plantea que los sistemas autónomos e inteligentes deben desarrollarse priorizando el bienestar humano, la dignidad, los derechos fundamentales y la sostenibilidad, por encima de la visión puramente técnica o instrumental de IA. Esta iniciativa resulta relevante para la gobernanza organizacional porque propone que las consideraciones éticas se incorporen desde el diseño y durante todo el ciclo de vida de la tecnología, orientando a empresas, gobiernos, y desarrolladores hacia prácticas de transparencia, responsabilidad y control humano (IEEE, 2019).

En materia de protección de datos, los artículos mencionan el GDPR europeo y la CCPA como referencias para tratamiento responsable de información personal y transaccional. Amarasuriya y Abston (2026) nos vinculan con compras responsables e IA en pymes, resaltando la necesidad de privacidad, minimización de datos, transparencia y seguridad. Coovadia y otros. (2025) agregan, la importancia de POPIA como marco de protección de información personal y del King IV Code como referente de gobierno corporativo aplicado a decisiones tecnológicas.

En el sector salud, Kim y otros. (2025) mencionan la Algorithmic Impact Assessment del gobierno de Canadá y el NIST AI RMF como herramientas de evaluación de riesgo. Sin embargo, destacan que los marcos generales no siempre

ofrecen instrucciones suficientes para establecer estructuras de gobernanza organizacional por ello proponen el uso de PPTO, qué garantiza capacidades en personas, procesos, tecnologías y operaciones. Esta iniciativa evidencia que los instrumentos internacionales requieren traducción institucional para hacer efectivos.

En el sector de finanzas Rahmatulh y Nozlan (2026) discuten la adopción de IA en instituciones y la función de aseguramiento de calidad para la seguridad, cumplimiento y transferencia. En este contexto aparecen referencias a lineamientos regulatorios del sector financiero de Malasia, la Personal Data Protection Act (PDPA), la gestión de riesgos tecnológicos (RMIT), el National Fraud Portal inppusado por Bank Negara Malaysia y debate sobre el cumplimiento transfronterizo. Estas iniciativas muestran que la regulación sectorial debe considerar datos sensibles, estabilidad financiera, sesgos, inclusión y seguridad.

Regulación ambiental y política pública, Wang, Wilchek y Batarseh (2026) muestran que la IA puede apoyar el cumplimiento normativo en agua mediante detección de violaciones, monitoreo, optimización de infraestructura y sistemas de decisión. Donde se mencionan instrumentos que muestra la regulación basada en datos y los requerimientos de infraestructura de información confiable. Sin embargo, ellos advierten que, pese a los avances técnicos, sólo una proporción limitada de investigaciones incorpora elementos de aseguramiento como explicabilidad, equidad o confiabilidad. Se destaca que la IA puede modernizar la regulación, pero también puede amplificar problemas de opacidad si no se integra con controles de calidad, evaluación de impacto y mecanismos de transparencia.

En el contexto de economías emergentes, Sarkambayeva y otros. (2026) analiza en el caso de Kazajistán y mencionan la relevancia de estrategias nacionales de IA, regulaciones de localización de datos, ciberseguridad, arquitecturas modulares y privacidad. Su discusión muestra que la regulación de IA no debe separarse de la infraestructura, talento, capacidades sectoriales y soberanía de datos. Para México esto empírica que la regulación debe dialogar con

transformación digital la ciber seguridad y las capacidades de cómputo, así como la capacitación empresarial y el uso responsable de datos.

5. Regulaciones, empresas y mercados emergentes.

Evidencia sobre pymes y mercados emergentes muéstrame la regulación debe ser proporcional a las capacidades empresariales. Boikanyo (2025) sostiene que la adopción de IA en pymes de mercados emergentes se ubica entre oportunidades de competitividad y riesgos de privacidad, sesgo, transparencia, exclusión digital y desplazamiento laboral. El estudio enfatiza que la ética no debe asegurarse al final del proceso, sino incorporarse en la estrategia de adopción desde el inicio, con liderazgo responsable y capacitaciones y participación de los Stakeholders. Por su parte Amarasuriya Abston (2026) proponen pensar en la adopción ética de la IA en compras de tres pilares: automatización inteligente y aumento de decisiones, ecosistemas éticos de compras y adaptación regulatoria dinámica. Esta perspectiva resulta útil para reconocer que la IA debe mejorar la capacidad humana de decisión, y no sustituirlas sin supervisión. Además, conecta la responsabilidad con cadena de suministro, proveedores, transparencia, sostenibilidad y cumplimiento de normas cambiantes.

Sagala y Ori (2024) sostienen que la transformación digital de pymes requiere partir de sus limitaciones, adoptar estrategias incrementales e invertir en educación y aprendizaje continuo. Si las obligaciones son demasiado complejas, muchas empresas no podrán cumplirlas si son demasiado laxas, se expandirán usos opacos. Una regulación efectiva debe considerar madurez digital, recursos disponibles y gradualidad, sin renunciar a la protección de derechos.

Omani y otros. (2022) muestra que los impulsores de adopción digital en pymes incluyen infraestructura tecnológica, herramientas digitales, innovación, regulación corporativa, habilidades y recursos financieros. Esto confirma que la regulación de IAs no puede limitarse a prohibiciones o sanciones, sino que debe crear condiciones de adopción responsable. Las empresas necesitan capacidades

técnicas, capital humano, financieros lineamientos sectoriales y herramientas de autodiagnóstico para implementar la IA de forma segura.

Bawa y Asongu (2026) aportan un enfoque relevante desde la sostenibilidad y la relación entre IA, métricas ESG y desempeño ambiental depende de la gobernanza de responsabilidad social corporativa y de la fortaleza institucional. Esto sugiere que, en contextos emergentes, la regulación debe promover tanto innovación como gobernanza corporativa. La IA puede contribuir a biodiversidad, economía circular y trazabilidad, pero sólo si las empresas cuentan con mecanismos para convertir información Algoritmica en decisiones responsables.

Coovadia y otros. (2025) muestran que la gobernanza corporativa de IA puede integrarse con códigos de buen gobierno y obligaciones de protección de información. Su marco identifica transparencia, sesgos de máquina, privacidad, IA benefiosa e IA responsable como elementos esenciales.

6. Vacíos normativos en México

México se encuentra en una etapa de construcción regulatoria en materia de IA. A la fecha, no existe una ley integral plenamente vigente dedicada exclusivamente a regular la IA en el país, aunque existen marcos legales aplicables de manera indirecta como la protección de datos personales, la propiedad intelectual, derechos del consumidor, responsabilidad civil, derechos laborales, ciber seguridad y de comercio electrónico (CMS, 2016). Esta situación genera incertidumbre para las empresas que desean implementar inteligencia artificial porque las obligaciones específicas sobre explicabilidad, clasificación de riesgos, auditoría algoritmica y supervisión humana no están claramente articuladas.

La discusión legislativa mexicana avanzado mediante iniciativas que buscan establecer una ley general para regular y fomentar el uso de la IA. En 2026 se reportó que el Senado analiza propuestas orientadas a crear un marco integral con estrategia nacional, clasificación de riesgos, instancias de coordinación y obligaciones para que nos desarrollen, fortalezcan o utilicen sistemas de IA

(Basham, 2026). Sin embargo, el tránsito de iniciativa a ley aplicable aún requiere definición institucional, capacidades de supervisión y claridad sobre obligaciones diferenciadas.

La principal brecha frente a Marcos como IA Act europeo es la ausencia de una clasificación normativa de sistemas por nivel de riesgo. Sin esa clasificación una empresa podría tratar de manera equivalente un asistente de redacción interna, un sistema de selección de personal, una herramienta de Pricing dinámico o un sistema de selección de análisis crediticio. La experiencia internacional muestra que la regulación debe diferenciar usos por impacto, autonomía, sensibilidad de datos y la posibilidad de afectar derechos (Leon, 2026; Stein, 2026).

Otra brecha importante es la falta de mecanismos de evaluación de impacto algorítmico. La experiencia canadiense con Algorithmic Impact Assessment, mencionada por Kim y otros. (2025), muestra que las herramientas de evaluación pueden ayudar a clasificar riesgos y definir medidas de mitigación. En México, una futura regulación podría considerar evaluaciones obligatorias para sistemas de alto impacto, especialmente en el empleo salud, educación, servicios financieros es la seguridad, movilidad y administración pública.

También existe una brecha de capacidades, las empresas pueden tener incentivos para adoptar IA generativa, analítica predictiva o automatización de procesos, pero no necesariamente cuentan con comités de gobernanza, responsables de datos, políticas de uso, procedimientos de auditoría o criterios para la selección de proveedores. Kim y otros. (2025) muestran que las políticas deben ser simples, implementables y proporcionales para no frenar la innovación. Esta lección es clave para México que necesita lineamientos que sean exigentes pero aplicables a organizaciones con distintos niveles de madurez,

La ausencia de regulación específica también incrementa el riesgo de adopción informal o shadow IA. Cuando el personal utiliza herramientas generativa sin autorización ni reglas internas, pueden generarse filtraciones de información, uso indebido de datos personales, errores no detectados, plagio, sesgos o dependencia de proveedores. Ozturkcan y Bozdağ (2025) advierten que la

exageración de capacidades y la opacidad pueden producir ciclos de desconfianza. Es por ello que las empresas necesitan políticas internas incluso antes de que exista una ley integral.

7. Hacia una agenda de gobernanza regulatoria para México

El análisis de los artículos revisados permite afirmar que la regulación mexicana de IA debería avanzar sobre tres planos complementarios. El primero en el plano normativo, orientado a definir principios, niveles de riesgo como obligaciones de transparencia, supervisión humana, gestión de datos, auditoría y responsabilidad. El segundo es el plano organizacional, orientado a qué empresas y entidades públicas implementen políticas, comités documentación, controles y capacitación y el tercero en el plano institucional, orientado a crear capacidades estatales que supervisen color de supervisión coordinación, evaluación técnica y actualización regulatoria.

Se debe evitar dos extremos el primero no regular y dejar que la IA avance mediante decisiones dispersas de mercado, lo cual aumentaría riesgos de privacidad, sesgos, discriminación y responsabilidad difusa y el segundo sería sobre regularse considerar capacidades de empresas lo cual podría frenar la innovación o fomentar complemento simbólico.

Las regulaciones e iniciativas internacionales ofrecen aprendizajes claros. OCDE y UNESCO aportan principios; NIST e ISO ofrecen gestión de riesgos y sistemas de gestión; el AI Act europeo introduce clasificación por riesgo; GDPR y CCPA refuerzan muestra privacidad; PPTO demuestra cómo traducir principios en organización; King IV y POPIA muestran integración con gobernanza corporativa; evidencia de lo importante la importancia de datos regulares; las estrategias nacionales que muestran la regulación deben vincularse con infraestructura y talento. México puede construir una ruta propia toman los referentes sin copiarlos mecánicamente.

Conclusiones

La regulación de la inteligencia artificial en empresas mexicanas debe entenderse como un problema de gobernanza. La literatura revisada muestra que la IA responsable exige transparencia, explicabilidad, privacidad, equidad, seguridad, supervisión humana, auditoría y gestión durante todo el ciclo de vida de la IA. Estos elementos no son accesorios éticos, sino condiciones para que la adopción genere confianza y valor sin comprometer derechos ni legitimidad organizacional.

Los marcos internacionales han avanzado hacia instrumentos más operativos la OCDE y UNESCO establecieron marcos éticos; NIST, ISO e IEEE tradujeron principios en gestión de riesgos y estándares; el AI Act europeo consolidó la clasificación por riesgo; NIST Generative IA Profile respondió a riesgos de IA generativa; GDPR, CCPA y POPIA reforzaron privacidad; y las herramientas sectoriales evidenciaron que la gobernanza requiere datos, monitoreo, auditoría y participación de actores. Este conjunto de iniciativas constituye un repertorio útil para México.

En México, el vacío normativo no significa ausencia total de normas relacionadas, sino falta de articulación específica para IA. La regulación vigente sobre datos personales, consumo, propiedad intelectual, trabajo y responsabilidad civil ofrecen puntos de apoyo, pero no resuelven por completo la explicabilidad algorítmica, la clasificación de riesgos, la auditoría, la trazabilidad, la supervisión humana o la responsabilidad distribuida entre desarrolladores, proveedores y usuarios. Por ello se requiere un marco flexible, gradual y basado en riesgos acompañado de guías para las empresas.

Finalmente, la adopción de IA no debe plantearse como una carrera tecnológica sin controles ni como una amenaza queda bloquearse. Debe asumirse como una oportunidad regulada de innovación responsable. Para ello las empresas necesitan políticas internas, capacitación, gobernanza de datos, evaluación de proveedores, transferencia frente a usuarios y mecanismos de revisión. El estado

por su parte, debe proveer reglas claras, coordinación institucional, incentivos y herramientas prácticas para que la IA se convierta en un instrumento de competitividad con responsabilidad social.

