PRIMERA EDICIÓN

e-Estado 2026

Inteligencia artificial, sociedad y Estado:

Desafíos jurídicos, éticos e institucionales

Pontificia Universidad JAVERIANA Bogotá

Observatorio de Gobierno y TIC

Escuela Javeriana de Gobierno y Ética Pública

e ESTADO

E-ESTADO

Primera Edición

REVISTA E-ESTADO
Es una publicación de la Escuela Javeriana de Gobierno y Ética Pública y el Observatorio de Gobierno y TIC de la Universidad Javeriana.

FRECUENCIA
Anual

AÑO DE EDICIÓN
2026

NÚMERO
Uno

UNIVERSIDAD JAVERIANA
LUIS FERNANDO MUNERA S.J.
Rector
LUIS JAVIER SARRALDE DELGADO S.J.
Secretario General

ESCUELA JAVERIANA DE GOBIERNO Y ÉTICA PÚBLICA
LUIS FERNANDO ÁLVAREZ LONDOÑO S.J.
Director
MARIANA PLAZAS JÁCOME
Coordinadora Académica y de Relacionamiento Estratégico e Internacional
NATALIA REYES FERNÁNDEZ
Profesional de Gestión Académica y Administrativa
LINA MARÍA RENDÓN ZAMBRANO
Auxiliar Administrativa

OBSERVATORIO DE GOBIERNO Y TIC
HÉCTOR JOSÉ GARCÍA SANTIAGO
Director
ANDRÉS EDUARDO SALCEDO CAMACHO
Coordinador

CONSEJO EDITORIAL
LUIS FERNANDO ÁLVAREZ LONDOÑO S.J.
MARIANA PLAZAS JÁCOME
HÉCTOR JOSÉ GARCÍA SANTIAGO
ANDRÉS EDUARDO SALCEDO CAMACHO

Correo electrónico:
observatoriogobtic@javeriana.edu.co

© Derechos reservados
El contenido de los artículos es de exclusiva responsabilidad de los autores. Los textos pueden reproducirse total o parcialmente citando la fuente.

ÍNDICE


Foto: Magnific

04 **Editorial**
Luis Fernando Álvarez Londoño, S.J.

08 **Prólogo**
Alexandra Falla Zerrate

INTELIGENCIA ARTIFICIAL, ESTADO, JUSTICIA Y GOBERNANZA

11 **01: Inteligencia artificial en la gestión pública: caso colombiano**
María Stella Arismendy y Fabián Andrés Mora

33 **02: Inteligencia artificial generativa y control fiscal en Colombia: análisis económico de la detección de detrimentos patrimoniales sistemáticos**
Carlos Pablo Márquez Escobar

56 **03: Responsabilidad civil derivada de daños causados por la inteligencia artificial y robots autónomos: estado del arte doctrinal, análisis de casos y desafíos regulatorios**
Héctor José García Santiago

87 **04: La inteligencia artificial, clave para la eficiencia del proceso laboral en el Perú**
Homero Jóndec Briones

108 **05: Prevención de lavado de activos con activos físicos y criptoactivos. Inteligencia financiera, debida diligencia, GAFI y GAFILAT**
Rubén Flores Dapkevicius

INTELIGENCIA ARTIFICIAL, SOCIEDAD, DEMOCRACIA Y CIUDADANÍA

140 **01: Algunas reflexiones sobre el impacto de la inteligencia artificial en la sociedad (y viceversa)**
José Luis Verdegay y David Pelta

162 **02: Entre el pesimismo y la esperanza. De cómo los valores democráticos pueden impedir los efectos más perversos de la inteligencia artificial**
Daniel Jove Villares

182 **03: Inteligencia artificial en el aula: desmontando mitos y trazando proyecciones informadas**
Francisco García

INTELIGENCIA ARTIFICIAL, DERECHO, CREATIVIDAD Y REGULACIÓN

214 **01: La IA, el derecho de autor y el futuro de la creatividad: notas desde la Feria Internacional del Libro de Panamá**
Andrés Izquierdo

222 **02: La inteligencia artificial como ¿artista?: del sentimiento al algoritmo, de los derechos de autor a la esencia del ser humano**
Andrés Eduardo Salcedo

252 **Análisis**
Percepciones sobre la inteligencia artificial. Usos, confianza, riesgos y apropiación social.

265 **Semillero de investigación**
Servicios ciudadanos digitales en Colombia: estado actual, desafíos y perspectivas para la política pública
Camilo Charris

Editorial
REVISTA E-ESTADO
Primera Edición

Con especial satisfacción presentamos la primera edición de la Revista e-Estado, publicación académica del Observatorio de Gobierno y TIC de la Escuela Javeriana de Gobierno y Ética Pública. El nacimiento de una Revista es siempre un acontecimiento institucional relevante: expresa una apuesta por la deliberación rigurosa, por la construcción de comunidad académica y por la intervención responsable de la Universidad en los debates que marcan su tiempo. En este caso, además, representa la consolidación de un proyecto intelectual que busca ofrecer un espacio estable para pensar críticamente la relación entre tecnología, derecho, ética pública, gobierno e instituciones.

La creación de esta Revista responde a una convicción de fondo: los procesos de transformación tecnológica ya no pueden comprenderse únicamente desde la ingeniería o desde la innovación empresarial. Su impacto se proyecta sobre la organización del Estado, el ejercicio de los derechos, la prestación de servicios públicos, la producción normativa, la educación, la administración de justicia, la cultura y, en general, sobre la forma misma en que las sociedades contemporáneas se gobiernan y se comprenden. Por ello, la reflexión académica sobre estos asuntos exige enfoques interdisciplinares, sentido público y una preocupación constante por las implicaciones éticas de toda innovación.

No es casual, entonces, que el eje temático de esta primera edición sea la inteligencia artificial, dado que pocas cuestiones sintetizan mejor los desafíos del presente. La inteligencia artificial se ha convertido en uno de los principales vectores de transformación de las instituciones, de los mercados, de los procesos decisorios y de la vida cotidiana. Sin embargo, su creciente incorporación en múltiples sectores no puede ser leída únicamente en clave de eficiencia o novedad: también obliga a preguntarnos por la transparencia, la responsabilidad, la equidad, la protección de los derechos fundamentales, la preservación de la autonomía humana y la orientación del cambio tecnológico hacia el bien común.

Desde la vocación humanista de la Pontificia Universidad Javeriana, esta discusión adquiere un sentido particular. Pensar la inteligencia artificial desde la ética pública y desde el gobierno implica reconocer que no toda capacidad tecnológica se traduce, sin más, en progreso social. La pregunta de fondo no es solamente qué puede hacerse con estas herramientas, sino qué debe hacerse con ellas, bajo qué límites, con qué salvaguardas y al servicio de qué fines. En otras palabras, se trata de insistir en que la innovación debe permanecer vinculada a la dignidad humana, a la justicia, al interés general y a la responsabilidad institucional.



Foto: Pontificia Universidad Javeriana

Esta primera edición reúne, en ese sentido, una muestra amplia y plural de aproximaciones al fenómeno, permitiendo abordar la inteligencia artificial desde la interdisciplinariedad y diversidad temática. Además de ello, uno de los rasgos más valiosos de la Revista es la nutrida participación internacional, con contribuciones de autores procedentes de distintos países y tradiciones académicas, lo que enriquece de manera sustantiva la conversación.

Junto a voces colombianas, el número incorpora reflexiones y análisis elaborados desde España, Perú y Uruguay, entre otros contextos, permitiendo contrastar experiencias, enfoques regulatorios, preocupaciones sociales y desarrollos sectoriales. Esta diversidad confirma que la inteligencia artificial constituye un asunto global, pero también que sus efectos deben ser pensados a la luz de realidades institucionales concretas.

Un primer bloque temático del número se concentra en la relación entre inteligencia artificial, Estado, gobierno y gestión pública. Allí se ubican contribuciones que examinan el papel de la inteligencia artificial en la gestión pública colombiana, su posible aprovechamiento en tareas de control fiscal, así como el estado actual y los desafíos de los servicios ciudadanos digitales en Colombia. En conjunto, estos trabajos muestran que la discusión sobre inteligencia artificial en el ámbito público no se agota en la incorporación de herramientas tecnológicas, sino que exige revisar capacidades estatales, marcos de gobernanza, calidad regulatoria, control institucional y formas de interacción entre la administración y la ciudadanía.

Un segundo bloque aborda los vínculos entre inteligencia artificial, derecho, democracia y valores públicos. En este grupo se inscriben reflexiones sobre el derecho de autor y el futuro de la creatividad, sobre la forma en que los valores democráticos pueden contener los efectos más problemáticos de la automatización, y sobre el impacto de la inteligencia artificial en la sociedad contemporánea. A ello se suma una aproximación filosófico-jurídica al problema de la inteligencia artificial como posible artista, así como un análisis sobre prevención de lavado de activos y criptoactivos en un entorno tecnológico cada vez más complejo. Se trata, en todos los casos, de textos que recuerdan que la discusión sobre inteligencia artificial es inseparable de las preguntas por la normatividad, la libertad, la cultura, la responsabilidad y la protección de bienes jurídicos colectivos.

Un tercer conjunto de aportes se orienta hacia los usos sectoriales y las aplicaciones específicas de la inteligencia artificial. Destacan aquí los trabajos sobre inteligencia artificial en el aula y sobre su potencial para mejorar la eficiencia del proceso laboral en el Perú, así como la reflexión sobre el control de vehículos y robots autónomos. Estos artículos permiten apreciar que la inteligencia artificial no irrumpe de manera abstracta, sino en escenarios concretos donde se redefinen prácticas profesionales, esquemas de responsabilidad, modelos pedagógicos y criterios de decisión. Tal constatación invita a evitar tanto el entusiasmo acrítico como el rechazo simplificador: cada ámbito requiere examen propio, regulación proporcionada y una deliberación informada sobre riesgos y oportunidades.

La presente edición incluye, además, dos contribuciones especiales que merecen destacarse por su valor dentro del proyecto académico del Observatorio. De una parte, el artículo sobre servicios ciudadanos digitales, resultado del trabajo del semillero del Observatorio, que pone de relieve la importancia de formar nuevas generaciones de investigadores capaces de vincular el análisis académico con problemas públicos concretos. Junto a ello, el artículo derivado de la encuesta realizada a ciudadanos sobre percepciones, riesgos y expectativas frente a la inteligencia artificial aporta una dimensión empírica particularmente valiosa: escuchar cómo la ciudadanía entiende, teme, proyecta o valora estas transformaciones es indispensable para cualquier agenda seria de gobierno y ética pública.

Como puede advertirse, el número no ofrece una visión única ni cerrada sobre la inteligencia artificial, sino que por el contrario, propone una conversación diversa que combina aproximaciones jurídicas, institucionales, sociales, filosóficas y aplicadas. Esa pluralidad no es accidental; responde a la convicción editorial de que los grandes debates contemporáneos deben abordarse desde múltiples perspectivas, especialmente cuando comprometen decisiones públicas y transformaciones profundas en la vida social. La Revista quiere ser, precisamente, un espacio de encuentro entre disciplinas, trayectorias y experiencias que, desde el rigor académico, contribuyan a comprender mejor los desafíos de nuestro tiempo.

Con esta primera edición, la Revista e-Estado aspira también a fortalecer el posicionamiento académico de la Escuela Javeriana de Gobierno y Ética Pública y de su Observatorio de Gobierno y TIC como escenarios de referencia para el estudio de temas de actualidad tecnológica, ética y de gobierno. Se trata de una apuesta que desborda la publicación de un solo número, puesto que cada edición buscará ofrecer un eje temático central y convocar, alrededor de él, debates capaces de articular reflexión crítica, pertinencia pública y calidad académica. En esa medida, la Revista se proyecta como una plataforma de diálogo y producción de conocimiento al servicio de una comprensión más profunda de los vínculos entre innovación tecnológica e instituciones democráticas.

Queremos expresar un agradecimiento especial a los autores que confiaron en esta iniciativa y contribuyeron con generosidad intelectual al lanzamiento de la revista, dado que su participación ha sido crucial para poder materializar este proyecto en una Revista cuyo número inaugural es robusto, plural y de gran actualidad. Agradecemos igualmente a quienes han acompañado los procesos editoriales y académicos necesarios para poner en marcha esta publicación.

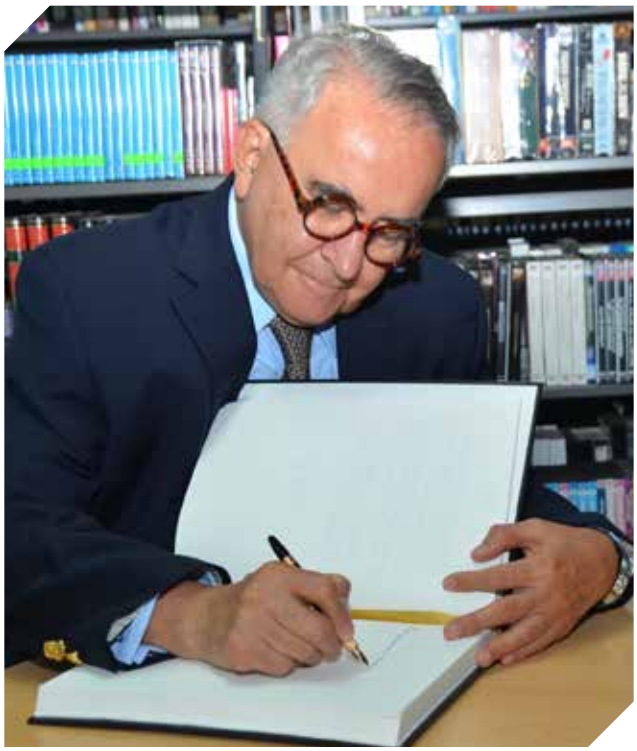


Foto: La Fundación Global Democracia y Desarrollo (Funglode)

Toda Revista nace del trabajo conjunto, de la paciencia institucional y de la convicción compartida de que el conocimiento riguroso sigue siendo indispensable para orientar responsablemente el cambio social.

Confiamos en que esta primera edición de la Revista e-Estado contribuya a enriquecer la discusión pública y académica sobre la inteligencia artificial, y que sirva también como una invitación a seguir pensando, con seriedad y sentido ético, el lugar de la tecnología en la construcción de nuestras sociedades. Presentamos este número con esperanza y con responsabilidad, conscientes de que reflexionar sobre la tecnología es, en buena medida, reflexionar sobre el tipo de Estado, de ciudadanía y de comunidad política que queremos promover.

Luis Fernando Álvarez Londoño, S.J.
Director de la Escuela Javeriana de Gobierno y Ética Pública

PRÓLOGO

Revista e-Estado · Primera edición

Cada transformación tecnológica trae consigo una promesa de futuro, pero también una responsabilidad colectiva: decidir para qué y para quién ponemos la tecnología en movimiento. La inteligencia artificial no es la excepción, porque redefine la manera de aprender, crear, trabajar, administrar justicia y relacionarnos con el Estado y ante los retos propios de estas capacidades necesitamos conocimiento riguroso, deliberación pública y una mirada ética capaz de orientar el cambio hacia el bienestar de las personas. Por ello celebro la aparición de la Revista e-Estado, publicación del Observatorio de Gobierno y TIC de la Escuela Javeriana de Gobierno y Ética Pública. Su primera edición no representa solamente el nacimiento de una revista académica, sino que constituye la apertura de un espacio de encuentro entre la investigación, las instituciones y la ciudadanía, donde la academia puede acompañar las políticas públicas del sector TIC desde distintas perspectivas. Este número inaugural asume con acierto uno de los debates decisivos de nuestro tiempo. Sus páginas examinan la inteligencia artificial desde la gestión pública, el control fiscal, la educación, la justicia, la democracia, los derechos de autor, la creación artística y los desafíos que plantean los criptoactivos. Esa diversidad recuerda que la transformación digital no es un asunto exclusivamente técnico, sino que en ella se juegan derechos, oportunidades, confianza institucional y la posibilidad de construir un Estado más cercano, eficiente e incluyente. Como Ministra de Tecnologías de la Información y las Comunicaciones, considero fundamental que la política pública dialogue de manera permanente con espacios académicos independientes y plurales. Una política TIC sólida debe aprender de la investigación, contrastar sus decisiones con la experiencia y mantener abiertas las preguntas sobre sus efectos sociales, así como reconocer que no habrá una transformación digital verdaderamente democrática sin educación: alfabetización digital para la ciudadanía, formación para docentes y servidores públicos, desarrollo de capacidades en los territorios y herramientas para que todas las personas puedan comprender, aprovechar y cuestionar las tecnologías que inciden en sus vidas.



Foto: Universidad Externado

No basta con ampliar el acceso a las herramientas, sino que es indispensable formar criterio para utilizarlas con seguridad, creatividad y responsabilidad. Revistas como e-Estado aportan a esa tarea al traducir debates complejos, conectar disciplinas y poner conocimiento de calidad al servicio de la conversación pública.

Invito a leer esta primera edición con curiosidad y espíritu crítico, proponiendo que sus artículos no sean únicamente respuestas, sino puntos de partida para nuevas preguntas, mejores decisiones y colaboraciones duraderas. Ojalá este primer número inaugure una trayectoria fecunda y consolide a la Revista e-Estado como un referente para pensar la tecnología desde el interés general, porque acompañar las políticas TIC también significa estudiarlas, debatirlas y ayudar a que estén siempre al servicio de las personas y de una sociedad más justa.

Alexandra Falla Zerrate
Ministra de Tecnologías de la Información y las Comunicaciones



Foto: Magnific

01 Inteligencia artificial en la gestión pública

E-ESTADO
Primera Edición

Inteligencia artificial en la gestión pública: caso colombiano

Artificial Intelligence in Public Management: The Colombian Case

Fecha de recepción: 14 de agosto de 2025
Fecha de aceptación: 18 de febrero de 2026

María Stella Arismendy Ramírez
Universidad de Córdoba, Colombia
mariaarismendyr@correo.unicordoba.edu.co
ORCID: https://orcid.org/0000-0001-6798-7012

Fabian Andrés Mora Soto
Universidad Pontificia Bolivariana, Colombia
mora.fsoto@gmail.com
ORCID: https://orcid.org/0009-0006-9406-4102

Cómo citar: Arismendy Ramírez, María Stella y Fabián Andrés Mora Soto. “Inteligencia artificial en la gestión pública: caso colombiano”. Revista e-Estado 1, n.° 1 (2026).



Foto: Magnific

RESUMEN

El presente artículo analiza el rol emergente de la inteligencia artificial (IA) en el ámbito de la gestión pública en Colombia, mediante una postura propositiva y crítica. A través de la valoración de informes elementales como el Documento CONPES 4144 de 2025,1 el manual Uso responsable de IA para política pública2 y el balance presentado por Vargas,3 se determinan progresos institucionales y normativos enfocados en el desenvolvimiento de una política de Estado en torno al uso de la IA. De forma concreta, se subrayan los compromisos para establecer fundamentos éticos y aparatos administrativos, así como estrategias de cooperación interinstitucional, con el objeto de asegurar una implementación regulada de estas herramientas digitales. No obstante, esta valoración resalta también un conjunto de retos recurrentes, como la ausencia de formación técnica en el territorio nacional, o la poca intervención ciudadana en el diseño de proyectos cimentados en IA y estrategias de evaluación de impacto algorítmico. Adicionalmente, se exponen los peligros vinculados con la opacidad algorítmica y la falta de equidad para la obtención de instrumentos digitales a nivel regional. Finalmente, se presentan recomendaciones para fortalecer la gobernanza algorítmica en el sector público, con énfasis en el respeto de los derechos fundamentales, la equidad territorial y la rendición democrática de cuentas.

Palabras clave: inteligencia artificial, gestión pública, gobernanza digital, política pública, ética algorítmica, Colombia

ABSTRACT

This academic reflection article examines the emerging role of artificial intelligence (AI) in public administration in Colombia from a critical and forward-looking perspective. Based on the analysis of key sources such as Document CONPES 4144 of 2025, the Manual for the responsible use of AI in public policy, and the assessment presented by Vargas, the paper identifies institutional and regulatory advances aimed at developing a national AI policy. Notably, it highlights efforts to establish ethical principles, governance frameworks, and inter-institutional coordination strategies to ensure the responsible adoption of these technologies. However, the article also identifies persistent challenges. Among these are the lack of technical capacity at the territorial level, the absence of mandatory mechanisms for algorithmic impact assessment, and limited public participation in the design and implementation of AI-based projects. Additionally, the paper warns of risks related to algorithmic opacity, the concentration of automated decision-making in sensitive sectors such as security, and unequal access to technological tools across regions. Finally, the article presents a series of policy recommendations to strengthen algorithmic governance in the public sector, emphasizing the protection of fundamental rights, territorial equity, and democratic accountability.

Keywords: Artificial Intelligence, Public Administration, Digital Governance, Public Policy, Algorithmic Ethics, Colombia

1 Departamento Nacional de Planeación (DNP), Documento CONPES 4144: Política Nacional de Inteligencia Artificial (Bogotá: DNP, 2025). **2** Banco Interamericano de Desarrollo (BID), Uso responsable de IA para política pública: manual de formulación de proyectos (Washington: BID, 2021). **3** Carlos Vargas, "Balance de la política pública de inteligencia artificial y transformación digital (2019-2024)", Revista Internacional del Instituto de Pensamiento Liberal 1, n.o 2 (2024): 261.

INTRODUCCIÓN

Colombia ha adoptado un enfoque proactivo para incorporar la IA en su aparato estatal. Desde el Documento CONPES 3975 de 2019[4] hasta la reciente adopción del Documento CONPES 4144 de 2025,[5] se ha construido una hoja de ruta que busca posicionar al país como líder regional en la adopción responsable de tecnologías emergentes. Este esfuerzo ha estado acompañado de una narrativa institucional que asocia la IA con el fortalecimiento de la competitividad, la eficiencia administrativa y la innovación pública. El país incluso fue uno de los primeros en América Latina en estructurar una política nacional de IA articulada con los principios de ética, gobernanza y derechos humanos.[6]

No obstante, como advierte Vargas,[7] la implementación real de esta política enfrenta múltiples desafíos institucionales, técnicos y éticos. Uno de los principales problemas ha sido la fragmentación de responsabilidades entre entidades públicas, lo cual genera duplicidad de esfuerzos, vacíos regulatorios y falta de coordinación interinstitucional. Además, la ausencia de estándares de transparencia y explicabilidad en los sistemas de decisión automatizada (SDA) pone en riesgo la legitimidad de las decisiones públicas cuando estas se delegan total o parcialmente a algoritmos.[8]

Aunado a lo anterior, se ha identificado que la mayoría de las iniciativas en IA no pasan por procesos de evaluación de impacto ético o social, ni cuentan con mecanismos de rendición de cuentas efectivos.[9] Tal como lo advierte Londoño[10] en su análisis sobre el despliegue institucional de la IA en Colombia, existe un desfase entre el diseño normativo y las capacidades reales de implementación, particularmente en los niveles subnacionales.

Las entidades territoriales no cuentan con suficiente infraestructura, talento humano ni marcos de referencia para integrar tecnologías avanzadas sin comprometer el cumplimiento de principios fundamentales como la equidad, la inclusión y la protección de datos personales.

En este contexto, la transformación digital del Estado colombiano mediante el uso de IA no puede entenderse únicamente como un proceso técnico o administrativo, dado que implica profundos alcances políticos, jurídicos y sociales que deben abordarse desde una gobernanza reflexiva, abierta y participativa. Como lo señala el BID,[11] el uso responsable de IA en política pública requiere que las decisiones automatizadas sean auditables, explicables y reversibles, elementos aún débiles en muchos sistemas implementados.

Por ello, esta reflexión propone analizar los avances normativos, los retos éticos y los riesgos institucionales que plantea la IA en la gestión pública colombiana. El objetivo no es solo valorar el impacto de las políticas actuales, sino también identificar propuestas que permitan fortalecer una gobernanza algorítmica centrada en los derechos humanos y en la rendición democrática de cuentas.

4 DNP, Documento CONPES 3975: Política Nacional para la Transformación Digital e Inteligencia Artificial (Bogotá: DNP, 2019). **5** DNP, Documento CONPES 4144. **6** BID, Uso responsable de IA…, 16-18. **7** Vargas, "Balance de la política pública…", 261. **8** Ibíd., 263. **9** Ibíd., 264.**10** Juan Londoño, "Inteligencia artificial y gestión pública en Colombia: análisis de su despliegue institucional", Revista de la Universidad del Zulia 14, n.o 42 (2023): 234-236. **11** BID, Uso responsable de IA..., 45-46.

01 INTELIGENCIA ARTIFICIAL Y GESTIÓN PÚBLICA

En el ámbito de la gestión pública, la IA alude al empleo de algoritmos y sistemas de computación que permiten asistir y automatizar operaciones para la ejecución de acciones, valoración de información, designación de recursos y planeación de políticas. De conformidad con la categorización acogida en este país, la IA se refiere a la simulación que se realiza alrededor del razonamiento humano a través de una computadora.[12] En esa medida, su aplicación tiene que asociarse tanto a criterios de eficiencia como a principios de transparencia, responsabilidad y justicia.

Para la gestión pública, este recurso cuenta con la facultad de aportar al robustecimiento de las funciones administrativas del Estado, la singularización de servicios públicos y la optimización de la participación de los ciudadanos. Sin embargo, su empleo debe valorarse de forma rigurosa, sobre todo por su incidencia legal, ética y social. De acuerdo con el BID,[13] se trata de cambiar la dinámica institucional según lineamientos de equidad, legitimidad y gobernanza, lo que trasciende el mero acto de digitalizar funciones.

Aunado a esto, la incidencia de la IA debe superar los términos de eficiencia, al enfocarse en su facultad para incentivar la confianza de los ciudadanos, asegurar derechos y promover el acceso justo al avance tecnológico. Ahora bien, esto supone que la IA, en materia pública, tiene que ser auditable, transparente y articulada con la participación civil, sobre todo cuando incide en decisiones que trastocan derechos como la salud, la educación y la justicia.[14]

1.1 AVANCES NORMATIVOS E INSTITUCIONALES: CASO COLOMBIANO

El Documento CONPES 4144[15] propone seis pilares elementales para el desenvolvimiento y aplicación de la IA: gobernanza, formación de talento, ética, desarrollo empresarial, ciencia y tecnología, y articulación institucional. De tal modo, mediante lo anterior se plantea incentivar los elementos adecuados para el empleo consciente de la IA en los procesos estatales, lo que supone el acatamiento de principios como sostenibilidad e innovación.

Asimismo, se plantea la formulación del Sistema Nacional de Inteligencia Artificial, que posibilite articular compromisos entre organismos privados, públicos y académicos. Igualmente, se propone el robustecimiento de las facultades territoriales para la aplicación de programas de IA, como capital para contrarrestar la inequidad digital entre distintos sectores de la nación.[16]

Al respecto, el BID[17] proporciona una metodología prolija que permite aplicar la herramienta en mención para el diseño de políticas públicas.

12 DNP, Documento CONPES 4144, 9. **13** BID, Uso responsable de IA..., 11. **14** BID, Uso responsable de IA..., 19. **15** DNP, Documento CONPES 4144, 13-14. **16** Ibíd., 17-18. **17** BID, Uso responsable de IA..., 31-46.

En efecto, este instrumento tiene en cuenta distintas fases de aplicación: diagnóstico, planificación, diseño, implementación y evaluación, además de que incluye principios de explicabilidad, participación civil y gestión de riesgos. Su modelo multidimensional y multifactorial resulta crucial para contrarrestar los retos que implica la gobernanza algorítmica en escenario de intervención ciudadana.

Por su parte, según Vargas,[18] y de conformidad con el balance ejecutado en su análisis, se han llevado a cabo alrededor de cien iniciativas de SDA en la nación, pero la mayoría de ellas no cuenta con una evaluación de impacto, mecanismos de monitoreo robustos o aparatos de rendición de cuentas. Esto revela una distancia considerable entre el diseño normativo y la capacidad de implementación efectiva en la administración pública.

Adicionalmente, Londoño[19] advierte que la IA en Colombia se ha concentrado en sectores como seguridad, vigilancia y asignación de subsidios, sin que exista una evaluación de sus efectos distributivos o de su conformidad con principios constitucionales como igualdad o debido proceso. Ante esta situación, cobra importancia que el país avance hacia una regulación más robusta, que incluya protocolos de evaluación algorítmica y directrices vinculantes sobre el uso de datos sensibles.

Tabla 1. Análisis normativo del Documento CONPES 4144[20]

Sección del documento	Contenido principal	Implicaciones clave para la administración pública
Introducción y objetivos	La IA se presenta como una herramienta versátil para resolver problemas sociales, económicos y ambientales, como la seguridad alimentaria o la pobreza. El objetivo principal es construir capacidades para la investigación, el desarrollo, la adopción y el empleo ético de la IA hasta 2030, con seis metas específicas: gobernanza ética, infraestructura, I+D, talento humano, mitigación de riesgos y promoción de su uso. Se alinea con estándares de la OCDE y la Unesco, priorizando valores humanos, equidad y sostenibilidad.	Obliga a una colaboración entre instituciones para alinear objetivos nacionales de desarrollo, destinando fondos a la participación multisectorial y a la modernización estatal, lo que impulsa una transformación integral del sector público.
Diagnóstico: desafíos y oportunidades	Se identifican barreras como gobernanza débil, cobertura de internet baja (36,1 % de hogares sin acceso), inversión limitada en I+D, desajuste de habilidades digitales (79.° a nivel global) y riesgos en privacidad y desempleo. Las oportunidades incluyen la automatización para lograr una mayor productividad y sus aplicaciones en sectores como la agricultura, la salud y la justicia, con puntajes positivos en índices globales, pero con lagunas en innovación.	Demanda inversiones en infraestructura y en capacitación para reducir brechas digitales, incorporando la IA en áreas prioritarias como la salud y la agricultura para fomentar la equidad y la eficiencia en la prestación de servicios públicos.

18 Vargas, "Balance de la política pública…", 263. **19** Londoño, "Inteligencia artificial y gestión pública…", 234-236. **20** DNP, Documento CONPES 4144.

Riesgos éticos y sociales	Incluye amenazas como ciberataques, la pérdida de empleos (26-38 % en riesgo), violaciones de privacidad (deepfakes), la incertidumbre en materia de propiedad intelectual, el aumento de desigualdades, los impactos ambientales (alto consumo energético), los efectos en salud mental y sesgos que afectan grupos vulnerables (étnicos, LGBTIQ+).	Requiere de marcos regulatorios sólidos y de vigilancia continua para salvaguardar los derechos, fomentando la inclusión y la confianza en el empleo público de la IA, con un énfasis en las evaluaciones de impacto para prevenir daños sociales.
Estrategias y acciones propuestas	Se dividen en seis ejes: (i) ética y gobernanza (modelo formal en 2025); (ii) datos e infraestructura (mejorar la conectividad); (iii) I+D (aumentar los fondos); (iv) capacidades y talento (integrar la IA en educación); (v) mitigación de riesgos (estudios laborales); (vi) adopción (fomentar su uso en entidades públicas). Incluye 106 acciones hasta 2030, lideradas por MinTIC y MinCiencias.	Es pertinente establecer lazos interinstitucionales, capacitación continua y modelos sectoriales, con el fin de aplicar el principio ético. De esta manera, se garantiza la regulación humana y se implementan estrategias basadas en la IA que atiendan las particularidades de la gestión pública.
Plan de implementación, monitoreo y presupuesto	Involucra 106 acciones en 2025-2030, con informes semestrales desde junio de 2025. El monitoreo se detalla en la sección 5.4. presupuesto total: COP 479 273 millones del presupuesto general, distribuidos por objetivos.	Implica sistemas de responsabilidad, planificación presupuestal y revisiones periódicas para una ejecución eficaz, permitiendo ajustes en la gestión pública basados en resultados medibles y transparentes.
Recomendaciones	Propone al CONPES aprobar la política y acciones específicas, como expedir normativas para el uso del espectro radioeléctrico, crear observatorios para monitorear impactos laborales y fomentar alianzas para I+D en IA. Incluye medidas para entidades como MinTIC, MinCiencias y Dapre, con énfasis en ética, infraestructura y mitigación de riesgos, asegurando un enfoque integral y sostenible.	Orienta la acción gubernamental hacia una implementación práctica, fortaleciendo la coordinación interinstitucional y la asignación de responsabilidades para garantizar que la política de IA se traduzca en beneficios concretos para la administración pública, promoviendo la equidad y la prevención de desigualdades

Fuente: DNP, Documento CONPES 4144: Política Nacional de Inteligencia Artificial (Bogotá: DNP, 2025).

Ahora bien, la presente valoración posibilita señalar que la aplicación de la IA en la gestión pública de Colombia supone una evolución significativa, al permitir que la administración esté orientada hacia el ciudadano. Esto se observa en la definición acogida por la nación y su potencial para robustecer capacidades del Estado en zonas más delicadas, como la personalización de servicios y la toma de decisiones. No obstante, es pertinente balancear este avance mediante una valoración minuciosa sobre sus incidencias ético-sociales, según lo expone el BID.[21]

Del mismo modo, el Documento CONPES 4144,[22] a través de los ejes señalados, establece un precedente normativo que incentiva que el Sistema Nacional de Inteligencia Artificial articule compromisos entre

21 BID, Uso responsable de IA..., 11. **22** DNP, Documento CONPES 4144.

diversos sectores; sin embargo, también expone brechas en la aplicación eficaz de esta herramienta. En ese sentido, diversas iniciativas de automatización no cuentan con evaluaciones de impacto que posibiliten un cambio digital según las políticas públicas de IA.[23] Tal discrepancia entre práctica y planificación destaca el interés por una gobernanza participativa que acople el control democrático y la transparencia. Así, se garantiza la optimización de procesos mediante la IA y la protección de derechos fundamentales a la hora de llevar a cabo políticas públicas actuales y posteriores.

En ese orden de ideas, el porvenir de la IA en materia de gestión estatal está en función de una implementación constante que resalte la inclusión territorial y la reducción de amenazas, tal como se evidencia en el análisis del Documento CONPES 4144,[24] en el que se subrayan los retos en términos de talento humano e infraestructura, al tiempo que se plantean estrategias que permitan promover la adopción consciente.

Avances como los destacados por Londoño[25] en sectores de seguridad y subsidios demandan regulaciones robustas, con protocolos de evaluación algorítmica que eviten desigualdades distributivas y que respeten principios constitucionales de igualdad. En última instancia, una IA pública ética no solo mide su impacto en eficiencia, sino en su capacidad para generar confianza ciudadana y equidad, transformando la lógica institucional hacia un modelo participativo que alinee innovación y bien común, tal como lo sugiere el enfoque multidimensional del BID.[26]

1.2 HACIA UNA REGULACIÓN LEGISLATIVA: EL PROYECTO DE LEY SOBRE IA EN COLOMBIA

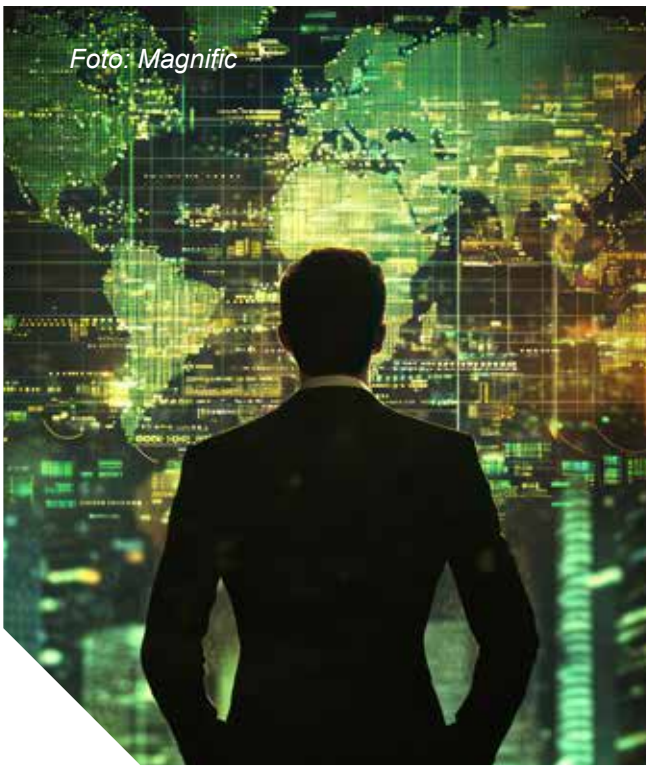
En el ecosistema normativo colombiano para la IA, el Proyecto de Ley 043 de 2025[27] representa un salto cualitativo que complementa y operacionaliza las directrices del Documento CONPES 4144,[28] al proponer un marco legal vinculante que busca equilibrar la innovación tecnológica con imperativos éticos en la gestión pública. En tal sentido, el proyecto en mención, residido en el Congreso, expone que la moderación de la IA debe transcender a un sistema adaptativo que posibilite el desenvolvimiento competitivo y sostenible de la gestión pública, lo que permite una perspectiva más clara de las políticas declarativas.

Aunado a esto, el Proyecto de Ley 043 de 2025[29] incentiva y aplica criterios de derechos humanos, teniendo en cuenta la productividad, de modo que se optimicen áreas críticas como la educación, la salud y la agricultura. Al crear una autoridad nacional de IA bajo el liderazgo del Ministerio de Ciencia, Tecnología e Innovación (MinCiencias), el texto legislativo fomenta una gobernanza intersectorial que articula entidades como el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), el Ministerio de Justicia (MinJusticia) y la

23 Vargas, "Balance de la política pública…", 263. **24** DNP, Documento CONPES 4144, 13-14.**25** Londoño, "Inteligencia artificial y gestión pública…", 234-236. **26** BID, Uso responsable de IA…, 31-46. **27** Congreso de la República de Colombia, Proyecto de Ley 043 de 2025: Por medio del cual se regula la inteligencia artificial en Colombia para garantizar su desarrollo ético, responsable, competitivo e innovador, y se dictan otras disposiciones (Bogotá: Congreso de la República de Colombia, 2025). **28** DNP, Documento CONPES 4144.**29** Congreso de la República de Colombia, Proyecto de Ley 043 de 2025, 1-5.

Superintendencia de Industria y Comercio, reconociendo que la fragmentación institucional —un desafío persistente en la administración pública colombiana— podría diluir los beneficios de la IA en regiones con brechas digitales marcadas, donde solo el 36,1 % de los hogares rurales accede a internet, según diagnósticos previos.[30]

Un elemento central de este proyecto es su clasificación técnica de sistemas de IA por niveles de riesgo —crítico, alto, limitado y bajo—, mediante un enfoque preventivo que aplica obligaciones proporcionales sin imponer barreras innecesarias a la innovación. Para sistemas de riesgo crítico o alto, como aquellos en vigilancia pública o asignación de subsidios, se exigen revisiones previas y medidas de transparencia estrictas, lo que mitiga potenciales vulneraciones a la privacidad o discriminaciones algorítmicas que podrían perpetuar desigualdades históricas en contextos como el colombiano.


Foto: Magnific

Esta categorización, adaptada de estándares internacionales, pero con énfasis local, argumenta que la gobernanza pública debe priorizar la explicabilidad y la trazabilidad ética, evitando opacidades que erosionen la confianza ciudadana en entidades estatales. Además, al habilitar sandbox regulatorios —espacios de prueba supervisados—, el proyecto permite experimentar soluciones innovadoras con acompañamiento ético, lo que podría impulsar aplicaciones en logística rural o industria farmacéutica, transformando desafíos territoriales en oportunidades de desarrollo sostenible sin comprometer derechos fundamentales.[31]

Reflexionando sobre sus implicaciones éticas, el proyecto integra un enfoque diferencial que obliga a los sistemas de IA a respetar diversidad por género, etnia, discapacidad y orientación sexual, promoviendo accesibilidad en lenguas indígenas y formatos inclusivos para evitar sesgos que amplifiquen exclusiones sociales. En el ámbito laboral, argumenta por la reconversión de competencias digitales, proponiendo programas de formación que fomenten la adaptación humana, preservando la dignidad en las relaciones laborales.[32] Al asignar responsabilidades diferenciadas a desarrolladores, comercializadores y usuarios, con mecanismos de supervisión gradual que priorizan orientación sobre sanciones, el texto legislativo defiende una responsabilidad que equilibre innovación con rendición de cuentas, extendiendo principios de no discriminación a todo el ciclo de vida de los sistemas.

30 Ministerio de Ciencia, Tecnología e Innovación (MinCiencias), ABC del Proyecto de Ley de Inteligencia Artificial en Colombia (Bogotá: MinCiencias, 2025), 2; DNP, Documento CONPES 4144, 25.**31** MinCiencias, ABC del Proyecto de Ley…, 3-4. **32** Congreso de la República de Colombia, Proyecto de Ley 043 de 2025, 20-22

Esto se alinea con la necesidad de proteger propiedad intelectual y datos sensibles, fomentando licencias abiertas donde sea viable, lo que podría fortalecer ecosistemas públicos colaborativos sin vulnerar derechos de autor o imagen.



En última instancia, este proyecto no solo cierra vacíos regulatorios del CONPES al promover colaboración internacional soberana y actualización penal limitada a conductas dolosas graves, sino que posiciona a Colombia como un actor proactivo en la gobernanza global de la IA, donde la regulación adaptativa podría habilitar transformaciones en la administración pública que prioricen el bien común.[33] Al exigir educación integral en IA desde niveles básicos y fortalecer capacidades académicas territoriales, el marco argumenta por una alfabetización digital inclusiva que empodere a poblaciones excluidas, evitando que la IA exacerbe divisiones y, en cambio, impulse un desarrollo ético que armonice competitividad con equidad social. En caso de que se apruebe, este proyecto podría representar un marco de referencia, al permitir que la IA constituya un catalizador para una administración equitativa y robusta en un escenario plural de cosmovisiones y estilos de vida.[34]

02 RIESGOS ÉTICOS Y SOCIALES DE LA IA EN LA GESTIÓN PÚBLICA

La reproducción de sesgos algorítmicos, sobre todo cuando los sistemas se capacitan mediante información desbalanceada o carentes de monitoreo humano, supone una amenaza en el empleo de IA.[35] Estos sesgos pueden reforzar desigualdades históricas y resultar en decisiones discriminatorias que afecten grupos vulnerables, particularmente en ámbitos como justicia, salud y seguridad. La falta de explicabilidad algorítmica, es decir, la incapacidad para entender cómo y por qué un algoritmo tomó una decisión, socava principios fundamentales del debido proceso y la transparencia estatal.

Según Vargas,[36] más del 50 % de los SDA implementados en Colombia manejan datos personales sensibles, lo que expone vulnerabilidades importantes en materia de privacidad, consentimiento informado y protección de derechos fundamentales. Aunque el país cuenta con legislación sobre protección de datos, esta aún no ha sido adaptada completamente a los desafíos que plantea la IA, como la posibilidad de reidentificación y la creación de perfiles automatizados con fines administrativos.

33 Ibíd., 30-32. **34** MinCiencias, ABC del Proyecto de Ley…, 5-6. **35** Londoño, “Inteligencia artificial y gestión pública…”, 237. **36** Vargas, “Balance de la política pública…”, 263.



Foto: Magnific

Del mismo modo, la Procuraduría General de la Nación[37] señaló las amenazas relacionadas con el empleo de la IA en funciones de vigilancia e inteligencia del Estado, al subrayar la preocupación de asegurar la proporcionalidad, la legalidad y el respeto por los derechos universales. En tal sentido, dichas advertencias cobran fuerza en un escenario donde se carece de control ciudadano y de facultad institucional para auditar sistemas algorítmicos.

Por otro lado, una indagación más honda, detectada en textos académicos mexicanos, plantea que, aunque la IA aporta a la resolución de problemáticas sociales con un alto grado de dificultad, también supone discusiones axiológicas importantes, como la opacidad en los procesos en la toma de decisiones, la discriminación automatizada y el interés de tener reglamentos actualizados que controlen su empleo en el ámbito público. Igualmente, a través de un enfoque ético, a nivel internacional, se expone que la solidaridad debe ser un principio neurálgico en su desenvolvimiento.[38]

Cabe señalar que Colombia tiene una problemática propia en torno a la infrarrepresentación algorítmica. Por ejemplo, zonas del casco rural, poblaciones indígenas y comunidades con niveles socioeconómicos muy bajos no cuentan con una representación adecuada en los datos empleados para capacitar modelos.[39] Como consecuencia, lo anterior podría llevar a políticas públicas insuficientes que no representan los intereses verdaderos de estas comunidades, ahondando en la inequidad estructural y mermando la legitimidad de los proyectos del Estado.

37 Procuraduría General de la Nación (PGN), Recomendaciones sobre el uso ético y legal de la inteligencia artificial en el sector público colombiano (Bogotá: PGN, 2023), 5. **38** Miguel Luengo, "Solidarity should be a core ethical principle of Artificial Intelligence", Nature Machine Intelligence 1, n.o 494 (2019): 1-2. **39** BID, Uso responsable de IA…, 67.

Así las cosas, el uso de IA puede exacerbar las asimetrías de poder entre el Estado y la ciudadanía. En ausencia de mecanismos claros de rendición de cuentas como auditorías obligatorias, derecho a apelación o explicabilidad, los ciudadanos afectados por decisiones automatizadas tienen pocas herramientas para impugnar o comprender esas decisiones.[40] Por eso, la gobernanza de la IA debe trascender los criterios técnicos y normativos para incorporar principios democráticos como la participación ciudadana informada, el control social efectivo y la transparencia institucional.

La adopción de la IA en la gestión pública, aunque prometedora, conlleva riesgos éticos y sociales que pueden socavar principios fundamentales como la transparencia, la equidad y la protección de derechos humanos. En Colombia, estos riesgos se agudizan por el contexto de desigualdades socioeconómicas, brechas digitales y un historial de conflictos internos, donde la IA podría reproducir o amplificar sesgos históricos. A continuación, se reflexiona sobre estos riesgos, incorporando casos reales en el país que evidencian su manifestación, basados en jurisprudencia, marcos éticos y aplicaciones prácticas en el sector público.

Uno de los riesgos más prominentes es la introducción de sesgos en los algoritmos de IA, derivados de datos de entrenamiento no representativos o prejuiciados, lo cual puede llevar a decisiones discriminatorias en la gestión pública, afectando grupos vulnerables como comunidades étnicas, indígenas o de bajos ingresos. En Colombia, donde el 36,1 % de la población rural carece de acceso a internet,[41] la IA entrenada en datos urbanos podría excluir perspectivas periféricas, lo que perpetúa las desigualdades.

Un caso real que ilustra este riesgo es el uso de sistemas de reconocimiento facial y biométrico en seguridad pública. En aplicaciones gubernamentales, como las implementadas por entidades de vigilancia ciudadana, se han identificado sesgos raciales inherentes a los algoritmos, que podrían discriminar contra poblaciones afrodescendientes o indígenas. Por ejemplo, en proyectos de IA para detección de comportamientos sospechosos mediante cámaras de seguridad, los modelos entrenados en datos globales (predominantemente del hemisferio norte) han mostrado tasas de error más altas en identificación de rostros no caucásicos, lo que podría resultar en perfiles injustos o detenciones erróneas.[42] Este problema se evidencia en reflexiones académicas sobre la IA, que apuntan a que la falta de datos integrados y de controles éticos podría exacerbar discriminaciones en servicios públicos como la movilidad o la asignación de recursos.[43] La Superintendencia de Industria y Comercio ha señalado la necesidad de evaluaciones de impacto previas, pero la capacidad limitada para sanciones agrava el riesgo.

Otro ejemplo proviene del sector anticorrupción, donde la IA se usa para predecir irregularidades en contratos públicos. Aunque útil, si los algoritmos se entrenan en datos históricos con sesgos (e.g., enfocados en ciertas regiones), podrían pasar por alto corrupción en áreas marginadas, reforzando inequidades territoriales.[44]

40 Ibíd., 70. **41** DNP, Documento CONPES 4144, 47. **42** Alex Camacho, Juan Gómez y Cristian Galindo, "Los desafíos éticos del uso de IA en la toma de decisiones gubernamentales en Colombia" (trabajo de especialización, Universidad EAN, Bogotá, 2024), 45. **43** Milton Ospina y Kelly Zambrano, "Gobierno digital e inteligencia artificial: una mirada al caso colombiano", Administración & Desarrollo 53, n.o 1 (2023): 50. **44** Camacho, Gómez y Galindo, "Los desafíos éticos", 52.



Foto: Magnific

La IA en la gestión pública maneja vastos volúmenes de datos ciudadanos, lo que plantea riesgos de privacidad y de manipulación. En Colombia, donde la protección de datos personales es regulada por la Ley 1581 de 2012,[45] el uso no consensuado de datos en sistemas de IA podría vulnerar derechos fundamentales, especialmente en contextos de vigilancia estatal.

Un caso emblemático es el empleo de la IA en biometría para el registro ciudadano y el control de acceso. En iniciativas gubernamentales, como las de identificación digital, se han reportado preocupaciones éticas sobre el almacenamiento y el procesamiento de datos biométricos sin suficientes salvaguardas, potencialmente expuestos a fugas o usos políticos.[46] Por ejemplo, en sistemas de chatbots para servicios públicos (p. ej., el agendamiento de citas), la recolección de datos sin transparencia podría facilitar manipulaciones, como en procesos electorales, en los que la IA podría influir en la opinión pública mediante análisis de redes sociales.[47] El marco ético para la IA en Colombia destaca estos riesgos, enfatizando la necesidad de consentimiento y auditorías, pero su implementación ha sido limitada en entidades locales.[48]

La opacidad de los algoritmos de IA, conocidos como cajas negras, complica la rendición de cuentas en la gestión pública, lo que erosiona la confianza ciudadana y viola principios de transparencia administrativa.

El caso más destacado en Colombia es la Sentencia T-323 de 2024,[49] que abordó el uso de ChatGPT 3.5 por un juez de circuito en una decisión judicial sobre derechos de un menor con autismo. El juez consultó la herramienta para elaborar argumentos legales, pero no reemplazó su razonamiento humano. La Corte determinó que no se violó el debido proceso, ya que la IA sirvió como un apoyo posterior a la decisión, pero resaltó riesgos éticos como alucinaciones (información falsa generada por la IA), sesgos y falta de transparencia. Se establecieron criterios éticos, incluyendo la no sustitución del juez natural, la verificación rigurosa y la protección de la privacidad, ya que herramientas como ChatGPT, propiedad de entidades privadas, podrían exponer datos sensibles. Este fallo evidencia riesgos en el sector judicial, parte de la gestión pública, donde la IA podría comprometer la independencia e imparcialidad, perpetuando discriminaciones históricas si no hay un control humano. La Corte ordenó al Consejo Superior de la Judicatura emitir guías para el uso de la IA en cuatro meses, destacando la urgencia de la regulación.

45 Congreso de la República de Colombia, Ley 1581 de 2012: Por la cual se dictan disposiciones generales para la protección de datos personales (Bogotá: Congreso de la República de Colombia, 2012). **46** Camacho, Gómez y Galindo, "Los desafíos éticos…", 60. **47** Ibíd., 65. **48** Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), Marco Ético para la Inteligencia Artificial en Colombia (Bogotá: MinTIC, 2021), 15-20. **49** Corte Constitucional de Colombia, Sentencia T-323 de 2024 (Bogotá: Corte Constitucional de Colombia, 2024).

De conformidad con lo anterior, el Consejo Superior de la Judicatura promulgó el Acuerdo PCSJA24-12243 de 2024,[50] que adopta lineamientos para el uso y el aprovechamiento respetuoso, responsable, seguro y ético de la IA en la rama judicial, cumpliendo con lo ordenado por la Corte Constitucional y alineándose con facultades constitucionales como las del numeral 3 del artículo 257 de la Constitución Política de Colombia[51] y con el literal a del numeral 1 del artículo 85 de la Ley 270 de 1996, modificada por la Ley 2430 de 2024.[52]

Este acuerdo no solo responde directamente a la exhortación de la sentencia, sino que representa un hito regulatorio al establecer un marco integral para toda la rama judicial, la cual abarca magistrados, jueces, empleados y dependencias, con el objetivo de maximizar beneficios como la mejora en el acceso a la justicia y la eficiencia en la gestión documental. Mientras mitiga riesgos éticos y legales inherentes a la IA, argumenta que su adopción debe priorizar la primacía de los derechos fundamentales y la no sustitución de la racionalidad humana, principios que extienden el fallo judicial a un nivel operativo y preventivo en la administración de justicia pública.

Argumentativamente, el acuerdo amplía el análisis de riesgos éticos identificados en la sentencia al incorporar definiciones técnicas precisas y principios operativos que guían el uso de la IA en contextos judiciales, reconociendo que herramientas como la IA generativa —definida como sistemas que se comunican en lenguaje natural y generan contenidos a partir de prompts o de instrucciones— poseen el potencial para optimizar tareas administrativas, pero implican vulnerabilidades como alucinaciones, sesgos y opacidad algorítmica, conceptos extraídos de estándares internacionales de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco), la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y la Comisión Europea para la Eficacia de la Justicia (Cepej).[53]

Por ejemplo, el artículo segundo del acuerdo define explícitamente los sesgos de IA como diferencias sistemáticas con los datos usados que pueden reforzar estereotipos o prejuicios lesivos para los ciudadanos. Esto deriva directamente de las preocupaciones de la Corte Constitucional sobre sesgos que podrían comprometer la imparcialidad en decisiones que afectan derechos fundamentales, como en casos de menores y de grupos vulnerables, o en aspectos como la raza, el género, la sexualidad, la etnia o la discapacidad.

En un análisis más extensivo, los principios enumerados en el artículo 3 del acuerdo —que incluyen la primacía de derechos fundamentales, la regulación ética, la supervisión humana y la prevención de riesgos— configuran un enfoque preventivo que trasciende la sentencia, al aplicar estos a toda la estructura judicial, argumentando que la IA debe servir como herramienta auxiliar sin suplantar el juicio humano, ya que cualquier decisión asistida de este tipo se atribuye íntegramente al funcionario (principio 3.6), lo que refuerza la responsabilidad en la gestión pública y mitiga la opacidad criticada en herramientas privadas como ChatGPT.

50 Consejo Superior de la Judicatura, Acuerdo PCSJA24-12243 de 2024: Por el cual se adoptan lineamientos para el uso y aprovechamiento respetuoso, responsable, seguro y ético de la inteligencia artificial en la rama judicial (Bogotá: Consejo Superior de la Judicatura, 2024), artículo 1. **51** Asamblea Nacional Constituyente, Constitución Política de Colombia (Bogotá: Legis, 1991), artículo 257. **52** Congreso de la República de Colombia, Ley 2430 de 2024: Por la cual se modifica la Ley 270 de 1996 Estatutaria de la Administración de Justicia y se dictan otras disposiciones (Bogotá: Congreso de la República de Colombia, 2024). **53** Consejo Superior de la Judicatura, Acuerdo PCSJA24-12243 de 2024, 2-3.



Foto: Magnific

Este marco se sustenta en una encuesta institucional realizada en julio de 2024, en la que el 30 % de los servidores judiciales reportó uso de IA generativa, pero el 91,68 % demandó lineamientos para abordar preocupaciones como la protección de datos y los sesgos, datos que el acuerdo cita para justificar su necesidad, destacando cómo la falta de regulación previa podría exacerbar desigualdades en el acceso a la justicia, particularmente en jurisdicciones territoriales con brechas digitales.54 Críticamente, esto argumenta por una gobernanza colaborativa (principio 3.5), involucrando mesas de trabajo con especialistas y alianzas como la formalizada con Unesco el 10 de octubre de 2024, para asegurar que la IA no amplifique divisiones sociales, lo que se alinea con recomendaciones globales que enfatizan la equidad y la inclusión en entornos públicos.

El Acuerdo PCSJA24-12243 de 202455 extiende su alcance al detallar usos permitidos y prohibidos, clasificándolos en categorías que requieren de una revisión detallada o de una supervisión especial, como la redacción asistida de textos administrativos o la clasificación de expedientes, pero prohíbe aplicaciones que involucren valoración probatoria o motivación de decisiones sustanciales, argumentando que estos usos podrían vulnerar el debido proceso, si no se verifican rigurosamente. Este principio resulta directamente influenciado por la Sentencia T-323 de 202456 y resalta la urgencia de control humano para evitar la perpetuación de discriminaciones históricas en la administración judicial.

Por ejemplo, el parágrafo primero permite herramientas de este tipo siempre que cumplan con los deberes éticos contemplados en el acuerdo, pero el artículo 5 restringe la IA generativa a aquellas provistas institucionalmente para usos contemplados, argumentando que versiones comerciales gratuitas —como las criticadas en la sentencia— incrementan riesgos de privacidad, ya que podrían apropiarse de datos personales y sensibles, violando la Ley 1581 de 2012[57] y el derecho fundamental de habeas data (artículo 15).

Este enfoque preventivo se complementa con deberes en el artículo 8, como abstenerse de introducir datos sensibles en herramientas externas o de presentar textos generados por IA como propios, lo que fortalece la integridad judicial y argumenta contra la erosión de la autoría humana en la gestión pública. También, las exigencias de transparencia (artículo 10) obligan a declarar explícitamente el uso de IA en providencias, lo que, a su vez, garantiza que las partes puedan ejercer efectivamente su derecho de contradicción y defensa al conocer los fundamentos de la decisión.[58] Este mecanismo representa un aspecto claramente positivo, pues eleva el estándar de responsabilidad y materializa el principio constitucional de publicidad y contradicción, preservando la legitimidad de las decisiones judiciales al exigir que el juez asuma plena responsabilidad por el contenido final.

54 Ibíd., 2. **55** Ibíd., artículo 4. **56** Corte Constitucional de Colombia, Sentencia T-323 de 2024.**57** Congreso de la República de Colombia, Ley 1581 de 2012, artículo 15. **58** Ibíd., artículos 8 y 10.

Ahora bien, el Acuerdo PCSJA24-12243 de 2024[59] establece directrices de implementación institucional, en las que se le exige a la Unidad de Transformación Digital e Informática realizar un análisis de impacto algorítmico, priorizando herramientas transparentes y auditables para mitigar sesgos y alucinaciones mediante la calidad de los datos, desde el argumento de que, sin estos controles, la IA[60] podría reproducir prejuicios en procesos repetitivos, como notificaciones o análisis de cargas de trabajo, perpetuando inequidades en un sistema judicial ya sobrecargado.

De manera crítica, lo anterior muestra una dificultad significativa: al tiempo que el acuerdo incentiva la neutralidad tecnológica e impide el vicio de depender de suministradores particulares, no encara de forma taxativa las incidencias ambientales o en la salud mental provenientes del empleo excesivo de la IA. Asimismo, el plan de formación, articulado con la Escuela Judicial Rodrigo Lara Bonilla, incluye áreas de aprendizaje en habilidades digitales, algo que permite destacar discrepancias sociodemográficas para formar a funcionarios, lo que se considera un avance notable para el empleo consciente de estas herramientas.[61]

Por otro lado, se designa a la Unidad de Transformación Digital como responsable del monitoreo a través de reportes continuos, lo que permite acoplar los lineamientos a los progresos en la tecnología, lo que extiende los mecanismos y las pruebas a toda la parte administrativa de lo público.[62] Asimismo, el acuerdo aplica la Sentencia T-323 de 2024[63] y fomenta una práctica axiológica de la justicia, al garantizar que la IA esté en función del bienestar general sin malograr la independencia humana, lo que también tiene la capacidad de reproducirse en diversos ámbitos de la sociedad para robustecer la democracia.

2.1 GOBERNANZA ÉTICA DE LA INTELIGENCIA ARTIFICIAL PÚBLICA

La gobernanza en consideración se plantea como una realidad necesaria en la gestión pública de la actualidad. Sin embargo, este mecanismo supone la generación de aparatos institucionales, sociales y jurídicos que permiten el desenvolvimiento y el empleo consciente de sistemas algorítmicos, sobre todo en escenarios donde la toma decisiones automatizada incide significativamente en los derechos y la integridad de los individuos. De acuerdo con la Unesco, es pertinente una gobernanza enfocada en el interés general y con enfoque en lo humano.[64]

Ahora bien, en el país, el Documento CONPES 4144[65] sentó los pilares para edificar una política pública fundada en la IA, pero con un matiz necesario en la equidad, la transparencia y la ética de esta herramienta. Sin embargo, este instrumento de política carece de fuerza normativa vinculante, por lo que su efectividad depende de la implementación voluntaria por parte de las entidades públicas. Como advierte el análisis del BID, "la ausencia de marcos legales claros puede derivar en decisiones opacas, discriminatorias o irresponsables".[66]

59 Consejo Superior de la Judicatura, Acuerdo PCSJA24-12243 de 2024, artículo 11. **60** DNP, Documento CONPES 4144, 35. **61** Consejo Superior de la Judicatura, Acuerdo PCSJA24-12243 de 2024, artículo 13. **62** Ibíd., artículo 15. **63** Corte Constitucional de Colombia, Sentencia T-323 de 2024. **64** Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco), Recommendation on the ethics of artificial intelligence (París: Unesco, 2021). **65** DNP, Documento CONPES 4144. **66** BID, Uso responsable de IA…, 34.

Aunado a lo anterior, la Unesco[67] propuso una serie de instrumentos que posibilitan valorar si una institución se encuentra capacitada para aplicar tecnología de IA de forma responsable. Estos recursos, empleados en naciones como Finlandia y Países Bajos, son puntos de partida globales para cuantificar el nivel de alineación entre los sistemas algorítmicos y los derechos humanos.

Cabe señalar que la formación institucional resulta ser una base capital en este tipo de gobernanza, dado que integra el desarrollo de habilidades éticas y técnicas en los órganos territoriales. El informe del BID[68] afirma que diversas instituciones del continente, incluyendo a Colombia, no cuentan con el personal capacitado, los medios digitales y los estándares axiológicos para el empleo de IA, lo que amenaza la legitimidad y la aplicación de tales sistemas. Por ello, la Unesco[69] sugiere incluir la alfabetización digital y la ética algorítmica como componentes esenciales en la formación de funcionarios públicos.

En este orden, la aplicación de auditorías algorítmicas y los mecanismos de evaluación de impacto resultan ser elementos críticos para ejercer la gobernanza ética. En efecto, es posible afirmar que las auditorías posibilitan detectar fallas, sesgos o potenciales violaciones a los derechos, antes de que el sistema comience a funcionar. De acuerdo con Mišić, Van Est y Kool,[70] la correcta gobernanza tiene que mezclar el elemento de "buen orden" y la justicia procedimental, al enfatizar en la equidad, el bienestar y la inclusión de la ciudadanía.[71]

Además, es imprescindible garantizar la participación ciudadana en el diseño, implementación y monitoreo de los sistemas de IA. Como señalan Dankloff et al.,[72] los procesos de desarrollo de IA en el sector público suelen excluir a los ciudadanos y grupos sociales afectados, lo que conduce a decisiones alejadas de sus necesidades reales. Este déficit democrático no solo reduce la legitimidad de los sistemas, sino que también aumenta el riesgo de error, sesgo y resistencia social.



Foto: Magnific

Adicionalmente, el Observatorio Global de Ética y Gobernanza de la IA de la Unesco recomienda que los gobiernos acojan políticas multifactoriales, en las cuales la academia, los ciudadanos y la empresa privada ocupen funciones definidas en el control y evaluación de la IA en el ámbito público,[73] planteamiento que resulta congruente con los principios de gobernanza digital inclusiva incentivados por la Unión Europea y acoplados por diversas democracias consolidadas en términos tecnológicos.

67 Unesco, Ethical impact assessment: a practical guide (París: Unesco, 2022); Unesco, AI readiness assessment methodology (París: Unesco, 2023). **68** BID, Uso responsable de IA…, 45. **69** Unesco, Ethical impact assessment. **70** Jana Mišić, Rinie van Est y Linda Kool, "Good governance of public sector AI: a combined value framework for good order and a good society", AI and Ethics 1, n.o 1 (2025): 1-15. **71** Unesco, AI readiness assessment methodology. **72** Mirthe Dankloff et al., "Analysing and organising human communications for AI fairness related decisions: use cases from the public sector", arXiv 1, n.o 1 (2024): 1-5. **73** Unesco, Global AI Ethics and Governance Observatory (París: Unesco, 2024).

La gobernanza ética de la IA en el sector público colombiano debe trascender los lineamientos generales para abarcar aplicaciones concretas en áreas sensibles como la seguridad y la inteligencia policial, en las que el despliegue de sistemas basados en IA plantea dilemas profundos sobre privacidad, discriminación y responsabilidad humana. Un ejemplo ilustrativo es el proceso de contratación del Sistema de Ciberinteligencia basado en IA para la Dirección de Inteligencia Policial (DIPOL), iniciado en 2019 por la Policía Nacional de Colombia, que buscaba adquirir una herramienta para el análisis de datos en redes sociales y bases públicas con el fin de generar perfiles de riesgo en operaciones contra el crimen organizado. Sin embargo, debido a las deficiencias en el cumplimiento de los requisitos técnicos por parte de los proponentes, el proceso fue declarado desierto el 27 de agosto de 202074 y actualmente se encuentra descontinuado.

Extendiendo este análisis al ámbito de la fiscalización, la Contraloría General de la República ha integrado la IA en su denominado Modelo Analítico Predictivo en Contratación, desarrollado en colaboración con la Universidad de Antioquia para analizar registros de procesos contractuales entre 2014 y 2025 mediante técnicas de IA y machine learning, con el objetivo de identificar riesgos antes de la adjudicación de contratos públicos.[75]

Según el informe publicado en El Colombiano,[76] este modelo va más allá de reconocer contratistas con antecedentes problemáticos, identificando patrones variables, atípicos y complejos. Sin embargo, dicho avance técnico plantea interrogantes éticas sobre la equidad algorítmica, ya que algoritmos entrenados en datos históricos podrían perpetuar sesgos territoriales, ignorando regiones marginadas con menor representación digital, lo que podría derivar en decisiones discriminatorias si no se incorporan mecanismos de corrección como los propuestos por la Unesco,[77] que aboga por evaluaciones de impacto ético para mitigar desigualdades.

En este contexto, la gobernanza ética requiere no solo de una formación institucional —como la alfabetización digital sugerida por la Unesco[78] para funcionarios públicos—, sino también de la participación ciudadana en el monitoreo de estos modelos. Tal y como Dankloff et al.[79] argumentan, la exclusión de grupos afectados reduce la legitimidad y aumenta el riesgo de error en sistemas públicos de IA. El enfoque de “buena sociedad” de Mišić, Van Est y Kool,[80] que prioriza el bienestar y la inclusión, se alinea aquí en la demanda que se hace a la Contraloría para que integre auditorías independientes que aseguren que el modelo no amplifique divisiones sociales, promoviendo así una administración pública más equitativa y responsable.

74 PGN, Sistema de ciberinteligencia basado en inteligencia artificial para la DIPOL (Bogotá: PGN, 2020). **75** “Inteligencia artificial para anticipar riesgos en la contratación pública”, El Colombiano, 26 de mayo de 2025, acceso: 12 de agosto de 2025. **76** Ibíd. **77** Unesco, Recommendation on the ethics…**78** Ibíd. **79** Dankloff et al., “Analysing and organising human communications…”, 5. **80** Mišić, Van Est y Kool, “Good governance of public sector AI”, 8. **81** Unesco, Global AI Ethics…

2.2 RECOMENDACIONES PARA UNA GOBERNANZA ALGORÍTMICA RESPONSABLE

El empalme entre los casos en mención y la gobernanza ética destaca que, en Colombia, la aplicación de IA en el ámbito de la administración pública tiene que poner en un primer plano las evaluaciones recurrentes, con el fin de articular la innovación con los derechos fundamentales. Así, se evita que instrumentos predictivos en fiscalización o seguridad continúen reproduciendo sesgos sociales. Sobre esto, la Unesco[81] plantea políticas multifactoriales para incluir a la ciudadanía en el monitoreo de las acciones públicas, como el Reglamento de IA, propuesto en 2024, en el que se definen las amenazas altas en términos de seguridad y justicia.

Teniendo en cuenta la realidad de amenazas sociales y éticas señaladas, emerge la preocupación de implementar proyectos accionables que consoliden la gobernanza ética en relación con prácticas particulares en la administración pública del país. Verbigracia, se sugiere que se institucionalicen las auditorías algorítmicas en cada uno de los sistemas de IA del Estado, para lo cual se pueden seguir los lineamientos de instrumentos de la Unesco,[82] con el objeto de identificar sesgos previos a su difusión. De este modo, se puede garantizar que las decisiones automatizadas en ámbitos como la salud o la seguridad no repliquen inequidades territoriales. Ahora bien, lo anterior supone formar a las instituciones subnacionales a través de recursos digitales y de estándares internacionales, con el fin de abordar las brechas detectas en el Documento CONPES 4144.[83]

En segundo lugar, promover mecanismos de participación ciudadana activa, como foros multiactor que involucren comunidades vulnerables en el diseño de algoritmos, mitigaría déficits democráticos y fomentaría la legitimidad, alineándose con principios de inclusión que permitan evitar la infrarrepresentación algorítmica en datos rurales o indígenas. Finalmente, fortalecer alianzas interinstitucionales para la rendición de cuentas, como observatorios independientes que monitoreen impactos en privacidad y empleo, aseguraría que la IA no exacerbe asimetrías de poder, transformando desafíos en oportunidades para una gestión pública más equitativa y humana.

Mediante lo anterior, se garantiza que haya una gobernanza ética consolidada, lo que reduce amenazas como la exhibición de datos privados en sistemas como el modelo acogido por la Contraloría y robustece la credibilidad pública. Esto posibilita que la IA esté al servicio de todos, sin denigrar el bienestar humano; por ende, es pertinente evolucionar hacia la edificación de un ambiente normativo y participativo fuerte que avale el respeto por la equidad entre territorios, los derechos fundamentales y la transparencia en la automatización de la esfera pública. De este modo, es posible consolidar una IA democrática y enfocada a las preocupaciones colectivas.

82 Unesco, Ethical impact assessment. **83** DNP, Documento CONPES 4144.

03 CONCLUSIONES

Examinar la IA en la gestión pública, mediante documentos institucionales como el Documento CONPES 4144,[84] así como en la efectuación de justicia y fiscalización, destaca un progreso que reconfigura la eficacia del Estado; de manera paralela, señala las deficiencias éticas, como opacidades y sesgos que ensanchan las brechas sociales.

Al reflexionar en torno a esta valoración, surge la necesidad de una gobernanza algorítmica que priorice la innovación como elemento inclusivo, con una exigencia al Estado de transparencia y de participación que permita transformar las amenazas en oportunidades de justicia territorial. En ese orden de ideas, el legado real de la IA consiste en su facultad para robustecer un Estado sólido, en el cual el recurso tecnológico esté en función de la colectividad, sin causar detrimento en la calidad de vida humana. De esta manera, Colombia se puede posicionar como un país que destaca por su administración pública humanizada y democrática.

No obstante, dicho avance no es a prueba de tensiones, dado que la réplica de sesgos algorítmicos, así como las otras amenazas señaladas, malogran a las comunidades vulnerables en un escenario de pluralidad territorial y socioeconómica. Al reflexionar en torno a los casos examinados, surgió el interés de balancear la innovación con salvaguardas éticas, con el fin de que la gobernanza algorítmica reduzca amenazas como la discriminación y la opacidad de la privacidad, e igualmente que promueva una administración incluyente que destaque la calidad de todos los ciudadanos.

84 Ibíd.



Foto: Magnific

Finalmente, el porvenir de la IA en el ámbito público de Colombia se plantea en función de un enfoque holístico afianzado en fundamentos democráticos, lo que supera la operatividad técnica. En esa medida, se reconoce que el valor real de esta herramienta radica en consolidar la legitimidad del Estado y la confianza de los civiles. Por consiguiente, en tanto se progrese hacia una regulación sólida y participativa, el país se puede ubicar como un modelo a seguir en términos de gobernanza ética entre los países vecinos, cuya base armoniza el potencial disruptivo de la IA con los derechos humanos, la rendición de cuentas y la equidad entre los territorios.

REFERENCIAS

Asamblea Nacional Constituyente. Constitución Política de Colombia. Bogotá: Legis, 1991.

Banco Interamericano de Desarrollo (BID). Uso responsable de IA para política pública: manual de formulación de proyectos. Washington: BID, 2021.

Camacho, Alex, Juan Gómez, y Cristian Galindo. Los desafíos éticos del uso de IA en la toma de decisiones gubernamentales en Colombia. Trabajo de especialización, Bogotá: Universidad EAN, 2024.

Congreso de la República de Colombia. Ley 1581 del 17 de octubre de 2012: Por la cual se dictan disposiciones generales para la protección de datos personales. Bogotá, 2012.

—.Ley 2430 del 9 de octubre de 2024: Por la cual se modifica la Ley 270 de 1996 Estatutaria de la Administración de Justicia y se dictan otras disposiciones. Bogotá, 2024.

—.Proyecto de Ley 043 de 2025: Por medio del cual se regula la inteligencia artificial en Colombia para garantizar su desarrollo ético, responsable, competitivo e innovador, y se dictan otras disposiciones. Bogotá, 2025.

Consejo Superior de la Judicatura. Acuerdo PCSJA24-12243 del 16 de diciembre de 2024: Por el cual se adoptan lineamientos para el uso y aprovechamiento respetuoso, responsable, seguro y ético de la inteligencia artificial en la rama judicial. Bogotá, 2024.

Corte Constitucional de Colombia. Sentencia T-323 del 2 de agosto de 2024. Sala Segunda de Revisión de la Corte Constitucional. M. P. Diana Fajardo Rivera. Bogotá, 2024.

Dankloff, Mirthe, Vanja Skoric, Giovanni Sileno, Sennay Ghebreab, Jacco Van Ossenbruggen y Emma Beauxis-Aussalet. "Analysing and organising human communications for AI fairness related decisions: use cases from the public sector". arXiv 1, n.º 1 (2024): 1-5.

Departamento Nacional de Planeación (DNP). Documento CONPES 3975: Política Nacional para la Transformación Digital e Inteligencia Artificial. Bogotá: DNP, 2019.

—.Documento CONPES 4144: Política Nacional de Inteligencia Artificial. Bogotá: DNP, 2025.