Referencias

- Amarasuriya, S., y Abston, K. A. (2026). Procurement 4.0: A Three-Pillar Framework for Ethical AI Adoption in Small Businesses. *Journal of Small Business Strategy*, 36(2), 27–37. Scopus. <https://doi.org/10.53703/001c.157792>
- Anwer, B. (2026). Trust in ChatGPT and Perceived Academic Writing Improvement: A TAM-based Quantitative Study in a ESL Context. *International Journal of Technology in Education and Science*, 10(1), 162–177. Scopus. <https://doi.org/10.46328/ijtes.5282>
- Bawa, S., y Asongu, S. A. (2026). Investigating How Firms in Emerging Economies Integrate AI-Driven Systems and ESG Metrics Into Biodiversity Conservation and Circular Economy Strategies. *Business Strategy and the Environment*, n/a(n/a). <https://doi.org/10.1002/bse.70547>
- Boikanyo, D. H. (2025). ETHICAL CONSIDERATIONS AND SOCIETAL IMPACTS OF AI ADOPTION IN SMEs WITHIN EMERGING MARKETS. *Business Management*, 2025(2), 25–46. Scopus. <https://doi.org/10.58861/tae.bm.2025.2.02>
- Chedrawi, C., Haddad, G., Tarhini, A., Osta, S., y Kazoun, N. (2025). Exploring the impact of responsible AI usage on users' behavioral intentions. *Journal of Innovation and Knowledge*, 10(6). Scopus. <https://doi.org/10.1016/j.jik.2025.100813>
- CMS. (2026). AI laws and regulations in Mexico. CMS Expert Guide. <https://cms.law/en/int/expert-guides/ai-regulation-scanner/mexico>
- Coovadia, H., Marx, B., Botha, I., y Gold, N. O. (2025). Building an ethical artificial intelligence corporate governance framework for the integration of emerging technologies into business processes. *South African Journal of Accounting Research*, 39(3), 286–316. Scopus. <https://doi.org/10.1080/10291954.2025.2523661>
- Future of Life Institute. (s. f.). El Explorador de la Ley de IA. Artificial Intelligence Act. Recuperado el 30 de abril de 2026, de <https://artificialintelligenceact.eu/es/ai-act-explorer/>

- Jaturat, N., Na-Nan, K., y Hu, B. (2026). Measuring AI responsibility: A cross-country validation of a multidimensional framework. *International Journal of Information Management Data Insights*, 6(1), 100388.
<https://doi.org/10.1016/j.ijime.2026.100388>
- Kim, J. Y., Hasan, A., Kueper, J., Tang, T., Hayes, C., Fine, B., Balu, S., y Sendak, M. (2025). Establishing organizational AI governance in healthcare: A case study in Canada. *npj Digital Medicine*, 8(1), 522. <https://doi.org/10.1038/s41746-025-01909-3>
- Kumar, P., Dwivedi, Y. K., y Anand, A. (2023). Responsible Artificial Intelligence (AI) for Value Formation and Market Performance in Healthcare: The Mediating Role of Patient's Cognitive Engagement. *Information Systems Frontiers*, 25(6), 2197–2220. Scopus. <https://doi.org/10.1007/s10796-021-10136-6>
- Leon, M. (2026). Lifecycle-Based Governance to Build Reliable Ethical AI Systems. *Systems Research and Behavioral Science*. Scopus.
<https://doi.org/10.1002/sres.70014>
- Low, J., y Zhang, X. (2026). Explainable Multi-View Modeling of AI-Driven Personalized Learning Adoption in TVET Systems. *Journal of Logistics, Informatics and Service Science*, 13(4), 125–135. Scopus. <https://doi.org/10.33168/JLISS.2026.0408>
- OECD (2025), *AI adoption by small and medium-sized enterprises: OECD discussion paper for the G7*, OECD Publishing, Paris, <https://doi.org/10.1787/426399c1-en>.
- Ozturkcan, S., y Bozdağ, A. A. (2025). Responsible AI in Marketing: AI Booming and AI Washing Cycle of AI Mistrust. *International Journal of Market Research*, 67(6), 696–722. Scopus. <https://doi.org/10.1177/14707853251379285>

- Rahmatullah, T., y Nozlan, N. N. (2026). Exploring AI Integration in Financial Institutions: Challenges, Compliance, and Quality Assurance's Role in Security. *Journal of Central Banking Law and Institutions*, 5(1), 175–204. Scopus.
<https://doi.org/10.21098/jcli.v5i1.445>
- Sagala, G., y Ori, D. (2024). Toward SMEs digital transformation success: A systematic literature review. *Information Systems and e-Business Management*, 22, 667–719. <https://doi.org/10.1007/s10257-024-00682-2>
- Sarkambayeva, S., Singh, S., Mukhanova, G., Amralinova, B., y Turegeldinova, A. (2026). Artificial Intelligence and Business Process Management: A Responsible Framework for Sustainable Transformation. *Emerging Science Journal*, 10(1), 448–475. <https://doi.org/10.28991/ESJ-2026-010-01-022>
- Stein, H. (2026). Toward a taxonomy of generative AI use cases in business contexts: Integrating complexity, risk, and strategy. *Journal of Computational Science*, 96, 102826. <https://doi.org/10.1016/j.jocs.2026.102826>
- The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (1st ed.). IEEE.
<https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html>
- Wang, Y., Wilchek, M., y Batarseh, F. A. (2026). From data to policy: A systematic review of AI in water regulations and compliance. *npj Clean Water*, 9(1), 33.
<https://doi.org/10.1038/s41545-026-00555-w>

Yigitcanlar, T., Liu, K., Senadheera, S., Marasinghe, R., David, A., Cheong, P. H., y
Corchado, J. (2026). Configurational pathways to smart city AI Adoption: Evidence
from local governments in Australia, Hong Kong, Saudi Arabia, Spain, and the
United States. *Cities*, 175, 107117. <https://doi.org/10.1016/j.cities.2026.107117>

CAPÍTULO XXXV

**REGULACIÓN DE LA IA EN MÉXICO EN LA
PRÁCTICA DOCENTE Y DE INVESTIGACIÓN
UNIVERSITARIA**

Regulation of AI in Mexico in university teaching and research practice

Arcadio Zebadúa Sánchez

zebadua@unach.mx

Universidad Autónoma de Chiapas, México

<http://orcid.org/0009-0002-4070-1525>

Zoily Mery Cruz Sánchez

zcruz@unach.mx

Universidad Autónoma de Chiapas, Mexico

<http://orcid.org/0009-0005-3222-007X>

Karla Beatriz García Arteaga

karla.garcia@unach.mx

Universidad Autónoma de Chiapas, México

<http://orcid.org/0009-0000-3649-0046>

Consuelo Guadalupe Morales Flores

consuelo.morales@unach.mx

Universidad Autónoma de Chiapas, México

<http://orcid.org/0009-0009-4741-6904>

Cruz Susana Estrada Castellanos

susana.estrada@unach.mx

Universidad Autónoma de Chiapas, México

<http://orcid.org/0009-0004-7699-5515>

Resumen

Se propone un análisis académico sobre la regulación de la inteligencia artificial (IA) en México, con énfasis en su uso en la práctica docente y en la investigación universitaria. Se parte de un problema central: las instituciones de educación

superior incorporan herramientas de IA, especialmente de IA generativa, con mayor rapidez que sus marcos normativos. Esta brecha genera oportunidades pedagógicas y científicas, pero también riesgos relacionados con la integridad académica, la protección de datos personales, la transparencia de los procesos, la discriminación algorítmica, la propiedad intelectual, la trazabilidad de evidencias y la responsabilidad institucional. Frente a ello, se plantea que la regulación universitaria no debe entenderse únicamente como prohibición o control, sino como gobernanza ética, proporcional, preventiva y orientada al aprendizaje.

Se fundamenta en una revisión documental de instrumentos internacionales y nacionales relevantes, entre ellos la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO (UNESCO, 2021), el Reglamento de Inteligencia Artificial de la Unión Europea (European Parliament and Council of the European Union, 2024), los Principios de Chapultepec (SECIHTI & ATDT, 2026), la evaluación RAM de la UNESCO para México (UNESCO, 2024), la legislación mexicana en materia de protección de datos personales (Cámara de Diputados, 2025) y recomendaciones universitarias recientes para el uso de IA generativa (GAIA-GEN UNAM, 2025).

Palabras clave: Inteligencia artificial; regulación; educación superior; ética académica; gobernanza institucional.

Abstract

This chapter analyzes the regulation of artificial intelligence (AI) from a socio-technical governance perspective, emphasizing the need to articulate organizational capabilities, international standards, rights protection, and practical compliance mechanisms, particularly in Mexican medium-sized enterprises. AI is not limited to being a technological tool; rather, it transforms decision-making processes, redistributes responsibilities, and creates new demands for transparency, privacy, security, and accountability.

The study highlights that AI governance involves supervising the entire lifecycle of intelligent systems, including risk management, explainability, bias mitigation, data protection, human oversight, and organizational accountability. It also emphasizes that medium-sized enterprises face particular challenges due to limitations in infrastructure, financing, technical capabilities, and regulatory compliance, making it necessary to establish proportional regulations adaptable to their level of digital maturity.

The text reviews the main international AI governance and regulatory initiatives, such as the OECD and UNESCO recommendations, the NIST AI Risk Management Framework, the European Union AI Act, ISO/IEC 42001 and 23894 standards, as well as privacy frameworks such as GDPR, CCPA, and POPIA. These initiatives demonstrate an evolution from general ethical principles toward operational models of risk management, auditing, and supervision.

In the Mexican context, the chapter identifies a regulatory gap resulting from the absence of a comprehensive AI law. Although there are legal provisions related to data protection, consumer rights, intellectual property, and civil liability, issues such as risk classification, algorithmic auditing, explainability, and human oversight are not yet specifically regulated. Finally, the chapter proposes advancing toward a regulatory governance agenda based on risk assessment, gradual implementation, institutional strengthening, and support for businesses, with the aim of promoting the responsible and trustworthy adoption of AI in Mexico.

Keywords: artificial intelligence, AI governance, technology regulation, socio-technical governance, medium-sized enterprises, risk management, algorithmic transparency, human oversight, data protection, responsible AI, artificial intelligence ethics, digital transformation, and algorithmic auditing.

Introducción

La regulación de la inteligencia artificial se ha convertido en una discusión prioritaria para los sistemas educativos, científicos y gubernamentales. En la educación superior, la IA ya no se limita a laboratorios especializados de computación o ingeniería; ahora aparece en plataformas de escritura asistida, sistemas de retroalimentación, buscadores académicos, traductores, generadores de imágenes, tutores conversacionales, gestores de referencias, programas de revisión de similitud, herramientas de apoyo metodológico, etc. Su presencia altera las formas de enseñar, aprender, evaluar, investigar y administrar procesos universitarios. Por ello, hablar de regulación de la IA en México desde la práctica docente y de investigación implica analizar no solo el derecho positivo vigente, sino también las decisiones institucionales que toman las universidades para orientar el uso responsable de estas tecnologías.

El punto de partida de este capítulo es que México vive una etapa de transición regulatoria. A diferencia de la Unión Europea, que aprobó un reglamento integral basado en categorías de riesgo, México aún no cuenta con una ley general de inteligencia artificial vigente de alcance nacional. Sin embargo, ello no significa ausencia normativa. La IA ya se encuentra alcanzada por marcos existentes: derechos humanos, protección de datos personales, transparencia, propiedad intelectual, responsabilidades administrativas, ética de la investigación, normas de integridad académica y lineamientos internos de instituciones educativas. Además, se observan iniciativas legislativas, declaraciones de principios y diagnósticos internacionales que buscan ordenar el debate nacional. La discusión mexicana se encuentra, por tanto, en una fase de construcción institucional, donde las universidades pueden desempeñar un papel de discusión y propuestas.

La docencia universitaria enfrenta un desafío concreto: ¿cómo permitir que estudiantes y profesores utilicen IA como herramienta de apoyo sin que se pierda la autoría intelectual, el aprendizaje significativo, la evaluación auténtica y la confianza académica? La respuesta no puede reducirse a detectar o sancionar el

uso de IA. La experiencia reciente muestra que los detectores automáticos son limitados, pueden producir falsos positivos y no resuelven el problema pedagógico de fondo. La regulación universitaria debe desplazar el centro de atención desde la sospecha hacia el diseño de actividades, la transparencia, la trazabilidad, la alfabetización crítica y la corresponsabilidad. En otras palabras, no basta con preguntar si se usó IA; es necesario preguntar para qué se usó, con qué límites, con qué datos, bajo qué criterios de validación y con qué evidencia del aprendizaje humano.

La investigación universitaria presenta una tensión semejante. La IA puede apoyar la búsqueda bibliográfica, la organización de datos, la traducción, la revisión de estilo, la visualización de información y la exploración preliminar de patrones. Sin embargo, también puede introducir sesgos, inventar referencias, exponer datos confidenciales, distorsionar resultados o dificultar la rendición de cuentas sobre los procedimientos. El problema se agrava cuando se trabaja con información sensible, entrevistas, datos personales, expedientes académicos, registros institucionales o comunidades en situación de vulnerabilidad. La investigación asistida por IA requiere lineamientos de consentimiento informado, anonimización (desvinculación de identidad), auditoría, resguardo de datos, declaración metodológica y revisión por comités de ética.

En el presente trabajo se sostiene que la regulación de la IA en la educación superior mexicana debe pensarse como un sistema de gobernanza multinivel. En el nivel internacional se encuentran recomendaciones, principios y estándares; en el nivel nacional, leyes e iniciativas; en el nivel institucional, políticas universitarias; y en el nivel de aula o proyecto de investigación, reglas operativas y decisiones pedagógicas concretas. La articulación entre esos niveles permite evitar dos extremos: la prohibición absoluta, que puede resultar impracticable e inhibir la innovación, y la adopción acrítica, que puede reproducir desigualdades, violar derechos o debilitar la calidad académica.

Con base en lo anterior, el objetivo es identificar los principales referentes normativos y éticos, así como, analizar los riesgos y oportunidades de la IA en docencia e investigación y proponer criterios de uso aceptable.

La pregunta orientadora es la siguiente: ¿qué elementos debe contener una regulación universitaria de IA en México para equilibrar innovación académica, protección de derechos, integridad científica y responsabilidad institucional? La respuesta que se propone no pretende agotar la complejidad del tema, sino ofrecer una base práctica para que las universidades, facultades, programas educativos y cuerpos académicos diseñen lineamientos contextualizados.