El Colombiano. "Inteligencia artificial para anticipar riesgos en la contratación pública". 26 de mayo de 2025.https://www.elcolombiano.com/informes-comerciales/las-marcas-hablan/contraloria-inteligencia-artificial-MI27494697 (último acceso: 12 de agosto de 2025).

Londoño, Juan. "Inteligencia artificial y gestión pública en Colombia: análisis de su despliegue institucional". Revista de la Universidad del Zulia 14, n.º 42 (2023): 234-236.

Luengo, Miguel. "Solidarity should be a core ethical principle of artificial intelligence". Nature Machine Intelligence 1, n.º 494 (2019): 1-2.

Ministerio de Ciencia, Tecnología e Innovación (MinCiencias). ABC del Proyecto de Ley de Inteligencia Artificial en Colombia. Bogotá: MinCiencias, 2025.

Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC). Marco Ético para la Inteligencia Artificial en Colombia. Bogotá: MinTIC, 2021.

Mišić, Jana, Rinie Van Est, y Linda Kool. "Good governance of public sector AI: a combined value framework for good order and a good society". AI and Ethics 1, n.º 1 (2025): 1-15.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco). AI readiness assessment methodology. París: Unesco, 2023.

—.*Ethical impact assessment: a practical guide. París: Unesco, 2022.*

—.*Global AI Ethics and Governance Observatory. París: Unesco, 2024.*

—.*Recommendation on the ethics of artificial intelligence. París: Unesco, 2021.*

Ospina, Milton, y Kelly Zambrano. "Gobierno digital e inteligencia artificial, una mirada al caso colombiano". Administración & Desarrollo 53, n.º 1 (2023): 31-55.

Procuraduría General de la Nación (PGN). Recomendaciones sobre el uso ético y legal de la inteligencia artificial en el sector público colombiano. Bogotá: PGN, 2023.

—.*Sistema de ciberinteligencia basado en inteligencia artificial para la DIPOL. Bogotá: PGN, 2020.*

Vargas, Carlos. "Balance de la política pública de inteligencia artificial y transformación digital (2019-2024)". Revista Internacional del Instituto de Pensamiento Liberal 1, n.º 2 (2024): 255-306.

ACERCA DE LOS AUTORES

Fabián Andrés **Mora Soto**

Candidato a Magíster en Derecho de la Universidad Pontificia Bolivariana (Colombia), Especialista en Derecho Público de la Universidad Pontificia Bolivariana (Colombia), Abogado de la Universidad Militar Nueva Granada (Colombia).

María Stella **Arismendy Ramírez**

Magíster en Derecho de la Universidad del Norte (Colombia), Especialista en Derecho Público de la Universidad Pontificia Bolivariana (Colombia), Abogada de la Universidad del Sinú (Colombia). Docente universitaria, de la Facultad de Ciencias Económicas, Jurídicas y Administrativas - Departamento de Ciencias Jurídicas de la Universidad de Córdoba (Colombia). Experiencia investigativa en áreas de derechos humanos, derecho internacional, derecho público y estudios socio-jurídicos; participación en publicaciones académicas colectivas.



Foto: Magnific

02

Inteligencia artificial generativa y control fiscal en Colombia

E-ESTADO
Primera Edición

Inteligencia artificial generativa y control fiscal en Colombia: análisis económico de la detección de detrimentos patrimoniales sistemáticos

Generative Artificial Intelligence and Fiscal Control in Colombia: An Economic Analysis of the Detection of Systematic Asset Depletion

Fecha de recepción: 8 de abril de 2026
Fecha de aceptación: 21 de julio de 2026

Pablo Márquez Escobar
Pontificia Universidad Javeriana, Colombia
pmarquez@javeriana.edu.co

Cómo citar: Márquez Escobar, Pablo. “Inteligencia artificial generativa y control fiscal en Colombia: análisis económico de la detección de detrimentos patrimoniales sistemáticos”.
Revista e-Estado 1, n.° 1 (2026).



Foto: Magnific

RESUMEN

El presente artículo examina, desde la perspectiva del análisis económico del derecho, la viabilidad jurídica y la eficiencia institucional del uso de inteligencia artificial generativa en el control fiscal colombiano. Partiendo del modelo de disuasión óptima, según el cual la eficacia de todo sistema de control depende fundamentalmente de la probabilidad de detección de la conducta irregular y del costo marginal de elevar dicha probabilidad, el artículo explica que la inteligencia artificial generativa opera precisamente sobre esa variable: reduce el costo marginal de la detección al permitir el procesamiento de volúmenes masivos de datos que exceden la capacidad humana de análisis, desplazando así el equilibrio hacia una mayor probabilidad de detección sin un incremento proporcional en los costos institucionales.

Palabras clave: inteligencia artificial generativa, control fiscal, análisis económico del derecho, detrimento patrimonial, disuasión óptima, costos de transacción, debido proceso, responsabilidad fiscal, Contraloría General de la República, regulación algorítmica

ABSTRACT

This article examines, from the perspective of the economic analysis of law, the legal feasibility and institutional efficiency of deploying generative artificial intelligence in Colombian fiscal oversight. Drawing on the optimal deterrence model, according to which the effectiveness of any enforcement system depends fundamentally on the probability of detection and the marginal cost of raising that probability, the article demonstrates that generative AI operates precisely on this variable: it reduces the marginal cost of detection by enabling the processing of massive data volumes that exceed human analytical capacity, thereby shifting the equilibrium toward a higher probability of detection without a proportional increase in institutional costs.

Keywords: *Generative Artificial Intelligence, Fiscal Oversight, Economic Analysis of Law, Patrimonial Detriment, Optimal Deterrence, Transaction Costs, Due Process, Fiscal Liability, Office of The Comptroller General, Algorithmic Regulation*

INTRODUCCIÓN

El análisis económico del derecho ha tenido diferentes perspectivas a lo largo del tiempo, pero hay algunas doctrinas que se mantienen vigentes, puesto que facilitan la interpretación de diferentes cuerpos normativos. El control fiscal no se escapa a las preguntas que se hacen normalmente en el análisis económico del derecho. La pregunta de investigación que anima el presente artículo puede formularse en términos simples: ¿es posible que la inteligencia artificial generativa reduzca los costos del control fiscal sin sacrificar las garantías del debido proceso? Sin embargo, antes de responderla, debemos resolver una cuestión previa de tipo metodológico: ¿desde qué perspectiva tiene sentido formular esa pregunta? Podríamos abordarla desde la dogmática del derecho público, limitándonos a examinar si el ordenamiento vigente habilita o prohíbe el uso de estas tecnologías. Con todo, esa vía sería insuficiente, pues dejaría intacta la cuestión de fondo, que no es de habilitación, sino de eficiencia: ¿por qué el control fiscal colombiano, a pesar de contar con un diseño normativo robusto, produce resultados de detección y de resarcimiento que están lejos del óptimo? Para responder esta segunda pregunta —que es, en rigor, la que da sentido a la primera— es necesario partir de un lugar que, a primera vista, podría parecer distante del derecho público colombiano: la teoría económica de la disuasión óptima.

En un trabajo seminal publicado en 1982, Polinsky y Shavell formalizaron una proposición que Gary Becker ya había adelantado en 1968 sobre el proceso de decisión que enfrenta cualquier persona cuando se halla frente a una norma. [1] Según Becker, el valor esperado de una sanción es una función de dos variables: la magnitud de la sanción impuesta y la probabilidad de que el infractor sea detectado y efectivamente sancionado. Así pues, un individuo racional decidirá incurrir en una conducta dañosa siempre que su ganancia exceda la sanción esperada, esto es, siempre que

$$g > p(f + \alpha x)$$

Donde g representa la ganancia, p la probabilidad de aprehensión, *f* la multa, *α* el equivalente monetario de la sanción por unidad de tiempo y x el término de la pena. [2]

Lo que interesa para los efectos de este artículo no es la pena de prisión —que en el contexto fiscal colombiano no opera directamente, sino a través del proceso penal conexo—, sino la estructura lógica del modelo. Si la probabilidad de detección es baja, la sanción esperada se reduce con independencia de la severidad de la sanción nominal. Una baja probabilidad de detección de las conductas que se pretenden disuadir es, desde la perspectiva económica, un sistema ineficiente que genera lo que Polinsky y Shavell denominaron infradisuasión. 3 La razón salta de bulto: ¿de qué sirve un régimen de responsabilidad fiscal cuyas sanciones son severas sobre el papel si la probabilidad de que el gestor fiscal infractor sea descubierto es, en la práctica, tan reducida que la sanción esperada resulta inferior a la ganancia derivada de la conducta irregular?

1 A. Mitchell Polinsky y Steven Shavell, "The Optimal Use of Fines and Imprisonment", NBER Working Paper, n.º 932 (1982). El modelo de Polinsky y Shavell se construye sobre el de Gary S. Becker, "Crime and Punishment: An Economic Approach", Journal of Political Economy 76, n.º 2 (1968). Conviene advertir que el modelo beckeriano, aunque fue formulado para el derecho penal, tiene una estructura lógica que trasciende esa rama del derecho y que resulta aplicable a todo sistema sancionatorio o resarcitorio en el que la conducta del agente pueda modelarse como una decisión racional sujeta a incentivos. **2** Polinsky y Shavell, "The Optimal Use of Fines and Imprisonment", 4. **3** Ibíd., 7-8.

Ahora bien, el modelo revela una asimetría de costos que resulta crucial para nuestro argumento. Polinsky y Shavell demostraron que, mientras las multas son socialmente poco costosas —pues constituyen una transferencia patrimonial del infractor al Estado—, el incremento de la probabilidad de detección es socialmente costoso, toda vez que exige inversión en recursos humanos, tecnológicos e institucionales para el aparato de vigilancia. [4] Esta conclusión tiene una implicación directa para las autoridades de inspección, de vigilancia y de control: el principal cuello de botella de cualquier sistema sancionatorio o resarcitorio no reside en la severidad de las sanciones previstas por el legislador, sino en la capacidad institucional para detectar las infracciones. En el lenguaje del modelo, el costo marginal de elevar p —la probabilidad de detección—, expresado como c'(p), es la variable que determina el nivel de disuasión efectivamente alcanzable.[5]

El control fiscal que ejerce la Contraloría General de la República no es la excepción a esta lógica. El control fiscal en Colombia, tal como fue reformado por el Acto Legislativo 4 de 2019 y desarrollado por el Decreto Ley 403 de 2020, opera bajo los principios de selectividad, oportunidad y eficiencia. [6] Con todo, el universo de sujetos vigilados comprende miles de entidades públicas y privadas que administran recursos públicos, con presupuestos que superan cientos de billones de pesos y millones de procesos de contratación. [7] La brecha entre el universo auditable y la capacidad operativa de la Contraloría genera, en la práctica, una probabilidad de detección (p) que está lejos del óptimo teórico. Los detrimentos patrimoniales que no se detectan —o que se detectan con años de rezago— constituyen una pérdida social neta que el sistema de responsabilidad fiscal, por más robusto que sea en su diseño normativo, no consigue resarcir.

Es precisamente en este punto en el que la inteligencia artificial generativa y el aprendizaje automático ofrecen una transformación que no es meramente incremental sino estructural. La implementación de herramientas de inteligencia artificial generativa en el proceso auditor tiene el potencial de reducir drásticamente el costo marginal de la detección —esto es, de disminuir c'(p)— permitiendo que, con la misma inversión presupuestal, la probabilidad de identificar detrimentos patrimoniales aumente de manera significativa. Como enseña el modelo de Polinsky y Shavell, el bienestar social se maximiza cuando el beneficio marginal de la detección iguala su costo marginal; [8] así pues, toda innovación que reduzca dicho costo desplaza el equilibrio hacia un nivel superior de disuasión sin necesidad de incrementar la severidad de las sanciones.

El objeto del presente artículo es, entonces, demostrar que la inteligencia artificial generativa constituye una herramienta que permite mejorar la eficiencia del control fiscal colombiano a través de la reducción de los costos de detección, sin comprometer las garantías fundamentales del debido proceso que rigen los asuntos de responsabilidad fiscal. Para ello, vamos a desarrollar el argumento en cuatro etapas. Primero, examinaremos las reglas del control fiscal en Colombia, con atención al mandato tecnológico introducido por la reforma de 2019. Segundo, analizaremos cómo la inteligencia artificial puede emplearse para detectar detrimentos patrimoniales sistemáticos y repetitivos, integrando herramientas de informática forense y de análisis de datos. Tercero, explicaremos por qué esta implementación facilitaría la gestión administrativa del control, reduciendo tiempos de investigación sin sacrificar la valoración probatoria que exige el proceso de responsabilidad fiscal. Finalmente, evaluaremos si el marco normativo vigente es suficiente o si se requiere algún tipo de reforma para regular adecuadamente el uso de la inteligencia artificial en este ámbito.

4 Ibíd., **5** La condición de primer orden que determina la probabilidad óptima establece que (e – py) dH(py)/dp = c'(p), donde e es el daño externo, y la riqueza del individuo y H(·) la distribución acumulada de ganancias. De esta expresión se sigue que toda reducción de c'(p) desplaza el equilibrio hacia una probabilidad de detección más alta, lo cual es exactamente lo que la inteligencia artificial promete hacer en el contexto del control fiscal. **6** Congreso de la República de Colombia, Acto Legislativo 04 de 2019, "por medio del cual se reforma el Régimen de Control Fiscal", Diario Oficial n.º 51.080, 18 de septiembre de 2019; Presidencia de la República de Colombia, Decreto Ley 403 de 2020, "por el cual se dictan normas para la correcta implementación del Acto Legislativo 04 de 2019", Diario Oficial n.º 51.258, 16 de marzo de 2020. **7** Contraloría General de la República, Informe de Gestión al Congreso 2023-2024 (Contraloría General de la República, 2024). **8** Polinsky y Shavell, "The Optimal Use of Fines and Imprisonment", 7-8.

Es necesario advertir que nuestra aproximación no es puramente dogmática. Siguiendo la tradición del análisis económico del derecho —que, como hemos desarrollado en otros escritos, concibe las instituciones legales no como elementos aislados del sistema económico, sino como variables inmersas en él [9] —, el presente trabajo integra herramientas de la economía de la información, la informática forense y el derecho comparado para construir un argumento, a la vez, teóricamente riguroso y operativamente viable.

01

LAS REGLAS DEL CONTROL FISCAL EN COLOMBIA: LA VARIABLE TECNOLOGÍAS DE LA INFORMACIÓN

Para comprender el alcance de la transformación que la inteligencia artificial puede introducir en el control fiscal, es indispensable partir del diseño institucional que lo rige en Colombia, entendiéndolo no como un accidente normativo, sino como una arquitectura constitucional que ha experimentado, en las últimas tres décadas, dos momentos de inflexión: la Constitución Política de 1991, que sustituyó el control previo por el posterior y selectivo, y el Acto Legislativo 04 de 2019, que reintrodujo modalidades de control en tiempo real sin regresar al modelo anterior. Ambos momentos responden a una misma preocupación que, en los términos del modelo de Polinsky y Shavell que expusimos en la introducción, puede formularse así: ¿cómo elevar la probabilidad de detección del detrimento patrimonial sin incurrir en costos institucionales que excedan el beneficio social de dicha detección?

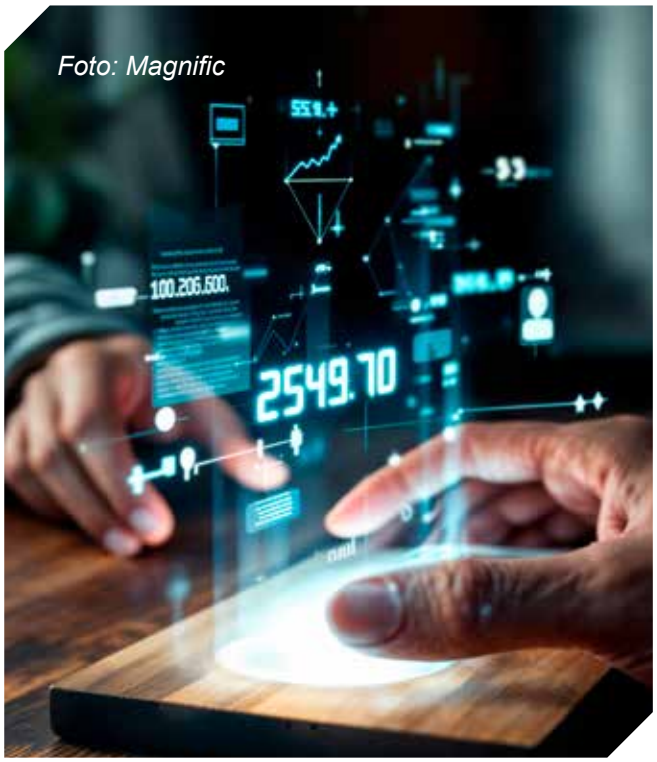


Foto: Magnific

El Título X de la Constitución Política, artículos 267 a 274, establece que la vigilancia y el control fiscal son funciones públicas que ejerce la Contraloría General de la República, la cual vigila la gestión fiscal de la administración y de los particulares o entidades que manejen fondos o bienes de la Nación. [10] El artículo 267, en su versión reformada por el Acto Legislativo 04 de 2019, dispone que el control fiscal se ejercerá en forma posterior y selectiva, y además de manera preventiva y concomitante, según sea necesario, para garantizar la defensa y protección del patrimonio público.

9 Varias de las perspectivas acá explicadas están en Carlos Pablo Márquez Escobar, Anotaciones sobre análisis económico del derecho, vol. I (Pontificia Universidad Javeriana, 2005). La premisa fundamental, que hemos defendido en otros contextos, es que todo el derecho es económico y que las instituciones jurídicas no pueden comprenderse cabalmente sin atender a los incentivos que generan y a los costos que imponen. 10 Constitución Política de Colombia, 1991, artículo 267, modificado por Acto Legislativo 04 de 2019.

La reforma de 2019 introdujo una cláusula que, para los efectos del presente artículo, reviste una importancia cardinal: el control fiscal se realizará "en tiempo real a través del seguimiento permanente de los ciclos, uso, ejecución, contratación e impacto de los recursos públicos, mediante el uso de tecnologías de la información, con sujeción al debido proceso y la audiencia del afectado". [11] Conviene detenernos en esta disposición. No se trata de una cláusula meramente programática ni de una habilitación discrecional: constituye un mandato constitucional expreso que, a la vez, habilita y obliga al uso de tecnologías de la información como instrumento del control fiscal. La Corte Constitucional, en la Sentencia C-140 de 2020, declaró la exequibilidad de la reforma y precisó que el control preventivo y concomitante no sustituye el principio de separación de poderes ni equivale al control previo eliminado en 1991, pues opera dentro de límites que impiden la coadministración. [12]

En el plano legal, tres cuerpos normativos configuran el núcleo del sistema. La Ley 42 de 1993, por la cual se organiza el sistema de control fiscal financiero y los organismos que lo ejercen, establece los principios rectores de eficiencia, economía, eficacia, equidad y valoración de costos ambientales, y define las modalidades de auditoría que puede ejercer la Contraloría. [13]

La Ley 610 de 2000 regula el proceso de responsabilidad fiscal, esto es, el conjunto de actuaciones administrativas adelantadas para determinar y establecer la responsabilidad de los servidores públicos y de los particulares que, en el ejercicio de la gestión fiscal o con ocasión de esta, causen por acción u omisión y en forma dolosa o gravemente culposa un daño al patrimonio del Estado. [14]

Los elementos de la responsabilidad fiscal son, entonces, tres: una conducta dolosa o gravemente culposa atribuible a un gestor fiscal, un daño patrimonial al Estado y un nexo causal entre ambos. Por su parte, el Decreto Ley 403 de 2020, dictado en ejercicio de las facultades extraordinarias conferidas por el propio Acto Legislativo 04 de 2019, desarrolla la reforma constitucional en sus diferentes dimensiones: regula la competencia prevalente de la Contraloría General, el control preventivo y concomitante, el seguimiento permanente a los recursos públicos y las modificaciones al proceso de responsabilidad fiscal, entre otras materias. [15] Creó, además, la Dirección de Información, Análisis y Reacción Inmediata (DIARI), que constituye el brazo operativo de la Contraloría para el análisis de datos y la generación de alertas tempranas.

Ahora bien, el mandato tecnológico introducido por la reforma de 2019 no opera en el vacío. Su implementación efectiva mediante herramientas de inteligencia artificial debe articularse con al menos tres regímenes legales preexistentes que, aunque no fueron concebidos específicamente para el control fiscal, condicionan la forma en que las tecnologías de la información pueden emplearse por las autoridades de inspección y control.

En primer lugar, la Ley 527 de 1999, por medio de la cual se define y reglamenta el acceso y el uso de los mensajes de datos, del comercio eletrónico y de las firmas digitales, constituye el fundamento normativo de la equivalencia funcional entre los documentos electrónicos y los documentos en soporte físico. [16]

11 Congreso de la República de Colombia, Acto Legislativo 4 de 2019, artículo 1. 12 Corte Constitucional de Colombia, Sentencia C-140 de 2020. La precisión de la Corte es importante porque disipa la objeción según la cual el control en tiempo real equivaldría a una forma velada de coadministración, lo cual habría comprometido la separación funcional que la propia Constitución establece. 13 Congreso de la República de Colombia, Ley 42 de 1993, artículos 8 a 13. 14 Congreso de la República de Colombia, Ley 610 de 2000, artículos 1 y 5.15 Presidencia de la República de Colombia, Decreto Ley 403 de 2020. 16 Congreso de la República de Colombia, Ley 527 de 1999.

Su artículo 5 dispone que no se negarán efectos jurídicos, validez o fuerza obligatoria a todo tipo de información por la sola razón de que esté en forma de mensaje de datos. Los artículos 10 y 11 establecen, respectivamente, la admisibilidad y la fuerza probatoria de los mensajes de datos y de los criterios para su valoración probatoria, atendiendo a la confiabilidad en la forma en que se haya generado, archivado o comunicado; la confiabilidad en la forma en que se haya conservado la integridad de la información; la forma en que se identifique a su iniciador, y cualquier otro factor pertinente. Esta ley es crucial para nuestro argumento por una razón que trasciende lo meramente procesal: si la Contraloría ha de emplear inteligencia artificial para analizar contratos, estados financieros, flujos de pago y demás registros de la gestión fiscal, los hallazgos derivados de ese análisis deben poder incorporarse válidamente al proceso de responsabilidad fiscal como material probatorio. La Ley 527 provee la base para ello, siempre que se garantice la integridad, la autenticidad y la confiabilidad de los datos procesados por los sistemas de inteligencia artificial, lo cual conecta directamente con la noción de cadena de custodia digital, que desarrollaremos más adelante.

En segundo lugar, la Ley 1341 de 2009, por la cual se definen principios y conceptos sobre la sociedad de la información y la organización de las tecnologías de la información y las comunicaciones, modificada sustancialmente por la Ley 1978 de 2019, establece el marco general para la modernización tecnológica del Estado. [17] Su artículo 2, numeral 8, consagra el principio de masificación del gobierno en línea, y, en conexión con el Decreto 1008 de 2018, que establece la política de Gobierno Digital, obliga a toda la administración pública a usar tecnologías de la información y las comunicaciones para la generación de valor público.

Para la Contraloría, esta normativa no es opcional: constituye una obligación institucional cuyo cumplimiento se mide en términos de interoperabilidad con otras entidades del Estado, acceso a bases de datos públicas y capacidad de procesamiento de información masiva.

La implementación de inteligencia artificial en el control fiscal no sería, bajo esta lectura, una innovación discrecional, sino la forma de articular adecuadamente el cumplimiento del mandato de modernización tecnológica que el legislador ha impuesto a la totalidad de la administración pública.

En tercer lugar —y quizás con la mayor complejidad jurídica—, la Ley Estatutaria 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales, establece los principios de legalidad, finalidad, libertad, veracidad, transparencia, acceso, circulación restringida, seguridad y confidencialidad que rigen el tratamiento de datos personales. [18] La tensión entre el imperativo de vigilancia fiscal y la protección del habeas data no es menor. En efecto, el análisis masivo de datos que la inteligencia artificial requiere para identificar patrones de detrimento patrimonial implica necesariamente el tratamiento de datos personales de servidores públicos, contratistas, beneficiarios de pagos y terceros vinculados a la gestión fiscal. Con todo, la propia Ley 1581 contempla excepciones que habilitan este tratamiento. Su artículo 10, literal a, autoriza el tratamiento sin consentimiento del titular cuando una ley o una autoridad competente así lo haya dispuesto. [19] El Acto Legislativo 04 de 2019 reforzó esta habilitación, al eliminar la oponibilidad de la reserva legal frente al acceso a información por parte de la Contraloría, lo cual, interpretado sistemáticamente con la Ley 1581, configura una autorización constitucional de tratamiento de datos personales para fines de control fiscal.

17 Congreso de la República de Colombia, Ley 1341 de 2009. 18 Congreso de la República de Colombia, Ley Estatutaria 1581 de 2012.19 Según la Ley 1581 de 2012, artículo 10, literal a, el tratamiento de datos personales por parte del responsable o del encargado del tratamiento no requerirá la autorización del titular cuando se trate de: a) información requerida por una entidad pública o administrativa en ejercicio de sus funciones legales o por orden judicial".

Claro está que lo anterior no exime del cumplimiento de los principios de finalidad, proporcionalidad y seguridad: el tratamiento de datos mediante inteligencia artificial debe circunscribirse a la finalidad de control fiscal, los datos deben ser pertinentes y no excesivos para dicha finalidad, y las medidas de seguridad de la información deben garantizar la confidencialidad de los datos tratados.

La convergencia de estos regímenes normativos permite formular una conclusión preliminar, que será desarrollada en las secciones siguientes. Colombia cuenta con un marco constitucional y legal que no solamente habilita, sino que ordena el uso de tecnologías de la información para el ejercicio del control fiscal. La Ley 527 provee la base probatoria para la incorporación de hallazgos generados por inteligencia artificial al proceso de responsabilidad fiscal. La Ley 1341, en articulación con la política de Gobierno Digital, impone a la Contraloría la obligación de modernizar sus procesos mediante tecnologías de la información y las comunicaciones. Y la Ley 1581, lejos de constituir un obstáculo insalvable, establece las condiciones bajo las cuales el tratamiento masivo de datos personales para fines de control fiscal resulta legítimo.

Lo que falta, entonces, no es habilitación normativa sino implementación efectiva. Vamos ahora a examinar cómo es esto posible y qué herramientas concretas lo permiten.

02

INTELIGENCIA ARTIFICIAL Y DETECCIÓN DE DETRIMENTOS PATRIMONIALES SISTEMÁTICOS

Establecido el marco normativo que habilita y ordena el uso de tecnologías de la información en el control fiscal, corresponde examinar la cuestión central de este artículo: ¿de qué manera concreta puede la inteligencia artificial emplearse para identificar detrimentos patrimoniales que, por su carácter sistemático y repetitivo, escapan a los métodos tradicionales de auditoría? Para responder esta pregunta, es necesario comprender primero la naturaleza del problema que enfrentan los órganos de control, y luego analizar las herramientas específicas que la informática forense y el análisis económico ponen a disposición del auditor.



Foto: Magnific

El detrimento patrimonial, tal como lo define el artículo 6 de la Ley 610 de 2000, es la lesión del patrimonio público representada en el menoscabo, disminución, perjuicio, detrimento, pérdida, uso indebido o deterioro de los bienes o recursos públicos, o a los intereses patrimoniales del Estado, producida por una gestión fiscal antieconómica, ineficaz, ineficiente, inequitativa e inoportuna, que en términos generales no se aplique al cumplimiento de los cometidos y de los fines esenciales del Estado. [20] Lo que interesa para nuestros efectos es que el detrimento patrimonial no siempre se manifiesta como un acto aislado y conspicuo de malversación. Con frecuencia —y esta observación es decisiva para lo que sigue—, las formas más costosas de daño al patrimonio público adoptan la forma de patrones repetitivos: sobreprecios sistemáticos en la contratación de bienes y servicios recurrentes, fraccionamiento contractual destinado a eludir los procedimientos de selección objetiva, pagos duplicados a un mismo proveedor por conceptos sustancialmente idénticos, o desviaciones persistentes entre los valores presupuestados y los efectivamente ejecutados en determinadas categorías de gasto. Estos patrones, precisamente por ser repetitivos y estar dispersos en volúmenes masivos de información, resultan difíciles de detectar mediante las técnicas convencionales de auditoría basadas en muestreo aleatorio y revisión documental.

Desde la perspectiva de la economía de la información, el problema puede formularse como una asimetría informativa entre el ente de control y los sujetos vigilados. El gestor fiscal posee información completa sobre sus propias decisiones de gasto, contratación e inversión, mientras que la Contraloría, aun con las facultades ampliadas que le confiere el Decreto Ley 403 de 2020, opera con información parcial, fragmentada y frecuentemente desactualizada. [21]

En los términos del modelo de Polinsky y Shavell, esta asimetría eleva el costo marginal de la detección —c'(p)—, porque cada unidad adicional de probabilidad de descubrimiento exige un esfuerzo desproporcionado de recopilación, de cruce y de análisis de datos. La inteligencia artificial opera precisamente sobre esta variable: al procesar volúmenes de información que exceden la capacidad humana de análisis, reduce la asimetría informativa y, con ella, el costo marginal de elevar la probabilidad de detección.

Conviene examinar, entonces, las herramientas específicas que la inteligencia artificial y la informática forense ofrecen para la detección de detrimentos patrimoniales sistemáticos. Nos detendremos en tres de ellas, que resultan particularmente relevantes para el control fiscal colombiano.

20 Congreso de la República de Colombia, Ley 610 de 2000, artículo 6. 21 Véase el Decreto Ley 403 de 2020, artículos 32 a 36. La asimetría informativa a la que nos referimos no es de índole contractual —como la que estudia la teoría de Akerlof sobre los mercados de limones—, sino institucional: se origina en la estructura misma de la relación entre el vigilante y el vigilado, en la que el segundo conoce su propia conducta con una granularidad que el primero no puede replicar, sino a un costo muy elevado.

La primera es la aplicación de la Ley de Benford al análisis de datos financieros públicos. Descubierta empíricamente por el astrónomo Simón Newcomb en 1881 y formalizada por el físico Frank Benford en 1938, esta ley establece que, en conjuntos de datos numéricos que surgen de procesos naturales o transacciones reales, el primer dígito significativo no se distribuye uniformemente, sino que sigue una distribución logarítmica: el dígito 1 aparece como primer dígito en aproximadamente el 30,1 % de los casos; el 2, en el 17,6 %, y así sucesivamente hasta el 9, que aparece en apenas el 4,6 %. [22] Cuando los datos financieros de una entidad pública presentan desviaciones estadísticamente significativas respecto de esta distribución, ello constituye un indicio de manipulación, fabricación o alteración de los registros. Rodrigues Da Cunha y Bugarin aplicaron esta técnica al análisis de los costos de renovación del estadio Maracanã en Brasil y encontraron que los 17 ítems que no se conformaban con la Ley de Benford correspondían al 71,54 % del sobreprecio total identificado por el Tribunal de Cuentas de la Unión. [23] Azevedo y otros desarrollaron un método de código abierto en el lenguaje de programación R que permite aplicar la Ley de Benford a grandes volúmenes de datos, y lo validaron con más de 13,4 millones de registros del programa Bolsa Familia. [24] El Comptroller and Auditor General de India publicó un documento de investigación sobre la aplicación de esta técnica en auditoría pública, con referencia expresa a las normas ISSAI 1315 e ISA 315 sobre identificación de riesgos de incorrección material. [25]

Lo relevante de la Ley de Benford para nuestro argumento es que constituye una herramienta de tamizaje que no requiere hipótesis previas sobre la naturaleza del fraude: opera como un detector de anomalías estadísticas que permite al auditor focalizar sus recursos de investigación en aquellos segmentos de datos que presentan mayor probabilidad de irregularidad, reduciendo así el costo marginal de la detección que el modelo de Polinsky y Shavell identifica como la restricción principal del sistema.

La segunda herramienta es el aprendizaje automático supervisado y no supervisado para la detección de patrones anómalos en la contratación pública. Los algoritmos de aprendizaje supervisado —bosques aleatorios, máquinas de vectores de soporte y redes neuronales— pueden entrenarse con datos históricos de procesos de responsabilidad fiscal previamente resueltos para identificar características predictivas de detrimento patrimonial en nuevos conjuntos de datos. [26] Los algoritmos no supervisados —análisis de agrupamiento y detección de anomalías mediante aislamiento de bosques— permiten identificar transacciones atípicas sin necesidad de datos etiquetados previamente. [27] Palmieri y Pérez, del Tribunal de Cuentas de la Provincia de Mendoza en Argentina, aplicaron el algoritmo de aislamiento de bosques a datos de auditoría contable y demostraron su capacidad para identificar registros fraudulentos con tasas de precisión superiores al 90 %. [28] El procesamiento de lenguaje natural, por su parte, permite analizar documentos contractuales no estructurados —pliegos de condiciones, informes de supervisión, actas de liquidación— para detectar cláusulas inusuales, omisiones sistemáticas o patrones de redacción que se correlacionan con prácticas irregulares. [29]

22 Frank Benford, “The Law of Anomalous Numbers”, Proceedings of the American Philosophical Society 78, n.º 4 (1938). 23 Fábio C. Rodrigues da Cunha y Maurício S. Bugarin, “Benford’s Law for Audit of Public Works: An Analysis of Overpricing in Maracanã Soccer Arena’s Renovation”, Economics Bulletin 35, n.º 2 (2015): 1168-1176. 24 Caio da Silva Azevedo et al., “A Benford’s Law Based Method for Fraud Detection Using R Library”, MethodsX 8, (2021). 25 Comptroller and Auditor General of India, “Using Benford’s Law in Audit”, Research Paper (2020). 26 Véase Diego Leocádio et al., “Artificial Intelligence in Auditing: A Conceptual Framework for Auditing Practices”, Administrative Sciences 14, n.º 10 (2024). 27 Véase Avinash Malladhi, “Artificial Intelligence and Machine Learning in Forensic Accounting”, SSRG International Journal of Computer Science and Engineering 10, n.º 7 (2023). 28 También véase Maximiliano Palmieri y Matías Pérez, “Inteligencia artificial, aprendizaje automático y controles automáticos aplicados a la auditoría contable” (ponencia presentada en el XXVII Congreso Nacional de Tribunales de Cuentas, Mendoza, Argentina, 2024). 29 Dice el autor. Zangana et al., “AI in Forensic Accounting: Detecting Anomalies and Financial Fraud Through Intelligent Systems”, en AI in Digital Forensics and Cybercrime Investigation (IGI Global, 2025).

La tercera herramienta, y la que más directamente conecta con la inteligencia artificial generativa, es el uso de modelos de lenguaje de gran escala para el análisis documental masivo y la generación de hipótesis de investigación. La experiencia más avanzada en este campo corresponde al Tribunal de Cuentas de la Unión de Brasil, que desde 2016 opera un ecosistema de sistemas de inteligencia artificial: ALICE analiza diariamente las publicaciones de licitaciones en el Diario Oficial de la Unión y alerta sobre irregularidades, SOFIA cruza textos de auditoría con hallazgos previos para orientar al auditor, MONICA monitorea todas las compras públicas federales y ADELE analiza datos corporativos de licitantes para detectar empresas de fachada. [30] Este sistema evitó pérdidas por 504 millones de reales en 2021. [31] En 2024, el Tribunal de Cuentas lanzó ChatTCU, una plataforma de inteligencia artificial generativa conectada a sus sistemas internos con más de 1400 usuarios, y la Organización para la Cooperación y el Desarrollo Económicos lo reconoció como la única institución de integridad con uso avanzado de inteligencia artificial generativa a nivel mundial. [32] La Government Accountability Office de los Estados Unidos, por su parte, publicó en 2021 un marco de rendición de cuentas para el uso de inteligencia artificial por entidades federales que identifica 31 prácticas clave organizadas en cuatro principios: gobernanza, datos, desempeño y monitoreo. 33 En 2025, esa misma oficina documentó que los casos de uso de inteligencia artificial en agencias federales se duplicaron de 571 a 1110 entre 2023 y 2024, con un incremento de nueve veces en los casos de inteligencia artificial generativa. [34] El Departamento del Tesoro de los Estados Unidos reportó que sus procesos de detección de fraude con aprendizaje automático previnieron y recuperaron más de 4 mil millones de dólares en el año fiscal 2024, frente a 652 millones en 2023. [35]



Foto: Magnific

30 Silva Neves y Menezes de Carvalho, “Artificial Ladies against Corruption: Searching for Legitimacy at the Brazilian Supreme Audit Institution”, Revista de Administração Pública 53, n.º 6 (2019): 1060-1079. 31 Por ejemplo, véase Agência FAPESP, “Artificial Intelligence for More Efficient Public Spending Audits”, FAPESP Research, 2022. 32 Portal Tribunal de Cuentas de la Unión, “TCU é única instituição com uso avançado de inteligência artificial generativa, segundo a OCDE”, Notícias, 2024. 33 U.S. Government Accountability Office, Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities, GAO-21-519SP (Government Accountability Office, 2021). 34 U.S. Government Accountability Office, Artificial Intelligence: Generative AI Use and Management at Federal Agencies, GAO-25-107653 (Government Accountability Office, 2025). 35 U.S. Department of the Treasury, “Treasury Announces Enhanced Fraud Detection Processes, Including Machine Learning AI, Prevented and Recovered Over $4 Billion in Fiscal Year 2024”, Press Release JY-2650, 2024.