Planteamiento del problema

La expansión de la IA generativa ha producido una aceleración normativa. Las universidades han tenido que responder a herramientas que pueden producir textos, códigos, imágenes, ejercicios, esquemas, resúmenes, traducciones y análisis preliminares en segundos. Esta capacidad modifica el sentido de tareas tradicionales: ensayos, reportes de lectura, revisiones bibliográficas, protocolos, exámenes, proyectos finales y artículos académicos. Una tarea que antes funcionaba como evidencia directa de lectura, comprensión o redacción ahora puede ser parcialmente automatizada. La consecuencia no es necesariamente negativa, pero sí exige rediseñar los criterios de evaluación y los acuerdos de integridad académica.

En la práctica docente, el problema regulatorio se manifiesta en varias preguntas recurrentes. ¿Puede un estudiante usar IA para redactar una introducción? ¿Debe declarar el uso de un asistente IA? ¿Se permite usar IA para traducir un artículo? ¿Es válido emplearla para generar preguntas de autoevaluación? ¿Puede un docente utilizar IA para elaborar rúbricas, reactivos o comentarios de retroalimentación? ¿Qué sucede si la herramienta procesa datos de estudiantes sin autorización? ¿Quién responde si la IA produce información falsa que se incorpora a material de clase? Estas preguntas muestran que la regulación debe distinguir

entre usos de bajo riesgo, usos permitidos con declaración, usos condicionados y usos prohibidos.

En la investigación, los dilemas son más complejos. Un investigador puede utilizar IA para organizar una matriz bibliográfica, depurar transcripciones, generar categorías iniciales o revisar coherencia argumentativa. Sin embargo, si utiliza una plataforma externa con datos no anonimizados (información personal que ha sido alterada, eliminada o generalizada para eliminar la relación entre un conjunto de datos y la persona física a la que pertenecen), de otro modo, podría comprometer la confidencialidad de participantes. Si emplea IA para generar citas, puede introducir referencias inexistentes. Si delega interpretación de resultados sin validación humana, puede debilitar la trazabilidad metodológica. Si usa IA para revisar propuestas de terceros, puede vulnerar confidencialidad académica. De ahí que la regulación tenga que incorporar principios de transparencia metodológica y responsabilidad humana indelegable.

La situación mexicana tiene además una dimensión institucional. Muchas universidades carecen de políticas específicas de IA o cuentan con orientaciones generales no vinculantes. En consecuencia, los criterios pueden variar entre profesores, asignaturas, programas y comités. Esa variabilidad genera incertidumbre para estudiantes y docentes. Mientras algunos cursos permiten el uso de IA sin mayores restricciones, otros lo consideran falta académica. En algunas investigaciones se permite apoyo en redacción, pero no se establece cómo declararlo. Esta heterogeneidad no es necesariamente incorrecta, pues las disciplinas tienen necesidades distintas; sin embargo, se requiere un piso común de derechos, responsabilidades, criterios de evaluación y protección de datos.

El reto no es solo técnico o legal, sino pedagógico. Regular la IA en la universidad implica decidir qué se entiende por aprendizaje, autoría, evidencia, desempeño, creatividad, originalidad, acompañamiento docente. La IA puede fortalecer la inclusión cuando ayuda a estudiantes con barreras lingüísticas, discapacidad, rezago académico o limitaciones de tiempo. También puede ampliar brechas si solo

algunos estudiantes tienen acceso a herramientas de pago o si no todos desarrollan competencias para evaluar sus resultados. Por ello, la regulación debe incorporar criterios de equidad, accesibilidad y justicia educativa.

El problema central que se aborda puede formularse de la siguiente manera: la adopción acelerada de IA en la docencia y la investigación universitaria en México ocurre en un contexto de regulación nacional incompleta, lineamientos institucionales desiguales y riesgos académicos emergentes. Esta situación exige una propuesta de gobernanza que traduzca principios éticos y normativos en prácticas concretas de aula, evaluación e investigación.

Marco normativo y de referencia

El marco de referencia para regular la IA en universidades mexicanas debe integrar instrumentos de distinto nivel. En el ámbito internacional, la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO constituye un referente central porque propone valores, principios y ámbitos de acción política para orientar el desarrollo y uso de sistemas de IA (UNESCO, 2021). Su importancia radica en que no se limita a una declaración abstracta, sino que vincula la ética con políticas públicas, derechos humanos, inclusión, medio ambiente, gobernanza de datos, educación, investigación y responsabilidad social. Para la universidad mexicana, este instrumento permite justificar que la IA debe emplearse al servicio de la dignidad humana, la justicia, la diversidad y el bien público.

El Reglamento de Inteligencia Artificial de la Unión Europea aporta otro elemento clave: la clasificación de sistemas de IA según niveles de riesgo (European Parliament and Council of the European Union, 2024). Aunque no es directamente aplicable a México, su lógica resulta útil para diseñar políticas institucionales. La idea de distinguir entre usos inaceptables, usos de alto riesgo, usos con obligaciones de transparencia y usos de riesgo limitado permite superar enfoques uniformes. En el ámbito educativo, ciertos usos podrían considerarse de mayor riesgo, por ejemplo, sistemas que deciden admisiones, asignan apoyos, evalúan automáticamente estudiantes, perfilan trayectorias académicas o procesan datos

sensibles. En cambio, herramientas para generación de ideas, apoyo lingüístico o planeación de clases pueden regularse con requisitos de transparencia, validación y supervisión humana.

En México, la legislación vigente en materia de protección de datos personales es indispensable para la regulación universitaria de la IA (Cámara de Diputados, 2025). Cuando una institución educativa utiliza herramientas de IA con datos de estudiantes, docentes, participantes de investigación o personal administrativo, debe considerar principios como licitud, consentimiento, información, calidad, finalidad, lealtad, proporcionalidad y responsabilidad. La IA puede ampliar el volumen, la velocidad y la complejidad del tratamiento de datos; por ello, los avisos de privacidad, las medidas de seguridad, la minimización de datos y la anonimización (proceso técnico e irreversible de modificar datos personales para que la identificación de una persona física sea imposible, garantizando la privacidad), se vuelven elementos centrales de cumplimiento. Este punto es especialmente importante en universidades públicas y privadas, pues cada una opera bajo marcos de protección de datos aplicables a sujetos obligados o particulares.

Los Principios de Chapultepec, presentados en México por autoridades nacionales en 2026, ofrecen una base ética emergente para el uso y desarrollo de IA (SECIHTI & ATDT, 2026). Aunque su fuerza jurídica puede diferir de una ley formal, su valor radica en establecer una orientación pública sobre buenas prácticas, responsabilidad, transparencia, derechos humanos y colaboración. En el contexto universitario, este tipo de declaraciones puede servir como puente entre estándares internacionales y lineamientos internos, especialmente mientras el país define un marco legislativo más completo.

La evaluación del estadio de preparación de la IA para México, desarrollada bajo la metodología RAM de la UNESCO, aporta un diagnóstico relevante (UNESCO, 2024). El documento reconoce que México enfrenta desafíos para consolidar una gobernanza ética y responsable de la IA, pero también afirma que la regulación

responsable puede ser compatible con innovación y crecimiento. Para las universidades, esta lectura es relevante porque muestra que la regulación no debe verse como obstáculo, sino como infraestructura de confianza. Una universidad que cuenta con políticas claras puede innovar con mayor legitimidad que una institución que adopta herramientas sin criterios públicos.

Las recomendaciones universitarias recientes, como las elaboradas por la UNAM para el uso educativo de IA generativa, son especialmente pertinentes porque traducen la discusión ética a dimensiones de trabajo académico: enseñanza, aprendizaje, evaluación, investigación y gestión educativa (GAIA-GEN UNAM, 2025). Estos instrumentos muestran que el debate universitario ya no puede limitarse a advertir riesgos; debe ofrecer orientaciones concretas para el diseño de actividades, la formación docente, la evaluación auténtica, la declaración de uso, la supervisión humana y la construcción de literacidad (conjunto de competencias, habilidades y actitudes para leer, escribir y procesar información en contextos socioculturales específicos) en IA.

A partir de este marco, el capítulo adopta una posición de regulación responsable. La regulación responsable se define aquí como el conjunto articulado de principios, normas, procedimientos, capacidades y mecanismos de supervisión que orientan el uso de IA para maximizar beneficios académicos, reducir daños previsibles, proteger derechos, mantener la integridad del conocimiento y asegurar responsabilidad humana. Esta definición permite ubicar la regulación universitaria en un punto intermedio entre derecho, ética, pedagogía y gestión institucional.

La IA en la práctica docente universitaria

La práctica docente es uno de los espacios donde la IA genera cambios más visibles. Un profesor puede emplear IA para diseñar secuencias didácticas, generar ejemplos, adaptar explicaciones, preparar reactivos, construir rúbricas, traducir materiales, simular casos, resumir textos, producir guías de estudio o elaborar retroalimentación inicial. Estas posibilidades pueden ahorrar tiempo y ampliar recursos de enseñanza. Sin embargo, el uso docente de IA exige una regla básica:

la responsabilidad pedagógica no puede delegarse en el sistema. El docente debe revisar, contextualizar, corregir y validar cualquier producto generado por IA antes de presentarlo a los estudiantes.

En la planeación didáctica, la IA puede apoyar la diversificación de actividades. Por ejemplo, una misma unidad temática puede transformarse en debate, estudio de caso, ejercicio de resolución de problemas, práctica guiada o evaluación diagnóstica. Esta capacidad es útil para docentes que atienden grupos heterogéneos. No obstante, los materiales generados deben revisarse para evitar errores, sesgos culturales, simplificaciones excesivas o referencias inexistentes. En carreras como ingeniería, derecho, administración, ciencias de la salud o educación, solo por citar algunos ejemplos, una respuesta técnicamente incorrecta puede inducir aprendizajes deficientes. La regulación institucional debe establecer que el uso docente de IA es auxiliar, no sustitutiva de la competencia profesional.

En la evaluación, el impacto es más profundo. Las tareas centradas únicamente en productos finales son vulnerables a la automatización. Por ello, se recomienda fortalecer evaluaciones parciales, orales, situadas y auténticas. Un ensayo puede complementarse con bitácora de lectura, defensa oral, reflexión sobre decisiones, evidencias de borradores, análisis de fuentes verificadas y declaración de uso de IA. Un proyecto puede evaluarse mediante avances, revisión de supuestos, aplicación contextual y discusión de limitaciones. Este enfoque no busca perseguir el uso de IA, sino recuperar evidencia del razonamiento del estudiante.

La transparencia es un principio operativo indispensable. Cada asignatura debería informar desde el inicio qué usos de IA están permitidos, condicionados o prohibidos. Los criterios deben ser claros y proporcionales. Por ejemplo, puede permitirse el uso de IA para lluvia de ideas, corrección de estilo o explicación de conceptos, siempre que se declare y no sustituya el análisis del estudiante. Puede condicionarse el uso para elaboración de marcos teóricos, exigiendo verificación de fuentes y trazabilidad. Puede prohibirse en exámenes individuales, producción de

datos de investigación no autorizada o generación completa de trabajos presentados como propios. La claridad evita arbitrariedades y reduce conflictos.

La alfabetización en IA debe considerarse parte de la responsabilidad docente. No basta con decir al estudiante que use o no use una herramienta; necesita aprender a formular preguntas, verificar resultados, identificar sesgos, distinguir entre fuentes confiables y respuestas plausibles, proteger datos personales y reconocer límites epistémicos. En este sentido, la regulación debe acompañarse de formación. Una política sin capacitación se convierte en requisito formal de bajo cumplimiento. La universidad debe ofrecer talleres, guías, ejemplos de declaraciones de uso, bancos de actividades y criterios comunes para comités académicos.

La IA también puede apoyar inclusión educativa. Esto se refiere a que estudiantes con dificultades de lectura, escritura, idioma o accesibilidad pueden beneficiarse de herramientas de apoyo. No obstante, la regulación debe evitar que la inclusión dependa de plataformas privadas costosas o de uso opaco de datos. Las universidades deben favorecer alternativas institucionales, software seguro, criterios de accesibilidad y acompañamiento docente. Una política sensible a la equidad debe reconocer que no todos los estudiantes tienen el mismo acceso tecnológico ni las mismas habilidades para evaluar respuestas generadas.

La regulación docente debe organizarse alrededor de cinco criterios: Propósito pedagógico, transparencia, supervisión humana, protección de datos y evaluación auténtica. Si el uso de la IA fortalece el aprendizaje, se declara, se revisa, no expone información indebida y permite valorar el desempeño humano, puede considerarse legítimo. Si sustituye el aprendizaje, oculta autoría, procesa datos sin consentimiento o automatiza decisiones relevantes sin revisión, debe restringirse o prohibirse.

La IA en la investigación universitaria

La investigación universitaria se beneficia de la IA en múltiples etapas. En la fase de exploración puede ayudar a identificar palabras clave, organizar literatura, traducir resúmenes o construir mapas conceptuales. En la fase de diseño puede

apoyar matrices de variables, esquemas de instrumentos, planes de análisis o simulaciones preliminares. En la fase de procesamiento puede ayudar a limpiar bases, clasificar documentos, transcribir audios o sugerir categorías. En la fase de comunicación puede mejorar redacción, tablas, visualizaciones y estructura argumentativa. Sin embargo, cada etapa implica riesgos distintos y requiere reglas específicas.

El primer criterio regulatorio es la transparencia metodológica. Cuando la IA se utiliza de forma sustantiva en una investigación, debe declararse. La declaración puede incluir nombre de la herramienta, versión o fecha de consulta, tipo de uso, alcance, limitaciones y procedimiento de verificación. No es necesario declarar usos mínimos de corrección ortográfica si no alteran contenido académico, pero sí debe declararse el apoyo en análisis, codificación, generación de hipótesis, traducción sustantiva, redacción de apartados o construcción de instrumentos. Esta práctica fortalece la consistencia y la confianza científica.

El segundo criterio es la confidencialidad. No deben ingresarse a plataformas externas datos personales, sensibles, inéditos o confidenciales sin autorización, evaluación de riesgos y medidas de seguridad. Esto incluye entrevistas, expedientes, calificaciones, historias clínicas, datos financieros, información de comunidades, contratos, propuestas de investigación no publicadas o documentos sujetos a revisión. La facilidad de copiar y pegar información en un sistema de IA puede ocultar que se está realizando un tratamiento de datos fuera del control institucional. Por ello, los comités de ética deben actualizar sus formatos de revisión para preguntar explícitamente si se usará IA y con qué datos.

El tercer criterio es la validación humana. La IA puede sugerir relaciones, clasificaciones o interpretaciones, pero la responsabilidad del análisis recae en el equipo investigador. En estudios cualitativos, por ejemplo, una herramienta puede apoyar la agrupación inicial de códigos, pero no sustituye la interpretación contextual, la reflexión, la saturación teórica, ni el diálogo con participantes. En estudios cuantitativos, puede apoyar secuencia de instrucciones o visualizaciones,

pero los supuestos estadísticos, la depuración y la interpretación deben ser revisados por especialistas. El principio de responsabilidad humana evita que la IA se convierta en autoridad epistemológica no examinada.

El cuarto criterio es la integridad de fuentes. La IA generativa puede producir referencias falsas o atribuciones incorrectas. En un capítulo académico, tesis o artículo, todas las fuentes deben verificarse en bases de datos, repositorios oficiales o publicaciones reales. El uso de IA para elaborar referencias debe acompañarse de revisión manual. La regulación institucional debe considerar como falta académica la presentación de citas inexistentes o no verificadas, aun cuando hayan sido generadas por una herramienta.

El quinto criterio es la autoría. Las herramientas de IA no deben figurar como autoras porque no asumen responsabilidad moral, legal ni académica por el contenido. Sin embargo, su uso puede reconocerse en notas metodológicas o apartados de agradecimiento técnico, según lineamientos editoriales. El equipo humano debe poder explicar y defender cada decisión metodológica y cada afirmación. La autoría implica responsabilidad, no solo producción de texto.

La regulación de IA en investigación también debe atender los sesgos. Los modelos se entrenan con datos que pueden reflejar desigualdades lingüísticas, geográficas, raciales, de género, socioeconómicas o disciplinares. En México, esto es crucial para investigaciones con comunidades indígenas, contextos rurales, población migrante, estudiantes de primera generación universitaria o regiones con menor conocimiento digital. Un sistema entrenado principalmente con datos de otros contextos puede generar interpretaciones inadecuadas. La investigación responsable debe documentar límites, contrastar con fuentes locales y evitar convertir la salida algorítmica en verdad neutral.

La investigación universitaria debe anticipar que los comités editoriales, revistas y congresos exigirán cada vez más declaraciones de uso de IA. Por ello, las universidades mexicanas necesitan preparar formatos estandarizados. Una declaración mínima podría señalar: 'Se utilizó IA generativa para apoyo en

corrección de estilo y organización preliminar de ideas. Todas las fuentes, datos, interpretaciones y conclusiones fueron verificadas y son responsabilidad exclusiva de las personas autoras'. Para usos más intensivos, la declaración debe ser más detallada.

Riesgos regulatorios prioritarios

La regulación universitaria de IA debe iniciar por una identificación de riesgos. El riesgo no se reduce a la posibilidad de plagio; incluye daños a derechos, errores epistémicos, dependencia tecnológica, desigualdad, opacidad y pérdida de confianza. Una política madura debe reconocer que no todos los usos tienen el mismo impacto. La proporcionalidad regulatoria exige concentrar mayores controles donde el daño potencial sea mayor.

El primer riesgo es la afectación a datos personales. En educación superior se procesan datos de identificación, trayectoria académica, desempeño, salud, situación socioeconómica, becas, discapacidad, datos laborales y datos de investigación. Si estos datos se introducen en herramientas sin control, pueden perderse garantías de confidencialidad. La regulación debe prohibir el ingreso de datos personales no anonimizados en plataformas no autorizadas y exigir evaluación previa para proyectos institucionales de valoración, tutoría automatizada o seguimiento académico.

El segundo riesgo es la discriminación algorítmica. Si una herramienta se utiliza para seleccionar, evaluar, recomendar o clasificar estudiantes, puede reproducir sesgos. Un sistema de alerta temprana de reprobación, por ejemplo, puede ser útil si se apoya de una intervención docente; pero puede ser dañino si etiqueta estudiantes sin transparencia, sin revisión humana o sin considerar condiciones contextuales. La regulación debe exigir claridad, derecho a revisión, participación de especialistas y prohibición de decisiones automatizadas sin supervisión en asuntos de alto impacto académico.

El tercer riesgo es la erosión de la integridad académica. La facilidad de generar trabajos completos puede incentivar simulación de aprendizaje. No obstante, la

respuesta regulatoria no debe centrarse exclusivamente en sanciones. Debe combinar rediseño de evaluación, declaración de uso, acompañamiento, cultura de honestidad y sanciones proporcionadas para casos de suplantación deliberada. La integridad académica se fortalece cuando las tareas son significativas, contextualizadas y difíciles de delegar por completo a una máquina.

El cuarto riesgo es la desinformación. Los modelos de IA pueden presentar respuestas convincentes pero falsas. En la docencia, esto puede introducir errores en materiales. En investigación, puede contaminar marcos teóricos o referencias. La regulación debe establecer obligación de verificación y preferencia por fuentes primarias, normatividad oficial, bases académicas y documentos institucionales. La IA no debe citarse como fuente de hechos sin verificación externa. El uso de la inteligencia humana debe ejercer un rol de intervención para validar las acciones que se generen a partir de la inteligencia artificial.

El quinto riesgo es la dependencia tecnológica. Si docentes, estudiantes e investigadores delegan procesos cognitivos centrales, pueden debilitar habilidades de lectura, escritura, razonamiento, cálculo, interpretación y argumentación, cualidades que son propias del aprendizaje significativo. La regulación no debe impedir el uso de IA, pero sí exigir que el aprendizaje humano permanezca en el centro. Esto implica actividades donde el estudiante explique, compare, defienda, aplique y reflexione sobre el resultado.

El sexto riesgo es la concentración y opacidad de plataformas. Muchas herramientas de IA son privadas, operan con modelos no transparentes y modifican sus condiciones de uso. Las universidades deben evitar que procesos académicos esenciales dependan de tecnologías sin una evaluación institucional. Se recomienda priorizar plataformas con mejores garantías de seguridad, accesibilidad, transparencia contractual y protección de datos. También conviene promover capacidades internas y repositorios institucionales de buenas prácticas.

El séptimo riesgo es la brecha de acceso. El uso de herramientas de pago puede generar ventajas injustas. Una política universitaria debe indicar que ninguna

actividad obligatoria debe requerir suscripciones particulares sin alternativa institucional o gratuita. Asimismo, debe capacitar a estudiantes en el uso crítico para que la IA no se convierta en privilegio de quienes ya tienen mayor capital digital.

El octavo riesgo es la responsabilidad difusa. Cuando una decisión académica se basa en IA, puede ser difícil determinar quién responde por errores. La regulación debe establecer que la responsabilidad final corresponde a la persona o instancia que decide usar la herramienta. La IA no exime al docente de revisar materiales, al investigador de validar resultados ni a la autoridad de justificar decisiones.

Tabla 1.

Matriz básica de riesgo y uso de IA en docencia e investigación universitaria.

Uso de IA	Nivel de riesgo	Criterio regulatorio	Condición mínima
Lluvia de ideas, ejemplos, explicación de conceptos	Bajo	Permitido con criterio docente.	No sustituye actividad evaluada; verificar exactitud.
Corrección de estilo y claridad de redacción	Bajo/medio	Permitido con declaración cuando sea trabajo evaluable.	La autoría del contenido sigue siendo humana.
Generación de estructura para ensayo o protocolo	Medio	Permitido con declaración y verificación.	Debe acompañarse de fuentes reales y elaboración propia.
Elaboración completa de trabajo presentado como propio	Alto	No permitido	Constituye suplantación del proceso académico.
Ingreso de datos personales de estudiantes en IA externa	Alto	No permitido salvo autorización institucional.	Riesgo de privacidad y tratamiento no controlado.
Análisis preliminar de datos anonimizados de investigación	Medio/alto	Permitido con protocolo y validación.	Debe documentarse herramienta, fecha, alcance y revisión humana.

Evaluación automática de estudiantes con efectos académicos	Alto	Condicionado a revisión humana y evaluación institucional.	Requiere transparencia, explicabilidad y derecho a revisión.
Fabricación de citas, datos o resultados	Inaceptable	Prohibido	Falta grave a la integridad académica y científica.

Nota: Elaboración propia

Lineamientos operativos para la docencia

Los lineamientos para docentes deben ser comprensibles y aplicables. Se propone que cada profesor incluya en su planeación o programa de asignatura una sección de uso de IA con tres categorías: permitido, permitido con declaración y no permitido. Esta clasificación debe vincularse con los resultados de aprendizaje. Si el objetivo es aprender a escribir un argumento, la IA puede usarse para retroalimentación, pero no para producir el argumento final sin intervención del estudiante. Si el objetivo es aprender a resolver un problema, la IA puede explicar pasos después de un intento propio, pero no sustituir la resolución individual en una evaluación.

Para trabajos escritos, se recomienda solicitar una declaración de uso de IA. Un ejemplo básico sería: 'Utilicé IA generativa para revisar redacción y claridad del texto; no se utilizó para generar contenido sustantivo ni referencias. Verifiqué toda la información incluida'. Un ejemplo más amplio sería: 'Utilicé IA generativa para generar un esquema preliminar y preguntas de revisión. Posteriormente elaboré el contenido, consulté fuentes académicas y verifiqué las referencias'. La declaración no debe ser castigo, sino práctica de transparencia.

En la evaluación, se recomienda incorporar evidencias de proceso. Los estudiantes pueden entregar borradores, notas de lectura, capturas de prompts relevantes, bitácoras de decisiones, matrices de fuentes, reflexiones sobre errores de la IA y defensa oral. Estos elementos permiten distinguir entre uso asistido y sustitución completa. Además, enseñan a usar la IA de manera crítica.

En cursos en línea o híbridos, la regulación requiere especial atención. Las actividades asincrónicas pueden ser apoyadas por IA; por ello, conviene diseñar tareas contextualizadas al entorno del estudiante, problemas locales, análisis de casos propios, vinculación con experiencias de campo y discusiones síncronas. En lugar de pedir respuestas genéricas, se pueden pedir análisis comparativos, aplicación a un contexto específico y justificación de decisiones.

El profesor también debe transparentar su propio uso de la IA. Si utiliza IA para generar rúbricas, ejemplos o retroalimentación, debe revisar y adaptar los resultados. En actividades evaluativas, no debe ingresar trabajos completos de estudiantes a plataformas externas sin autorización institucional, especialmente si contienen datos personales. La regulación debe proteger tanto al estudiante como al docente.

Los lineamientos deben prever consecuencias. Cuando un estudiante usa IA de forma no autorizada en una evaluación, se debe aplicar el reglamento académico correspondiente. Sin embargo, antes de sancionar, la institución debe haber informado las reglas claras. La proporcionalidad es importante: no es lo mismo corregir estilo sin declararlo que entregar un trabajo completo generado por IA, fabricar datos o inventar fuentes.

Lineamientos operativos para la investigación

En investigación, los lineamientos deben incorporarse desde el protocolo. Todo proyecto debe responder si utilizará IA, para qué actividades, con qué datos, qué riesgos existen y cómo se mitigarán. Esta pregunta debe estar presente en seminarios de tesis, proyectos de cuerpos académicos, convocatorias internas y comités de ética. La declaración del uso de la IA así como la revisión previa evita correcciones tardías cuando ya se procesaron datos de forma indebida.

Se propone clasificar el uso de IA en investigación en cuatro niveles. El nivel uno corresponde a apoyo editorial menor: ortografía, estilo, traducción preliminar o formato. El nivel dos corresponde a apoyo conceptual o documental: generación de esquemas, preguntas, matrices o búsqueda exploratoria. El nivel tres corresponde

a apoyo metodológico: codificación preliminar, análisis de datos, generación de guion, clasificación de documentos o visualización. El nivel cuatro corresponde a procesos de alto impacto: decisiones sobre participantes, análisis de datos sensibles, revisión de propuestas confidenciales, generación de resultados o automatización de dictámenes. Los niveles tres y cuatro requieren mayor documentación y revisión ética.

Los proyectos cualitativos requieren especial cuidado. Las entrevistas y transcripciones contienen información contextual que puede identificar personas aun sin nombres. Antes de usar IA para transcripción, codificación o resumen, se debe revisar la política de privacidad de la herramienta y obtener consentimiento. Siempre que sea posible, se deben anonimizar datos, usar herramientas institucionales o locales y conservar registro de decisiones analíticas.

Los proyectos cuantitativos también enfrentan riesgos. Una IA puede sugerir pruebas estadísticas incorrectas o interpretar correlaciones como causalidad. Por ello, el investigador debe validar supuestos, revisar código, documentar transformaciones y conservar bases originales. La IA puede ser asistente técnico, pero no sustituto de formación metodológica.

En publicaciones, se recomienda incluir una declaración de uso de IA cuando haya contribuido a procesos sustantivos. Las revistas y editoriales académicas avanzan hacia reglas de transparencia. La universidad puede anticiparse con un formato institucional. Esta declaración debe diferenciar entre apoyo lingüístico, apoyo analítico y generación de contenido. También debe indicar que las personas autoras verificaron información, fuentes y resultados.