Colombia no parte de cero en esta materia. La Contraloría General de la República lanzó en 2025 un modelo analítico predictivo en contratación, desarrollado con la Universidad de Antioquia, que se nutre de más de 13 mil fuentes de datos conectadas a la DIARI. [36] Esta dirección ha emitido más de 3800 alertas de riesgo fiscal mediante cuatro decenas de modelos analíticos que procesan más de 30 mil millones de registros. [37] El caso Hidroituango constituye un precedente significativo: la Contraloría empleó inteligencia artificial y analítica de datos para evaluar tres millones de registros en cuatro días, y la Auditoría General de la República validó el procedimiento como confiable y verificable. [38] Estos desarrollos demuestran que la infraestructura tecnológica básica ya existe. Lo que falta —como argumentaremos en la sección siguiente— es la sistematización de estos esfuerzos en un proceso institucional que permita a los funcionarios emplear la inteligencia artificial generativa de manera rutinaria y con resultados oportunos.

La convergencia de estas tres herramientas —la Ley de Benford como instrumento de tamizaje estadístico, el aprendizaje automático como mecanismo de detección de patrones y la inteligencia artificial generativa como plataforma de análisis documental— configura lo que podríamos denominar un sistema de detección de detrimentos patrimoniales asistido por inteligencia artificial. Su implementación no requiere una revolución tecnológica, sino un proceso de investigación, adecuadamente diseñado, que permita transformar la gestión administrativa del control fiscal de manera estructural. En los términos del modelo económico que hemos venido desarrollando, este sistema opera simultáneamente sobre las dos variables que determinan la eficacia del control: reduce el costo marginal de la detección, c’(p), al automatizar el procesamiento de volúmenes masivos de datos, y disminuye la asimetría informativa entre el ente de control y los sujetos vigilados, al permitir que la Contraloría acceda a patrones de comportamiento que serían invisibles para el análisis humano convencional. El resultado es un desplazamiento del equilibrio hacia una mayor probabilidad de detección sin un incremento proporcional en los costos institucionales del control.

36 Véase “Contraloría lanza modelo predictivo con inteligencia artificial para vigilar contratación pública”, Portafolio, 2025. 37 “DIARI emitió 3800 alertas de riesgo fiscal en Colombia”, Ámbito Jurídico, 2025. 38 “Inteligencia artificial que usó la Contraloría para emitir fallo en caso Hidroituango es válida: Auditoría General”, Semana, 2021.

03

EFICIENCIA EN LA GESTIÓN DEL CONTROL FISCAL: COSTOS DE TRANSACCIÓN, CAPACITACIÓN Y DEBIDO PROCESO

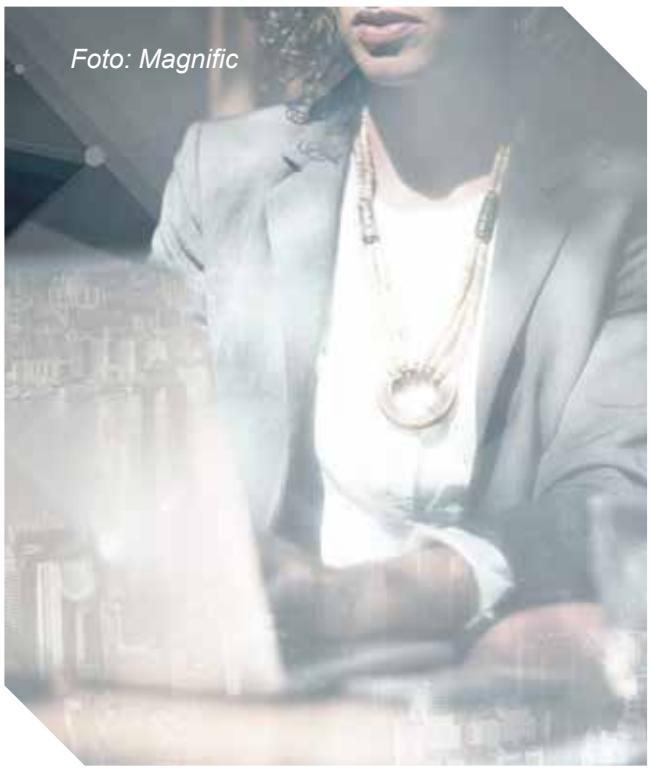
Hasta el momento hemos establecido dos premisas que conviene recapitular brevemente antes de avanzar. La primera es que Colombia posee un marco constitucional y legal que habilita y que ordena el uso de tecnologías de la información en el control fiscal. La segunda es que existen herramientas concretas de inteligencia artificial capaces de detectar detrimentos patrimoniales sistemáticos que escapan a los métodos convencionales de auditoría.

Corresponde ahora examinar una tercera cuestión que es, en rigor, la que determina la viabilidad práctica de la propuesta: ¿por qué la implementación de inteligencia artificial generativa facilitaría la gestión administrativa del control fiscal, y en qué condiciones ello puede lograrse sin menoscabo de las garantías fundamentales del debido proceso?.

Para abordar esta cuestión, resulta útil recurrir nuevamente a las categorías del análisis económico del derecho, particularmente a la noción de costos de transacción desarrollada por Ronald Coase y refinada posteriormente por Oliver Williamson. En su formulación más elemental, los costos de transacción son aquellos en que incurren las partes de una relación institucional para identificar, negociar, ejecutar y hacer cumplir un arreglo determinado. [39] En el contexto del control fiscal, los costos de transacción del proceso de responsabilidad fiscal comprenden los costos de identificación del presunto detrimento patrimonial, los costos de recopilación y procesamiento de la evidencia, los costos de análisis jurídico y contable de los hallazgos, los costos de notificación y vinculación de los presuntos responsables, y los costos de tramitación del proceso hasta su culminación en un fallo con o sin responsabilidad fiscal. Cada una de estas etapas consume tiempo, humano calificado y capacidad institucional.

Así las cosas, la magnitud de estos costos de transacción no es despreciable. El proceso de responsabilidad fiscal ordinario regulado por la Ley 610 de 2000 contempla una fase de investigación que puede extenderse por un término no mayor de tres meses, prorrogable por un término igual, seguida de una fase de juicio fiscal con un término de traslado para alegatos de diez días y un término para decidir de treinta días. [40] El procedimiento verbal introducido por la Ley 1474 de 2011, que buscaba acelerar el trámite, opera en una sola audiencia, pero requiere que la investigación previa haya alcanzado un grado de madurez probatoria suficiente para formular auto de apertura e imputación. [41]

En la práctica, los procesos de responsabilidad fiscal se extienden con frecuencia por periodos que superan los dos años, y la tasa de recuperación efectiva del detrimento patrimonial es, según la doctrina especializada, considerablemente inferior al monto total declarado como daño patrimonial. [42] Cada mes adicional que transcurre entre la ocurrencia del detrimento y su detección representa no solo una pérdida financiera directa —por cuanto los recursos públicos desviados generan rendimientos en cabeza de quien se beneficia indebidamente de ellos—, sino también una erosión de la capacidad probatoria, toda vez que los documentos se deterioran, los testigos olvidan y las trazas digitales pueden ser eliminadas o alteradas.



39 Ronald H. Coase, “The Nature of the Firm”, Economica 4, n.º 16 (1937): 386-405; Oliver E. Williamson, The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting (Free Press, 1985). Sobre costos de transacción y su aplicación al análisis de instituciones públicas, remitimos al lector a las reflexiones que hemos desarrollado en nuestros trabajos previos sobre la relación entre derecho y economía. 40 Ley 610 de 2000, artículos 41, 50 y 52. 41 Ley 1474 de 2011, artículos 97 a 105. 42 Tatiana Ordóñez Vásquez, “La responsabilidad fiscal en Colombia y la imposibilidad del resarcimiento patrimonial”, IUSTA 1, n.º 26 (2016): 140. La brecha entre el daño declarado y el efectivamente resarcido es uno de los indicadores más elocuentes de la ineficiencia del sistema, y obedece en buena medida a los costos de transacción que venimos analizando.

La inteligencia artificial generativa opera sobre varios de estos costos de transacción de manera simultánea. En la fase de identificación del detrimento, las herramientas de tamizaje estadístico y aprendizaje automático que describimos en la sección anterior permiten reducir el tiempo de detección de meses o años a días u horas. Se afirma que el caso Hidroituango —en el que la Contraloría procesó tres millones de registros en cuatro días— no constituye una excepción, sino una muestra de lo que es técnicamente posible cuando se dispone de la infraestructura adecuada.
En la fase de recopilación y procesamiento de evidencia, los modelos de lenguaje de gran escala permiten analizar miles de contratos, adiciones, otrosíes, informes de supervisión y actas de liquidación en tiempos que serían físicamente imposibles para un equipo humano, por numeroso que fuera. En la fase de análisis, la inteligencia artificial generativa puede asistir al funcionario en la elaboración de matrices de hallazgos, en la identificación de precedentes jurisprudenciales relevantes y en la cuantificación del daño patrimonial mediante modelos comparativos.

Por ejemplo, el Departamento del Tesoro de los Estados Unidos ofreció una cuantificación elocuente de este efecto, ya que sus procesos de detección de fraude con aprendizaje automático pasaron de prevenir y recuperar 652 millones de dólares en 2023 a más de 4 mil millones en 2024, un incremento superior a seis veces, que no obedeció a un aumento proporcional en el número de funcionarios sino a la optimización tecnológica del proceso. [43]

Ahora bien, la reducción de costos de supervisión mediante inteligencia artificial no opera de manera automática. Requiere una condición que con frecuencia se omite en las discusiones sobre transformación digital: la capacitación adecuada de los funcionarios que habrán de emplear estas herramientas. Y es que un modelo de lenguaje de gran escala, por sofisticado que sea, no sustituye el juicio profesional del auditor ni la valoración jurídica del abogado que sustancia el proceso de responsabilidad fiscal. Lo que hace es amplificar la capacidad analítica del funcionario, permitiéndole procesar volúmenes de información que de otro modo serían inabarcables y focalizar su atención en los hallazgos que presentan mayor probabilidad de constituir detrimento patrimonial. Esta distinción entre amplificación y sustitución es decisiva para nuestro argumento, pues de ella depende la compatibilidad de la propuesta con las garantías del debido proceso, cardinales en la legislación colombiana.

La Organización para la Cooperación y el Desarrollo Económicos publicó en 2026 un estudio titulado Building an AI-Ready Public Workforce, en el que se identifican tres grupos de funcionarios públicos con necesidades diferenciadas de formación en inteligencia artificial: los empleados generales, que requieren alfabetización básica sobre las capacidades y limitaciones de la tecnología; los líderes y directivos, que deben comprender las implicaciones estratégicas y éticas de su implementación, y los profesionales digitales, que necesitan competencias técnicas avanzadas para diseñar, implementar y supervisar los sistemas. [44] El estudio documenta que la formación disponible tiende a concentrarse en el tercer grupo, dejando brechas significativas en los dos primeros, lo cual genera un riesgo de adopción tecnológica sin comprensión institucional.

Un experimento de campo aleatorizado realizado en 2025 con cien auditores profesionales de una firma global encontró que aquellos con acceso a un modelo de lenguaje de gran escala mejoraron significativamente la calidad de sus resultados y redujeron el tiempo

43 U.S. Department of the Treasury, “Treasury Announces Enhanced Fraud Detection Processes”. 44 Organisation for Economic Co-operation and Development (OECD), Building an AI-Ready Public Workforce (2026).

de ejecución de las tareas, con mayores ganancias para los usuarios más activos y experimentados. [45] Este hallazgo sugiere que la inteligencia artificial no reemplaza la experiencia, sino que la potencia, lo cual refuerza nuestra tesis de la amplificación.

Lo anterior conduce a la cuestión que, desde la perspectiva del derecho público, resulta más delicada: la compatibilidad entre el uso de inteligencia artificial y las garantías del debido proceso en el proceso de responsabilidad fiscal. El artículo 29 de la Constitución Política consagra el derecho al debido proceso en toda clase de actuaciones judiciales y administrativas, lo cual comprende, entre otras garantías, el derecho a un proceso público sin dilaciones injustificadas, el derecho de defensa, el derecho a presentar pruebas y a controvertir las que se alleguen en contra, y el principio de legalidad de las formas propias de cada juicio. [46] La Corte Constitucional, en la Sentencia SU-620 de 1996, precisó que el proceso de responsabilidad fiscal tiene naturaleza administrativa y no sancionatoria, pues su finalidad es resarcitoria: busca obtener la reparación del patrimonio público y no castigar al responsable. [47] Con todo, la jurisprudencia ha sido enfática en que, aun tratándose de un proceso de naturaleza resarcitoria, las garantías del debido proceso deben observarse en su integridad.

La pregunta relevante, entonces, no es si la inteligencia artificial puede usarse en el control fiscal —la Constitución, como hemos demostrado, lo ordena expresamente—, sino en qué papel. Y aquí la distinción que hemos venido construyendo a lo largo de este artículo resulta determinante. La inteligencia artificial, en el diseño que proponemos, opera como instrumento de investigación y no como mecanismo de decisión. Esto significa que los hallazgos generados por los algoritmos de tamizaje estadístico, por los modelos de aprendizaje automático y por los sistemas de inteligencia artificial generativa constituyen insumos para la labor del auditor, no conclusiones que sustituyan su valoración.

Del mismo modo en que un dictamen pericial contable o un informe de policía judicial constituyen medios de prueba que el operador jurídico debe valorar conforme a las reglas de la sana crítica, los resultados producidos por la inteligencia artificial deben ser sometidos al escrutinio profesional del funcionario que sustancia el proceso de responsabilidad fiscal. El funcionario conserva la obligación de verificar la pertinencia, la conducencia y la utilidad de la información generada por la máquina, de contrastarla con otras fuentes probatorias, de garantizar al presunto responsable la oportunidad de conocer y controvertir los hallazgos, y de fundamentar su decisión en una valoración integral del acervo probatorio.

Esta arquitectura —inteligencia artificial como instrumento, funcionario como decisor— no solamente es compatible con el debido proceso, sino que, si se nos permite la paradoja, puede fortalecerlo. Un sistema que detecta detrimentos patrimoniales con mayor celeridad reduce las dilaciones que afectan tanto al Estado como al presunto responsable. Un sistema que procesa la totalidad de los datos relevantes, en lugar de una muestra aleatoria, reduce el riesgo de falsos positivos y de investigaciones infundadas. Y un sistema que documenta de manera trazable los parámetros, los algoritmos y los datos empleados para generar cada hallazgo facilita el ejercicio del derecho de contradicción, pues permite al presunto responsable conocer con precisión la metodología que condujo a su vinculación al proceso. En los términos del modelo económico que hemos venido empleando, la inteligencia artificial no solo reduce el costo marginal de la detección, sino que, al mejorar la calidad de la evidencia, eleva la probabilidad de que los fallos con responsabilidad fiscal sean confirmados en sede judicial, lo cual refuerza el efecto disuasorio del sistema en su conjunto.

45 Markus Jezierski et al., “The Effect of Large Language Models on Professional Auditing”, SSRN Working Paper, n.º 6095166 (2025).
46 Constitución Política de Colombia, 1991, artículo 29. 47 Corte Constitucional de Colombia, Sentencia SU-620 de 1996.

04
LA NECESIDAD DE REFORMA: REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN EL CONTROL FISCAL

Las secciones precedentes han demostrado que la inteligencia artificial generativa puede reducir los costos de detección de detrimentos patrimoniales sin comprometer las garantías del debido proceso, siempre que opere como instrumento de investigación y no como mecanismo de decisión. Resta por examinar una última cuestión que condiciona la sostenibilidad institucional de esta propuesta: ¿es el marco normativo vigente suficiente para regular el uso de inteligencia artificial en el control fiscal, o se requiere algún tipo de reforma?

La respuesta exige distinguir entre dos planos que no deben confundirse. En el plano de la habilitación, el marco vigente es suficiente. Como hemos expuesto en la sección segunda, el Acto Legislativo 04 de 2019 ordena expresamente que el control fiscal se realice mediante el uso de tecnologías de la información. La Ley 527 de 1999 provee la base probatoria. La Ley 1341 de 2009 impone la obligación de modernización tecnológica. Y la Ley Estatutaria 1581 de 2012, interpretada sistemáticamente con la eliminación de la oponibilidad de reserva legal frente a la Contraloría, habilita el tratamiento de datos personales para fines de control fiscal. No existe, pues, un vacío de habilitación que impida a la Contraloría emplear inteligencia artificial en el ejercicio de sus funciones constitucionales.

Empero, en el plano de la regulación técnica, el panorama es distinto. El ordenamiento colombiano carece de estándares específicos que regulen las condiciones bajo las cuales los hallazgos generados por sistemas de inteligencia artificial pueden incorporarse válidamente al proceso de responsabilidad fiscal. Esta ausencia no es un problema menor. Si un fallo con responsabilidad fiscal se fundamenta parcialmente en hallazgos producidos por un algoritmo de aprendizaje automático, el presunto responsable tiene derecho a conocer los parámetros del algoritmo, los datos con los que fue entrenado, las métricas de precisión y error, y la trazabilidad completa del proceso que condujo al hallazgo. Sin estándares que regulen estas exigencias, cada proceso de responsabilidad fiscal que emplee inteligencia artificial queda expuesto a un debate probatorio sobre la confiabilidad de la herramienta que puede dilatar el trámite y, en el peor de los casos, viciar la actuación administrativa.

El Documento CONPES 3975 de 2019, que estableció la Política Nacional para la transformación digital e inteligencia artificial, definió 14 principios orientadores y una inversión estimada de 121 619 millones de pesos, pero no reguló específicamente el control fiscal ni estableció estándares técnicos para el uso de inteligencia artificial en actuaciones administrativas con efectos jurídicos sobre los administrados. [48] El Documento CONPES 4144 de 2025 actualizó esta política con seis ejes estratégicos, más de cien acciones concretas y una inversión de 479 mil millones de pesos hasta 2030, pero tampoco abordó las especificidades del control fiscal ni la cuestión de la validez probatoria de los hallazgos generados por inteligencia artificial. [49]

48 Departamento Nacional de Planeación, Documento CONPES 3975 de 2019, “Política Nacional para la Transformación Digital e Inteligencia Artificial” (Departamento Nacional de Planeación, 2019).
49 Departamento Nacional de Planeación, Documento CONPES 4144 de 2025, “Política Nacional de Inteligencia Artificial” (Departamento Nacional de Planeación, 2025).



Foto: Magnific

Colombia ocupa el puesto 51 de 83 países en el Global AI Index y el puesto 55 de 188 en el Government AI Readiness Index, lo cual refleja una capacidad institucional intermedia que requiere fortalecimiento. [50] El Marco Ético para la Inteligencia Artificial en Colombia, publicado por el Departamento Administrativo de la Presidencia de la República en 2021, constituye un instrumento de derecho blando que define diez principios éticos, pero carece de fuerza vinculante y no establece obligaciones específicas para los organismos de control. [51]

En el ámbito legislativo, el Proyecto de Ley 043 de 2025 Senado, acumulado con el Proyecto 324 de 2025 Cámara, constituye el principal esfuerzo de regulación integral de la inteligencia artificial en Colombia. El proyecto propone una regulación basada en riesgos con cuatro niveles, siguiendo el modelo del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, conocido como EU AI Act, y designa al Ministerio de Ciencia, Tecnología e Innovación como Autoridad Nacional de Inteligencia Artificial. [52] Con todo, el proyecto aborda la inteligencia artificial en términos generales y no resuelve las especificidades del control fiscal: no regula la validación de algoritmos para hallazgos fiscales, ni la explicabilidad de las decisiones automatizadas en el contexto del proceso de responsabilidad fiscal, ni la cadena de custodia digital de los insumos y productos de la inteligencia artificial, ni los estándares de auditoría algorítmica que deben aplicarse a los sistemas empleados por la Contraloría.

El derecho comparado ofrece referentes que pueden orientar la reforma necesaria. El marco de 31 prácticas de la Government Accountability Office de los Estados Unidos proporciona un modelo de rendición de cuentas para el uso de inteligencia artificial por entidades públicas que puede adaptarse al contexto institucional colombiano. [53] La estrategia de inteligencia artificial del Comptroller and Auditor General de India, publicada en 2025, cubre la totalidad del ciclo de auditoría con principios de ética, privacidad de datos, gobernanza de modelos y desarrollo de capacidades. [54] La guía de la Organización Internacional de Entidades Fiscalizadoras Superiores sobre la conducción de actividades de auditoría con analítica de datos establece estándares para la integración de herramientas tecnológicas en los procesos de auditoría pública. [55] Y el documento de COSO publicado en febrero de 2026, titulado Achieving Effective Internal Control Over Generative AI, mapea las capacidades de la inteligencia artificial generativa al marco integrado de control interno con ocho tipos de capacidades y consideraciones específicas de control. [56]

50 Tortoise Media, Global AI Index 2024; Oxford Insights, Government AI Readiness Index 2024, citados en CONPES 4144 de 2025. 51 Departamento Administrativo de la Presidencia de la República, Marco Ético para la Inteligencia Artificial en Colombia (Departamento Administrativo de la Presidencia de la República, 2021). 54 Comptroller and Auditor General of India, Artificial Intelligence Strategy Framework (2025). 55 INTOSAI Working Group on Big Data, Guidance on Conducting Audit Activities with Data Analytics (2022). 56 The Treadway Commission, Achieving Effective Internal Control Over Generative AI (2026). 52 Congreso de la República de Colombia, Proyecto de Ley 043 de 2025 Senado - 324 de 2025 Cámara, “por medio de la cual se establecen los principios, derechos, deberes, gobernanza y regulación de la inteligencia artificial en Colombia”. Véase Reglamento (UE) 2024/1689, con el que se establecen normas armonizadas en materia de inteligencia artificial. 53 U.S. Government Accountability Office, Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities, GAO-21-519SP.

Con fundamento en estos referentes, sostenemos que la reforma necesaria no es de carácter estructural, sino procedimental. No se requiere una nueva ley orgánica de control fiscal ni una modificación del proceso de responsabilidad fiscal en sus elementos sustanciales. Lo que se requiere es un instrumento normativo —que bien podría adoptar la forma de una resolución reglamentaria orgánica de la Contraloría General, en ejercicio de la potestad reglamentaria que le confiere el artículo 268, numeral 12, de la Constitución Política— que establezca al menos las siguientes condiciones para el uso de inteligencia artificial en el control fiscal: estándares de validación algorítmica que definan los requisitos de precisión, de exhaustividad y de reproducibilidad que deben cumplir los modelos de inteligencia artificial antes de ser empleados en actuaciones de control fiscal; protocolos de cadena de custodia digital que garanticen la integridad, autenticidad y trazabilidad de los datos procesados y de los hallazgos producidos, en concordancia con los criterios de la Ley 527 de 1999; estándares de explicabilidad que aseguren que todo hallazgo generado por inteligencia artificial pueda ser expresado en términos comprensibles para el presunto responsable y para el juez administrativo que eventualmente revise el fallo de responsabilidad fiscal; requisitos de auditoría algorítmica que permitan verificar periódicamente el desempeño, los sesgos y la vigencia de los modelos empleados; y protocolos de capacitación obligatoria para los funcionarios que empleen herramientas de inteligencia artificial en el ejercicio del control fiscal.

Esta propuesta tiene la ventaja de no requerir trámite legislativo, ya que puede ser implementada por la Contraloría General mediante el ejercicio de su propia potestad reglamentaria, lo cual permite una adopción más ágil y una revisión periódica conforme evolucione la tecnología. Al mismo tiempo, provee la seguridad jurídica requerida por parte de los sujetos vigilados para que el uso de inteligencia artificial en el control fiscal no se convierta en una fuente de litigiosidad que, paradójicamente, eleve los costos de transacción que la tecnología pretende reducir.

05 CONCLUSIONES

Volvamos a la pregunta con la que iniciamos este escrito: ¿es posible que la inteligencia artificial generativa reduzca los costos del control fiscal sin sacrificar las garantías del debido proceso? El argumento desarrollado en las secciones precedentes permite responder afirmativamente, pero con una precisión que resulta indispensable para evitar tanto el optimismo tecnológico ingenuo como el escepticismo institucional paralizate.

La primera proposición que hemos sustentado es de naturaleza teórica y se inscribe en la tradición del análisis económico del derecho. Las normas son precios, y el control fiscal no es la excepción: su eficacia depende fundamentalmente de la probabilidad de detección del detrimento patrimonial. Como demostraron Polinsky y Shavell, el principal obstáculo para elevar dicha probabilidad es el costo marginal de la detección. En un sistema donde el universo auditable excede con creces la capacidad operativa del ente de control, la probabilidad de detección es subóptima y el resultado es una infradisuasión que erosiona la función resarcitoria del proceso de responsabilidad fiscal.

La segunda proposición es de naturaleza jurídica. Colombia posee un marco constitucional y legal que no solamente habilita, sino que ordena el uso de tecnologías de la información en el control fiscal. El Acto Legislativo 04 de 2019 estableció un mandato expreso en este sentido, la Ley 527 de 1999 provee la base probatoria, la Ley 1341 de 2009 impone la obligación de modernización tecnológica y la Ley Estatutaria 1581 de 2012, interpretada sistemáticamente, habilita el tratamiento de datos personales para fines de control fiscal bajo los principios de finalidad, proporcionalidad y seguridad. La tercera proposición es de naturaleza empírica. Existen herramientas concretas de inteligencia artificial, cuya eficacia ha sido documentada tanto en el ámbito internacional como en la propia experiencia colombiana. La Ley de Benford, el aprendizaje automático y la inteligencia artificial generativa, empleadas de manera convergente, configuran un sistema de detección que reduce simultáneamente el costo marginal de la detección y la asimetría informativa entre el ente de control y los sujetos vigilados. Los resultados cuantitativos son elocuentes: basta considerar el salto de 652 millones a más de 4 mil millones de dólares en un solo año fiscal en el Departamento del Tesoro de los Estados Unidos para comprender la magnitud de la transformación que está en juego.

La cuarta proposición es de naturaleza normativa. La implementación de herramientas de inteligencia artificial en el control fiscal es compatible con las garantías del debido proceso, siempre que se respete la distinción entre la inteligencia artificial como instrumento de investigación y como mecanismo de decisión. El funcionario conserva la inmediación, la obligación de fundamentar su decisión en una valoración integral del acervo probatorio y el deber de garantizar al presunto responsable la oportunidad de conocer y de controvertir los hallazgos —lo cual debe incluir la posibilidad de examinar los algoritmos empleados para la detección—.

Así pues, la reforma necesaria no es estructural, sino procedimental: estándares de validación algorítmica, cadena de custodia digital, explicabilidad y auditoría algorítmica, todo lo cual puede ser implementado por la Contraloría mediante resolución reglamentaria orgánica sin necesidad de trámite legislativo.

No se requiere, en suma, una transformación radical del sistema de control fiscal, sino algo más modesto y, acaso por ello, más realizable: un proceso de investigación asistido por inteligencia artificial generativa que, integrado a la gestión administrativa de la Contraloría y sujeto a los estándares técnicos y jurídicos que hemos descrito, permita elevar la probabilidad de detección de detrimentos patrimoniales sistemáticos a un costo institucional razonable. La tecnología existe, el marco jurídico la habilita, la experiencia internacional demuestra su eficacia y la experiencia colombiana confirma que no partimos de cero. Lo que resta es la decisión institucional de implementarla y la voluntad académica de seguir pensando, desde el análisis económico del derecho, las condiciones bajo las cuales esa implementación puede ser eficiente sin dejar de ser conforme a derecho.

REFERENCIAS

Agência FAPESP. "Artificial Intelligence for More Efficient Public Spending Audits". FAPESP Research, 2022.

Becker, Gary S. "Crime and Punishment: An Economic Approach". Journal of Political Economy 76, n.º 2 (1968).

Benford, Frank. "The Law of Anomalous Numbers". Proceedings of the American Philosophical Society 78, n.º 4 (1938).

Coase, Ronald H. "The Nature of the Firm". Economica 4, n.º 16 (1937): 386-405

Comptroller and Auditor General of India. "Using Benford's Law in Audit". Research Paper (2020).

Comptroller and Auditor General of India. Artificial Intelligence Strategy Framework. 2025.

Congreso de la República de Colombia, Acto Legislativo 04 de 2019, "por medio del cual se reforma el Régimen de Control Fiscal". Diario Oficial n.º 51.080, 18 de septiembre de 2019.

Congreso de la República de Colombia, Ley 1341 de 2009, "por la cual se definen principios y conceptos sobre la sociedad de la información y la organización de las Tecnologías de la Información y las Comunicaciones ?TIC?, se crea la Agencia Nacional de Espectro y se dictan otras disposiciones". Diario Oficial n.º 47426, 30 de julio de 2009.

Congreso de la República de Colombia, Ley 1474 de 2011, "por la cual se dictan normas orientadas a fortalecer los mecanismos de prevención, investigación y sanción de actos de corrupción y la efectividad del control de la gestión pública". Diario Oficial n.º 48128, julio 12 de 2011.

Congreso de la República de Colombia, Ley 42 de 1993, "Sobre la organización del sistema de control fiscal financiero y los organismos que lo ejercen". Diario Oficial n.º .40732, 27 de enero de 1993.

Congreso de la República de Colombia, Ley 527 de 1999, "Por medio de la cual se define y reglamenta el acceso y uso de los mensajes de datos, del comercio electrónico y de las firmas digitales, y se establecen las entidades de certificación y se dictan otras disposiciones". Diario Oficial n.º 43673, 21 de agosto de 1999.

Congreso de la República de Colombia, Ley 610 de 2000, "por la cual se establece el trámite de los procesos de responsabilidad fiscal de competencia de las contralorías". Diario Oficial n.º 44133, 18 de agosto de 2000.

Congreso de la República de Colombia, Ley Estatutaria 1581 de 2012, "por la cual se dictan disposiciones generales para la protección de datos personales". Diario Oficial n.º 48587, 18 de octubre de 2012.

Congreso de la República de Colombia, Proyecto de Ley 043 de 2025 Senado - 324 de 2025 Cámara, "por medio de la cual se establecen los principios, derechos, deberes, gobernanza y regulación de la inteligencia artificial en Colombia"
Constitución Política de Colombia, 1991.

"Contraloría lanza modelo predictivo con inteligencia artificial para vigilar contratación pública". Portafolio, 2025.

Contraloría General de la República, Informe de Gestión al Congreso 2023-2024. Contraloría General de la República, 2024.

Corte Constitucional de Colombia, Sentencia C-140 de 2020. Expediente D-13517. Magistrado ponente: José Fernando Reyes Cuartas. 6 de mayo de 2020.

Corte Constitucional de Colombia, Sentencia SU-620 de 1996. Expediente T-84714. Magistrado ponente: Antonio Barrera Carbonell. 13 de noviembre de 1996.

Da Silva Azevedo, Caio, Rodrigo Franco Gonçalves, Vagner Luiz Gava y Mauro de Mesquita Spinola. "A Benford's Law Based Method for Fraud Detection Using R Library". MethodsX 8, (2021).

Departamento Administrativo de la Presidencia de la República. Marco Ético para la Inteligencia Artificial en Colombia. Departamento Administrativo de la Presidencia de la República, 2021.

Departamento Nacional de Planeación. Documento CONPES 3975 de 2019, "Política Nacional para la Transformación Digital e Inteligencia Artificial". Departamento Nacional de Planeación, 2019.

Departamento Nacional de Planeación. Documento CONPES 4144 de 2025, "Política Nacional de Inteligencia Artificial". Departamento Nacional de Planeación, 2025.

"DIARI emitió 3800 alertas de riesgo fiscal en Colombia". Ámbito Jurídico, 2025.

"Inteligencia artificial que usó la Contraloría para emitir fallo en caso Hidroituango es válida: Auditoría General". Semana, 2021.

INTOSAI Working Group on Big Data. Guidance on Conducting Audit Activities with Data Analytics. 2022.

Jezierski, Markus, Sascha Kaltenpoth, Oliver Müller y Barbara E. Weißenberger. "The Effect of Large Language Models on Audit Quality: Causal Evidence from a Randomized Field Experiment". SSRN Working Paper, n.º 6095166 (2025). https://ssrn.com/abstract=6095166

Leocádio, Diogo, Luís Malheiro y João Reis. "Artificial Intelligence in Auditing: A Conceptual Framework for Auditing Practices". Administrative Sciences 14, n.º 10 (2024).

Malladhi, Avinash. "Artificial Intelligence and Machine Learning in Forensic Accounting". SSRG International Journal of Computer Science and Engineering 10, n.º 7 (2023).

Márquez Escobar, Carlos Pablo. Anotaciones sobre análisis económico del derecho, vol. I. Pontificia Universidad Javeriana, 2005.

Neves, Silva y Menezes de Carvalho. "Artificial Ladies against Corruption: Searching for Legitimacy at the Brazilian Supreme Audit Institution". Revista de Administração Pública 53, n.º 6 (2019): 1060-1079.

Ordóñez Vásquez, Tatiana. "La responsabilidad fiscal en Colombia y la imposibilidad del resarcimiento patrimonial". IUSTA 1, n.º 26 (2016): 140.

Organisation for Economic Co-operation and Development (OECD). Building an AI-Ready Public Workforce. Organisation for Economic Co-operation and Development (OECD), 2026.

Palmieri, Maximiliano, y Matías Pérez, "Inteligencia artificial, aprendizaje automático y controles automáticos aplicados a la auditoría contable". Ponencia presentada en el XXVII Congreso Nacional de Tribunales de Cuentas, Mendoza, Argentina, 2024.

Polinsky, Mitchell, y Steven Shavell. "The Optimal Use of Fines and Imprisonment". NBER Working Paper, n.º 932 (1982).

Portal Tribunal de Cuentas de la Unión. "TCU é única instituição com uso avançado de inteligência artificial generativa, segundo a OCDE". Notícias, 2024.

Presidencia de la República de Colombia, Decreto Ley 403 de 2020, "por el cual se dictan normas para la correcta implementación del Acto Legislativo 04 de 2019", Diario Oficial n.º 51.258, 16 de marzo de 2020.

Unión Europea. Reglamento (UE) 2024/1689 "por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial)". Diario Oficial de la Unión Europea L 2024/1689, 12 de julio de 2024. https://eur-lex.europa.eu/legal-content/EN-ES/TXT/?-from=EN&uri=CELEX%3A32024R1689

Rodrigues da Cunha, Fábio C., y Maurício S. Bugarin. "Benford's Law for Audit of Public Works: An Analysis of Overpricing in Maracanã Soccer Arena's Renovation". Economics Bulletin 35, n.º 2 (2015): 1168-1176.

The Treadway Commission. Achieving Effective Internal Control Over Generative AI. 2026.

U.S. Department of the Treasury. "Treasury Announces Enhanced Fraud Detection Processes, Including Machine Learning AI, Prevented and Recovered Over $4 Billion in Fiscal Year 2024". Press Release JY-2650, 2024.

U.S. Government Accountability Office. Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities, GAO-21-519SP. Government Accountability Office, 2021.

U.S. Government Accountability Office. Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities. GAO-21-519SP.

U.S. Government Accountability Office. Artificial Intelligence: Generative AI Use and Management at Federal Agencies, GAO-25-107653. Government Accountability Office, 2025.

Williamson, Oliver E. The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting. Free Press, 1985.

Zangana, Hewa Majeed, Rindah Febriana Suryawati, Firas Mahmood Mustafa y Anik Vega Vitianingsih. "AI in Forensic Accounting: Detecting Anomalies and Financial Fraud Through Intelligent Systems". En AI in Digital Forensics and Cybercrime Investigation. IGI Global, 2025.

ACERCA DE LOS AUTORES

Carlos Pablo **Márquez Escobar**

Abogado de la Pontificia Universidad Javeriana, con una maestría en economía de la Pontificia Universidad Javeriana, cuenta con otra Maestría en Derecho de Harvard Law School y un Doctorado en Derecho de University of Oxford.

Fue director ejecutivo de la Comisión de Regulación de Comunicaciones, es profesor de la catedra de derecho financiero en la Pontifica Universidad Javeriana, Actualmente es socio de ECIJA Colombia, es Editor de Oxford University Press y Miembro de la Junta de SoyYo.



Foto: Magnific



03

Responsabilidad civil derivada de daños causados por la inteligencia artificial y robots autónomos:

E-ESTADO
Primera Edición

Responsabilidad civil derivada de daños causados por la inteligencia artificial y robots autónomos: estado del arte doctrinal, análisis de casos y desafíos regulatorios

Civil Liability for Damage Caused by Artificial Intelligence and Autonomous Robots: A Review of Legal Scholarship, Case Studies, and Regulatory Challenges

Fecha de recepción: 7 de abril de 2026
Fecha de aceptación: 22 de julio de 2026

Héctor José García Santiago
Pontificia Universidad Javeriana, Colombia
hjgsabogado@gmail.com

Cómo citar: García Santiago, Héctor José. "Responsabilidad civil derivada de daños causados por la inteligencia artificial y robots autónomos: estado del arte doctrinal, análisis de casos y desafíos regulatorios".
Revista e-Estado 1, n.° 1 (2026).