La propiedad intelectual es otro punto crítico. El uso de IA para generar imágenes, textos o gráficos puede ocasionar incertidumbre sobre derechos. En materiales de investigación y docencia, conviene utilizar insumos con licencias claras, evitar reproducción de contenido protegido y documentar la procedencia de imágenes o datos. Si se generan materiales con IA, debe indicarse cuando sea relevante y verificar que no se infrinjan derechos de terceros.

Los comités de ética deben recibir capacitación específica. No basta con evaluar consentimiento y confidencialidad bajo criterios tradicionales; deben comprender qué hace una herramienta, qué datos procesa, dónde se almacenan, si se usan para entrenamiento, qué riesgos de acción de vincular datos anonimizados con información externa para reconocer al individuo existen y cómo se supervisarán resultados. La regulación universitaria debe fortalecer capacidades de estos comités.

Tabla 2.

Lista de cotejo para comités de ética e investigación ante proyectos con uso de IA.

Pregunta de revisión	Finalidad regulatoria
¿Se usará IA en alguna fase del proyecto?	Identificar alcance real del uso.
¿Qué herramienta, versión o proveedor se utilizará?	Permite trazabilidad y evaluación de condiciones de uso.
¿Se procesarán datos personales, sensibles o confidenciales?	Activa revisión de privacidad, consentimiento y seguridad.
¿Los datos estarán anonimizados?	Reduce riesgo de identificación o exposición.
¿La IA influirá en resultados o interpretación?	Exige declaración metodológica y validación humana.
¿Cómo se verificarán fuentes, códigos o salidas?	Asegura integridad y reproducibilidad.
¿Se informará el uso de IA en publicaciones?	Fortalece transparencia editorial.
¿Existe plan de resguardo de datos y evidencias?	Permite auditoría y rendición de cuentas.

Nota: Elaboración propia

Discusión: regulación, autonomía universitaria y justicia social

La regulación de la IA en la educación superior mexicana debe dialogar con la autonomía universitaria. Las universidades tienen capacidad para definir sus procesos académicos, pero esa autonomía implica responsabilidad. No pueden esperar pasivamente a que exista una ley nacional integral para actuar. La

autonomía permite construir políticas contextualizadas y de participación, siempre alineadas con derechos humanos, protección de datos y estándares éticos.

Un punto delicado es evitar que la regulación se convierta en mecanismo de vigilancia excesiva. La preocupación por plagio o fraude puede llevar a prácticas invasivas, como monitoreo permanente, detectores poco confiables o revisión indiscriminada de trabajos en plataformas externas. Una regulación responsable debe proteger la integridad académica sin vulnerar privacidad ni presunción de buena fe. La confianza institucional se construye con claridad, diálogo y proporcionalidad. Se debe evitar una sensación de monitoreo, vigilancia, castigo, sino más bien de confiabilidad y de promoción de prácticas éticas.

La justicia social también debe estar en el centro. En México, las desigualdades educativas y tecnológicas son profundas. Una política de IA que presupone acceso universal a dispositivos, conectividad y suscripciones puede profundizar brechas. En universidades públicas, rurales o regionales, la regulación debe contemplar alternativas de acceso, laboratorios, capacitación y criterios de no discriminación. La IA puede apoyar inclusión si se gobierna adecuadamente; puede excluir si se deja al mercado.

La regulación también debe reconocer diversidad disciplinaria. En ingeniería, la IA puede apoyar cálculo, modelación y programación; en humanidades, análisis textual y escritura; en salud, búsqueda de evidencia; en derecho, análisis normativo; en administración, simulación de escenarios. Cada área requiere criterios propios. Sin embargo, todas comparten mínimos: transparencia, verificación, protección de datos y responsabilidad humana.

Otro tema es la tensión entre innovación y cumplimiento. Algunas personas temen que regular frene la creatividad. Este capítulo plantea lo contrario: la regulación responsable permite innovar con confianza. Cuando existen reglas claras, docentes e investigadores saben qué pueden hacer, qué deben documentar y qué no deben hacer. La ausencia de reglas favorece improvisación, inequidad y conflictos.

La discusión para el caso de México debe evitar copiar modelos sin adaptación. El enfoque europeo de riesgo es útil, pero la educación superior mexicana tiene características propias: autonomía universitaria, diversidad institucional, brechas digitales, marcos de protección de datos diferenciados, necesidades regionales y desafíos de financiamiento. Por ello, la propuesta aquí presentada combina principios internacionales con criterios operativos para el aula e investigación.

Finalmente, la regulación debe entenderse como proceso dinámico. Las herramientas de IA cambian rápidamente; aparecen nuevos modelos, capacidades multimodales, agentes autónomos y sistemas integrados a plataformas educativas. Una política aprobada una vez puede quedar obsoleta. Las universidades deben adoptar mecanismos de revisión continua, aprendizaje institucional y participación de la comunidad.

Conclusiones

La regulación de la IA en México para la práctica docente y de investigación universitaria es una necesidad inmediata. Aunque el país aún construye un marco jurídico integral, como muestran las iniciativas y debates recientes sobre una ley nacional de IA (Basham, Ringe y Correa, 2026), las universidades ya enfrentan decisiones concretas. La IA se usa para enseñar, aprender, evaluar, investigar, escribir, traducir, analizar y gestionar. La pregunta no es si la universidad debe regular, sino cómo hacerlo de manera proporcional, ética y pedagógicamente significativa.

El análisis realizado permite sostener que la regulación universitaria debe articular cuatro dimensiones. La primera es jurídica: protección de datos, derechos humanos, propiedad intelectual y responsabilidad. La segunda es ética: transparencia, justicia, no discriminación, interpretabilidad y supervisión humana. La tercera es pedagógica: evaluación auténtica, aprendizaje activo, autoría, alfabetización y evidencias de proceso. La cuarta es institucional: políticas, comités, capacitación, herramientas autorizadas y actualización normativa.

En la docencia, el eje debe ser el aprendizaje humano. La IA puede apoyar, explicar, retroalimentar y ampliar recursos, pero no debe sustituir los procesos cognitivos que la universidad busca formar. En la investigación, el eje debe ser la integridad científica. La IA puede asistir, pero no asumir responsabilidad por datos, métodos, fuentes o conclusiones. En ambos casos, la transparencia y la validación humana son principios irrenunciables.

Consideramos que se puede construir una regulación universitaria de IA que no sea reactiva, sino formativa; no solo disciplinaria, sino orientada al bien público; no únicamente tecnológica, sino centrada en derechos y aprendizaje. Las instituciones de educación superior pueden convertirse en laboratorios de gobernanza responsable, mostrando que la innovación y la regulación no son fuerzas opuestas, sino condiciones complementarias para una inteligencia artificial al servicio de la sociedad.

Referencias

- Basham, Ringe y Correa. (2026, 26 de febrero). Ley Nacional para Regular el Uso de la Inteligencia Artificial. <https://basham.com.mx/ley-nacional-para-regular-el-uso-de-la-inteligencia-artificial/>
- Cámara de Diputados del H. Congreso de la Unión. (2025). Ley Federal de Protección de Datos Personales en Posesión de los Particulares. Diario Oficial de la Federación. Última reforma consultada en versión vigente. <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
- Comisión Europea. (2024). La Comisión pone en marcha un plan de acción contra el ciberacoso para proteger a los jóvenes en línea. European Commission. <https://digital-strategy.ec.europa.eu/es>
- European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Grupo Académico de Inteligencia Artificial Generativa en Educación, UNAM. (2025). Recomendaciones para el uso educativo de la inteligencia artificial generativa en la UNAM. Universidad Nacional Autónoma de México. https://iagen.unam.mx/recursos/Recomendaciones_IAGEN_UNAM_2025.pdf
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2021). Recomendación sobre la ética de la inteligencia artificial. UNESCO. <https://www.unesco.org/es/legal-affairs/recommendation-ethics-artificial-intelligence>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2024). México: evaluación del estadio de preparación de la inteligencia artificial. UNESCO. <https://www.unesco.org/es/articles/mexico-evaluacion-del-estadio-de-preparacion-de-la-inteligencia-artificial>
- Organización para la Cooperación y el Desarrollo Económicos. (2019). OECD principles on artificial intelligence. OECD. <https://oecd.ai/en/ai-principles>
- Secretaría de Ciencia, Humanidades, Tecnología e Innovación y Agencia de Transformación Digital y Telecomunicaciones. (2026). Principios de Chapultepec. Declaración de ética y buenas prácticas para el uso y desarrollo de la inteligencia artificial. Gobierno de México. https://secihtl.mx/wp-content/uploads/2026/02/Principios_Chapultepec_F.pdf

World Intellectual Property Organization. (2024). WIPO conversation on intellectual property and artificial intelligence. WIPO. https://www.wipo.int/about-ip/en/artificial_intelligence/

INTELIGENCIA ARTIFICIAL GENERAL: ENTRE LA EXPANSIÓN DE CAPACIDADES Y LOS LÍMITES DE SU REGULACIÓN

General Artificial Intelligence: Between the expansion of capacities and the
limits of their regulation

Oscar Lozano Carrillo¹⁸

exato@azc.uam.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0001-6166-3033>

Aylin Torres Ramírez¹⁹

aytora@azc.uam.mx

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0000-0001-6300-1666>

Cinthia Noemi Zacatenco Arellano²⁰

noemizacatencoa@gmail.com

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

<https://orcid.org/0009-0000-7515-8310>

RESUMEN

La Inteligencia Artificial General (IAG) representa un horizonte tecnológico caracterizado por la expansión progresiva de capacidades cognitivas artificiales, con implicaciones profundas en los ámbitos social, económico y político. Este capítulo analiza la relación entre el desarrollo de la IAG y los límites de su regulación, partiendo de la premisa de que los marcos normativos actuales resultan

¹⁸ Profesor-Investigador del departamento de administración de la Universidad Autónoma Metropolitana Unidad Azcapotzalco.

¹⁹ Profesora Visitante del del departamento de administración de la Universidad Autónoma Metropolitana Unidad Azcapotzalco.

²⁰ Licenciada en administración por la Universidad Autónoma Metropolitana Unidad Azcapotzalco.

insuficientes frente a una tecnología dinámica, compleja y de alcance global. A partir de una revisión teórica, se examinan las principales dimensiones de la expansión de capacidades de la inteligencia artificial, así como los riesgos emergentes asociados, entre ellos la concentración de poder, la opacidad algorítmica, los sesgos y la dependencia tecnológica. Asimismo, se identifican los vacíos normativos que dificultan su gobernanza, destacando la fragmentación regulatoria, el predominio de instrumentos no vinculantes y la falta de mecanismos efectivos de rendición de cuentas. El capítulo también analiza los principales marcos internacionales de gobernanza de la inteligencia artificial y sus limitaciones estructurales. Finalmente, se propone un enfoque de regulación integral basado en la gobernanza adaptativa, la participación social, la alfabetización en inteligencia artificial y la incorporación de principios de justicia social y equidad. Se concluye que la regulación de la IAG es posible únicamente mediante enfoques multidimensionales y dinámicos, capaces de articular dimensiones jurídicas, éticas, educativas y sociales, orientando el desarrollo tecnológico hacia el bienestar colectivo.

Palabras clave: Inteligencia Artificial General; regulación tecnológica; gobernanza de la IA; riesgos emergentes; justicia social

ABSTRACT

Artificial General Intelligence (AGI) represents a technological horizon characterized by the progressive expansion of artificial cognitive capabilities, with profound implications in the social, economic, and political spheres. This chapter analyzes the relationship between the development of AGI and the limits of its regulation, based on the premise that current regulatory frameworks are insufficient in the face of a dynamic, complex, and globally impactful technology. Through a theoretical review, the chapter examines the main dimensions of the expansion of artificial intelligence capabilities, as well as the emerging risks associated with them, including the concentration of power, algorithmic opacity, bias, and technological

dependency. Likewise, it identifies the regulatory gaps that hinder its governance, highlighting regulatory fragmentation, the predominance of non-binding instruments, and the lack of effective accountability mechanisms.

The chapter also analyzes the main international frameworks for artificial intelligence governance and their structural limitations. Finally, it proposes a comprehensive regulatory approach based on adaptive governance, social participation, artificial intelligence literacy, and the incorporation of principles of social justice and equity. It concludes that the regulation of AGI is only possible through multidimensional and dynamic approaches capable of articulating legal, ethical, educational, and social dimensions, guiding technological development toward collective well-being.

Keywords: Artificial General Intelligence; technological regulation; AI governance; emerging risks; social justice.

INTRODUCCIÓN

En las últimas décadas, la inteligencia artificial (IA) ha transitado de ser una promesa tecnológica para convertirse en una infraestructura fundamental que atraviesa múltiples dimensiones de la vida social, económica y política. Este proceso ha estado marcado por una expansión progresiva de capacidades, impulsada por avances en aprendizaje automático, disponibilidad de datos y poder computacional. En este contexto, ha cobrado relevancia el concepto de Inteligencia Artificial General (IAG), entendido como un horizonte en el que los sistemas artificiales podrían alcanzar niveles de desempeño cognitivo comparables o superiores a los humanos en una amplia variedad de tareas (Goertzel & Pennachin, 2007).

Aunque la IAG aún no se ha materializado plenamente, los desarrollos recientes en modelos de propósito general y sistemas multimodales han reavivado el debate sobre sus implicaciones y, particularmente, sobre la posibilidad de su regulación. A diferencia de tecnologías anteriores, la inteligencia artificial se caracteriza por su

capacidad de adaptación, su carácter transversal y su integración en infraestructuras críticas, lo que amplifica tanto sus beneficios potenciales como sus riesgos. En este sentido, la expansión de capacidades de la IA plantea interrogantes fundamentales sobre los límites del control humano, la distribución del poder tecnológico y la protección de derechos fundamentales (Floridi y otros., 2018).

El debate sobre la regulación de la inteligencia artificial se sitúa, por tanto, en una tensión estructural entre la velocidad de la innovación tecnológica y la capacidad de respuesta de los marcos normativos. Mientras que los sistemas de IA evolucionan de manera acelerada y en muchos casos impredecible, los procesos regulatorios tienden a ser más lentos, fragmentados y reactivos. Esta brecha ha dado lugar a la emergencia de riesgos que no pueden ser abordados únicamente mediante enfoques tradicionales, como la opacidad algorítmica, los sesgos, la concentración de poder en actores tecnológicos y la creciente dependencia de sistemas automatizados (Russell, 2019).

Frente a estos desafíos, diversos organismos internacionales han impulsado marcos de gobernanza orientados a establecer principios éticos y lineamientos para el desarrollo responsable de la IA. Sin embargo, estos esfuerzos han mostrado limitaciones importantes, particularmente en términos de su carácter no vinculante, su implementación desigual y la ausencia de mecanismos efectivos de coordinación global (UNESCO, 2021; OECD, 2019). En consecuencia, la pregunta sobre si es posible regular la inteligencia artificial, y especialmente su forma más avanzada, no admite respuestas simples, sino que exige una revisión crítica de los enfoques existentes.

En este contexto, el presente capítulo tiene como objetivo analizar la Inteligencia Artificial General desde una perspectiva que articula la expansión de sus capacidades con los límites de su regulación. Para ello, se examinan, en primer lugar, las características que definen este horizonte tecnológico y las implicaciones derivadas de su desarrollo. Posteriormente, se abordan los riesgos emergentes asociados a la inteligencia artificial, así como los vacíos normativos que dificultan

su gobernanza. A continuación, se analizan los principales marcos internacionales de regulación y los límites estructurales que enfrentan. Finalmente, se propone un enfoque de gobernanza integral que incorpore dimensiones jurídicas, éticas, educativas y sociales, con el objetivo de contribuir a la construcción de modelos regulatorios más robustos y orientados al bienestar colectivo.

Más allá de su dimensión técnica, la discusión sobre la Inteligencia Artificial General remite a una cuestión fundamental: cómo orientar el desarrollo tecnológico en función de valores sociales y principios de justicia. En este sentido, la regulación de la inteligencia artificial no debe entenderse únicamente como un mecanismo de control, sino como una herramienta para garantizar que la expansión de sus capacidades contribuya al fortalecimiento de la democracia, la equidad y el bienestar humano.

1. EXPANSIÓN DE CAPACIDADES: ¿QUÉ IMPLICA LA INTELIGENCIA ARTIFICIAL GENERAL?

La Inteligencia Artificial General (IAG) se refiere a un tipo de sistema capaz de realizar una amplia variedad de tareas cognitivas con un nivel de desempeño comparable o superior al humano, lo que la distingue de la inteligencia artificial estrecha, diseñada para funciones específicas. Aunque la IAG aún no se ha materializado plenamente, su conceptualización ha cobrado relevancia debido a los avances recientes en modelos de aprendizaje automático que muestran capacidades cada vez más amplias, flexibles y adaptativas. En este sentido, la IAG no debe entenderse como un estado alcanzado, sino como un horizonte de desarrollo tecnológico que orienta la evolución de los sistemas actuales (Goertzel & Pennachin, 2007).

La transición desde sistemas especializados hacia modelos más generalizados ha sido impulsada por el desarrollo de arquitecturas capaces de operar en múltiples dominios. En particular, los llamados *modelos fundacionales* han introducido un cambio significativo, al permitir que un mismo sistema pueda ser adaptado a

diversas tareas sin requerir rediseños estructurales. Esta capacidad de generalización representa un punto intermedio entre la inteligencia artificial estrecha y la IAG, ya que amplía el rango de aplicación de los sistemas y reduce las barreras entre diferentes ámbitos de uso (Bommasani y otros., 2021). Como resultado, la inteligencia artificial ha dejado de ser una herramienta especializada para convertirse en una infraestructura transversal con impacto en múltiples sectores.

La expansión de capacidades de la inteligencia artificial se manifiesta en diversas dimensiones que reconfiguran la relación entre humanos y tecnología. En primer lugar, destaca la capacidad de generalización, entendida como la habilidad de aplicar conocimientos adquiridos en un contexto a situaciones nuevas. Este rasgo, fundamental para la inteligencia humana, comienza a observarse de manera incipiente en sistemas avanzados, lo que amplía su potencial de aplicación. En segundo lugar, se observa un aumento en los niveles de autonomía, ya que los sistemas pueden operar con menor intervención humana, tomando decisiones basadas en patrones complejos de datos. Esta autonomía, si bien incrementa la eficiencia, también plantea interrogantes en términos de control y responsabilidad (Russell, 2019).

Otra dimensión clave es la escalabilidad, que se refiere a la capacidad de los sistemas para mejorar su rendimiento a medida que aumentan los recursos computacionales y la cantidad de datos disponibles. Estudios recientes han demostrado que el desempeño de ciertos modelos sigue patrones predecibles de mejora cuando se incrementan estos factores, lo que sugiere una tendencia hacia sistemas cada vez más potentes (Kaplan y otros., 2020). A esto se suma la multimodalidad, que permite integrar distintos tipos de información, como texto, imágenes y audio, ampliando las posibilidades de interacción y aplicación. Finalmente, la capacidad generativa de los sistemas actuales, que les permite producir contenido original, ha transformado ámbitos como la educación, la cultura y la economía, al redefinir los procesos de creación y producción de conocimiento.

Estas transformaciones no solo tienen implicaciones técnicas, sino también epistemológicas y sociales. La inteligencia artificial, en su forma actual, está modificando las condiciones bajo las cuales se produce, valida y distribuye el conocimiento. Como señalan Floridi y otros. (2018), la IA debe entenderse como una tecnología infraestructural que reorganiza los ecosistemas informacionales, afectando la forma en que las sociedades interpretan y utilizan la información. En el ámbito educativo, esto se traduce en cambios en la autoría del conocimiento, en la forma de aprendizaje y en la relación entre estudiantes y fuentes de información, lo que plantea desafíos importantes en términos de autonomía cognitiva y pensamiento crítico.

Un aspecto central en la expansión de capacidades es su concentración en actores específicos, lo que genera asimetrías significativas a nivel global. El desarrollo de sistemas avanzados de inteligencia artificial requiere recursos que no están equitativamente distribuidos, como infraestructura computacional, grandes volúmenes de datos y capital financiero. Esto ha favorecido la consolidación de un número reducido de empresas y países como líderes en el desarrollo de estas tecnologías, lo que plantea riesgos en términos de dependencia tecnológica y concentración de poder. Como advierten Brynjolfsson y McAfee (2017), estas dinámicas pueden amplificar desigualdades existentes y limitar la capacidad de otros actores para participar en la construcción del ecosistema tecnológico.

Finalmente, es importante reconocer que la expansión de capacidades de la inteligencia artificial se desarrolla en un contexto de alta incertidumbre. No existe consenso sobre cuándo ni cómo podría alcanzarse una forma de inteligencia artificial general, lo que dificulta la anticipación de sus implicaciones. Esta incertidumbre no solo es técnica, sino también social y política, ya que implica desconocer los efectos que estos sistemas podrían tener en distintos ámbitos de la vida colectiva. Como señala Bostrom (2014), uno de los principales desafíos asociados a la IAG es precisamente la dificultad de prever sus impactos, lo que plantea límites importantes para su regulación.

En conjunto, la Inteligencia Artificial General debe entenderse como un proceso en construcción, caracterizado por la expansión progresiva de capacidades que ya están transformando múltiples dimensiones de la vida social. Esta expansión no solo redefine el potencial de la tecnología, sino que también introduce nuevas tensiones en relación con su gobernanza, al plantear desafíos que exceden los marcos regulatorios tradicionales. En este sentido, comprender la naturaleza de esta expansión constituye un paso fundamental para analizar los riesgos emergentes y los límites de su regulación.

2. RIESGOS EMERGENTES EN EL DESARROLLO DE LA INTELIGENCIA ARTIFICIAL GENERAL

El desarrollo de sistemas avanzados de inteligencia artificial, particularmente aquellos que se aproximan al horizonte de la Inteligencia Artificial General (IAG), ha dado lugar a una serie de riesgos emergentes que desafían los marcos tradicionales de análisis y regulación. Estos riesgos no solo se derivan de fallos técnicos, sino de la interacción compleja entre tecnología, estructuras sociales y dinámicas económicas, lo que los convierte en fenómenos sistémicos. En este sentido, diversos autores coinciden en que la IA no debe ser entendida únicamente como una herramienta, sino como una infraestructura que reconfigura relaciones de poder, producción de conocimiento y toma de decisiones (Floridi y otros., 2018).

Uno de los riesgos más significativos es la concentración de poder tecnológico. El desarrollo de sistemas avanzados de IA requiere recursos computacionales, acceso a grandes volúmenes de datos y capacidades técnicas altamente especializadas, lo que ha favorecido la consolidación de un número reducido de actores, principalmente grandes corporaciones tecnológicas y algunos Estados. Esta concentración no solo tiene implicaciones económicas, sino también políticas, ya que otorga a estos actores una influencia desproporcionada sobre el desarrollo, implementación y regulación de la tecnología. Como advierten Brynjolfsson y

McAfee (2017), estas dinámicas pueden amplificar desigualdades existentes y generar nuevas formas de dependencia tecnológica a nivel global.

Otro riesgo central es la opacidad de los sistemas de inteligencia artificial, especialmente aquellos basados en modelos complejos de aprendizaje profundo. Estos sistemas suelen operar como “cajas negras”, lo que dificulta comprender cómo se toman las decisiones y, por tanto, limita la posibilidad de establecer mecanismos efectivos de rendición de cuentas. Esta falta de transparencia plantea problemas importantes desde el punto de vista ético y jurídico, ya que impide identificar responsabilidades en casos de error o daño. En este contexto, la literatura ha subrayado la necesidad de desarrollar modelos más explicables y mecanismos de auditoría algorítmica (Doshi-Velez & Kim, 2017).

A ello se suma el problema de los sesgos algorítmicos, que pueden reproducir y amplificar desigualdades sociales existentes. Dado que los sistemas de IA aprenden a partir de datos históricos, estos pueden incorporar prejuicios relacionados con variables como género, raza o nivel socioeconómico. Diversos estudios han documentado cómo estos sesgos afectan ámbitos como el empleo, la justicia penal y el acceso a servicios, lo que plantea serias preocupaciones en términos de equidad y derechos humanos (Barocas y otros., 2019). En el contexto de una posible IAG, estos problemas podrían escalar significativamente debido a la mayor autonomía y alcance de los sistemas.

Asimismo, emergen riesgos relacionados con la dependencia tecnológica y la delegación cognitiva. A medida que los sistemas de IA adquieren mayor capacidad para realizar tareas complejas, existe el peligro de que los individuos y las instituciones deleguen en ellos funciones fundamentales de análisis, निर्णय y producción de conocimiento. Esto puede derivar en una disminución de la autonomía cognitiva y en una dependencia creciente de sistemas que no siempre son comprendidos por sus usuarios. Como señala Russell (2019), uno de los principales desafíos de la IA avanzada es garantizar que los sistemas permanezcan

alineados con los intereses humanos y no sustituyan de manera acrítica la toma de decisiones humanas.

Otro aspecto relevante es la aparición de riesgos sistémicos, derivados de la integración de la IA en infraestructuras críticas como sistemas financieros, redes energéticas, sistemas de salud o plataformas de información. En estos contextos, un fallo o mal funcionamiento puede tener efectos amplificados, generando impactos a gran escala. La interconexión de sistemas aumenta la vulnerabilidad frente a errores, ataques o usos indebidos, lo que exige una reconsideración de los enfoques tradicionales de seguridad tecnológica (Bostrom, 2014).

Finalmente, es importante considerar los riesgos asociados a la incertidumbre sobre el desarrollo futuro de la IAG. Dado que no existe consenso sobre cuándo ni cómo podría alcanzarse este tipo de inteligencia, la regulación se enfrenta a un escenario de incertidumbre radical. Esta situación dificulta la anticipación de riesgos y la formulación de políticas efectivas, lo que refuerza la necesidad de adoptar enfoques precautorios y adaptativos. Como señalan Erdélyi y Goldsmith (2020), la gobernanza de la IA debe ser capaz de operar en contextos de alta incertidumbre, combinando regulación formal con mecanismos flexibles de supervisión.

En conjunto, estos riesgos evidencian que la problemática de la inteligencia artificial no puede abordarse únicamente desde una perspectiva técnica. Por el contrario, requiere un enfoque integral que considere sus dimensiones sociales, económicas y políticas. La emergencia de estos desafíos plantea la necesidad de repensar los marcos regulatorios existentes y avanzar hacia modelos de gobernanza más inclusivos, transparentes y orientados al bienestar social.

3. LOS VACÍOS NORMATIVOS EN LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL

Uno de los principales desafíos en la gobernanza de la inteligencia artificial, particularmente en el contexto de sistemas avanzados que se aproximan a la

Inteligencia Artificial General (IAG), es la existencia de vacíos normativos significativos. Estos vacíos no solo reflejan la novedad de la tecnología, sino también la dificultad estructural de los sistemas jurídicos para adaptarse a entornos caracterizados por la rapidez del cambio tecnológico, la complejidad sistémica y la incertidumbre. En este sentido, la regulación de la IA se enfrenta a una brecha persistente entre la evolución tecnológica y la capacidad normativa de los Estados, lo que limita la efectividad de las políticas públicas en este ámbito (Erdélyi & Goldsmith, 2020).