Foto: Magnific

RESUMEN

El desarrollo de la inteligencia artificial y de los robots autónomos plantea importantes desafíos jurídicos en materia de atribución de responsabilidad por los daños que estos sistemas pueden causar al actuar con cierto grado de autonomía y aprendizaje. Este artículo analiza el problema de la responsabilidad jurídica derivada del comportamiento de agentes artificiales, revisando el estado del arte doctrinal, las principales propuestas teóricas sobre imputación de responsabilidad y algunos casos reales relacionados con tecnologías autónomas. A partir de este análisis, se concluye que los modelos tradicionales de responsabilidad civil resultan limitados frente a los riesgos propios de la autonomía tecnológica, por lo que se propone avanzar hacia esquemas regulatorios híbridos que combinen responsabilidad objetiva, supervisión humana y mecanismos de control como la transparencia y la trazabilidad algorítmica, con el fin de garantizar una adecuada protección jurídica frente al uso creciente de sistemas de inteligencia artificial.

Palabras clave: Responsabilidad civil; inteligencia artificial; robots autónomos; responsabilidad por productos defectuosos; derecho y tecnología; regulación de la IA.

ABSTRACT

The development of artificial intelligence and autonomous robots poses significant legal challenges regarding liability for damages these systems may cause when acting with a degree of autonomy and learning. This article analyzes the problem of legal liability arising from the behavior of artificial agents, reviewing the current state of legal scholarship, the main theoretical proposals on liability attribution, and some real-world cases related to autonomous technologies. Based on this analysis, it concludes that traditional models of civil liability are limited in addressing the risks inherent in technological autonomy.

Therefore, it proposes moving towards hybrid regulatory frameworks that combine strict liability, human oversight, and control mechanisms such as transparency and algorithmic traceability, in order to guarantee adequate legal protection against the increasing use of artificial intelligence systems.

Keywords: *Civil liability; artificial intelligence; autonomous robots; product liability; law and technology; AI regulation.*

01

CONTEXTUALIZACIÓN DEL AVANCE DE LA INTELIGENCIA ARTIFICIAL EN LA ROBÓTICA AUTÓNOMA.

La creciente autonomía de los sistemas de inteligencia artificial (IA) plantea desafíos jurídicos en materia de responsabilidad civil, especialmente en casos donde los robots autónomos causan daños sin intervención humana directa. Este artículo analiza el marco normativo actual, los desafíos de atribución de responsabilidad y propone soluciones regulatorias basadas en ejemplos reales y tendencias globales.

Los robots autónomos están revolucionando sectores como el transporte, la salud y la industria. Según Grand View Research, se estima que el mercado de la IA alcanzará los 1,81 billones de dólares para 2030, creciendo a una tasa compuesta anual del 37,3% entre 2023 y 2030 [1] . Sin embargo, este avance plantea dilemas sobre quién debe asumir la responsabilidad en caso de fallos. En la consulta pública que acompañó la evaluación de impacto de la Comisión Europea sobre IA, el 63% de quienes respondieron apoyó alguna forma de adaptación de las reglas nacionales de responsabilidad a la inteligencia artificial [2] .

El desarrollo de sistemas de inteligencia artificial y robots autónomos representa uno de los desafíos más complejos para el derecho contemporáneo. A diferencia de tecnologías tradicionales, los sistemas de inteligencia artificial poseen capacidades de aprendizaje automático, adaptación al entorno y toma de decisiones independientes. Estas características generan interrogantes fundamentales en torno a la atribución de responsabilidad cuando tales sistemas causan daños.

Las nuevas tecnologías traen consigo importantes beneficios para el ser humano, de hecho se crean para mejorar la calidad de vida de las personas, permitiendo por ejemplo, sustituir la mano de obra humana en actividades riesgosas por máquinas. Pero al tiempo que traen importantes beneficios, se generan nuevos riesgos por el uso de esta tecnología. Los robots autónomos tienen una alta capacidad de generar interacciones directas con el ser humano, lo que puede ocasionar una posible injerencia y violación de derechos de las personas, cuando se lleve a cabo un acto abusivo o no controlado por parte de la máquina y que pueda desencadenar en un daño cierto para una o varias personas [3] . Con la evolución de la tecnología están naciendo nuevas formas de lesionar los derechos, nuevas formas de dañar al ser humano y por qué no, posibles nuevos derechos que deben protegerse. De lo que se trata es de elaborar normas que hagan frente a estas circunstancias, para encontrar un equilibrio. El asunto es que la tecnología llegó para quedarse y para irrumpir en los aspectos más íntimos del ser humano, como menciona Rebollo, estamos ante una invasión inevitable que nos envuelve pero que al mismo tiempo necesitamos y deseamos [4] .

1 Redacción Portafolio, "Mercado de IA crecería más del 37 % entre 2023 y 2030", consultado el 3 de diciembre de 2025, https://www.portafolio.co/economia/inteligencia-artificial-su-mercado-creceria-entre-2023-y-2030-582998 2 Comisión Europea, Impact Assessment Report Accompanying the Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence, SWD(2022) 319 final, 28 de septiembre de 2022, 23, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022SC0319 3 María del Carmen Núñez Zorrilla, Inteligencia artificial y responsabilidad civil: régimen jurídico de los daños causados por robots autónomos con inteligencia artificial (Madrid: Reus, 2019). 4 Lucrecio Rebollo Delgado, Derechos fundamentales y protección de datos (Madrid: Dykinson, 2004), 55–56 y 63.

La tecnología es la muestra de la grandeza de la raza humana, aplicada generalmente a resolver problemas prácticos que aquejan al ser humano. Como lo indica Castilla y Alonso [5] , la tecnología no es otra cosa que la prolongación de los brazos y el corazón de los hombres, estando al servicio de estos.

En tiempos recientes, una de las preocupaciones que aquejan al legislador contemporáneo es la industria de los robots con inteligencia artificial. La preocupación parte de querer establecer una base normativa que permita una acción conjunta humano-máquina que brinde a este binomio un campo de acción que realmente permita impactar positivamente la vida de las personas mediante el uso de robótica con inteligencia artificial. Se trata de garantizar seguridad tecnológica y jurídica que promueva el uso de esta tecnología en beneficio de los seres humanos.

Proyectar respuestas jurídicas para esta nueva relación humano-máquina donde la IA aterriza con fuerza, comporta una necesidad crucial en tiempos modernos, tanto para promover su uso como para consolidar un marco de responsabilidad civil claro y preciso en esta materia.

Para la industria de la robótica es determinante este marco jurídico de responsabilidad civil, tanto contractual como extracontractual, a fin de realizar inversiones que permitan lanzar nuevos productos novedosos de robótica e IA, como por ejemplo el Robot humanoide Optimus Gen de Tesla, diseñado para realizar tareas industriales y domésticas.

Este robot incorporó inteligencia artificial, visión computacional y movilidad avanzada, permitiéndole manipular objetos, así como desplazarse de manera autónoma; o el “Go2 Robot Dog” de Unitree Robotics, uno de los desarrollos más innovadores de la robótica móvil.

Existe una clara y marcada preocupación para que se establezcan altos estándares de seguridad, sobre todo por las actuales características de la robótica, que está destinada a generar nuevas formas de interacción humano-máquina y con ello nuevas formas de ocasionar daño.



Foto: Magnific

5 Adolfo Castilla y María Cruz Alonso, La sociedad española ante las nuevas tecnologías: actitudes y grados de receptividad (Madrid: Fundesco, 1987), 69–70.

02

INTELIGENCIA ARTIFICIAL Y AUTONOMÍA EN ROBOTS

Los sistemas robóticos son cada vez más autónomos, y lo seguirán siendo al punto que operarán sin guía y sin ser teleoperados. La autonomía, fomentada por la inteligencia artificial, seguirá potenciándose y los robots tendrán margen de acción y decisión de manera autónoma, según su propio aprendizaje. La autonomía de los robots permitirá que su uso por parte de los seres humanos sea cada vez más extendido y beneficie a la mayoría de seres humanos, lo cual, al mismo tiempo, impone retos estructurales, entre ellos, la revisión del sistema normativo de responsabilidad civil.

Esta responsabilidad está directamente vinculada con la capacidad de control sobre la máquina. A su vez, esta capacidad de control depende en gran medida de la complejida del ambiente donde se desenvuelva el robot.

En un ambiente dinámico y desafiante el control se torna complejo; sin embargo en un ambiente estable y poco cambiante, el control se torna más seguro y confiable [6] . La capacidad de aprendizaje de las máquinas se ha potencializado gracias a la inteligencia artificial, lo cual a su vez ocasiona que la respuesta pueda ser impredecible en algunos casos, porque el robot con inteligencia artificial aprenderá y actuará en entornos tan diversos, que hará más difícil anticipar por parte del diseñador y programador, las respuestas que tendrá la máquina expuesta a diversos estímulos y aprendizajes, y con ello también queda de presente la dificultad de tomar precauciones y previsiones necesarias por parte del programador [7] .

Los robots autónomos están diseñados para operar con distintos niveles de autonomía, desde sistemas supervisados hasta aquellos con toma de decisiones independiente, lo cual propicia escenarios de automatización. Según un informe de McKinsey Global Institute se estimó que, bajo determinados escenarios, hasta el 30% de las horas trabajadas en la economía estadounidense podría automatizarse para 2030 [8] .

Para comprender mejor lo anterior, se diferencia entre automatización e inteligencia adaptativa: en la automatización tradicional se siguen instrucciones preprogramadas sin capacidad de aprendizaje (ejemplo: robots industriales); en la IA adaptativa, en cambio, se utiliza aprendizaje automático y redes neuronales para mejorar sus decisiones con el tiempo, como pueden ser los robots y vehículos autónomos.

La tecnología de IA, tanto a nivel de automatización como adaptativo, se está expandiendo y llegando cada vez a más ámbitos. Se estima que el mercado de vehículos con funciones autónomas alcanzará el 25% de las ventas mundiales en 2035 según Boston Consulting Group [9] ; en medicina el tamaño del mercado mundial de robots quirúrgicos se valoró en 11,06 mil millones de dólares en 2025 según Fortune Business Insights 10 y en el sector bancario según PWC se estima que la adopción plena de la IA podría generar una mejora de 15 puntos porcentuales en el índice de eficiencia [11] .

6 Sergio Alejandro Moriello, “Los robots inteligentes autónomos son la nueva generación”, The Economy Journal, https://www.theeconomyjournal.com/texto-diario/mostrar/595081/robots-inteligentes-autonomos- nueva-generacion. **7** Erica Palmerini, “Robótica y derecho: sugerencias, confluencias, evoluciones en el marco de una investigación europea”, Revista de Derecho Privado, n.º 32 (2017): 53–97, https://doi.org/10.18601/01234366.n32.03. **8** Kweilin Ellingrud et al., “Generative AI and the Future of Work in America”, McKinsey Global Institute, 26 de julio de 2023, https://www.mckinsey.com/mgi/our-research/generatve-ai-and-the-fut re-of-work-in-america. **9** Xavier Mosquet et al., Revolution in the Driver’s Seat: The Road to Autonomous Vehicles (Boston Consulting Group, 2015), 19. **10** Fortune Business Insights, “AI-Based Surgical Robots Market Size, Share & Industry Analysis”, https://www.fortunebusinessinsights.com/ai-based-surgical-robots-market-106330. **11** PwC, “The Future of Banking: How AI Is Reshaping the Industry”, https://www.pwc.com/us/en/industries/financial-services/library/how-ai-is-reshaping-banking.html.

Cada uno de estos sectores ha evolucionado a través de la IA, pero al mismo tiempo han generado problemas: en 2018, un coche autónomo de Uber atropelló y mató a un peatón en Arizona, aparentemente porque el software tuvo problemas para reconocer y clasificar adecuadamente la presencia de la víctima y poder frenar [12] ; el sistema da Vinci, utilizado en cirugía, ha sido objeto de demandas por errores en procedimientos [13] ; en 2019, un sistema de IA usado en hospitales de EE.UU. discriminó a pacientes afroamericanos al asignarles menor prioridad médica [14] y en el sector financiero algoritmos de asignación de crédito como los de Goldman Sachs fueron cuestionados por decisiones discriminatorias en la asignación de crédito [15] .

La pérdida del control por parte del diseñador y programador de la máquina con sistema inteligente, plantea un desafío jurídico apremiante; esto es, a quién debe atribuírsele la responsabilidad por el daño causado por el robot en el marco de su inteligencia y autonomía [16] .

Los robots humanoides son parte de este desafío a nivel técnico y jurídico, más aún teniendo en cuenta que se fabrican con distintas finalidades. En el mercado se encuentran disponibles robots humanoides cuyo propósito es ser el compañero sentimental del propietario y otros como el Atlas de Boston Dynamics puede correr, saltar, hacer parkour y manipular objetos con gran equilibrio gracias a su visión computarizada e IA avanzada. Otros ejemplos interesantes son el Tesla-Optimus, un robot humanoide diseñado para realizar tareas domésticas e industriales, usando tecnología derivada de los sistemas de conducción autónoma de Tesla; el Figure AI–Figure 01, pensado para trabajar junto a humanos en fábricas y logística y el Engineered Arts de Ameca, famoso por sus expresiones faciales hiperrealistas, orientado a la interacción social, demostraciones tecnológicas y conversaciones con IA.

Naturalmente, a mayores funcionalidades se abre el espectro de posibles circunstancias de responsabilidad. En este contexto, de robótica, inteligencia artificial y autonomía de las máquinas, la creación de un marco jurídico idóneo que regule estas relaciones y la responsabilidad ante la ocurrencia de un daño causado por estas máquinas autónomas, representa, como se ha manifestado, un aspecto crucial. Los derechos que pueden verse afectados, prima facie, son los siguientes:

1. Derecho a la vida (Art. 11 C.P).
2. Derecho a no recibir tratos crueles, inhumanos o degradantes (Art. 12 C.P).
3. Derecho a la libertad y a la no discriminación (Art. 13 C.P).
4. Derecho a la intimidad personal y familiar y al buen nombre (Art. 15 C.P).
5. Derecho al libre desarrollo de la personalidad (Art. 16 C.P).
6. Derecho a la libertad de conciencia (Art. 18 C.P).
7. Derecho a la libertad de cultos (Art. 19 C.P).
8. Derecho a libertad de expresar y difundir su pensamiento y opiniones (Art. 20 C.P).
9. Derecho a la honra (Art. 21 C.P).
10. Derecho a la libertad de circulación (Art. 24 C.P).

12 National Transportation Safety Board, Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, Arizona, March 18, 2018, NTSB/HAR-19/03 (Washington, D. C.: National Transportation Safety Board, 2019), 17–19 y 59, https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1903.pdf 13 Intuitive Surgical, Inc., Annual Report 2024, Form 10-K (2025), apartados "Risk Factors" y "Legal Proceedings". 14 Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations", Science 366, n.º 6464 (2019): 447–453, https://doi.org/10.1126/science.aax2342.15 New York State Department of Financial Services, Report on Apple Card Investigation (marzo de 2021). 16 María del Carmen Núñez Zorrilla, "Los nuevos retos de la Unión Europea en la regulación de la responsabilidad civil por los daños causados por la inteligencia artificial", Revista Española de Derecho Europeo, n.º 66 (2018): 9–53.

Todos estos derechos deben ser respetados y protegidos en el marco de la robótica y la inteligencia artificial en un nuevo escenario en el que la IA desafía y se impone al ser humano en la mayoría de las áreas en las que hasta ahora éste gozaba de hegemonía, como en su momento lo indicó el Parlamento Europeo en su Resolución del 16 de febrero de 2017.

El legislador, entonces, está abocado a elaborar una regulación de avanzada que impulse el desarrollo de la robótica y su autonomía en un marco de respeto por los derechos fundamentales de las personas, en un entorno sostenible y congruente donde el foco sea el bienestar de la humanidad [17].

03

RESPONSABILIDAD CIVIL EN EL MARCO JURÍDICO ACTUAL

3.1. Contextualización del problema jurídico de la responsabilidad civil en sistemas de IA

El derecho tradicional, en todas sus ramas, se ha construido bajo el entendimiento colectivo que las conductas reguladas son realizadas por seres humanos o por instrumentos bajo su control. En el presente caso, el de los robots autónomos, se cambia esa premisa, en la medida que las conductas a regular no son realizadas por seres humanos, sino por robots autónomos, lo que claramente dificulta, -y de ahí la necesidad de regulación-, precisar a quién debe atribuírsele esa responsabilidad por el daño causado por la máquina.

La problemática central consiste en determinar si los marcos jurídicos actuales de responsabilidad civil son suficientes para abordar los daños causados por robots autónomos o si, por el contrario, es pertinente desarrollar nuevas categorías jurídicas adaptadas a la robótica y la inteligencia artificial.

La propuesta de Directiva sobre responsabilidad extracontractual en materia de inteligencia artificial de 2022 [18] (actualmente retirada) indicaba que los sistemas de alto riesgo, como robots autónomos en salud y transporte, deberían cumplir con requisitos estrictos de supervisión y pruebas de seguridad, proponía mecanismos de exhibición de pruebas y, bajo determinadas condiciones, una presunción refutable del nexo causal cuando se acreditara una conducta culposa consistente, entre otros supuestos, en el incumplimiento de deberes jurídicos aplicables.La inteligencia artificial de los robots y su capacidad de aprendizaje autónomo, sumado a la dificultad de control sobre estos sistemas, en particular en ambientes complejos, hace que sea difícil determinar la causa del daño o error y en consecuencia hace compleja la atribución de responsabilidad, generado extensos debates sobre si la responsabilidad recae en el fabricante, el operador de la máquina o incluso el programador del software.

Existen varios casos documentados que han motivado las reflexiones en torno a la responsabilidad civil de sistemas de IA. Un ejemplo es el caso Benavides v. Tesla Inc. [19], donde un conductor de un Tesla Model S activó el sistema ”Autopilot” y, mientras buscaba su celular, el vehículo atravesó una intersección y atropelló a dos personas; una murió y la otra quedó gravemente herida. La familia de las víctimas demandó a Tesla argumentando que el sistema permitía usos inseguros, exagerando las capacidades “autónomas” y que el vehículo no frenó pese a detectar obstáculos, comportando una falla del vehículo.

17 Núñez Zorrilla, Inteligencia artificial y responsabilidad civil. **18** Comisión Europea, Propuesta de Directiva relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial, COM(2022) 496 final, 28 de septiembre de 2022, arts. 3–4, https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=-CELEX:52022PC0496. **19** Benavides v. Tesla, Inc., n.º 1:21-cv-21940-BB, (Tribunal Federal para el Distrito Sur de Florida, 19 de febrero de 2026), 2–3 y 8.

En 2025, un jurado de Florida declaró a Tesla parcialmente responsable y ordenó pagar alrededor de 240 millones de dólares a la familia. En este caso en particular, la responsabilidad fue compartida por el conductor y el diseñador del sistema autónomo (Tesla), abriendo la puerta para demandas similares contra los fabricantes por fallas en los sistemas autónomos.

Otro caso de responsabilidad fue el del sistema quirúrgico Intuitive Surgical, que enfrentó demandas por lesiones internas, quemaduras y fallos durante las cirugías [20] . En varios de los hechos se logró determinar entrenamiento insuficiente, uso incorrecto y falta de advertencias a los pacientes. Este contexto generó preguntas legales tales como: ¿el error fue del médico? ¿del software? ¿del fabricante de la máquina? ¿de la máquina? ¿del hospital?

Los sistemas autónomos están creando una zona gris de responsabilidad civil, toda vez que los regímenes de responsabilidad están diseñados sobre el actuar de las personas, sin embargo, hoy en día, con el auge de la tecnología y la unión de la robótica y la inteligencia artificial, las conductas ya no son exclusivas de los seres humanos.

En un accidente automovilístico, si un conductor choca es generalmente su culpa, por imprudencia suya o falla mecánica del vehículo que es usualmente atribuida a este y puede conllevar a una responsabilidad civil y/o penal.

Hoy en día, con los vehículos autónomos, la responsabilidad ya no es exclusiva del conductor y puede extenderse al fabricante, al programador, al diseñador de la IA, a la empresa operadora, al proveedor de los sensores, entre otros. De manera que empiezan a acuñarse nuevos entornos de responsabilidad, generando varias formas de responsabilidad como la algorítmica, la culpa distribuida, o el riesgo sistémico de automatización.

Esta complejidad en el régimen de responsabilidad parte de la fragmentación de la cadena en los productos desarrollados con IA, que se observa a partir de los intervinientes: el equipo de desarrolladores que crea el programa, la persona que los dirige, la empresa que es dueña de los derechos patrimoniales del software y se anuncia como fabricante, aquel que comercializa el producto y, por último, el usuario, aunque también se pueden agregar más actores a la cadena dependiendo del producto, como los integradores, fabricantes de componentes, entre otros.

Cuando se habla de robots autónomos, la cadena es todavía más compleja, pues se suma el fabricante del hardware (el cuerpo del robot) dado que una falla puede derivarse tanto del software como del hardware, así como de componentes que permitan la comunicación entre estos dos sistemas.

20 Intuitive Surgical, Annual Report 2024, apartados "Risk Factors" y "Legal Proceedings".



Foto: Magnific

La responsabilidad recaería en primera medida sobre el fabricante, quien es el garante del producto en el momento en que decide lanzarlo al mercado, Al igual que sucede con cadenas de producción de computadores, celulares o vehículos convencionales, el fabricante, en principio, no puede excusarse en que su equipo de diseño o su proveedor de partes o herramientas fue el causante de la falla para exonerarse de toda responsabilidad.

Sin embargo, cuando se trata de sistemas de IA, además de la fragmentación indicada anteriormente surgen otras preguntas en torno a la responsabilidad. En sistemas de IA, a partir de una idea inicial se construye el código de software por parte de los expertos técnicos para empezar a darle vida y ahí está el primer problema: quien está realizando el desarrollo de software en ocasiones no es una persona humana, sino otro sistema de IA. De hecho, motores abiertos al público como Claude o ChatGPT están avanzando en esas capacidades de desarrollo de software, lo cual abre una primera pregunta: ¿empresas como Anthropic podrían ser responsables de fallas en el código de software atribuibles a Claude? Siguiendo con la cronología, una vez el sistema se ha desarrollado es momento de entrenarlo, a partir de bases de conocimientos, ejercicios controlados, casuística, entre otros; la complejidad y tipo de entrenamiento dependerá del sistema a crear. Esta fase es crucial, dado que el sistema funcionará a partir del entrenamiento brindado para identificar todos los supuestos de hecho y recibir instrucciones sobre qué debe hacerse en cada uno de ellos.

Al respecto se puede pensar, por ejemplo, en un jugador de ajedrez. Al jugador se le forma en mover las fichas, tácticas y estrategias, pero una vez empieza una partida real, las posibilidades y supuestos que pueden acontecer son prácticamente infinitos.

Entre más partidas juegue, gane y pierda, el jugador irá aprendiendo sobre sus aciertos, errores y alternativas para enriquecer su sistema de juego.

En el ejemplo dado, la IA puede aprender de manera similar que este jugador de ajedrez, empezando por ese entrenamiento básico y luego practicando a partir de escenarios reales. Sin embargo, a pesar de que se tenga la misma base (reglas del juego, tácticas, aperturas y otros elementos universales), su comportamiento va a depender del entrenamiento dado, pues probablemente no jugará igual si se entrena con partidas de jugadores principiantes a entrenarse con partidas del nivel de gran maestro.

Además de ello, el entrenamiento no solamente consiste en darle a la IA un gran volumen de bases de datos. Se requieren reglas básicas para el análisis, metodología para examinar la información recibida, definición de prioridades, entre otros.

De esta forma, el entrenamiento dependerá tanto de las bases de conocimiento, como de la metodología para analizar esas bases y, posteriormente, de las pruebas que se realicen sobre el sistema de IA para evaluar si luego del aprendizaje está adoptando el resultado esperado. En estos casos, las fallas pueden provenir del proveedor de las bases de conocimiento, del diseño del entrenamiento o del propio sistema, al asimilar y/o aprender de forma inadecuada o distinto a lo esperado.

Luego de que se crea el sistema, se debe integrar con el hardware; en el caso de un vehículo autónomo, ya no se trata solamente de que el sistema de IA conozca con exactitud cada señal de tráfico y cómo comportarse frente a cada una de ellas, sino que el sistema deberá ser capaz de aplicar todo ese conocimiento para conducir un vehículo real.

Esto implica una integración de sistemas donde cualquier falla de la IA en su parte de procesamiento puede llevar a que el vehículo tenga un accidente e idéntico resultado se dará cuando la falla proviene del hardware o sistema del vehículo, por ejemplo, cuando interpreta mal la orden del sistema de IA o, de manera más sencilla, en aquellos casos en que falla el sistema de frenado, la caja de cambios y otros inconvenientes mecánicos que se presentan actualmente en los vehículos tradicionales.

Cuando todas estas piezas engranan y los controles son satisfactorios (explicado de forma muy sintetizada), el robot o vehículo autónomo puede salir al mercado y al final llegará al usuario. Al igual que con los vehículos tradicionales, los vehículos autónomos también dependen de las instrucciones dadas por el usuario.

En un panorama tradicional, se puede fabricar el vehículo perfecto, pero si el usuario no sabe manejar, inevitablemente causará un accidente de tránsito. En el marco de los sistemas de IA puede pasar lo mismo, pero la evaluación de responsabilidad se vuelve más compleja. Para entender mejor esto se puede pensar en los sistemas de ChatGPT o Claude. El sistema de IA que utiliza cada persona es el mismo, pero dependiendo de las configuraciones del usuario, pueden dar respuesta a una pregunta de forma totalmente distinta, dependiendo de la base de conocimientos personalizada que se tenga, las instrucciones recibidas y las preferencias del usuario.

Ello implica que ciertos sistemas de IA (no todos) sigan aprendiendo, aun cuando están con el usuario final, de tal forma que este usuario puede influir directamente en cómo funciona y responde ese robot y vehículo autónomo, participando de forma activa en la "maduración" del producto.

Esta larga cadena de creación de sistemas de IA implica que en caso de un daño sea más complejo el régimen de atribución de responsabilidades. Por ejemplo, en el caso de un robot autónomo cuyo usuario le dio instrucciones para cargar un bebé y lo deja caer, ¿Quién respondería? ¿El usuario por darle órdenes al robot a pesar de que no estaba diseñado para la labor de cargar bebés? ¿El fabricante, por no implementar controles más estrictos que de plano le prohibieran al robot hacer ese tipo de labores? ¿El fabricante del hardware del robot, por no diseñar piezas y funcionalidades que impidieran que el robot dejara caer al bebé?

Justamente por este tipo de preguntas y dificultades es que en distintos pronunciamientos, como el informe del grupo de expertos sobre responsabilidad y nuevas tecnologías del 2019 [21] se resaltó que las tecnologías emergentes dificultan la identificación de responsables y por ello se deben realizar ajustes selectivos y un estudio específico de la carga probatoria en cada caso.

Desde una perspectiva jurídica, la responsabilidad civil descansa sobre varios presupuestos esenciales. El primero es la existencia de un daño cierto, entendido como la afectación de un interés legítimo de la víctima, ya sea de carácter patrimonial o extrapatrimonial. El segundo corresponde a la conducta atribuible al presunto responsable, la cual puede manifestarse mediante una acción o una omisión. El tercer elemento es el nexo causal, es decir, la relación directa entre la conducta desplegada y el daño ocasionado. Finalmente, se requiere un factor de imputación, que en el régimen tradicional se identifica con la culpa o el dolo, aunque en determinados eventos la ley admite esquemas de responsabilidad objetiva [22] .

21 Expert Group on Liability and New Technologies, Liability for Artificial Intelligence and Other Emerging Digital Technologies (Luxemburgo: Publications Office of the European Union, 2019), https://doi.org/10.2838/573689. 22 Javier Tamayo Jaramillo, Tratado de responsabilidad civil, t. I, 2.ª ed., 8.ª reimp. (Bogotá: Legis, 2015), 188.

Cuando se habla de inteligencia artificial y robots autónomos, esta teoría tradicional de la responsabilidad civil enfrenta un desafío significativo. Históricamente, los sistemas de responsabilidad se han construido sobre la idea de que una persona natural o jurídica realiza una conducta que produce un daño y, por tanto, debe responder por sus consecuencias.

Sin embargo, los sistemas de inteligencia artificial avanzados pueden adoptar decisiones con un alto grado de autonomía, aprendiendo de su entorno y ejecutando acciones que no siempre son previsibles para sus desarrolladores, fabricantes o usuarios. Esta situación plantea la discusión acerca de si el robot o sistema autónomo debe ser considerado únicamente como una herramienta bajo control humano o si podría reconocerse algún tipo de estatus jurídico especial para efectos de atribución de responsabilidad.

Diversos autores sostienen que los modelos clásicos basados exclusivamente en la culpa resultan insuficientes para afrontar los riesgos derivados de la autonomía tecnológica.Según Calo [23] , la capacidad de aprendizaje de los sistemas inteligentes dificulta la identificación del sujeto responsable cuando se produce un daño. Por su parte, Abbott [24] señala que la creciente autonomía de la inteligencia artificial obliga a replantear las categorías jurídicas tradicionales, pues la conducta dañosa puede derivarse de procesos de decisión que no fueron programados de manera específica. Asimismo, Bertolini [25] considera que el derecho deberá evolucionar hacia modelos de imputación más flexibles, combinando mecanismos de responsabilidad objetiva, seguros obligatorios y fondos de compensación para proteger eficazmente a las víctimas.

Cuando se trata de relaciones de consumo, además de los problemas de responsabilidad también pueden surgir otras inquietudes. En el marco del análisis de la Propuesta de Directiva del Parlamento europeo y del Consejo del 28 de septiembre de 2022 (COM/2022/495 que posteriormente derivó en la Directiva (UE) 2024/2853), se identificaron dificultades para encajar el régimen de protección al consumidor frente a los sistemas de IA, tales como la falta de una definición de lo que es IA, la dependencia de la IA con ciertos servicios como los datos para cumplir sus funciones, sus continuas modificaciones y avances, sus distintas formas de comercialización y el hecho de que pueden comercializarse como un servicio [26] .

23 Ryan Calo, "Robotics and the Lessons of Cyberlaw", California Law Review 103, n.º 3 (2015): 513–563. 24 Ryan Abbott, The Reasonable Robot: Artificial Intelligence and the Law (Cambridge: Cambridge University Press, 2020). 25 Andrea Bertolini, Artificial Intelligence and Civil Liability: A European Perspective (Bruselas: Parlamento Europeo, 2025), PE 776.426, https://doi.org/10.2861/0075079. 26 Inmaculada Herbosa Martínez, "Encaje de los sistemas de IA en la definición de producto en la legislación de productos defectuosos: análisis de la legislación vigente con la vista puesta en la Propuesta de Directiva del Parlamento Europeo y del Consejo de 28 de septiembre de 2022 (COM/2022/495)", InDret, n.º 3 (2024): 52–98, https://doi.org/10.31009/InDret.2024.i3.02.


Foto: Magnific

Expuesto el panorama general, se realizará un breve repaso por la legislación colombiana y europea en materia de responsabilidad civil e IA. Este resumen también tendrá en cuenta la normativa de derecho del consumo, dado que en muchas ocasiones los sistemas de IA pueden ocasionar daños en el marco de una relación de consumo.

3.2 El régimen de responsabildad civil y la normativa sobre IA en Colombia

En el ordenamiento colombiano, el fundamento ético de la responsabilidad civil se relaciona con el principio romano alterum non laedere, que significa no causar daño a otro. Este postulado ha sido complementado por los valores constitucionales de dignidad humana, solidaridad y reparación integral, los cuales orientan la interpretación contemporánea de las normas sobre responsabilidad. En consecuencia, la obligación de indemnizar surge como un mecanismo de justicia correctiva destinado a proteger a quien ha sufrido un perjuicio injustificado.

La responsabilidad civil se sustenta en el principio jurídico según el cual toda persona que cause un daño a otra está obligada a repararlo. Este criterio encuentra respaldo tanto en las disposiciones del Código Civil como en el desarrollo doctrinal y jurisprudencial que ha buscado garantizar la protección efectiva de los derechos de las víctimas. Su finalidad principal es restablecer, en la mayor medida posible, la situación que existía antes de la ocurrencia del perjuicio [27] .

En Colombia no existe una legislación específica en torno a la responsabilidad por vehículos o robots autónomos, por lo cual es necesario recurrir a la normativa general en materia de responsabilidad y de protección al consumidor, sin perjuicio de la necesidad que existe en el país para avanzar en una legislación sobre IA que contemple estos casos de uso.

Lo anterior no quiere decir que en Colombia haya orfandad regulatoria en caso de un daño ocasionado por un robot autónomo, sino que se debería aplicar la normativa general y en consecuencia habrá ciertos desafíos derivados de la naturaleza especial de este tipo de responsabilidad.

Hecha esta claridad, la primera norma a tener en cuenta es el Código Civil, que consagra el régimen de responsabilidad general. Los artículos 1602 y 1603 consagran las reglas contractuales según las cuales el contrato “es una ley para los contratantes” y “deben ejecutarse de buena fe”. Estas disposiciones son la base para la exigibilidad de cualquier vínculo contractual.

En términos de responsabilidad extracontractual, los artículos 2341 y siguientes indican las reglas de responsabilidad por los delitos y las culpas, estableciendo el deber general de indemnización de aquel que por delito o culpa genere un daño a otro, sin perjuicio de la pena que consagre la legislación para el efecto.

27 Corte Suprema de Justicia, Sala de Casación Civil, Sentencia SC506-2022, 17 de marzo de 2022, M. P. Hilda González Neira, rad. 63001-31-03-001-2015-00095-02, 21–22

También se señalan otras reglas relevantes:

a. el principio de solidaridad cuando dos o más personas sean responsables;
b. se genera responsabilidad cuando el daño puede imputarse a malicia o negligencia;
c. la reparación puede reducirse cuando la víctima se expuso al daño imprudentemente.
d. el deber de indemnizar por un daño incluye no solo las acciones de la persona, sino "del hecho de aquellos que estuvieren a su cuidado", salvo que aun con "el cuidado que su respectiva calidad les confiere y prescribe, no hubieren podido impedir el hecho".

Debido a las funcionalidades que pueden tener los robots autónomos, cabe la posibilidad de que su funcionamiento encuadre dentro de una actividad peligrosa. En este contexto son importantes los siguientes lineamientos jurisprudenciales:

- Cuando se trata de actividades peligrosas, no basta con demostrar la diligencia del guardián, sino que se debe demostrar causa extraña; así mismo, cuando se alega como causa exclusiva del daño la conducta de la víctima, debe demostrarse que la intervención de la víctima fue el único factor que produjo el daño [28].

- Dicha postura, ratificada posteriormente, permite establecer que en mas ctividades peligrosas hay un régimen de responsabilidad objetiva [29].

- Debe aclararse que aunque opere una presunción ligada a una actividad peligrosa, la víctima siempre debe acreditar el daño y el nexo causal [30].

- Toda actividad peligrosa es riesgosa, pero no toda actividad riesgosa es peligrosa (salvamento de voto) [31].

Por otro lado, si la responsabilidad se enmarca en una relación de consumo bajo la cual sea aplicable la Ley 1480 de 2011, el usuario podrá ser considerado consumidor y, en consecuencia, tendrá derechos tales como el de recibir productos de calidad, recibir información, el derecho a la seguridad e indemnidad, entre otros.

De acuerdo con el artículo sexto de dicha norma, el productor debe asegurar la idoneidad, calidad y seguridad de los bienes y servicios que ofrezca. En caso de que este deber sea incumplido se genera una responsabilidad solidaria del productor y proveedor por la garantía, además de una responsabilidad ante las autoridades competentes.

El artículo décimo indica que los productores y proveedores son responsables solidarios de la garantía legal y en determinados casos de la garantía suplementaria. Habrá exoneración de la responsabilidad de la garantía cuando haya circunstancias de fuerza mayor o caso fortuito, hecho de un tercero, uso indebido del bien o cuando el consumidor no atienda las instrucciones de uso, mantenimiento o instalación, siempre que las mismas estén indicadas en el manual.

En el caso de daños ocasionados por productos defectuosos serán responsables de forma solidaria el productor y el expendedor, pero podrán exonerarse, entre otras causales, cuando los daños ocurran por culpa exclusiva del afectado, por hecho de un tercero o cuando *"en el momento en que el producto fue puesto en circulación, el estado de los conocimientos científicos y técnicos no permitía descubrir la existencia del defecto".*

28 Corte Suprema de Justicia, Sala de Casación Civil, Sentencia SC002-2018, 12 de enero de 2018, M. P. Ariel Salazar Ramírez, rad. 11001-31-03-027-2010-00578-01. 29 Corte Suprema de Justicia, Sala de Casación Civil, Sentencia SC4420-2020, 17 de noviembre de 2020, M. P. Luis Armando Tolosa Villabona, rad. 68001-31-03-010-2011-00093-01. 30 Corte Suprema de Justicia, Sala de Casación Civil, Sentencia SC065-2023, 27 de marzo de 2023, M. P. Hilda González Neira, rad. 05001-31-03-005-2010-00259-01, 36 y 40. 31 Luis Armando Tolosa Villabona, salvamento de voto a la Sentencia SC4204-2021 de la Corte Suprema de Justicia, Sala de Casación Civil, 22 de septiembre de 2021, rad. 05001-31-03-003-2004-00273-02, 98.