Uno de los vacíos más evidentes es el desfase temporal entre innovación y regulación. Las tecnologías de inteligencia artificial evolucionan a un ritmo acelerado, impulsadas por avances en capacidad computacional, disponibilidad de datos y modelos algorítmicos, mientras que los procesos legislativos suelen ser más lentos, deliberativos y reactivos. Esta asimetría genera escenarios en los que los sistemas tecnológicos se implementan ampliamente antes de que existan marcos regulatorios adecuados para gobernarlos. Como resultado, muchas regulaciones actuales tienen un carácter correctivo más que preventivo, lo que reduce su capacidad para anticipar riesgos emergentes (OECD, 2019).

Otro vacío normativo importante es la fragmentación de los marcos regulatorios a nivel global. Actualmente, no existe un sistema internacional unificado que regule el desarrollo y uso de la inteligencia artificial. En su lugar, coexisten múltiples enfoques nacionales y regionales, que varían en términos de alcance, obligatoriedad y principios rectores. Por ejemplo, mientras que la Unión Europea ha avanzado hacia un enfoque regulatorio vinculante basado en niveles de riesgo, otros países han optado por estrategias más flexibles o basadas en principios éticos. Esta diversidad genera inconsistencias normativas, dificulta la cooperación internacional y puede dar lugar a fenómenos de “arbitraje regulatorio”, en los que los actores tecnológicos operan en jurisdicciones con menores restricciones (UNESCO, 2021).

A esta fragmentación se suma el predominio de instrumentos de “soft law” o regulación no vinculante, como códigos de ética, principios voluntarios y guías de buenas prácticas. Si bien estos instrumentos han sido fundamentales para establecer estándares iniciales en la gobernanza de la IA, su carácter no obligatorio limita su capacidad de enforcement. Estudios comparativos han identificado la proliferación de marcos éticos sin mecanismos claros de implementación, lo que genera una brecha entre los principios declarados y las prácticas reales (Floridi y otros., 2018; Corrêa y otros., 2022). En este sentido, uno de los vacíos más relevantes es la falta de mecanismos efectivos para traducir principios éticos en obligaciones jurídicas concretas.

Otro aspecto crítico es la indefinición en torno a la responsabilidad jurídica en sistemas de inteligencia artificial. A medida que los sistemas adquieren mayor autonomía, se vuelve más complejo determinar quién es responsable en caso de errores, daños o decisiones perjudiciales. ¿Debe atribuirse responsabilidad a los desarrolladores, a los usuarios, a las organizaciones que implementan los sistemas o al propio sistema? La falta de claridad en este ámbito genera inseguridad jurídica y dificulta la protección de los derechos de las personas afectadas. Este problema se intensifica en el contexto de la IAG, donde los sistemas podrían operar con niveles de autonomía aún mayores (Erdélyi & Goldsmith, 2020).

Asimismo, existe un vacío significativo en relación con la protección de derechos fundamentales en entornos automatizados. Aunque muchos marcos regulatorios reconocen principios como la equidad, la no discriminación y la transparencia, su implementación práctica sigue siendo limitada. En particular, los sistemas de IA pueden afectar derechos como la privacidad, la igualdad de trato y el acceso a oportunidades, sin que existan mecanismos claros para prevenir o remediar estos impactos. Como señalan Barocas y otros. (2019), la falta de herramientas regulatorias robustas para abordar la discriminación algorítmica constituye uno de los principales desafíos en este campo.

Finalmente, es importante destacar la ausencia de enfoques regulatorios que integren de manera efectiva la dimensión social, educativa y participativa de la inteligencia artificial. La mayoría de los marcos actuales se centran en aspectos técnicos o económicos, dejando en segundo plano la necesidad de involucrar a la ciudadanía, fortalecer la alfabetización digital y promover formas de gobernanza inclusivas. Este vacío es particularmente relevante en contextos como América Latina, donde las desigualdades estructurales pueden amplificarse con la adopción de tecnologías no reguladas de manera adecuada.

En conjunto, estos vacíos normativos evidencian que la regulación de la inteligencia artificial no es simplemente una cuestión de crear nuevas leyes, sino de repensar los marcos institucionales existentes. La complejidad y el carácter transversal de la IA requieren enfoques más integrales, capaces de articular dimensiones jurídicas, éticas y sociales. Solo a través de este tipo de aproximaciones será posible avanzar hacia una gobernanza efectiva que responda a los desafíos planteados por la expansión de capacidades tecnológicas.

4. MARCOS INTERNACIONALES DE GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL

Ante la creciente expansión de la inteligencia artificial y sus implicaciones globales, diversos organismos internacionales han impulsado la construcción de marcos de gobernanza orientados a establecer principios, normas y lineamientos para su desarrollo y uso responsable. Estos esfuerzos reflejan el reconocimiento de que la IA no puede ser regulada únicamente a nivel nacional, dado su carácter transnacional y su impacto transversal en múltiples sectores. Sin embargo, estos marcos presentan también limitaciones importantes en términos de alcance, coherencia y capacidad de implementación efectiva.

Uno de los referentes más influyentes en este ámbito son los Principios de Inteligencia Artificial de la Organización para la Cooperación y el Desarrollo Económicos (OECD), adoptados en 2019. Estos principios establecen que los

sistemas de IA deben ser innovadores y confiables, respetar los derechos humanos, promover la transparencia y operar bajo marcos de responsabilidad clara (OECD, 2019). Además, enfatizan la necesidad de que los sistemas estén centrados en el ser humano y contribuyan al bienestar social. Su relevancia radica en que han sido adoptados por numerosos países y han servido como base para el desarrollo de políticas públicas en distintas regiones, consolidándose como un estándar internacional de referencia.

En una línea complementaria, la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO (2021) constituye uno de los instrumentos más amplios en términos de alcance global, al haber sido adoptado por más de 190 Estados miembros. Este marco introduce un enfoque explícitamente centrado en valores como la dignidad humana, la equidad, la inclusión, la sostenibilidad y la justicia social. A diferencia de otros instrumentos, la propuesta de la UNESCO incorpora dimensiones sociales y culturales más amplias, reconociendo la necesidad de considerar los contextos locales y las desigualdades estructurales en la gobernanza de la IA (UNESCO, 2021).

A pesar de su relevancia, ambos marcos comparten una característica fundamental: su naturaleza de instrumentos no vinculantes o “soft law”. Esto implica que, si bien establecen principios normativos importantes, su implementación depende de la voluntad de los Estados y de los actores involucrados. Como señalan Floridi y otros. (2018), la proliferación de principios éticos en el ámbito de la IA ha contribuido a generar consenso discursivo, pero no necesariamente ha logrado traducirse en mecanismos efectivos de regulación y control. Esta situación pone de manifiesto una de las principales tensiones en la gobernanza de la IA: la distancia entre los estándares normativos y su aplicación práctica.

En paralelo a estos instrumentos globales, se han desarrollado enfoques regulatorios más concretos a nivel regional, siendo el caso más destacado el de la Unión Europea, que ha impulsado un modelo basado en la clasificación de sistemas de IA según niveles de riesgo. Este enfoque busca establecer obligaciones

diferenciadas dependiendo del impacto potencial de los sistemas, lo que representa un avance hacia una regulación más específica y operativa. Sin embargo, incluso este modelo enfrenta desafíos relacionados con su implementación, la definición de categorías de riesgo y su adaptación a tecnologías en constante evolución (Veale & Borgesius, 2021).

Otro aspecto relevante en la gobernanza internacional de la IA es la creciente necesidad de coordinación multilateral. Dado que los sistemas de IA operan a través de fronteras y plataformas globales, la falta de armonización normativa puede generar inconsistencias y vacíos que dificultan su regulación efectiva. En este contexto, diversos estudios han señalado la importancia de avanzar hacia mecanismos de cooperación internacional que permitan establecer estándares comunes, compartir buenas prácticas y evitar fragmentaciones regulatorias (Tallberg y otros., 2023).

No obstante, la gobernanza global de la IA también enfrenta limitaciones derivadas de asimetrías geopolíticas y económicas. Los países con mayor capacidad tecnológica tienden a influir de manera más significativa en la definición de estándares y marcos regulatorios, lo que puede dejar en desventaja a regiones como América Latina. Esta situación refuerza la necesidad de desarrollar enfoques que no solo sean globales, sino también inclusivos y sensibles a contextos locales, incorporando perspectivas diversas en la toma de decisiones.

En síntesis, los marcos internacionales de gobernanza de la inteligencia artificial representan un avance importante en la construcción de principios compartidos y en la articulación de respuestas colectivas frente a los desafíos tecnológicos. Sin embargo, su carácter no vinculante, la fragmentación normativa y las desigualdades estructurales limitan su efectividad. Estos elementos evidencian la necesidad de transitar hacia modelos de gobernanza más robustos, capaces de integrar dimensiones jurídicas, éticas y sociales, y de responder de manera más efectiva a los retos planteados por la expansión de capacidades de la IA.

5. ¿ES POSIBLE REGULAR LA INTELIGENCIA ARTIFICIAL GENERAL? LÍMITES Y CONDICIONES DE POSIBILIDAD

La posibilidad de regular la Inteligencia Artificial General (IAG) constituye uno de los debates más complejos en la actualidad, no solo por la naturaleza aún hipotética de esta tecnología, sino por las limitaciones estructurales de los marcos regulatorios existentes. A diferencia de tecnologías previas, la IAG implicaría niveles de autonomía, adaptabilidad y capacidad de generalización que desafían las categorías tradicionales del derecho y la política pública. En este sentido, la pregunta no es únicamente si es posible regularla, sino bajo qué condiciones y con qué enfoques podría lograrse una regulación efectiva en un contexto de alta incertidumbre.

Uno de los principales obstáculos es la incertidumbre tecnológica. Dado que la IAG aún no se ha desarrollado plenamente, resulta difícil anticipar con precisión sus capacidades, aplicaciones y riesgos. Esta incertidumbre limita la capacidad de diseñar regulaciones específicas, ya que el derecho tiende a basarse en fenómenos relativamente estables y conocidos. Como señala Bostrom (2014), la dificultad de prever el comportamiento de sistemas altamente avanzados implica que cualquier intento de regulación debe enfrentarse a escenarios desconocidos, lo que reduce la eficacia de enfoques normativos tradicionales basados en la previsión y el control.

A esta incertidumbre se suma la complejidad sistémica de la inteligencia artificial, que se manifiesta en su integración en múltiples sectores sociales, económicos y políticos. La IA no es una tecnología aislada, sino una infraestructura transversal que interactúa con sistemas complejos como mercados financieros, sistemas educativos, plataformas digitales y procesos democráticos. Esta interconexión dificulta la aplicación de regulaciones sectoriales, ya que los efectos de la tecnología no se limitan a un ámbito específico. Como advierten Floridi y otros. (2018), la gobernanza de la IA requiere enfoques sistémicos capaces de abordar sus múltiples dimensiones y sus efectos indirectos.

Otro límite importante es la naturaleza global de la inteligencia artificial. Los sistemas de IA son desarrollados, desplegados y utilizados en contextos transnacionales, lo que plantea desafíos significativos para la regulación basada en jurisdicciones nacionales. La ausencia de un marco regulatorio global coherente dificulta la aplicación de normas y favorece la fragmentación normativa. En este contexto, Erdélyi y Goldsmith (2020) argumentan que la regulación efectiva de la IA requiere mecanismos de cooperación internacional que permitan coordinar esfuerzos y establecer estándares comunes, aunque reconocen las dificultades políticas y geopolíticas que esto implica.

Asimismo, la regulación de la IAG enfrenta el desafío de la asimetría entre la capacidad de los Estados y la de los actores tecnológicos. Las grandes corporaciones tecnológicas no solo concentran recursos y conocimiento, sino que también operan con una velocidad de innovación que supera la capacidad de respuesta de las instituciones públicas. Esta situación genera un desequilibrio que limita la capacidad de los Estados para supervisar, controlar y regular efectivamente el desarrollo de la IA. Como señalan Brynjolfsson y McAfee (2017), esta brecha puede traducirse en una pérdida de soberanía tecnológica y en una dependencia creciente de actores privados.

No obstante, reconocer estos límites no implica concluir que la regulación de la IAG sea imposible. Más bien, sugiere la necesidad de repensar los enfoques regulatorios tradicionales. En lugar de modelos rígidos y reactivos, diversos autores proponen avanzar hacia formas de regulación adaptativa, basadas en principios flexibles, mecanismos de supervisión continua y evaluación constante de riesgos (OECD, 2019). Este tipo de enfoques permite responder de manera más dinámica a la evolución tecnológica y reduce el desfase entre innovación y regulación.

En este sentido, también cobra relevancia la idea de una gobernanza multinivel y multiactor, que involucre no solo a los Estados, sino también a organismos internacionales, sector privado, comunidad académica y sociedad civil. La complejidad de la IA requiere procesos de toma de decisiones más inclusivos,

capaces de incorporar diversas perspectivas y conocimientos. La recomendación de la UNESCO (2021) enfatiza precisamente la importancia de promover la participación social y la inclusión en la gobernanza de la inteligencia artificial, como condición para asegurar su orientación hacia el bienestar colectivo.

Finalmente, es necesario reconocer que la regulación de la IAG no puede limitarse a instrumentos jurídicos formales. Dado su impacto en la producción de conocimiento, la toma de decisiones y la vida cotidiana, la gobernanza de la IA debe integrar también dimensiones educativas, éticas y culturales. Esto implica fortalecer la alfabetización en inteligencia artificial, promover una ciudadanía crítica y desarrollar capacidades institucionales que permitan comprender y evaluar el funcionamiento de estos sistemas. Como sugiere Russell (2019), el desafío no es solo controlar la tecnología, sino garantizar que permanezca alineada con los valores y objetivos humanos.