En materia jurisprudencial existen varios pronunciamientos relevantes en materia de protección al consumidor:

- El producto no es apto para el consumo cuando no puede satisfacer las expectativas del consumidor y es defectuoso cuando puede afectar la salud o la integridad física del consumidor [32] .

- Las obligaciones de seguridad que se incluyen en los contratos con los consumidores constituyen una obligación de resultado [33] .

- La Corte Constitucional analizó el anterior régimen de protección al consumidor, a través de la Sentencia C-1141 de 2000, donde estableció que cuando el productor controla el diseño, el proceso productivo y coloca el producto en circulación, también puede ser demandado en virtud de la acción directa, pues el artículo 78 de la Constitución impide su exclusión [34] .

Aunque no se cuenta con una legislación específica de IA, en Colombia cursó el Proyecto de Ley 043 de 2025 Senado y 324 de 2025 Cámara, el cual buscaba regular el desarrollo de la IA en Colombia con un enfoque, ético, responsable, competitivo, innovador y respetuoso de los derechos fundamentales; sin embargo actualmente esta iniciativa legislativa está archivada y se debe esperar una nueva versión del proyecto para analizarla. Entre los aspectos más relevantes a resaltar del proyecto de ley, se encuentran la regulación basada en niveles de riesgo, -partiendo del AI Act-, la protección de los derechos humanos, reiterando la importancia de respetar la privacidad, igualdad, no discriminación, debido proceso, transparencia y protección de datos personales.


Foto: Magnific

En el mismo sentido, uno de los aspectos más relevantes se centra en la transparencia y explicabilidad, de manera que se pueda tener la traza de las decisiones automatizadas. La supervisión humana obligatoria es también otro aspecto relevante, esto es, la capacidad de intervención y el control sobre decisiones críticas. Aunque todavía no tan detalladas como en Europa, el proyecto limitaba usos peligrosos de la IA, por ejemplo la manipulación de personas, la discriminación algorítmica, así como la vigilancia biométrica abusiva.

La creación de una Autoridad Nacional de IA especializada y liderada por el Ministerio de Ciencia, Tecnología e innovación (MinCiencias), es el punto central de la gobernanza institucional que planteaba el proyecto de ley. También se buscaba el impulso económico y la soberanía tecnológica del país en materia de IA, impulsando a las empresas y startups, fomentando el talento y fortaleciendo la investigación a fin de reducir la dependencia tecnológica externa.

32 Corte Suprema de Justicia, Sala de Casación Civil, Sentencia de 30 de abril de 2009, M. P. Pedro Octavio Munar Cadena, rad. 25899-31-03-002-1999-00629-01. 33 Ibíd. 34 Corte Constitucional, Sentencia C-1141 de 2000, 30 de agosto de 2000, M. P. Eduardo Cifuentes Muñoz

Como última referencia normativa se encuentre el CONPES 4144 de 2025 de IA, que aunque no es una norma vinculante si marca un derrotero acerca de la necesidad de desarrollar capacidades específicas de IA y un esquema de gobernanza que permita afrontar los retos de esa tecnología.

En este contexto, el régimen normativo actual está llamado a evolucionar. Más que reconocer una responsabilidad autónoma del robot en el corto plazo, la tendencia jurídica internacional apunta a fortalecer los mecanismos de atribución de responsabilidad a fabricantes, operadores, propietarios y desarrolladores de sistemas inteligentes, sin descartar futuras reformas que permitan construir nuevas categorías jurídicas adaptadas a la realidad tecnológica. El objetivo central continúa siendo garantizar la reparación integral del daño, preservar la seguridad jurídica y fomentar la innovación responsable en una sociedad cada vez más automatizada.

En conclusión, aunque Colombia tiene una normativa general de responsabilidad contractual y extracontractual, la responsabilidad civil en máquinas autónomas requiere regulaciones específicas que consideren la autonomía y capacidad de aprendizaje de estos sistemas. Las regulaciones sobre la IA en Europa han aportado un punto de partida para generar un marco regulatorio de responsabilidad civil por los daños causados por robots autónomos en Colombia.

3.3. Normativa europea sobre IA y responsabilidad

Una base normativa muy importante a nivel mundial y que puede ser referente para el caso colombiano es el AI Act o Reglamento Europeo de Inteligencia Artificial (UE 2024/1689), expedido en 2024 y con una implementación gradual, teniendo como objetivo principal el mejorar el funcionamiento del mercado promoviendo la adopción de la IA centrada en el ser humano y adoptando un enfoque regulatorio basado en el riesgo.

El Reglamento consagra, entre otras, normas armonizadas para la comercialización, la puesta en servicio y el uso de los sistemas de IA en la Unión; prohibiciones de determinadas prácticas de IA; requisitos específicos para los sistemas de IA de alto riesgo y obligaciones para los operadores de dichos sistemas; normas de transparencia armonizadas para determinados sistemas de IA y normas sobre supervisión, vigilancia, gobernanza y ejecución del mercado.



Foto: Magnific

Las prohibiciones marcan un punto de partida trascendental: por ejemplo la prohibición de usar técnicas subliminales que escapen a la conciencia de una persona o técnicas deliberadamente manipuladoras o engañosas, o técnicas que exploten cualquiera de las vulnerabilidades de una persona física o de un grupo específico de personas debido a su edad, discapacidad o una situación social o económica específica, haciéndoles tomar una decisión que de otro modo no habrían tomado causando un daño significativo, se convierte en un insumo valioso para fijar un nuevo y moderno marco normativo que contemple las nuevas situaciones de daño gracias a esa autonomía de los robots autónomos.

El AI Act también establece una categorización de los sistemas de IA de acuerdo con su nivel de riesgo, indica el deber de contar con un sistema de gestión de riesgos para ciertos casos, el registro automático de eventos, la vigilancia efectiva por personas físicas y el diseño y desarrollo de sistemas con un nivel adecuado de precisión, solidez y ciberseguridad.

Más que un régimen de indemnización o de responsabilidad pura, este instrumento debe ser concebido e interpretado como un régimen preventivo y, por ende, su objetivo es fijar reglas para que no se produzca un daño. Debe aclararse que el incumplimiento del AI Act puede ser relevante para valorar la diligencia, la ilicitud o la seguridad esperable, pero no determina automáticamente la responsabilidad civil, cuyos presupuestos deben probarse conforme al régimen aplicable.

Además del AI Act en la Unión Europea se emitió la Directiva 2024/2853 del 23 de octubre de 2024 sobre productos defectuosos, la cual contempla específicamente el software como producto. Dentro de esta reglamentación se destaca la definición de producto defectuoso, entendido como aquel que *"no ofrezca la seguridad que una persona tiene derecho a esperar o que se exige en virtud del Derecho de la Unión o nacional".*

El criterio de seguridad, basado en lo que una persona tiene "derecho a esperar" generará, sin duda, múltiples caminos interpretativos que se irán decantando cuando esta Directiva sea transpuesta por los Estados miembros, a partir de diciembre de 2026, pues no siempre puede quedar muy claro que es lo que una persona tiene derecho a esperar, aunque la misma Directiva brinda unas pautas al establecer que la seguridad exigible debe establecerse atendiendo al conjunto de circunstancias del producto, incluidas su finalidad, características, uso razonablemente previsible, capacidad de aprendizaje, interconexiones y requisitos de seguridad y ciberseguridad.

Así mismo, deben empezarse a evaluar los términos y condiciones, exclusiones y limitaciones impuestas por el mismo fabricante de forma más exhaustiva. El hecho de que sea una tecnología en construcción, cuyo uso implica riesgos conocidos por el usuario y cuya utilización puede implicar, por sí misma, un cierto nivel de riesgo, pueden ser elementos que maticen la expresión del "derecho a esperar".

Por otro lado, la Resolución del Parlamento Europeo del 20 de octubre de 2020 (2020/2014) [35] propone recomendaciones como un modelo escalonado de responsabilidad, estableciendo responsabilidad objetiva para operadores de alto riesgo y un régimen de culpa para otros sistemas.

35 Parlamento Europeo, Resolución de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial, 2020/2014(INL), https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_ES.html.

Por su parte, el Reino Unido expidió en 2024 una legislación enfocada en vehículos autónomos [36] , donde se separan las funciones y responsabilidades de la empresa autrizada y del usuario. En el capítulo 1 se establece que los vehículos autónomos deben hacer viajes con seguridad (viajar con un nivel de seguridad aceptable) y legalmente (con baja probabilidad de cometer una infracción de tránsito), garantizando que se desempeñen con un nivel de seguridad equivalente o superior al de conductores humanos cuidadosos y competentes.

La regulación del Reino Unido también exige el nombramiento de entidades autorizadas responsables de la conducción autónoma. Estas entidades deben ser de buena reputación y situación financiera, además de encontrarse en capacidad de cumplir cualquier requisito de autorización solicitado por la normativa. La autorización de determinado vehículo automatizado puede suspenderse o modificarse, entre otras circunstancias, si se comete una infracción de tránsito. Por otro lado, el régimen establece que las entidades que autorizan a los vehículos automatizados pueden cometer un delito, entre otras circunstancias, cuando se emite información falsa o engañosa sobre aspectos relevantes para la seguridad y el delito se agrava si hubo un incidente peligroso o se causa la muerte o lesiones graves.



Foto: Magnific

3.4. Consideraciones sobre el régimen probatorio

La Ley 1480 de 2011 consagra en su artículo 20 la regla general de responsabilidad, según la cual el productor y el expendedor son solidariamente responsables por los daños que se causen a raíz de los defectos de sus productos. Esta determinación de la responsabilidad exige que el afectado pruebe sus elementos, a saber: a) el defecto del bien; b) la existencia del daño y c) el nexo causal. Este estándar aplicado a robots autónomos suele imponer una carga de la prueba considerable, pues no siempre va a ser fácil probar el defecto del bien, ni su nexo causal con el daño ocasionado.

Aunque la misma norma establece la presunción del defecto cuando se vulnera un reglamento técnico, dicha presunción no solamente exige que ese reglamento técnico exista, sino el conocimiento de sus disposiciones y el entendimiento del defecto.

En la Sentencia C 472 de 2020 se reconoció que el régimen de derecho del consumo requiere reglas especiales de responsabilidad, fundado en hechos como el de haberse puesto en circulación un producto defectuoso. Esta misma Sentencia también reconoció que puede haber una carga probatoria difícil para el consumidor por complejidades técnicas o de otra índole [37].

Justamente por situaciones como estas es que en la Sentencia C 086-2016 se indicó que en determinados casos la potestad del juez de distribuir dinámicamente la carga de la prueba de acuerdo con lo establecido en el Código General del Proceso no es solo una facultad, sino un deber funcional.

36 Reino Unido, Automated Vehicles Act 2024, c. 10, partes 1–2, especialmente secs. 2 y 6, https://www.legislation.gov.uk/ukpga/2024/10/contents. 37 Corte Constitucional, Sentencia C-472 de 2020, 5 de noviembre de 2020, M. P. Diana Fajardo Rivera, consideraciones 14 y 21 y parte resolutiva

La Directiva Europea sobre productos defectuosos (2024/2853) tiene el mismo estándar de la carga de la prueba, donde el afectado debe demostrar el defecto, el daño y el nexo causal, pero si se cumple alguna de estas condiciones el defecto se presume: la falta de exhibición de las pruebas pertinentes que disponga, cuando ello sea exigible; el incumplimiento de requisitos obligatorios de seguridad o cuando se demuestre que el daño fue causado por un mal funcionamiento manifiesto del producto durante un uso razonablemente previsible o en circunstancias normales.

Así mismo, esta Directiva Europea establece que el nexo causal se presume si se comprueba el defecto del producto y el daño es compatible con dicho defecto; en otras palabras, cuando sea razonable inferir que el defecto encontrado sí pudo ocasionar el daño. Lo anterior permite establecer que en Colombia, cuando se trate de productos y sistemas de considerable complejidad como los robots autónomos, probar los tres elementos de la responsabilidad puede ser complejo para un consumidor promedio, más aún entendiendo que no hay estándares técnicos uniformemente aceptados en el país.

Aunque se aplique la carga dinámica de la prueba consagrada en el Código General del Proceso, el consumidor o afectado con el daño queda en una posición compleja, pues en muchos casos quedará en manos del fabricante, productor o distribuidor para poder probar tanto la falla del sistema como el nexo causal y generalmente esta prueba deberá ir acompañada de un peritaje técnico especializado que permita explicar al juez el funcionamiento del sistema de IA bajo estudio junto con los elementos técnicos necesarios para que posteriormente se pueda realizar el análisis de responsabilidad desde el punto de vista jurídico.

3.5. Propuestas sobre un régimen de responsabilidad civil en sistemas de IA

Según el monitor de incidentes de IA de la OCDE [38], desde el año 2020 hasta diciembre de 2025 se encontraron aproximadamente 12720 incidentes y peligros, mostrando una tendencia importante de ascenso, pues mientras en diciembre de 2021 se detectaron 120, en diciembre de 2025 el número subió hasta 453. Por otro lado, en 2023 Deloitte realizó una encuesta con más de 2.800 líderes empresariales y tecnológicos, de 16 países y seis sectores, incluido el financiero, donde el 78% de las personas encuestadas manifestó que hacía falta mayor regulación global para gestionar la adopción responsable y generalizada de la IA generativa [39] .

Estas estadísticas y percepciones, sumadas al contexto general indicado a lo largo del texto, evidencian que los sistemas de IA y los robots autónomos requieren reglas especializadas de responsabilidad que brinden claridad a los participantes en la cadena de creación y comercialización del producto y también protejan de forma adecuada a las víctimas de los daños causados por este tipo de sistemas.

38 OECD. AIM: AI Incidents and Hazards Monitor. https://oecd.ai/en/incidents. 39 Deloitte, "New Deloitte Survey Finds Expectations for Gen AI Remain High, but Many Are Feeling Pressure to Quickly Realize Value While Managing Risks", 15 de enero de 2024, https://www.deloitte.com/-cbc/en/about/press-room/gen-ai-survey.html.

La IA plantea un reto para el derecho, ya que el concepto tradicional de responsabilidad requiere identificar un sujeto responsable y establecer reglas para la imputación de responsabilidad, frente a lo cual existen tres enfoques principales:

- Responsabilidad objetiva (riesgo creado): Aplicada en casos de productos defectuosos, como vehículos autónomos.

- Responsabilidad subjetiva (culpa o negligencia): Se requiere probar un error humano en el diseño o uso.

- Responsabilidad proporcional (varios agentes involucrados): Divide la responsabilidad entre fabricante, programador y usuario. Además del AI Act, el régimen de Estados Unidos adopta un enfoque basado en la negligencia y en la regulación sectorial (NHTSA para vehículos autónomos) [40] .

De manera general, la tendencia actual apunta hacia una mayor supervisión y posibles seguros obligatorios para cubrir los riesgos asociados, exigiendo más transparencia sobre las capacidades reales de los sistemas autónomos, a través de medidas como el registro de logs de decisiones algorítmicas, así como la supervisión humana efectiva, con límites claros al marketing sobre autonomía y bajo el cumplimiento de estándares de seguridad superiores a los humanos [41] .

De hecho, como se verá a continuación, la existencia de seguros obligatorios es una de las medidas más aceptadas para garantizar la adecuada reparación de las víctimas. En Colombia el artículo 1127 del Código de Comercio regula el seguro de responsabilidad, estableciendo a cargo del asegurador la obligación de indemnizar los perjuicios patrimoniales causados por el asegurado.

No obstante, es relevante tener en cuenta que la existencia de un seguro permite que haya un garante en el pago de la indemnización, pero no exonera al afectado de probar el defecto, el daño y el nexo causal. Por esta razón no basta con establecer un seguro obligatorio, sino que se deben identificar reglas específicas en materia de responsabilidad, pues de lo contrario las víctimas del daño seguirán sin contar con un camino más sencillo para obtener su indemnización.

Anteriormente ya se ha advertido acerca de la necesidad de un régimen de responsabilidad especial y otra de las justificaciones para ello es que en el caso de la IA etapas como el desarrollo, la implementación y el uso no están claramente diferenciadas, sino que responden a procesos de cooperación y cocreación. Además de ello puede haber problemas en la imputación de la responsabilidad, bien sea porque no se puede identificar al actor responsable (dado que es la tecnología la que toma las decisiones), hay un error al identificar al responsable, puede haber varios actores responsables o hay varios regímenes legales aplicables, ante lo cual se puede plantear un régimen de responsabilidad compartida [42] .

Además, autores como Darío Parra y Ricardo Concha [43] resaltan la existencia de la “brecha de responsabilidad”, la cual se genera cuando se cumplen dos condiciones:

40 National Highway Traffic Safety Administration, “Automated Vehicles for Safety”, https://www.nhtsa.gov/vehicle-safety/automated-vehicles-safety. 41 Núñez Zorrilla, Inteligencia artificial y responsabilidad civil. 42 Bart Custers, Henning Lahmann y Benjamyn I. Scott, “From Liability Gaps to Liability Overlaps: Shared Responsibilities and Fiduciary Duties in AI and Other Complex Technologies”, AI & Society 40 (2025): 4035–4050, https://doi.org/10.1007/s00146-024-02137-1. 43 Darío Parra Sepúlveda y Ricardo Concha Machuca, “Responsabilidad civil por actos de robots autónomos en la Unión Europea: ¿hacia la personalidad electrónica?”, Civilistica.com 11, n.º 3 (2022): 1–31.

a) El sistema autónomo lleva a cabo alguna acción sin que una persona humana sea responsable de dicha acción y sus consecuencias y b) La ausencia de responsabilidad se deriva del nivel de autonomía logrado por la IA.

La existencia de esta brecha de responsabilidad abre la puerta a una pregunta importante que debe ser solucionada por un nuevo régimen normativo sobre sistemas de IA: ¿Los daños que causen de forma totalmente autónoma los sistemas de IA exoneran de responsabilidad al fabricante y los demás actores de la cadena? Además de las complejidades técnico-jurídicas, se encuentran otros debates de carácter económico y comercial, donde regulaciones en extremo rígidas con fabricantes y productores en materia de responsabilidad pueden desincentivar el desarrollo de nuevas tecnologías.

Ante este panorama comercial se ha planteado la posibilidad de pasar de un modelo exclusivamente basado en responsabilidad civil a una gestión financiera de pérdidas, valorando la existencia de sistemas de indemnización sin culpa cuando no exista evidencia de negligencia, imprudencia o falta de pericia por parte del fabricante. Ello implica que coexista un régimen de responsabilidad con sistemas de indemnización [44] .

Más allá de la perspectiva comercial, en el marco de los proyectos de regulación europea se recalcó la necesidad de explorar legislaciones que permitan mantener un equilibrio entre el desarrollo de la tecnología y la protección de derechos subjetivos [45] .

Miquel Martín-Casals[46], con base en lo que en su momento era el proyecto de regulación de la IA en la Unión Europea, también enfatizó en que es importante aumentar la seguridad de los sistemas de IA para generar una mayor confianza sobre su uso y garantizar la indemnización de los daños que se produzcan. Además indicó que las reglas de responsabilidad civil deben complementarse con las normativas preventivas y de seguridad.

Martin Ebers [47] examina críticamente distintas propuestas, entre ellas la creación de un fondo o sistema de seguro obligatorio y que al crear un régimen de responsabilidad propia del robot, se podría permitir incluso que el mismo robot pueda ser cuestionado o interpelado, así como entregarle un patrimonio que sirva para que responda por los daños que cause.

En el marco de las relaciones de consumo también pueden ser de utilidad los siguientes criterios, examinados por Inmaculada Herbosa Martínez al analizar la legislación sobre productos defectuosos y la propuesta de Directiva COM(2022) 495 [48] :

44 Emiliano Marchisio, "In Support of 'No-Fault' Civil Liability Rules for Artificial Intelligence", SN Social Sciences 1 (2021), art. 54, https://doi.org/10.1007/s43545-020-00043-z. 45 Manuel Ortiz Fernández, "La «adaptación» del derecho de daños a la inteligencia artificial: la propuesta de Directiva sobre responsabilidad civil", IDP. Revista de Internet, Derecho y Política, n.º 40 (2024): 1–12, https://doi.org/10.7238/idp.v0i40.419696. 46 Miquel Martín-Casals, "Las propuestas de la Unión Europea para regular la responsabilidad civil por los daños causados por sistemas de inteligencia artificial", InDret, n.º 3 (2023): 55–100, https://doi.org/10.31009/InDret.2023.i3.02. 47 Martin Ebers, "La utilización de agentes electrónicos inteligentes en el tráfico jurídico: ¿necesitamos reglas especiales en el Derecho de la responsabilidad civil?", InDret, n.º 3 (2016): 1–22, https://indret.com/wp- content/uploads/2018/05/1245.pdf. 48 Herbosa Martínez, "Encaje de los sistemas de IA en la definición de producto", 65–66 y 77–86.



Foto: Magnific

- La capacidad de aprendizaje del sistema debe ser tenida en cuenta para evaluar la seguridad del producto, teniendo en cuenta que un comportamiento impredecible del sistema no exonera de forma automática al fabricante, pero, al mismo tiempo, tampoco puede ser un defecto que genere responsabilidad en todos los casos.

- El fabricante puede responder por los daños causados a partir de datos defectuosos.

- El fabricante debe responder cuando el daño se haya causado a raíz de una actualización o mejora.

Al analizar la propuesta de Directiva sobre responsabilidad en materia de IA, COM(2022) 496, también se ha resaltado que conforme se vayan compenetrando más la IA y el internet de las cosas, cada vez más bienes y servicios tendrán sistemas de IA, lo cual implica retos en el régimen de responsabilidad. Por otro lado, también se señaló que la normativa preventiva o basada en el riesgo debe ser complementada cuando se trata de una responsabilidad extracontractual, pues en estos casos el daño ya se ha producido [49].

Todos los elementos indicados anteriormente son ideas que permiten empezar a construir un régimen de responsabilidad pero que, al mismo tiempo, señalan advertencias en torno al avance tecnológico (como la compenetración con el internet de las cosas) y la necesidad de conciliar intereses, como las reglas de responsabilidad y los incentivos para seguir generando nueva tecnología o el desarrollo tecnológico y la protección de los derechos subjetivos.

Diversos autores también han planteado reglas específicas en torno a cómo debería ser un esquema de responsabilidad civil en sistemas de IA. Por ejemplo, en una investigación de Tatiana Zabala y Carla Gómez [50] se examinaron estudios teóricos elaborados entre 2019 y 2022 sobre responsabilidad civil y ética en el desarrollo de inteligencia artificial; a partir de dichos estudios se recomendaron las siguientes características para la regulación de la responsabilidad civil de la IA:

- No limitar la compensación por daños causados por aplicaciones de IA.
- Tener como base la responsabilidad objetiva.
- Establecer un régimen de seguros obligatorios y fondos de compensación.

49 Guido Noto La Diega y Leonardo C. T. Bezerra, "Can There Be Responsible AI without AI Liability? Incentivizing Generative AI Safety through Ex-Post Tort Liability under the EU AI Liability Directive", International Journal of Law and Information Technology 32 (2024), eaae021, https://doi.org/10.1093/ijlit/eaae021. 50 Tatiana Dulima Zabala Leal y Carla Angélica Gómez Macfarland, "La responsabilidad civil y la ética en la inteligencia artificial: una revisión sistemática de las ideas del período 2018–2023", IUSTA, n.º 60 (2024): 66–93, https://doi.org/10.15332/25005286.9964.

Además de lo anterior, la investigación señala que al regular la IA se deben considerar los derechos afectados, los principios bioéticos y el equilibrio entre ventajas sociales y económicas frente a los riesgos potenciales en distintas áreas. Otro estudio [51] enfocado en la responsabilidad civil de la IA en la atención sanitaria indica la necesidad de analizar caso a caso la responsabilidad, pues ello depende del grado de autonomía del robot y de si la acción fue realizada por un robot supervisado o sin supervisión.

Los autores exploran tres soluciones principales: el otorgamiento de personalidad jurídica a los robots, la incorporación de "cajas negras" para que recopile suficientes datos sobre el sistema con fines probatorios y, por último, la existencia de seguros. En todo caso, concluyen que un régimen de responsabilidad en estos casos requiere de un mayor estudio y de una construcción interdisciplinaria. Abigail Quesada [52] empieza indicando que la IA tiene como particularidad que puede generar daños sin intervención directa del ser humano, lo cual implica un examen especial desde el punto de vista de la responsabilidad que requiere un estudio particular y una adecuación funcional de las categorías existentes sobre imputación del daño.

Puntualmente la autora plantea una "arquitectura escalonada de imputación", basada en la distinción entre los defectos del sistema, el incumplimiento de deberes normativos y la materialización de riesgos estructurales inherentes a su funcionamiento autónomo. De ello se desprenden las siguientes reglas:

- Si el daño se ocasionó por un defecto de seguridad, opera el régimen de responsabilidad objetiva, de acuerdo con la normativa de la Unión Europea.
- Si el daño nace por el despliegue, configuración o uso del sistema no se opta por un régimen de culpa objetiva o subjetiva, sino por una culpa normativa, que se basa en el incumplimiento de deberes legales y técnicos.
- La responsabilidad se endilga con base en la "imputación funcional del riesgo. algorítmico", en virtud de la cual se identifica a las personas que ocupan posiciones más relevantes en la creación, organización o explotación del riesgo.

En el estudio de Juana Flórez y Cristian Díaz [53] se indica que el régimen colombiano tiene una base jurídica consistente para ser aplicada en daños ocasionados por robots, pero se presentan las siguientes dificultades: identificar la causalidad de un actuar humano sobre la respuesta del robot; la identificación de la persona sobre la cual recaen las consecuencias de la imputación jurídica (es decir, si es un guardián, un productor, etc.); el daño transfronterizo cuando una vez las personas estén identificadas no se encuentren en el país; el denominado "riesgo de desarrollo", debido a la imprevisibilidad de las respuestas de un robot por su aprendizaje autónomo y tal vez una de las problemáticas más interesantes: la toma de decisiones éticas o morales por parte del robot y que a raíz de estas decisiones éticas o moralmente aceptables se genere un daño.

51 Ahmed Eldakak et al., "Civil Liability for the Actions of Autonomous AI in Healthcare: An Invitation to Further Contemplation", Humanities and Social Sciences Communications 11 (2024), art. 305, https://doi.org/10.1057/s41599-024-02806-y. 52 Abigail Quesada Páez, "La atribución de responsabilidad civil por actos de sistemas autónomos de inteligencia artificial: desafíos para el derecho privado", InDret, n.º 1 (2026): 193–226, https://doi.org/10.31009/InDret.2026.i1.05. 53 Juana Flórez Peláez y Cristian Andrés Díaz Díez, Imputación de daños causados por robots con inteligencia artificial: conceptos aplicables de la responsabilidad civil y del Estado (Centro de Estudios Regulatorios, 23 de junio de 2022), https://www.cerlatam.com/wp-content/uploads/2022/06/CER.EAFIT_.-Imputacion-de-danos-causados-por-robots-con-inteligencia-artificial.pdf.

Los autores, con base en normas existentes en la Unión Europea, proponen explorar las siguientes opciones de regulación:

- Establecer un régimen de seguros que sea obligatorio para ciertas categorías de robots, como los automóviles autónomos.

- Establecer un fondo de compensación para reparar a las víctimas cuando no haya un seguro y permitir que fabricantes, programadores, propietarios y usuarios puedan beneficiarse de un régimen de responsabilidad limitada si contribuyen a ese fondo.

- Sin perjuicio de lo anterior, dentro de las propuestas también se contempla endurecer el régimen de responsabilidad de propietarios y productores hacia la responsabilidad objetiva.

En Colombia el régimen de responsabilidad civil debe adecuarse a los desafíos tecnológicos de la IA y, así mismo, el derecho de consumo no resuelve interrogantes como la noción de producto, la identificación de responsables y el alcance del daño. También es necesario que se identifiquen deberes de conducta específicos de los desarrolladores de IA y elaborar un marco normativo que garantice la seguridad jurídica, la protección de las víctimas y un desarrollo tecnológico responsable [54].

Como se indicó anteriormente, el hecho de que no exista en Colombia un régimen específico de responsabilidad civil e IA no significa que los daños causados por estos sistemas se encuentren exentos de reparación, pero sí existen dificultades, sobre todo porque no se sabe bien cómo operan estos sistemas. Así mismo, también se plantea que el régimen de responsabilidad no debe ser objetivo en todos los casos, pues no todos los sistemas de IA presentan el mismo nivel de riesgo [55] .

En Chile también se ha identificado que su legislación actual no regula la responsabilidad civil de sistemas de IA. En el marco de esa legislación se abren problemáticas derivadas acerca de cuándo puede ser calificado un sistema inteligente como defectuoso, pues si se trata de daños complejos (provocados por procesos de autoaprendizaje) puede que el daño se genere porque el sistema operó de forma más eficiente [56].

Lo anterior confirma que, en efecto existe una necesidad de generar una legislación específica para crear reglas de responsabilidad de sistemas de IA; ello no significa eliminar el régimen de responsabilidad actual, sino adecuarlo a las particularidades y desafíos de las nuevas tecnologías.

54 Olenka Woolcott Oyague, Karen Isabel Cabrera Peña y Daniela Navarro Reyes, "Hacia una adaptación de la responsabilidad civil a la inteligencia artificial y su proyección en Colombia", Novum Jus 20, n.º 1 (2026): 259–287, https://doi.org/10.14718/NovumJus.2026.20.1.9. 55 Luisa Fernanda Jiménez Mahecha, "La responsabilidad contractual y extracontractual en el laberinto de la inteligencia artificial: perspectiva colombiana", Revista e-Mercatoria 24, n.º 2 (2025): 193–220, https://doi.org/10.18601/16923960.v24n2.07. 56 Fernanda Torres Mac-Pherson, Javier Salgado Alonso y Mirko Villarroel Contreras, "Responsabilidad civil por daños provenientes de ilícitos derivados del uso de la inteligencia artificial: régimen aplicable y desafíos de seguridad para su desarrollo", Revista Chilena de Derecho y Tecnología 14 (2025), e74613, https://doi.org/10.5354/0719-2584.2025.74613.

04

REFLEXIONES SOBRE LA POSIBILIDAD DE OTORGAR PERSONALIDAD JURÍDICA A LOS ROBOTS

La Corte Constitucional, en la sentencia T-241 de 2018, a partir de su jurisprudencia, señaló que la personalidad jurídica es un derecho fundamental autónomo que se materializa mediante los atributos de la personalidad, a saber: (i) el nombre; (ii) la capacidad; (iii) el estado civil; (iv) el domicilio; (v) la nacionalidad; y (vi) el patrimonio [57].

Al dotar de personalidad jurídica a los robots bajo esta misma concepción, se les estaría dotando de capacidad para que respondan por sí mismos frente a terceros por los daños que causen, incluso en escenarios judiciales, circunstancia especialmente útil cuando se trate de sistemas totalmente autónomos y, así mismo, les permitiría contar con un patrimonio que garantice el pago a las víctimas.

Sin embargo, esto también significaría que tendrían un nombre, un domicilio, una nacionalidad e incluso un estado civil, todos ellos atributos que normalmente se relacionan con los seres humanos, máxime porque se derivan de convenciones y relaciones sociales propias de la especie humana.

Por esta razón, prima facie se tiende a rechazar la idea de dotar de personalidad jurídica a los robots; sin embargo, desde diversos estudios se han dado elementos que justifican esta iniciativa o que la matizan para entregar una personalidad limitada o especial a los robots.

57 Corte Constitucional, Sentencia T-241 de 2018, 26 de junio de 2018, M. P. Gloria Stella Ortiz Delgado, apartado sobre el derecho a la personalidad jurídica 58 Lawrence B. Solum, "Legal Personhood for Artificial Intelligences", North Carolina Law Review 70, n.º 4 (1992): 1231–1287, https://scholarship.law.unc.edu/nclr/vol70/iss4/4/.



Foto: Pixabay

Lawrence Solum[58], en 1992, abordó la personalidad artificial e indagó en qué circunstancias un sistema podía tener derechos constitucionales; a partir de ello se puede concluir que cada facultad que se otorgue a la IA requiere una justificación propia según el interés protegido.

Más adelante, Pagallo [59] distingue al robot como instrumento de una interacción humana del robot concebido como agente en la práctica jurídica, aspecto relevante en materias como la ejecución automatizada de determinadas acciones. A partir de esta distinción ya se establecía la necesidad de analizar la problemática de acuerdo a las características del robot.

La Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, 2015/2103(INL) establece como una idea exploratoria y "a largo plazo" la posibilidad de generar una personalidad jurídica específica para los robots. Sin embargo, el Dictamen del Comité Económico y Social Europeo [60] indica que la creación de una personalidad electrónica puede generar riesgos morales, así como liberar de forma injustificada a fabricantes o usuarios de su propia responsabilidad.

Expertos como Marta Pérez [61] indican que han sido los avances de la IA los que han abierto la puerta a discutir la posibilidad de otorgarles a estos sistemas tecnológicos una personalidad jurídica. Empero, manifiesta que esta posibilidad realmente refleja una distorsión sociológica de lo que es la IA, esto es, una tecnología al servicio del ser humano.

La autora también señala que el otorgamiento de una personalidad jurídica a entes no humanos es algo excepcional [62] y que esta excepcionalidad implica que en el presente caso no sea necesario, pues la responsabilidad civil se puede regular a través de un régimen jurídico adecuado. A partir de ello la autora indica que un régimen basado en la responsabilidad civil y no en la personalidad jurídica es lo que permite "reivindicar la centralidad de la persona como razón de ser del Derecho (humanismo)" [63] . En el mismo sentido, también se ha indicado que la solución a los dilemas de la responsabilidad civil de la IA no es dotar al robot de derechos, pues ello va en contravía del régimen internacional, ni tampoco atribuirles unos "semiderechos", pues detrás de cada robot y sistema de IA siempre estará la mente del ser humano [64] . Por el contrario, en la investigación de Tatiana Zabala y Carla Gómez [65] citada anteriormente se recomienda que al regular la responsabilidad civil se analice la posibilidad de otorgar personalidad jurídica a aplicaciones de IA.

Otro estudio de Andrea Bertolini y Francesca Episcopo [66] se pregunta si los robots y la IA pueden ser considerados como sujetos jurídicos, en virtud de que son muy avanzadas tecnológicamente. Los autores proponen separar el ámbito ontológico (donde no se observa la necesidad de otorgar la calificación de sujeto jurídico debido a que el régimen legal actual ya permite imputar a una persona los efectos jurídicos generados por la conducta de otro) del ámbito funcional (donde pueden encontrarse casos de aplicaciones de IA en particular -no de toda la IA- donde sea mejor atribuirle al sistema derechos y obligaciones, en vez de hacerlo solo a las personas que están detrás de dicho sistema).

59 Ugo Pagallo, The Laws of Robots: Crimes, Contracts, and Torts (Dordrecht: Springer, 2013), https://doi.org/10.1007/978-94-007-6564-1. 60 Comité Económico y Social Europeo, "Dictamen sobre la inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad", 31 de mayo de 2017, apartado 3.33, CELEX 52016IE5369, https://eur-lex.europa.eu/legal- content/ES/-TXT/?uri=CELEX:52016IE5369. 61 Marta Pérez Escolar, "Personalidad jurídica e inteligencia artificial: fundamentos de asimilaciones imposibles", InDret, n.º 3 (2025): 39–66, https://doi.org/10.31009/InDret.2025.i3.02. 62 Por ejemplo, en Colombia se declaró al Río Atrato como sujeto de derechos en la sentencia T-622 de 2016, dictada por la Corte Constitucional (M.P. Jorge Iván Palacio Palacio). 63 Pérez Escolar, "Personalidad jurídica e inteligencia artificial", 63. 64 Francisca Ramón Fernández, "Robótica, inteligencia artificial y seguridad: ¿cómo encajar la responsabilidad civil?", Diario La Ley, n.º 9365, 25 de febrero de 2019, 1–13. 65 Zabala Leal y Gómez Macfarland, "La responsabilidad civil y la ética en la inteligencia artificial" 66 Andrea Bertolini y Francesca Episcopo, "Robots and AI as Legal Subjects? Disentangling the Ontological and Functional Perspective", Frontiers in Robotics and AI 9 (2022), art. 842213, https://doi.org/10.3389/frobt.2022.842213.

A partir de ello los autores concluyen que la atribución directa de responsabilidad a un sistema de IA debe analizarse caso por caso y aclaran que esta atribución no implica que la máquina sea merecedora de derechos individuales.

En otro estudio [67] se establece que dotar de personalidad jurídica a los robots en ámbitos sanitarios trae ventajas relacionadas con una mayor claridad en la imputación de la responsabilidad, pero también tiene desventajas, como el hecho de que al robot se le debe dotar de un patrimonio propio, la posibilidad de que médicos y hospitales sigan respondiendo por las acciones de robot en virtud de la responsabilidad del hecho ajeno, la posibilidad de responsabilidad por parte del médico cuando haya negligencia y porque se podría levantar el velo de responsabilidad jurídica del robot, de forma similar a como se levanta el velo corporativo.

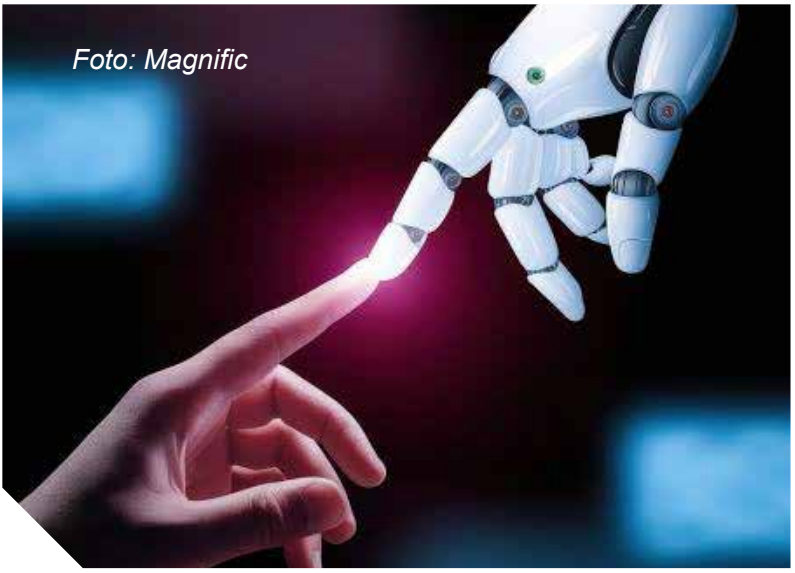
Los autores señalan dos puntos cruciales: el primero es que el derecho de daños puede diluirse si solo el robot responde, porque lo que busca el régimen jurídico es que las personas eviten y reduzcan las conductas riesgosas; el segundo es que aun cuando se otorgue personalidad jurídica, ello no siempre tendrá un efecto claro en los litigios, pues habrá que probar el nexo de causalidad e incluso puede que se dificulte más la disputa con un actor procesal más dentro del mismo.

En concordancia con lo anterior, Bryson, Diamantis y Grant [68] advierten que conceder personalidad jurídica a robots o sistemas de IA sin procedimientos efectivos de representación permitiría que tanto empresas como personas se protejan detrás de esa personalidad jurídica de la IA.

Por último, un estudio realizado en Polonia hizo un análisis a partir de doctrina y entrevistas estructuradas con auditores, para identificar argumentos a favor y en contra de dotar de personalidad jurídica a los sistemas de IA. A partir de los datos recopilados, los autores proponen reconocer la personalidad jurídica de una forma limitada y sui generis [69]

Dentro de dicho estudio se identificaron como argumentos a favor de la personalidad jurídica la autonomía decisoria de los sistemas de IA, la imposibilidad de encasillar a la IA en categorías jurídicas existentes y la evolución tecnológica, mientras que algunos argumentos en contra o como límites al reconocimiento de la personalidad jurídica son la ausencia de intenciones, deseos o intereses propios de la IA, la dificultad para hacer cumplir las obligaciones a una IA, el riesgo de exonerar a seres humanos responsables del daño y el hecho de que si bien las empresas están representadas por seres humanos, esta analogía no funciona de la misma forma en sistemas de IA Esta propuesta permitiría que ciertos sistemas de IA puedan ser tratados como cuasipersonas con derechos y obligaciones previamente definidos, una representación humana, el deber de sistemas de IA de actuar de forma comprensible para los humanos y responsabilidad compartida con personas humanas.

67 Eldakak et al., "Civil Liability for the Actions of Autonomous AI in Healthcare", 5. 68 Joanna J. Bryson, Mihailis E. Diamantis y Thomas D. Grant, "Of, for, and by the People: The Legal Lacuna of Synthetic Persons", Artificial Intelligence and Law 25 (2017): 273–291, https://doi.org/10.1007/s10506-017-9214-9. 69 Piotr Staszkiewicz et al., "Artificial Intelligence Legal Personality and Accountability: Auditors' Accounts of Capabilities and Challenges for Instrument Boundary", Meditari Accountancy Research 32, n.º 7 (2024): 120–146, https://doi.org/10.1108/MEDAR-10-2023-2204.


Foto: Magnific

05 CONCLUSIONES

Desde el punto de vista jurídico, es difuso aún hoy en día, determinar con precisión la responsabilidad civil ante un daño causado por un robot autónomo. La dificultad para determinar la responsabilidad del programador, el fabricante o el usuario, aunada a la falta de personalidad jurídica de la IA y el riesgo de regulación excesiva que puede frenar la innovación, comporta un verdadero desafío para el legislador.

Existe una clara dificultad para determinar la relación de causalidad entre el agente autónomo y el daño, aunado al problema del "agente autónomo" sin personalidad jurídica y la falta de regulación específica en muchos países. Todo ello presenta para el legislador un gran desafío que en la actualidad se encuentra en proceso de discusión y creación.

Se plantea entonces, la necesidad de tener una regulación autónoma, especializada, que regule de manera íntegra la relación entre los robots y los seres humanos desde el punto de vista de la responsabilidad.

Aunque el AI Act es un avance muy importante y sirve como parámetro regulatorio, no es en sí mismo un sistema completo de responsabilidad civil y tampoco puede trasladarse de forma mecánica a otras legislaciones como la colombiana.

Desde el punto de vista académico, parece haber cierto consenso frente a ideas como la existencia de seguros obligatorios, de fondos de reparación, mayores claridades sobre la responsabilidad de los fabricantes y deberes más estrictos de seguridad.

Otros elementos, como la responsabilidad objetiva en ciertos casos y la creación de reglas diferenciales según la función del robot, sus capacidades y la ejecución de actividades peligrosas, sirven de complemento para empezar a construir este régimen especial.

Sin embargo, ideas como la personalidad jurídica de los robots todavía no cuentan con consenso amplio, dado que trae muchas problemáticas que deben ser abordadas adecuadamente: qué implica esa personalidad jurídica, qué derechos tendrá el robot, cuáles son los límites, entre otros.

Al respecto, el hecho de que haya robots de compañía con los cuales los seres humanos puedan entablar una relación afectiva ya muestra que la definición de personalidad jurídica trae múltiples retos, no solo jurídicos, sino también filosóficos, morales, antropológicos y éticos, para lo cual se debe empezar por abordar como se trata al robot en relaciones sociales, -como lo hizo David Gunkel [70] -, evaluar las capacidades de los sistemas actuales y examinar como la relación con robots puede afectar las relaciones humanas, modificar comportamientos, entre otras implicaciones.

70 David J. Gunkel, Robot Rights (Cambridge, MA: MIT Press, 2018).

En suma, el debate en torno a un régimen de responsabilidad civil de sistemas de IA y robots autónomos parece haber alcanzado algunos consensos, pero todavía quedan aspectos pendientes por resolver para crear una propuesta normativa completa, que logre el equilibrio entre la innovación y los derechos de las personas y que al mismo tiempo tenga la vocación de permanencia sin importar los cambios que puedan surgir por los avances tecnológicos que se presentan día a día.

En todo caso, el tiempo para la construcción de este régimen de responsabilidad civil se agota. Hoy en día, en el mundo ya están circulando vehículos autónomos y las personas pueden comprar robots autónomos con distintos fines, lo cual abre paso a la posibilidad de generar daños, escenario en el cual se requiere de legislaciones y jueces preparados para poder analizar y abordar estos casos.

REFERENCIAS

Abbott, Ryan. The Reasonable Robot: Artificial Intelligence and the Law. Cambridge: Cambridge University Press, 2020.

Bertolini, Andrea. Artificial Intelligence and Civil Liability: A European Perspective. PE 776.426. Bruselas: Parlamento Europeo, 2025. https://doi.org/10.2861/0075079.

Bertolini, Andrea, y Francesca Episcopo. "Robots and AI as Legal Subjects? Disentangling the Ontological and Functional Perspective". Frontiers in Robotics and AI 9 (2022), art. 842213. https://doi.org/10.3389/frobt.2022.842213.

Bryson, Joanna J., Mihailis E. Diamantis y Thomas D. Grant. "Of, for, and by the People: The Legal Lacuna of Synthetic Persons". Artificial Intelligence and Law 25 (2017): 273–291. https://doi.org/10.1007/s10506-017-9214-9.

Calo, Ryan. "Robotics and the Lessons of Cyberlaw". California Law Review 103, n.º 3 (2015): 513–563.

Castilla, Adolfo, y María Cruz Alonso. La sociedad española ante las nuevas tecnologías: actitudes y grados de receptividad. Madrid: Fundesco, 1987.

Custers, Bart, Henning Lahmann y Benjamyn I. Scott. "From Liability Gaps to Liability Overlaps: Shared Responsibilities and Fiduciary Duties in AI and Other Complex Technologies". AI & Society 40 (2025): 4035–4050. https://doi.org/10.1007/s00146-024-02137-1.

Deloitte. "New Deloitte Survey Finds Expectations for Gen AI Remain High, but Many Are Feeling Pressure to Quickly Realize Value While Managing Risks". 15 de enero de 2024. https://www.deloitte.com/cbc/en/about/press-room/gen-ai-survey.html.

Ebers, Martin. "La utilización de agentes electrónicos inteligentes en el tráfico jurídico: ¿necesitamos reglas especiales en el Derecho de la responsabilidad civil?". InDret, n.º 3 (2016): 1–22. https://indret.com/wp-content/uploads/2018/05/1245.pdf.

Eldakak, Ahmed, Abdulla Alremeithi, Emad Dahiyat, Moatasem El-Gheriani, Hassan Mohamed y Mohammad Ibrahim Abdulrahim Abdulla. "Civil Liability for the Actions of Autonomous AI in Healthcare: An Invitation to Further Contemplation". Humanities and Social Sciences Communications 11 (2024), art. 305. https://doi.org/10.1057/s41599-024-02806-y.

Ellingrud, Kweilin, Saurabh Sanghvi, Gurneet Singh Dandona, Anu Madgavkar, Michael Chui, Olivia White y Paige Hasebe. "Generative AI and the Future of Work in America". McKinsey Global Institute, 26 de julio de 2023. https://www.mckinsey.com/mgi/our-research/generative-ai-and-the-future-of-work-in-america.

Expert Group on Liability and New Technologies. Liability for Artificial Intelligence and Other Emerging Digital Technologies. Luxemburgo: Publications Office of the European Union, 2019. https://doi.org/10.2838/573689.

Solum, Lawrence B. "Legal Personhood for Artificial Intelligences". North Carolina Law Review 70, n.º 4 (1992): 1231–1287. https://scholarship.law.unc.edu/nclr/vol70/iss4/4/.

Flórez Peláez, Juana, y Cristian Andrés Díaz Díez. Imputación de daños causados por robots con inteligencia artificial: conceptos aplicables de la responsabilidad civil y del Estado. Centro de Estudios Regulatorios, 23 de junio de 2022. https://www.cerlatam.com/wp-content/uploads/2022/06/CER.EAFIT_.-Imputacion-de-danos-causados-por-robots-con-inteligencia-artificial.pdf.

Fortune Business Insights. "AI-Based Surgical Robots Market Size, Share & amp; Industry Analysis". https://www.fortunebusinessinsights.com/ai-based-surgical-robots-market-106330.

Gunkel, David J. Robot Rights. Cambridge, MA: MIT Press, 2018.

Herbosa Martínez, Inmaculada. "Encaje de los sistemas de IA en la definición de producto en la legislación de productos defectuosos: análisis de la legislación vigente con la vista puesta en la Propuesta de Directiva del Parlamento Europeo y del Consejo de 28 de septiembre de 2022 (COM/2022/495)". InDret, n.º 3 (2024): 52–98. https://doi.org/10.31009/InDret.2024.i3.02.

Intuitive Surgical, Inc. Annual Report 2024. Form 10-K. 2025.

Jiménez Mahecha, Luisa Fernanda. "La responsabilidad contractual y extracontractual en el laberinto de la inteligencia artificial: perspectiva colombiana". Revista e-Mercatoria 24, n.º 2 (2025): 193–220. https://doi.org/10.18601/16923960.v24n2.07.

Marchisio, Emiliano. "In Support of 'No-Fault' Civil Liability Rules for Artificial Intelligence". SN Social Sciences 1 (2021), art. 54. https://doi.org/10.1007/s43545-020-00043-z.

Martín-Casals, Miquel. "Las propuestas de la Unión Europea para regular la responsabilidad civil por los daños causados por sistemas de inteligencia artificial". InDret, n.º 3 (2023): 55–100. https://doi.org/10.31009/InDret.2023.i3.02.

Moriello, Sergio Alejandro. "Los robots inteligentes autónomos son la nueva generación". The Economy Journal. https://www.theeconomyjournal.com/texto-diario/mostrar/595081/robots-inteligentes-autonomos-nueva-generacion.

Rebollo Delgado, Lucrecio. Derechos fundamentales y protección de datos. Madrid: Dykinson, 2004.

Mosquet, Xavier, Thomas Dauner, Nikolaus Lang, Michael Rüßmann, Antonella Mei-Pochtler, Rakshita Agrawal y Florian Schmieg. Revolution in the Driver's Seat: The Road to Autonomous Vehicles. Boston Consulting Group, 2015.

National Transportation Safety Board. Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, Arizona, March 18, 2018. NTSB/HAR-19/03. Washington, D. C.: National Transportation Safety Board, 2019. https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1903.pdf.

New York State Department of Financial Services. Report on Apple Card Investigation. Marzo de 2021.

Noto La Diega, Guido, y Leonardo C. T. Bezerra. "Can There Be Responsible AI without AI Liability? Incentivizing Generative AI Safety through Ex-Post Tort Liability under the EU AI Liability Directive". International Journal of Law and Information Technology 32 (2024), eaae021. https://doi.org/10.1093/ijlit/eaae021.

Núñez Zorrilla, María del Carmen. Inteligencia artificial y responsabilidad civil: régimen jurídico de los daños causados por robots autónomos con inteligencia artificial. Madrid: Reus, 2019.

Núñez Zorrilla, María del Carmen. "Los nuevos retos de la Unión Europea en la regulación de la responsabilidad civil por los daños causados por la inteligencia artificial". Revista Española de Derecho Europeo, n.º 66 (2018): 9–53.

Obermeyer, Ziad, Brian Powers, Christine Vogeli y Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations". Science 366, n.º 6464 (2019): 447–453. https://doi.org/10.1126/science.aax2342.

Ortiz Fernández, Manuel. "La «adaptación» del derecho de daños a la inteligencia artificial: la propuesta de Directiva sobre responsabilidad civil". IDP. Revista de Internet, Derecho y Política, n.º 40 (2024): 1–12. https://doi.org/10.7238/idp.v0i40.419696.

Pagallo, Ugo. The Laws of Robots: Crimes, Contracts, and Torts. Dordrecht: Springer, 2013. https://doi.org/10.1007/978-94-007-6564-1.

Palmerini, Erica. "Robótica y derecho: sugerencias, confluencias, evoluciones en el marco de una investigación europea". Revista de Derecho Privado, n.º 32 (2017): 53–97. https://doi.org/10.18601/01234366.n32.03.

Parra Sepúlveda, Darío, y Ricardo Concha Machuca. "Responsabilidad civil por actos de robots autónomos en la Unión Europea: ¿hacia la personalidad electrónica?". Civilistica.com 11, n.º 3 (2022): 1–31.

Pérez Escolar, Marta. "Personalidad jurídica e inteligencia artificial: fundamentos de asimilaciones imposibles". InDret, n.º 3 (2025): 39–66. https://doi.org/10.31009/InDret.2025.i3.02.

PwC. "The Future of Banking: How AI Is Reshaping the Industry". 16 de octubre de 2025. https://www.pwc.com/us/en/industries/financial-services/library/how-ai-is-reshaping-banking.html.

Quesada Páez, Abigail. "La atribución de responsabilidad civil por actos de sistemas autónomos de inteligencia artificial: desafíos para el derecho privado". InDret, n.º 1 (2026): 193–226. https://doi.org/10.31009/InDret.2026.i1.05.

Ramón Fernández, Francisca. "Robótica, inteligencia artificial y seguridad: ¿cómo encajar la responsabilidad civil?". Diario La Ley, n.º 9365, 25 de febrero de 2019, 1–13.

Staszkiewicz, Piotr, Jarosław Horobiowski, Anna Szelągowska y Agnieszka Maryla Strzelecka. "Artificial Intelligence Legal Personality and Accountability: Auditors' Accounts of Capabilities and Challenges for Instrument Boundary". Meditari Accountancy Research 32, n.º 7 (2024): 120–146. https://doi.org/10.1108/MEDAR-10-2023-2204.

Tamayo Jaramillo, Javier. Tratado de responsabilidad civil. T. I. 2.ª ed., 8.ª reimp. Bogotá: Legis, 2015.

Torres Mac-Pherson, Fernanda, Javier Salgado Alonso y Mirko Villarroel Contreras. "Responsabilidad civil por daños provenientes de ilícitos derivados del uso de la inteligencia artificial: régimen aplicable y desafíos de seguridad para su desarrollo". Revista Chilena de Derecho y Tecnología 14 (2025), e74613. https://doi.org/10.5354/0719-2584.2025.74613.

Woolcott Oyague, Olenka, Karen Isabel Cabrera Peña y Daniela Navarro Reyes. "Hacia una adaptación de la responsabilidad civil a la inteligencia artificial y su proyección en Colombia". Novum Jus 20, n.º 1 (2026): 259–287. https://doi.org/10.14718/NovumJus.2026.20.1.9.

Zabala Leal, Tatiana Dulima, y Carla Angélica Gómez Macfarland. "La responsabilidad civil y la ética en la inteligencia artificial: una revisión sistemática de las ideas del período 2018–2023". IUSTA, n.º 60 (2024): 66–93. https://doi.org/10.15332/25005286.9964.

ACERCA DE LOS AUTORES

Héctor José **García Santiago**

Abogado de la Pontificia Universidad Javeriana con especialización en Hacienda Pública, con estudios de maestría en Derecho Administrativo de la Universidad del Rosario y magíster en Derecho Informático y Nuevas Tecnologías de la Información y las Comunicaciones de la Universidad Externado de Colombia; con estudios en Gobierno Abierto y Gobierno Electrónico, de la Universidad Nacional de La Plata, Universidad Tres de Febrero y Universidad Complutense de Madrid, experto en procesos de transformación digital del Massachusetts Institute of Technology (MIT). Director Académico del Observatorio Javeriano de Gobierno y TIC y miembro activo del equipo de expertos en gobierno electrónico, contratación y seguridad digital del Banco Interamericano de Desarrollo (BID) y del Banco Mundial (BM).



Foto: Magnific

04

La Inteligencia artificial, clave para la eficiencia del proceso laboral en el Perú

E-ESTADO
Primera Edición

La inteligencia artificial, clave para la eficiencia del proceso laboral en el Perú

Artificial Intelligence, Key to Work Process Efficiency in Perú

Fecha de recepción: 29 de septiembre de 2025
Fecha de aceptación: 18 de febrero de 2026

Homero Pracedes Jondec Briones
Universidad Nacional de Trujillo y
Universidad César Vallejo, Perú

hjondecb@hotmail.com
ORCID: https://orcid.org/0000-0001-9691-2722

Cómo citar: Jondec Briones, Homero Pracedes. “La inteligencia artificial, clave para la eficiencia del proceso laboral en el Perú”. Revista e-Estado 1, n.° 1 (2026).



Foto: Magnific

RESUMEN

Este artículo reflexiona sobre la IA como un factor clave en la eficiencia del proceso laboral en el Perú, a partir del plazo razonable como garantía de la tutela jurisdiccional efectiva y el debido proceso. Se propone analizar la implementación de la IA en la administración de justicia en el Perú, la naturaleza social del proceso laboral y cómo mejorar su eficiencia mediante la tecnología, con la finalidad de contribuir en la eficacia y eficiencia del servicio de justicia para los ciudadanos. Desde una perspectiva teórica, práctica y crítica, se describe la implementación de la IA en la administración de justicia en el Perú, la naturaleza social del proceso laboral y la necesidad de su eficiencia, así como el imperativo de aplicar IA en la solución de los casos sencillos durante el proceso laboral, para alcanzar mayor eficiencia. Nuestras conclusiones están referidas a la implementación rudimentaria de la IA en la administración de justicia en el Perú, las características esenciales que determinan la naturaleza social del proceso laboral y la necesidad de su eficiencia para garantizar la tutela de los derechos sociales, alimentarios e indisponibles de los trabajadores; además, nuestra propuesta de resolución de los casos fáciles o sencillos del proceso laboral mediante la aplicación de la IA, busca motivar en los académicos y en los operadores de la justicia laboral, la participación y la creatividad en el uso de la IA para mejorar la eficiencia del servicio de justicia a los ciudadanos.

Palabras clave: administración de justicia, derecho a la justicia, inteligencia artificial, proceso

ABSTRACT

This article reflects on AI as a key factor for the efficiency of the labor process in Peru, based on the reasonable timeframe as a guarantee of effective jurisdictional protection and due process. It aims to analyze the implementation of AI in the Administration of Justice in Peru, the social nature of the labor process, and how to improve the efficiency of the labor process through the application of AI, with the aim of contributing to the effectiveness and efficiency of the justice service to citizens. From a theoretical, practical, and critical perspective, the article describes the implementation of AI in the Administration of Justice in Peru, the social nature of the labor process, and the need for its efficiency, as well as the imperative of applying AI in the solution of easy or simple cases in the labor process to achieve its efficiency. Our conclusions refer to the rudimentary implementation of AI in the Administration of Justice in Peru, the essential characteristics that determine the social nature of the labor process and the need for its efficiency to guarantee the protection of social, food and non-disposable rights of workers and, our proposal for resolving easy or simple cases of the labor process through the application of AI, seeks to motivate academics and operators of Labor Justice, their participation and creativity in the use of AI to improve the efficiency of justice services to citizens.

Keywords: Administration of Justice, Right to Justice, Artificial Intelligence, Process

INTRODUCCIÓN

La gobernabilidad, de conformidad con el Diccionario de política de Bobbio y Matteucci[1] es la relación de gobernantes y gobernados, y la gobernabilidad democrática se vincula a conceptos de legitimidad, estabilidad y eficacia.[2] En América Latina, los gobernantes tienen problemas de legitimidad respecto a la ciudadanía; esto se refleja en las encuestas de aceptación, así como también en los problemas de inestabilidad política permanente, lo cual afecta la eficacia de las políticas públicas que elaboran los gobiernos para solucionar los problemas sociales específicos y satisfacer las necesidades ciudadanas.[3]

Esta problemática se ve reflejada en el derecho fundamental de los ciudadanos a la tutela jurisdiccional efectiva que debe otorgar el Estado a través de un debido proceso, lo cual, en el caso peruano, está reconocido en los artículos 138 y 139, inciso 3, de la Constitución de 1993. En específico, la deslegitimación de la administración de justicia en el caso peruano, seguramente, al igual que en la mayoría de los países latinoamericanos, está referida al plazo razonable que deben durar los procesos judiciales, en especial, el proceso laboral.

El plazo razonable forma parte del derecho fundamental al debido proceso que, en el caso peruano, no está expresamente reconocido en los artículos 138 y 139 de la Constitución de 1993; pero, indudablemente, está implícito en el debido proceso que se encuentra recogido en el inciso 3 del artículo 139. Ha sido definido por el Tribunal Constitucional Peruano (TC), refiriendo que “tiene como finalidad impedir que los acusados permanezcan durante largo tiempo bajo acusación y asegurar que su tramitación se realice prontamente. En consecuencia, el derecho a que el proceso tenga un límite temporal entre su inicio y fin, forma parte del núcleo mínimo de derechos reconocido por el sistema internacional de protección de los derechos humanos, y, por tanto, no puede ser desconocido”. (Exp. N° 618-2005-HC/TC, fundamento 10) .En ese contexto, podemos anotar, por ejemplo, que la Convención Americana sobre los Derechos Humanos lo reconoce en el artículo 8.1., sosteniendo que “toda persona tiene derecho a ser oída, con las debidas garantías y dentro de un plazo razonable”; el Pacto Internacional de Derechos Civiles y Políticos lo hace en el inciso 3 del artículo 9, disponiendo que “toda persona detenida o presa […] tendrá derecho a ser juzgada dentro de un plazo razonable”; y, el Convenio para la Protección de los Derechos Humanos y de las Libertades Fundamentales de 1950, en su artículo 6, inciso 1, establece que “toda persona tiene derecho a que su causa sea oída equitativa, públicamente y dentro de un plazo razonable”.

Para mantenernos dentro de los estándares de los derechos humanos y fundamentales, en especial, respecto al plazo razonable y la eficiencia de la administración de justicia; en el Perú, el 2 de noviembre del año 2016, se suscribió el Acuerdo Nacional por la Justicia con participación del Poder Judicial, Ministerio Público, Consejo Nacional de la Magistratura (hoy Junta Nacional de Justicia), Academia de la Magistratura y Ministerio de Justicia y Derechos Humanos. El objetivo de este acuerdo era “concretar los esfuerzos de coordinación y colaboración entre las instituciones del sistema de justicia con el fin último de brindar un mejor servicio de justicia (eficiente y eficaz)

1 Norberto Bobbio y Nicola Matteucci, Diccionario de Política (México: Siglo XXI, 1998). **2** Virginia Rodríguez, La gobernabilidad: el debate de un concepto cambiante, Más Poder Local, n.o 7 (2011): 48-49. **3** Andrea Ancira, Gobernabilidad democrática en América Latina (México D. F., México: Centro de Investigación y Docencias Económicas [CIDE], 2003). **4** Providencia disponible en: https://tc.gob.pe/jurisprudencia/2005/00618-2005-HC.html. **5** Poder Judicial del Perú. https://www.minjus.gob.pe/wp-content/uploads/2017/08/04_ACUERDO-NACIONAL-POR-LA-JUSTICIA-ayuda-memoria-agosto-2017.pdf



Foto: Magnific

En ese marco, se adoptaron medidas transversales, dirigidas a dinamizar la acción de la justicia y proveer a las personas usuarias del sistema una justicia de calidad. Era imperativo recuperar la confianza en la ciudadanía y, para ello, se tomaron medidas a corto plazo que impactaran en el servicio. Adicionalmente, se plantearon acciones que buscaban reforzar las acciones coordinadas entre las instituciones. [6]

Dentro de estas últimas, se planteó la acción de reforma institucional, la cual tenía los siguientes objetivos: i) revisar del marco normativo y el funcionamiento de las instituciones del sistema de justicia; e ii) interoperar, es decir, actuar coordinadamente y no solo implementar soluciones tecnológicas, esto incluye también impulsar iniciativas previas, como el Expediente Judicial Electrónico, la Carpeta Fiscal, la Ficha Única del Magistrado y el Plan de Interoperabilidad en materia penal.[7] Como puede inferirse, para lograr la interoperabilidad como objetivo de la reforma institucional propuesta por el Acuerdo Nacional por la Justicia en el Perú, se recurrió a la aplicación de la (IA), "habitualmente, en los procesadores de texto y en los buscadores de jurisprudencia", [8] pero esto, como es de suponer, no ha contribuido a lograr el fin último de dicho Acuerdo Nacional por la Justicia, establecido en el Punto 2 de la Ayuda memoria, referido a "brindar un mejor servicio de justicia (eficiente y eficaz) a los ciudadanos".

Ante esta problemática; nos interesa revisar la eficiencia de la administración de justicia laboral en el Perú, en concreto, del proceso laboral, que tiene como finalidad la solución de los conflictos y controversias que surgen de la prestación del trabajo, es decir, de la relación entre el trabajador y el empleador. En ese contexto, el trabajo como derecho fundamental tiene una función social que, en el caso del Perú, y de acuerdo con lo que prescribe el artículo 22 de la Constitución de 1993, está referida a lograr el bienestar del trabajador y su familia, así como, a su realización como persona. Por tanto, como refería Pasco,[9] el salario o remuneración, elemento esencial del contrato de trabajo, es la medida económica del trabajo y, por consiguiente, tiene carácter alimentario y, de su percepción oportuna y completa, depende la subsistencia del trabajador y de su familia.

De acuerdo con lo señalado, y recordando nuevamente a Pasco, el desequilibrio o la desigualdad económica que caracteriza la relación laboral se refleja en el conflicto laboral y compromete valores de distinta jerarquía patrimonial para el trabajador y el patrono. Mientras el patrono asiste en el proceso al debate que puede conllevar a una disminución de su margen de utilidad o, en el peor de los casos, a la pérdida de su capital; el trabajador pone en juego la satisfacción de sus derechos alimentarios, es decir, asiste al proceso por subsistencia, en definitiva, por sobrevivencia o por vida.

6 Ibíd. **7** Ibíd. **8** Jordi Nieva Fenoll, Inteligencia artificial y proceso judicial (Madrid: Marcial Pons, 2018), pp 23-24. **9** Mario Pasco Cosmópolis, Fundamentos de derecho procesal del trabajo (Lima: AELE, 1997).

Por tanto, necesitamos que el proceso laboral sea eficiente y que cumpla con la finalidad y con la función social que le corresponde en un Estado constitucional, social y democrático de derecho, pues, como decía Couture, "la naturalidad de nuestra justicia lenta nos envuelve y nos domina. Los magistrados apenas la perciben porque su falta de contacto con los interesados los priva del espectáculo de su desesperación […] en el procedimiento el tiempo es algo más que oro: es justicia.[10] Creemos sin vacilación que, dado el grado de desarrollo de la IA y su aplicación en diversos aspectos de la vida del ser humano, esta es una herramienta esencial para ayudar en la eficiencia del proceso laboral. Por consiguiente, vamos a reflexionar sobre su uso, teniendo como guía la siguiente pregunta: ¿de qué manera se debe implementar la IA en el proceso laboral para que sea eficiente?

01

DESARROLLO CÓMO HEMOS IMPLEMENTADO LA IA EN LA ADMINISTRACIÓN DE JUSTICIA EN EL PERÚ

La justicia, como pilar fundamental del Estado constitucional, social y democrático de derecho, está presente en los Objetivos de Desarrollo Sostenible (ODS) y en la Agenda 2030 planteada por las Naciones Unidas; concretamente, el objetivo 16, contempla el acceso a la justicia para todos en condiciones de igualdad y el fortalecimiento de las instituciones para que se encaminen a un servicio más eficiente para el ciudadano, es decir, con transparencia.

En concordancia con ello, en el Perú se adoptó el Acuerdo Nacional por la Justicia, cuyas acciones se iniciaron el 2 de noviembre de 2016, con la suscripción de un documento por los representantes y titulares de las instituciones que conforman el Sistema de Administración de Justicia, el Poder Judicial, el Consejo Nacional de la Magistratura (hoy Junta Nacional de Justicia), el Ministerio Público, la Academia de la Magistratura y el Ministerio de Justicia y Derechos Humanos.

El Acuerdo Nacional por la Justicia busca incidir con acciones concretas en cuatro ejes temáticos: reforma institucional, lucha contra la corrupción, acceso a la justicia y formación, capacitación y selección de magistrados. De estos ejes, el que nos interesa es el que tiene que ver con la reforma institucional, específicamente, el documento que desarrolla una estrategia para el cambio en el modelo de la gestión, enfocado en procesos por resultados, con lineamientos de corto, mediano y largo plazo, orientado a brindar un servicio interoperable y de calidad a la ciudadanía.

En el contexto de la reforma institucional, se hace referencia a que esta medida debe ser abordada mediante la implementación de proyectos de gestión del cambio que se desarrollen mediante un enfoque multidimensional y corporativo. El primero, busca asegurar el éxito de la reforma al interior de cada institución para lograr la modernización de los sistemas de gestión de cada una de ellas; el segundo, es necesario para asegurar la facilitación de la interoperabilidad de los procesos de las instituciones que conforman el circuito, logrando la modernización del sistema en su conjunto.

10 Eduardo J. Couture, Fundamentos del derecho procesal civil, 3.a ed. (Buenos Aires: Roque Depalma Editor, 1962), 100.

Los proyectos de gestión han puesto el énfasis en las dimensiones del cambio, entre las que se encuentra el uso intensivo de herramientas y de técnicas de gestión y el uso intensivo de tecnologías de la información y del conocimiento. La primera dimensión, consiste en implementar estrategias y programas que aseguren el uso intensivo y constante de herramientas y técnicas de gestión moderna en las entidades que forman parte del sistema. La segunda, está referida al abordaje de la automatización y de la sistematización de los procesos de gestión e intercambio de información entre las entidades que lo conforman, mediante el uso intensivo de tecnologías de la información y del conocimiento.

Se agrega que, otros desafíos que deben enfrentarse en esta dimensión, son la falta de soporte y de infraestructura tecnológica y la ausencia de una estructura institucional que facilite la interoperabilidad desde un enfoque de procesos y de atención al ciudadano.

La interoperabilidad será entonces un resultado del proceso de reforma, que facilitará el acceso del ciudadano al sistema de justicia, brindándole información oportuna, de calidad y comparable. Iniciativas importantes en este aspecto son: la implementación del Expediente Judicial Electrónico, la Carpeta Fiscal Electrónica, la Ficha Única del Magistrado y, el Plan de Interoperabilidad en Materia Penal.

Como es de suponer, la inclusión de la IA en el Sistema de Administración de Justicia en el Perú, ha tenido una aplicación intensiva en la interoperabilidad de los procesos de las instituciones que lo conforman, para lograr su modernización y facilitar el acceso del ciudadano. La evidencia de lo expuesto se encuentra en el documento del Acuerdo Nacional por la Justicia, en el cual se establece que se debe intervenir en las áreas o dimensiones del cambio, como las de tecnología de la información y conocimiento, que mejorarán los procesos de gestión de gobierno electrónico, gobierno móvil, gobierno digital, gobierno abierto; la infraestructura de datos espaciales; la inclusión digital; la infraestructura tecnológica, entre otros aspectos.

También se busca incidir en la interoperabilidad como resultado del proceso de reforma, tomando en cuenta iniciativas ya existentes, como el Expediente Judicial Electrónico, la Carpeta Fiscal Electrónica, la Ficha única del Magistrado y el Plan de Interoperabilidad en Materia Penal, reconociendo esos esfuerzos y sus avances y, sobre todo, integrándolos en beneficio del sistema de administración de justicia.

Lo expuesto se puede evidenciar actualmente en la página web del Poder Judicial Peruano,[11] en la que están consignados todos los servicios que esta entidad presta a los ciudadanos y que lógicamente han sido implementados con IA; por ejemplo, el módulo de sentido de decisiones judiciales, los casilleros digitales de jueces y juezas, la autoridad nacional de control del poder judicial, la mesa de partes virtual, la consulta de expedientes judiciales, las notificaciones electrónicas, los sistemas de remates virtuales, la jurisprudencia, los plenos jurisdiccionales, las quejas, las denuncias, etc.

Sin embargo, la pregunta de este primer eje de reflexión es: ¿el uso de la IA que hemos implementado en el sistema de administración de justicia peruano contribuye a la eficacia y eficiencia del servicio de justicia a los ciudadanos?, en concreto, ¿esto garantiza el plazo razonable que deben durar los procesos judiciales?

Sinceramente, consideramos que a lo que ha contribuido esta implementación es a mejorar y modernizar, en parte, la gestión de los despachos judiciales y, en cierta medida, de

11 Para acceder a este dominio, consúltese https://www.pj.gob.pe

manera indirecta, puede haber aportado a la reducción de algunos actos procesales que son repetitivos o que han podido ser digitalizados o procesados electrónicamente, pero que, por su escasa importancia o relativa incidencia en el desarrollo del proceso judicial, no son suficientes ni determinantes para reducir los tiempos o costos procesales.

Evidencia de lo expresado en la realidad peruana y, vinculado a la duración de los procesos, es lo que hemos conocido recientemente, de forma pública, como consecuencia de una ley de amnistía para procesados y presos por delitos de lesa humanidad, por hechos ilícitos realizados en la época del terrorismo que azotó nuestro país durante los años 80 y 90; justamente, para amnistiar a militares procesados penalmente durante veinte a treinta años y que aún no han sido sentenciados de forma definitiva. Esto que sucede en materia penal, indudablemente, también se ve reflejado en los demás tipos de procesos en el poder judicial peruano.

En ese contexto, es bueno meditar sobre la aplicación de la IA al proceso laboral, entendiendo, como refiere Nieva,[12] que, aun cuando vivimos absolutamente rodeados de IA, su aplicación en el campo es todavía rudimentaria, muy por debajo del uso social que se hace de estas tecnologías en la actualidad, porque el mundo del derecho está lamentablemente ligado a la tradición.

Actualmente, en los juzgados, lo habitual es ver únicamente la aplicación de IA en los procesadores de texto, en buscadores de jurisprudencia y como una herramienta de "corta y pega" en la elaboración de las sentencias. Sin embargo, necesitamos seguir avanzando creativamente y usar la IA para la generación de documentos reflexivos que nos permitan solucionar casos sencillos, en los que la prueba es documentada, sencilla y previsible, de tal manera que los algoritmos de la IA los puedan analizar, para emitir respuestas acertadas.

12 Nieva Fenoll, Inteligencia artificial y proceso judicial.

02
EL PROCESO LABORAL COMO PROCESO SOCIAL POR EXCELENCIA

El proceso laboral, de acuerdo con la teoría general del proceso, comprende la jurisdicción que tienen los jueces especializados de trabajo, para resolver los conflictos laborales de naturaleza jurídica, garantizando el derecho a la tutela jurisdiccional efectiva y el debido proceso que asiste a los trabajadores y empleadores, dentro de un conjunto de garantías procesales de tutela de derechos fundamentales en un Estado Constitucional de Derecho, valiéndose para ello del proceso como instrumento para resolver la controversia legal y restablecer la paz social en justicia.