En conjunto, estos elementos permiten afirmar que la regulación de la Inteligencia Artificial General es posible, pero bajo condiciones específicas: requiere enfoques flexibles, cooperación internacional, participación social y una integración de dimensiones jurídicas, éticas y educativas. Más que una regulación definitiva, se trata de construir procesos de gobernanza dinámicos, capaces de adaptarse a la evolución de una tecnología en constante transformación.

6. HACIA UNA REGULACIÓN INTEGRAL DE LA INTELIGENCIA ARTIFICIAL: PROPUESTA PARA UNA GOBERNANZA SOCIALMENTE RESPONSABLE

Frente a los límites identificados en los marcos regulatorios actuales y los desafíos derivados de la expansión de capacidades de la inteligencia artificial, particularmente en el horizonte de la Inteligencia Artificial General (IAG), resulta necesario avanzar hacia un enfoque de regulación que supere las aproximaciones tradicionales centradas exclusivamente en lo jurídico. En este sentido, se propone una gobernanza integral de la inteligencia artificial, entendida como un conjunto articulado de mecanismos normativos, éticos, educativos y participativos orientados

a garantizar que el desarrollo tecnológico se alinee con el bienestar social y los principios de justicia.

Uno de los pilares fundamentales de esta propuesta es la regulación adaptativa. Dado el carácter dinámico y evolutivo de la inteligencia artificial, los marcos regulatorios deben diseñarse con capacidad de ajuste continuo, incorporando mecanismos de evaluación periódica, revisión normativa y aprendizaje institucional. Este enfoque permite reducir el desfase entre innovación y regulación, y ha sido promovido por organismos internacionales como la OECD (2019), que enfatizan la necesidad de políticas flexibles capaces de responder a contextos de incertidumbre tecnológica.

En segundo lugar, se plantea la necesidad de fortalecer una gobernanza multinivel y multiactor. La regulación de la inteligencia artificial no puede recaer exclusivamente en los Estados, sino que requiere la participación coordinada de diversos actores, incluyendo organismos internacionales, sector privado, comunidad académica y sociedad civil. Esta perspectiva reconoce que la IA es una tecnología transversal cuyos impactos afectan múltiples ámbitos de la vida social, por lo que su regulación debe construirse a partir de procesos inclusivos y deliberativos. La UNESCO (2021) ha subrayado la importancia de promover la participación social como un elemento central en la gobernanza ética de la inteligencia artificial.

Un tercer componente clave es la incorporación de la dimensión educativa como eje de regulación. Tradicionalmente, la regulación tecnológica se ha centrado en normas jurídicas y estándares técnicos, dejando en segundo plano el papel de la educación. Sin embargo, en el caso de la inteligencia artificial, resulta fundamental promover una alfabetización crítica en IA, que permita a los ciudadanos comprender, evaluar y cuestionar el funcionamiento de estos sistemas. Esta dimensión no solo contribuye a reducir la dependencia tecnológica, sino que también fortalece la capacidad de la sociedad para participar activamente en la toma de decisiones relacionadas con la tecnología. Como señala Russell (2019), el

control efectivo de la IA no depende únicamente de su diseño técnico, sino también de la capacidad humana para comprender y supervisar su funcionamiento.

Asimismo, esta propuesta incorpora un enfoque basado en la justicia social y la equidad, reconociendo que la inteligencia artificial puede amplificar desigualdades existentes si no se regula adecuadamente. En este sentido, la gobernanza de la IA debe orientarse a garantizar el acceso equitativo a sus beneficios, prevenir la discriminación algorítmica y promover el desarrollo de tecnologías que respondan a las necesidades de diversos contextos sociales. Este enfoque es consistente con los principios establecidos por la UNESCO (2021), que destacan la importancia de la inclusión y la equidad como valores centrales en el desarrollo de la inteligencia artificial.

En línea con lo anterior, se propone integrar elementos de la economía social y solidaria en la regulación de la inteligencia artificial. Este enfoque implica considerar la tecnología no solo como un producto de mercado, sino como un bien con implicaciones colectivas, cuya gobernanza debe orientarse al interés público. Desde esta perspectiva, se pueden promover modelos alternativos de desarrollo tecnológico, como plataformas cooperativas, infraestructuras digitales abiertas y proyectos de innovación social, que contribuyan a democratizar el acceso y la toma de decisiones en torno a la IA. Como señalan Floridi y otros. (2018), la construcción de una “buena sociedad de la IA” requiere no solo regulación, sino también la promoción de valores y prácticas orientadas al bien común.

Otro elemento relevante es el fortalecimiento de mecanismos de transparencia y rendición de cuentas. La complejidad de los sistemas de IA exige el desarrollo de herramientas que permitan auditar su funcionamiento, identificar posibles sesgos y garantizar la responsabilidad de los actores involucrados. Esto incluye la implementación de evaluaciones de impacto, auditorías algorítmicas y estándares de explicabilidad, que permitan hacer visibles los procesos de toma de decisiones automatizados. La literatura ha destacado que la transparencia es una condición

necesaria, aunque no suficiente, para una gobernanza efectiva de la inteligencia artificial (Doshi-Velez & Kim, 2017).

Finalmente, se plantea la necesidad de avanzar hacia una cooperación internacional más robusta, capaz de articular esfuerzos entre países y regiones. Dado el carácter global de la inteligencia artificial, ningún Estado puede regularla de manera aislada. Esto implica no solo la adopción de principios comunes, sino también la construcción de mecanismos institucionales que permitan coordinar políticas, compartir información y enfrentar desafíos transnacionales. Como señalan Tallberg y otros. (2023), la gobernanza global de la IA es un proceso en construcción que requiere voluntad política, cooperación y reconocimiento de las asimetrías existentes.

En conjunto, esta propuesta sugiere que la regulación de la inteligencia artificial, especialmente en el contexto de la IAG, debe entenderse como un proceso integral que articule múltiples dimensiones. Más que establecer un conjunto cerrado de normas, se trata de construir un ecosistema de gobernanza capaz de adaptarse a la evolución tecnológica, promover la participación social y garantizar que el desarrollo de la inteligencia artificial contribuya al bienestar colectivo. En este sentido, la regulación no debe concebirse como un límite al progreso tecnológico, sino como una herramienta para orientarlo hacia fines socialmente deseables.

CONCLUSIONES

El análisis desarrollado a lo largo de este capítulo permite afirmar que la Inteligencia Artificial General (IAG), más que una realidad plenamente alcanzada, constituye un horizonte tecnológico cuya expansión progresiva de capacidades ya está reconfigurando estructuras sociales, económicas y cognitivas. En este contexto, la pregunta sobre su regulación no puede abordarse en términos tradicionales, ya que los marcos normativos existentes resultan insuficientes frente a una tecnología caracterizada por la incertidumbre, la complejidad y su naturaleza transnacional.

Los riesgos emergentes identificados —concentración de poder, opacidad algorítmica, sesgos, dependencia cognitiva y vulnerabilidades sistémicas—

evidencian que la inteligencia artificial no es un fenómeno meramente técnico, sino un proceso profundamente político y social, cuyos efectos inciden en la distribución del poder, la producción de conocimiento y el ejercicio de derechos fundamentales. En este sentido, los vacíos normativos no deben interpretarse únicamente como carencias regulatorias, sino como manifestaciones de una tensión estructural entre la velocidad de la innovación tecnológica y la capacidad de respuesta de las instituciones.

Asimismo, el análisis de los marcos internacionales de gobernanza muestra avances importantes en la construcción de principios compartidos, pero también revela limitaciones significativas derivadas de su carácter no vinculante, su fragmentación y las asimetrías globales que condicionan su implementación. Estos elementos refuerzan la idea de que la regulación de la inteligencia artificial no puede depender exclusivamente de instrumentos jurídicos tradicionales ni de iniciativas aisladas, sino que requiere enfoques más integrales y coordinados.

En este contexto, la pregunta central —si es posible regular la Inteligencia Artificial General— encuentra una respuesta matizada: sí es posible, pero no bajo los paradigmas regulatorios convencionales. La regulación efectiva de la IA exige repensar las formas de gobernanza, transitando hacia modelos adaptativos, participativos y multidimensionales. Esto implica reconocer que la regulación no es un acto puntual, sino un proceso dinámico que debe evolucionar junto con la tecnología.

La propuesta planteada en este capítulo apunta precisamente en esa dirección, al integrar dimensiones jurídicas, éticas, educativas y sociales en un enfoque de gobernanza orientado al bienestar colectivo. En particular, la incorporación de la educación como eje regulador, el fortalecimiento de la participación social y la articulación con principios de justicia y equidad constituyen elementos clave para construir una regulación más robusta y legítima. De igual manera, la inclusión de perspectivas provenientes de la economía social y solidaria abre la posibilidad de

imaginar formas alternativas de desarrollo tecnológico, menos centradas en la lógica de mercado y más orientadas al bien común.

No obstante, avanzar hacia este tipo de gobernanza implica enfrentar desafíos significativos, entre ellos la necesidad de fortalecer las capacidades institucionales, reducir las asimetrías tecnológicas y promover una cooperación internacional más efectiva. En particular, resulta fundamental incorporar las voces y experiencias de regiones como América Latina, que históricamente han tenido una participación limitada en la definición de estándares tecnológicos globales, pero que enfrentan de manera directa los impactos de estas tecnologías.

En última instancia, la discusión sobre la regulación de la Inteligencia Artificial General trasciende el ámbito técnico y jurídico para situarse en el terreno de las decisiones colectivas sobre el futuro de la sociedad. No se trata únicamente de controlar una tecnología emergente, sino de definir qué tipo de relación queremos establecer con ella y bajo qué principios queremos orientar su desarrollo. En este sentido, la regulación de la inteligencia artificial debe entenderse como una herramienta para garantizar que la expansión de sus capacidades no se traduzca en una ampliación de desigualdades, sino en una oportunidad para fortalecer la justicia social, la democracia y el bienestar humano.

Así, más que una limitación al progreso, la regulación de la inteligencia artificial constituye una condición necesaria para su legitimidad y sostenibilidad. El desafío no radica en frenar la innovación, sino en orientarla de manera consciente, crítica y responsable, asegurando que los beneficios de la tecnología sean compartidos y que sus riesgos sean gestionados de forma equitativa. Solo a través de este enfoque será posible construir una gobernanza de la inteligencia artificial capaz de responder a los desafíos del presente y a las incertidumbres del futuro.

Referencias

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities*. MIT Press.

Bommasani, R., Hudson, D. A., Adeli, E., y otros. (2021). *On the opportunities and risks of foundation models*. arXiv.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Brynjolfsson, E., & McAfee, A. (2017). *The business of artificial intelligence*. Harvard Business Review.

Corrêa, N. K., Galvão, C., Santos, J. W., y otros. (2022). *Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance*. arXiv.

Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. arXiv.

Erdélyi, O. J., & Goldsmith, J. (2020). *Regulating artificial intelligence: Proposal for a global solution*. arXiv.

Floridi, L., Cowls, J., Beltrametti, M., y otros. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707.

Goertzel, B., & Pennachin, C. (2007). *Artificial general intelligence*. Springer.

Kaplan, J., McCandlish, S., Henighan, T., y otros. (2020). *Scaling laws for neural language models*. arXiv.

OECD. (2019). *OECD Principles on Artificial Intelligence*.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Tallberg, J., y otros. (2023). *The global governance of artificial intelligence*. arXiv.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(

LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL



La inteligencia artificial está transformando profundamente las estructuras sociales, económicas y jurídicas a nivel global. Ante este panorama, la regulación de estas tecnologías emergentes se convierte en un desafío central para los Estados, las instituciones y la comunidad internacional.



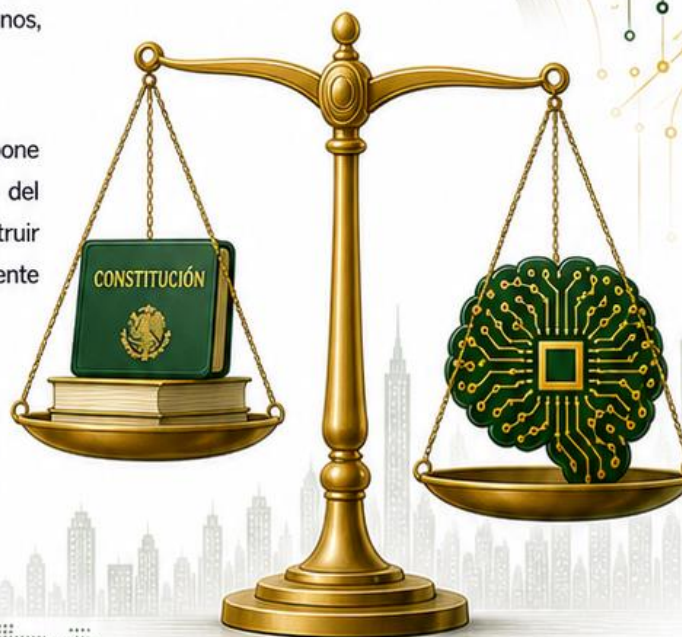
La presente obra colectiva reúne diversas perspectivas académicas y profesionales en torno a los principales retos de la regulación de la inteligencia artificial, abordando temas como la ética algorítmica, la protección de derechos humanos, la gobernanza digital y la responsabilidad jurídica.



A través de un enfoque interdisciplinario, este libro propone un análisis crítico que sitúa a la persona en el centro del desarrollo tecnológico, destacando la necesidad de construir marcos normativos que garanticen un uso justo, transparente y responsable de la inteligencia artificial.



Esta obra está dirigida a estudiantes, investigadores, profesionales del derecho y tomadores de decisiones interesados en comprender y contribuir a la regulación de la inteligencia artificial en México y el mundo.



DERECHOS
HUMANOS



ÉTICA
ALGORÍTMICA



GOBERNANZA
DIGITAL



RESPONSABILIDAD
JURÍDICA



IMPACTO
SOCIAL