Foto: Pixabay

La naturaleza social del proceso está dada por la naturaleza del derecho laboral, que involucra tanto a trabajadores como empleadores, en tanto está referido a un derecho humano y fundamental que tiene una función social primordial, referida a la subsistencia de los trabajadores y su familia, así como medio de realización de la persona del trabajador; pues, la única forma que tenemos para realizarnos en la vida como personas o seres humanos, es a través del trabajo. Por tanto, éste, como derecho social, se basa en la dignidad del trabajador y está reconocido por el Estado como un derecho fundamental, en el artículo 22 de la Constitución del Estado de 1993 en el Perú.

Sin embargo, lo que caracteriza la relación entre el trabajador y el empleador es la desigualdad que existe entre ambos, la misma que es de naturaleza económica. Hay un sujeto que tiene poder económico, en tanto, dueño del capital, de los medios y los procedimientos de producción o del servicio que se otorga; y, otro, que no tiene ese poder económico, pero que tiene algo equivalente como factor de la producción: su trabajo, el cual pone a disposición del empleador para que este obtenga frutos o renta con la prestación personal otorgada.

En ese contexto, la forma de equiparar a los desiguales es creando otra desigualdad de tipo jurídico. Esa desigualdad que hemos creado para compensar la desigualdad económica y que caracteriza la relación laboral, es el derecho al trabajo; por tanto, se trata de un derecho compensador, equilibrador o protector, desde su origen; ese es su ADN.

Por ello, coincidimos en afirmar que el derecho al trabajo es el derecho social por excelencia, porque es el derecho que más se acerca a la vida.

Ahora, el proceso laboral es el instrumento, método o herramienta que sirve para solucionar los conflictos y controversias que derivan de la relación entre el trabajador y el empleador; relación que, por naturaleza y en atingencia a la asimetría económica que lo caracteriza, es sumamente conflictiva en atención a los intereses contrapuestos de cada uno de los sujetos que intervienen en la relación laboral.

La desigualdad económica entre el trabajador y el empleador también se refleja en el proceso laboral y, de hecho, esta es una de sus características fundamentales. Dicha desigualdad se manifiesta, por ejemplo, en la pretensión expresada en el proceso, la misma que para el empleador tiene significado patrimonial, cifrado en el valor económico de los derechos sociales postulados por el trabajador que, de ser declarados fundados, implica para el empleador una disminución en sus utilidades y, en el peor de los escenarios, la pérdida de su capital. En cambio, el trabajador asiste por subsistencia, debido al carácter alimentario de la remuneración y de los beneficios que la Constitución y la ley establecen a su favor, por tanto, el trabajador asiste al proceso laboral por vida.

Además, dicha desigualdad genera una desigualdad de tipo probatorio en el trabajador; pues, toda la prueba que se genera como consecuencia de la prestación del trabajo y que, por tanto, la acredita, se queda y permanece en el centro de trabajo. En consecuencia, cuando el trabajador asiste al proceso laboral y quiere aportar prueba para acreditar algún hecho vinculado a la prestación del trabajo, tiene que arrancarlo de ese medio hostil que es el centro de trabajo; por ello, se le considera un *hiposuficiente probatorio.*

Por último, la desigualdad económica también causa una desigualdad de tipo sicológico en el trabajador; pues, en tanto subsiste la relación de subordinación o dependencia, elemento esencial de naturaleza jurídica del contrato de trabajo, el trabajador se ve limitado para reclamar el pago de los beneficios sociales que el empleador puede estar incumpliendo, debido al temor de las represalias que este pueda adoptar como consecuencia del reclamo efectuado. Por ello, el plazo de prescripción de los derechos laborales se computa a partir de la fecha en que cesa o culmina la prestación del trabajo.

Dentro de ese contexto y para superar las desigualdades anotadas, tratamos a través del proceso laboral de compensarlas, de tal manera, que ambos sujetos, trabajador y empleador, asistan en igualdad de condiciones al proceso. Por tanto, es la norma la que establece la equiparación o compensación y, en consecuencia, el juez laboral debe seguir siendo imparcial e independiente. Como diría Pasco,[13] *asépticamente imparcial*

.

Ahora, ¿cómo hacemos para compensar?. Sencillamente, estructuramos el proceso laboral con características tanto del principio dispositivo como del principio inquisitivo, generando un proceso que denominamos "proceso dispositivo atenuado".[14] Dicho proceso, se caracteriza porque las características dispositivas dan a las partes ciertas facultades de disposición del proceso y, las inquisitivas, están destinadas a tutelar la imperatividad e indisponibilidad de los derechos laborales del trabajador; por ejemplo, se redistribuye la carga probatoria, facilitando o aligerando la que corresponde al trabajador, se crean presunciones a favor de este y se otorgan acentuados poderes de dirección formal y material del proceso al juez laboral.

En ese contexto, con la finalidad que el proceso laboral cumpla la función social que hemos reseñado, en el Perú, en el año 2010, se puso en vigencia una nueva norma procesal laboral contenida en la Ley No 29497, la misma que se ha denominado Nueva Ley Procesal del Trabajo (NLPT). De acuerdo con dicha norma, apostamos por la estructura de un proceso laboral oral, es decir, un proceso con audiencias. En el proceso ordinario laboral se trató de privilegiar la conciliación como forma de conclusión anticipada del mismo y se adicionó una audiencia de conciliación previa a la audiencia de juzgamiento. Además de apostar por la oralidad, apostamos también por la tecnología, pues la audiencia de juzgamiento se graba en audio y video. Por último, establecimos como regla que la sentencia se debe emitir en la audiencia de juzgamiento, luego de terminados los alegatos de clausura o de cierre y de un receso de sesenta minutos; como excepción, justificada en la complejidad o la magnitud del caso, se podía emitir la sentencia cinco días después de realizada la audiencia de juzgamiento.

13 Pasco Cosmópolis, Fundamentos de derecho procesal del trabajo. **14** Paul Paredes, Prueba y presunciones en el proceso laboral (Lima: ARA Editores, 1997).



Foto: Red de Comunicación Regional - Perú

Cuando fue publicada la norma, nuestro más distinguido procesalista, el maestro Juan Monroy Gálvez, vaticinó:

"Estamos ante una ley que es buena y original. Lamentablemente en aquello que es buena no es original y en aquello que es original no es buena […]. Su medianía consiste en creer que algunas palabras mágicas —como celeridad, oralidad, tecnología, entre otras—, pueden convencer a los ingenuos de siempre que se han producido cambios sustanciales. El eterno mito de la ley en los Estados emergentes". [15]

Han pasado casi quince años, y ya tenemos en circulación un nuevo proyecto de Código Procesal del Trabajo, publicado por el Ministerio de Justicia y Derechos Humanos, mediante la Resolución Ministerial No 0103-2024-JUS, del 30 de abril de 2024. En una entrevista que se ha consignado en la exposición de motivos de dicho proyecto, se formuló la siguiente pregunta: ¿considera usted necesario que se introduzcan modificaciones en la nueva Ley Procesal del Trabajo para darle mayor celeridad? El 88 % de los entrevistados respondió que sí (Figura 1).

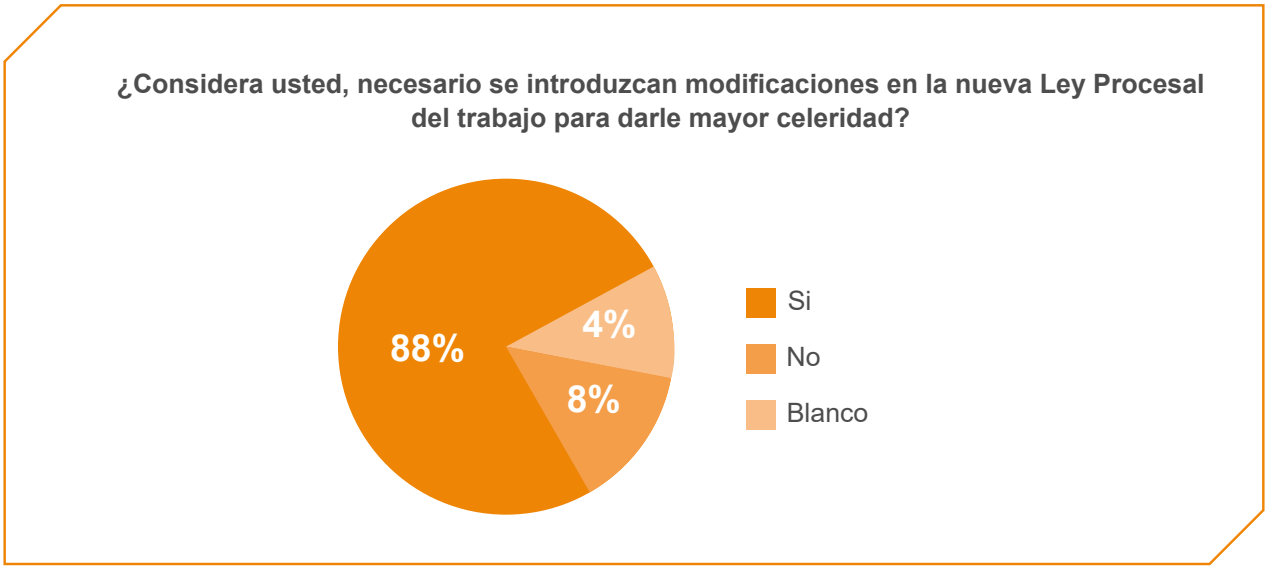


Figura 1. Necesidad de modificar la Nueva Ley Procesal del Trabajo
Fuente: Proyecto del Código Procesal del Trabajo. Ministerio de Justicia y Derechos Humanos. 2024.

15 Juan Monroy Gálvez, Comentarios a la Nueva Ley Procesal del Trabajo (Lima: Gaceta Jurídica, 2010), 103-141.

Otro aspecto importante que se debe considerar, es que al inicio de la entrevista se formularon dos preguntas para conocer la opinión de los entrevistados respecto a la ley procesal vigente. La primera, para establecer si consideraban necesaria la reforma del actual proceso laboral, y, la segunda, para que indicaran las razones que fundamenten su afirmación. Las respuestas fueron contundentes: primero, respecto a la necesidad de reformar el actual proceso laboral, 98 % la consideraba necesaria; y, segundo, respecto a las razones que fundamentan esa necesidad, los entrevistados consideraron que son mayormente las cuantías mínimas de los procesos laborales y, que los jueces no están preparados (Figuras 2 y 3).

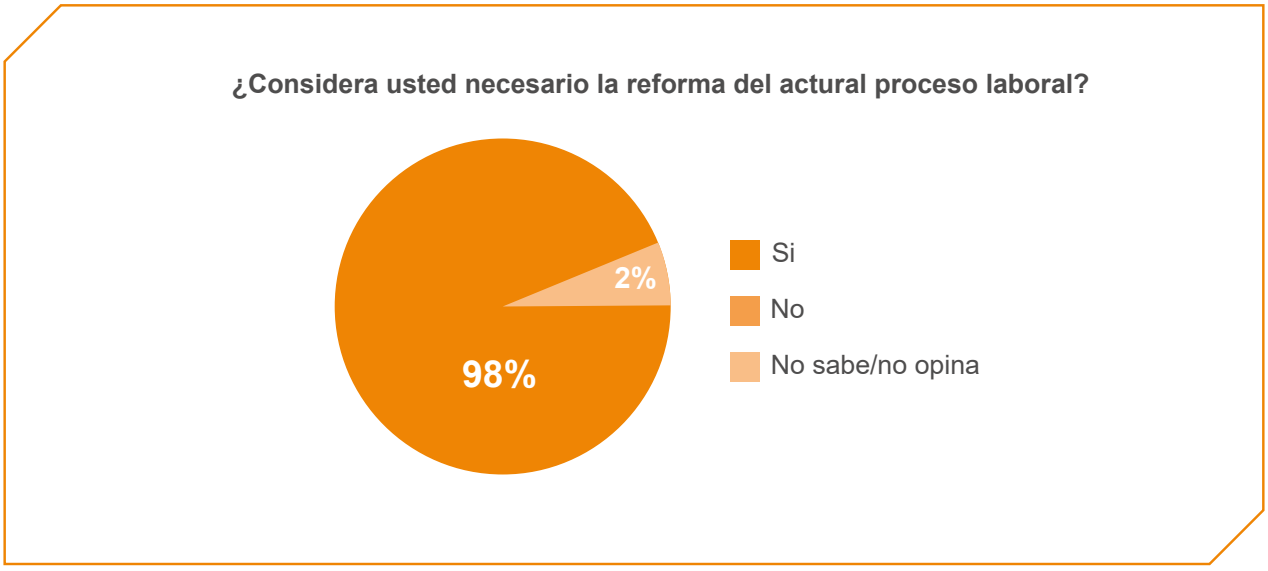


Figura 2. Necesaria la reforma del actual proceso laboral
Fuente: Proyecto del Código Procesal del Trabajo. Ministerio de Justicia y Derechos Humanos. 2024.

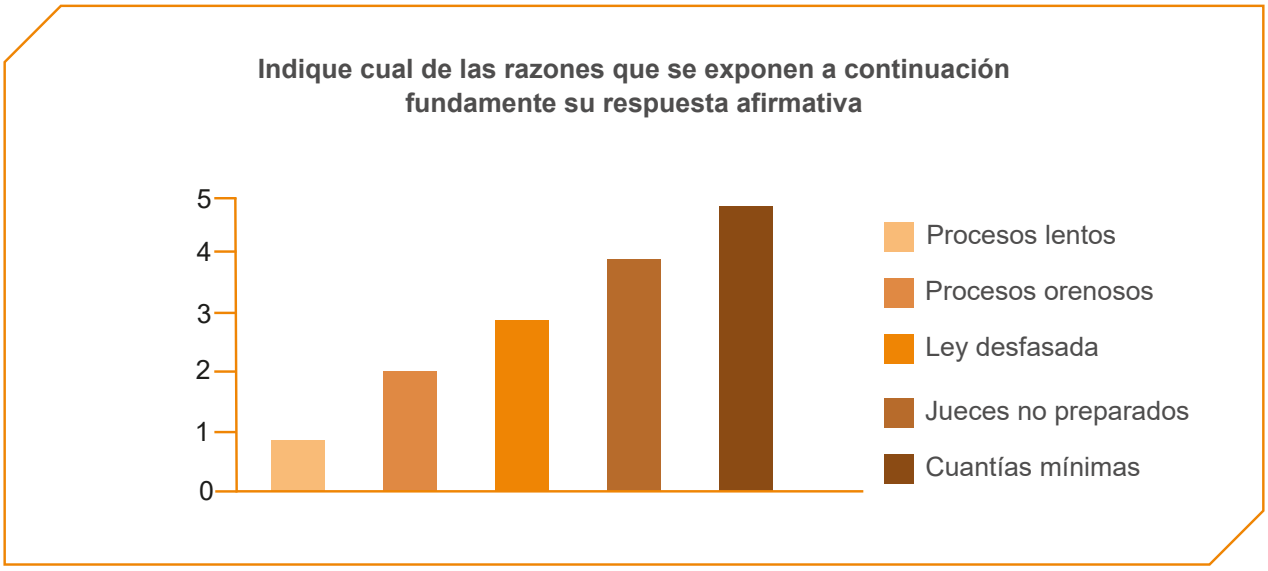


Figura 3. Fundamentos de la necesaria reforma del actual proceso laboral
Fuente: Proyecto del Código Procesal del Trabajo. Ministerio de Justicia y Derechos Humanos. 2024.

Para terminar, también consideramos que la denominada Nueva Ley Procesal del Trabajo No 29497 ha sido modificada en dos oportunidades; la primera, mediante la Ley No 31699 de 1 de marzo de 2023, denominada "Ley que optimiza el recurso de casación laboral en la Nueva Ley Procesal del Trabajo", porque ha restringido el acceso a dicho recurso extraordinario; la segunda, por la Ley No 32155 del 7 de noviembre de 2024, que ha ampliado la competencia por materia de los Juzgados de Paz Letrados Laborales.

Como conclusión de este eje de reflexión, debemos sostener que debido a la naturaleza social del proceso laboral, a la naturaleza alimentaria de los derechos laborales y a la función social que debe cumplir el proceso laboral en un Estado constitucional, social y democrático de derecho, es imperativo que este sea eficaz y eficiente. Solo en esa medida se convertirá en un instrumento de justicia social, en tanto permita que el trabajador pueda vivir con dignidad.

03

EL USO DE LA IA COMO FACTOR CLAVE PARA LOGRAR LA EFICIENCIA DEL PROCESO LABORAL

La problemática de la Nueva Ley Procesal del Trabajo

1. Vamos a tratar de determinar cuáles son los aspectos problemáticos de la Nueva Ley Procesal del Trabajo (NLPT) durante sus quince años de aplicación. Para ello, tomamos como referencia los criterios expuestos en nuestro libro,[16] considerando como primera problemática la morosidad o lentitud de los procesos. No hemos podido obtener datos respecto a la demora de los procesos ordinarios o abreviados laborales que se han estructurados con la NLPT; al efecto, la justicia especializada laboral ni el poder judicial peruano tiene data que nos permita conocer cuál es la duración de los procesos laborales vigentes.
2. El Equipo Técnico Institucional de Implementación de la Nueva Ley Procesal del Trabajo del Poder Judicial (ETIINLPT), encargado de todo lo relativo a la implementación y funcionamiento adecuado de los órganos de la jurisdicción laboral que aplican la NLPT, no considera este tipo de información en la poca estadística que genera. Sus data estadística está referida a determinar la carga procesal de los órganos jurisdiccionales, de acuerdo con el número total de expedientes ingresados en cada año, el número de los que han sido resueltos en ese año, para determinar luego la cantidad de expedientes pendientes por resolver.

Sin embargo, basados en la investigación elaborada por Vásquez,[17] asumimos los criterios obtenidos con la finalidad de determinar el plazo en el que se resuelven los procesos laborales de menor cuantía; es decir, los que son de competencia de los Juzgados de Paz Letrados Laborales y que se tramitan como proceso abreviado, llegando a la conclusión de que estos se resuelven en un plazo de hasta tres y cinco veces mayor del plazo establecido por la NLPT

16 Homero Jondec Briones, Propuesta para un proceso laboral eficiente (Trujillo: Industria Gráfica ABC SAC, 2022). **17** K. Vásquez, "La oralidad en el proceso laboral ¿Es necesaria en los procesos de menor cuantía?", Soluciones Laborales, N° 14 (Lima: Gaceta Jurídica), 165-181.

3. Para comprender la morosidad, debemos precisar que los gestores de la norma consideraron que el plazo de duración, en promedio, de un proceso laboral con la NLPT debía ser de ocho a doce meses para contar con una sentencia definitiva. En realidad, según el sentir de los operadores de justicia a la fecha, un proceso laboral con la NLPT en el Perú demora de cuatro a seis años para tener sentencia definitiva. Para confirmar lo expuesto, presentamos una pregunta de la entrevista que se consigna en la exposición de motivos del proyecto del nuevo Código Procesal del Trabajo, al que hemos hecho alusión anteriormente. La pregunta se refiere a la duración aproximada entre la interposición de la demanda y la sentencia definitiva. La respuesta evidencia que el 56 % de entrevistados considera que la duración aproximada es de más de tres años (Figura 4).

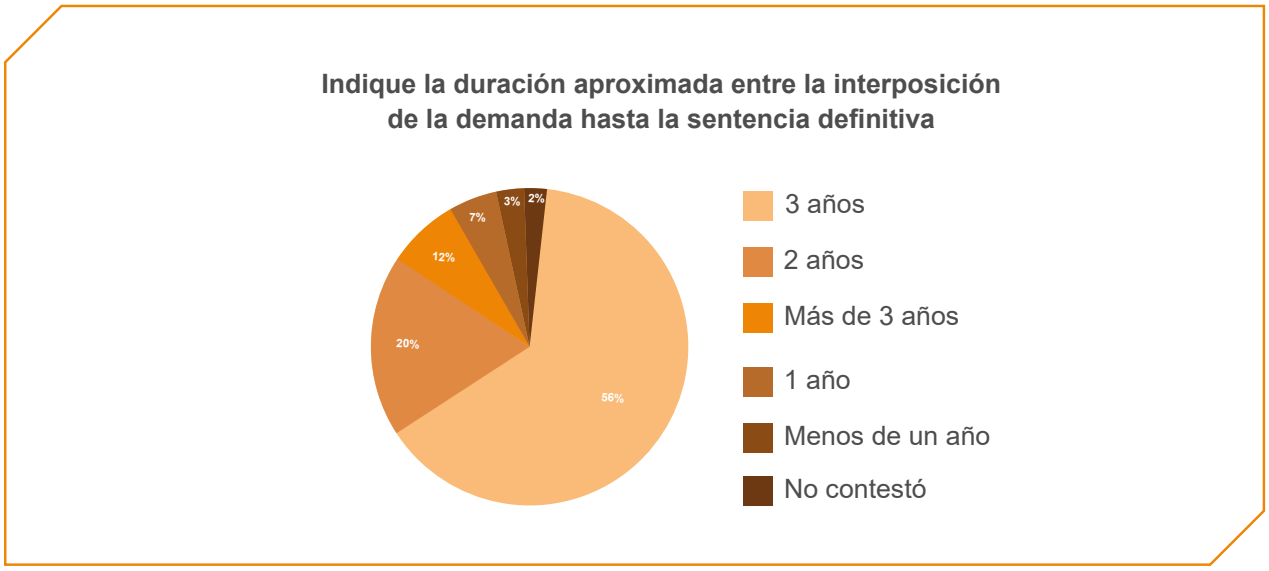


Figura 4. Duración del proceso ordinario laboral con la NLPT

Fuente: Proyecto del Código Procesal del Trabajo. Ministerio de Justicia y Derechos Humanos. 2024.

Otra problemática que afecta a la NLPT es la inoperancia de la conciliación como forma de conclusión anticipada del proceso laboral. Si bien hemos querido privilegiarla, por ser una característica del proceso laboral la promoción de la solución del proceso mediante la conciliación, dicho postulado no se ha hecho tangible con la vigencia de la NLPT, pese a que el proceso ordinario laboral se ha estructurado estableciendo una audiencia exclusiva para ello y previa a la audiencia de juzgamiento.

No ternemos datos actualizados respecto al porcentaje de procesos laborales que culminan por conciliación, porque, como reiteramos, la justicia especializada laboral ni el poder judicial los generan; sin embargo, vamos a recurrir nuevamente a la entrevista realizada en la Exposición de Motivos del Proyecto del Código Procesal del Trabajo publicado por el Ministerio de Justicia y Derechos Humanos el 30 de abril del año 2024. Al respecto, se preguntó a los entrevistados cuántos de sus procesos concluyó por conciliación. La respuesta del 52 % de los entrevistados fue de 1 a 5 procesos (Figura 5).

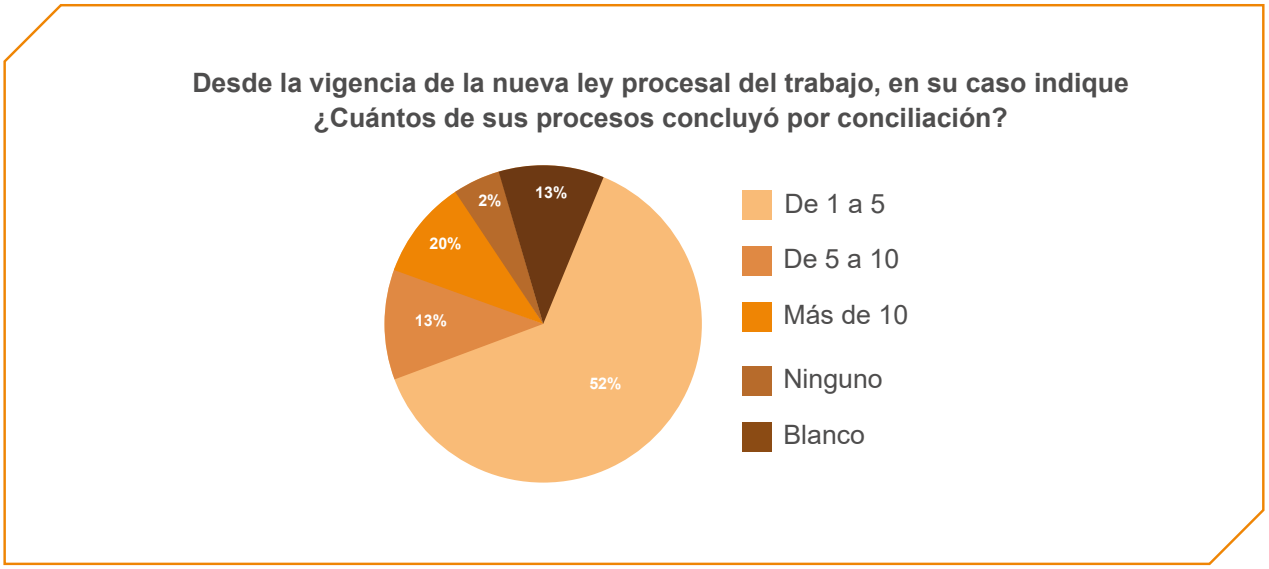


Figura 5. Procesos conciliados durante la vigencia de la NLPT
Fuente: Proyecto del Código Procesal del Trabajo. Ministerio de Justicia y Derechos Humanos. 2024.

Luego se les preguntó si la conciliación solucionaba los procesos judiciales en la actualidad. El 40 % de los entrevistados respondió que no (Figura 6).

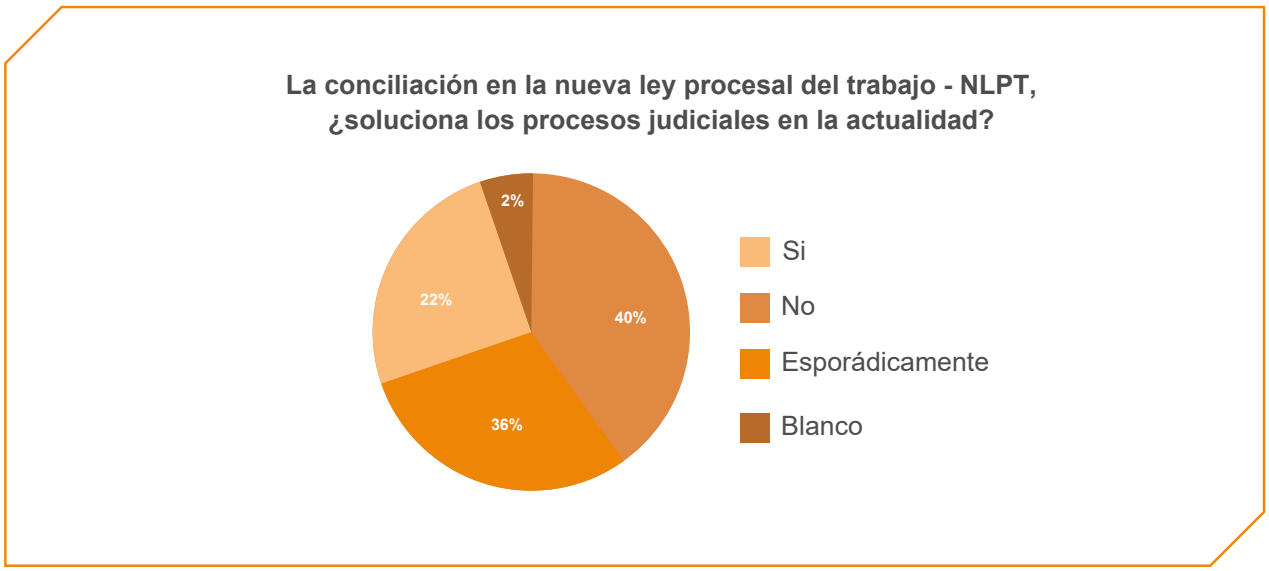


Figura 6. La conciliación como forma de conclusión anticipada del proceso laboral en la NLPT
Fuente: Proyecto del Código Procesal del Trabajo. Ministerio de Justicia y Derechos Humanos. 2024.

También se les preguntó si se deben unificar las audiencias de conciliación y de juzgamiento del proceso ordinario laboral en una sola audiencia. El 64 % de los entrevistados respondió que sí (Figura 7).

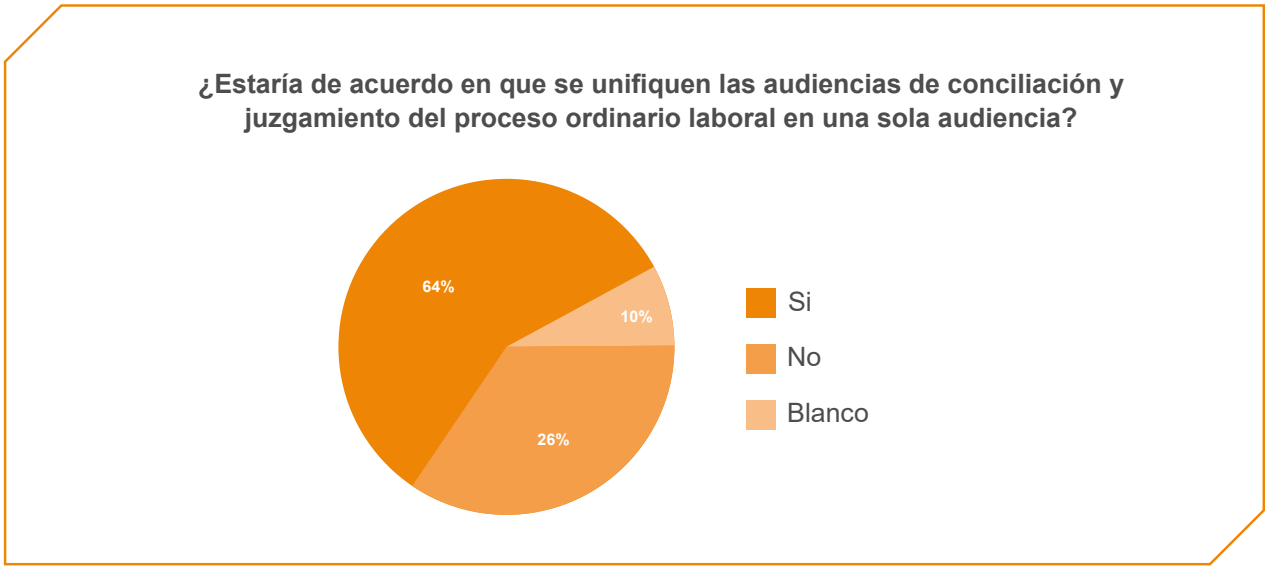


Figura 7. Unificación de las audiencias de conciliación y juzgamiento en el proceso laboral ordinario de la NLPT
Fuente: Proyecto del Código Procesal del Trabajo. Ministerio de Justicia y Derechos Humanos. 2024.

De la evidencia expuesta, podemos concluir que la audiencia de conciliación establecida en la NLPT no cumple con los postulados propuestos en la norma. Dicha audiencia se ha convertido en un mero acto procesal sin mayor importancia e incidencia; como decíamos en nuestra publicación anterior,[18] mantener una audiencia previa y privilegiada para insistir en la promoción de la conciliación como forma de conclusión anticipada del proceso laboral, es inútil y es una de las causas de su morosidad.

Este no es el espacio ni la oportunidad para exponer las causas que consideramos han propiciado el fracaso de la conciliación como forma de conclusión anticipada del proceso laboral; sin embargo, queremos dejar precisado que a nuestro entender, esta problemática tiene que ver con el diseño estructural de la norma, con la regulación supuestamente original pero no adecuada que se le ha dado a la conciliación, generando consecuencias funestas en la relación jurídico procesal. En efecto, como consecuencia de lo anotado, se ha generado toda una problemática con el saneamiento en el proceso laboral, con la rebeldía procesal y, con la eficacia misma de la conciliación, en tanto, se realiza sin que el demandado haya presentado su escrito de contestación a la demanda, lo cual recién procede cuando fracasa la conciliación.

Dentro de este contexto, tal como refieren los entrevistados en la Exposición de Motivos del Proyecto del Código Procesal del Trabajo, es necesario unir la audiencia de conciliación con la audiencia de juzgamiento en el proceso ordinario laboral, de tal manera que el proceso ordinario laboral tenga una audiencia única al igual que el proceso abreviado laboral. En dicho acto estelar, se debe realizar primero el saneamiento procesal; luego se deben formular los alegatos de inicio o de apertura por las partes; posteriormente, se debe intentar la conciliación; y, si esta no prospera,

18 Jondec Briones, Homero (2022). Propuestas para un proceso laboral eficiente. (Trujillo, Perú, Industrias Gráficas ABC SAC).

el juez debe proceder a determinar los puntos controvertidos; luego, admitir los medios de prueba ofrecidos por las partes; a continuación, se pueden formulan cuestiones probatorias; posteriormente, se actúan los medios de prueba admitidos; finalmente, se procede con los alegatos de cierre o de clausura por las partes; y, por último, el juez debe pronunciar el fallo de la sentencia a viva voz.

Para terminar con esta problemática de la NLPT y de forma breve, voy a referirme al papel que ha cumplido el recurso de casación en el proceso laboral. En nuestra realidad y de acuerdo con lo verificado en la práctica, el recurso extraordinario de casación en el proceso laboral nos ha generado más problemas que soluciones. Nuestra Corte Suprema de Justicia está en deuda con la justicia laboral en el Perú, porque no ha cumplido con establecer, en forma eficiente, precedentes jurisdiccionales en materia laboral.

Podemos sostener que la Corte Suprema de Justicia de la República del Perú no ha cumplido con su función dikelógica, menos con la función nomofiláctica propia de un órgano jurisdiccional de cierre o supremo, a través del recurso de extraordinario de casación, el cual tiene por finalidad garantizar la seguridad jurídica y la uniformidad jurisprudencial. Por el contrario, el recurso de casación en el proceso laboral es un factor dilatorio que agravia el plazo razonable, por cuanto, su trámite en la práctica, dura entre dos y tres años. Esto contribuye a la dilación del proceso laboral y agravia el derecho fundamental de los trabajadores a la tutela jurisdiccional efectiva; su centralismo y funcionamiento, son sinónimo de dilación y onerosidad respecto al derecho social de los trabajadores.

04

NUESTRA PROPUESTA DE EFICIENCIA DEL PROCESO LABORAL VINCULADA A LA INTELIGENCIA ARTIFICIAL

Antes de enunciar y argumentar nuestra propuesta, es fundamental referirnos a dos características fundamentales del proceso laboral que nos van a servir de sustento. La primera, es la naturaleza de la controversia jurídica que se genera en el proceso laboral; la segunda, la simplicidad y la sencillez del mismo.


Foto: Magnific

Respecto a la naturaleza de la controversia jurídica en el proceso laboral. Todos están de acuerdo que en un proceso laboral no se oponen derechos contra derechos; es decir,

en el proceso laboral, frente a un derecho de naturaleza social que propone un trabajador, el empleador no opone o controvierte otro derecho que a él le reconoce otra fuente normativa. La controversia jurídica en el proceso laboral está referida al derecho del trabajador y el cumplimiento de la obligación correspondiente por el empleador. Por consiguiente, en el proceso laboral se oponen derechos (del trabajador) contra obligaciones (del empleador).

Este esquema marca la peculiaridad de la controversia jurídico-laboral, por tanto, se trata de casos sencillos que demandan soluciones rápidas y eficientes, que garanticen la naturaleza social, alimentaria e indisponible de los derechos subjetivos de los trabajadores.

Respecto a la simplicidad y la sencillez del proceso laboral. En congruencia con la característica anterior, el proceso laboral debe ser ágil, rápido, formalista en lo imprescindible y debe facilitar el acceso a la tutela jurisdiccional efectiva.[19] El proceso laboral, de acuerdo con su finalidad, debe ir a lo esencial, a lo sustancial, y debe limitar el formalismo procesal. El proceso laboral debe ser de un rito sumarísimo y de una simplicidad extrema. El proceso debe ser simple, claro, directo tanto en lo estructural como en lo normativo. La sencillez debe reflejarse así mismo en la norma procesal.[20]

Sobre la base de estas características fundamentales, formulamos a continuación nuestra propuesta:

I. En cuanto a la competencia por materia en el proceso laboral, debemos regirnos por la teoría de los casos fáciles o sencillos y los casos difíciles o complejos; es decir, debemos distribuir la competencia por materia de los órganos jurisdiccionales laborales, teniendo en cuenta la complejidad, dificultad o sencillez de los casos judiciales, de acuerdo con las pretensiones expresadas en la demanda.

   Para ello, brevemente debemos tener en cuenta la teoría de Ronald Dworkin,[21] quién refería que todo problema jurídico tiene una única solución, aún en los casos difíciles, y que el deber del juez es descubrir los derechos de las partes, en vez de inventar derechos nuevos, por esta razón, había que tomar los derechos en serio.

Asimismo, hay que hacer alusión a Barak,[22] quien refiere que la opinión aceptada es que la mayoría de los casos que llegan a los tribunales no son casos difíciles; solo una pequeña parte de todas las decisiones de un sistema se encuentra en la esfera de los casos difíciles.

En ese contexto, nuestra propuesta es que, en el proceso laboral, se consideren casos fáciles o sencillos a los que tengan las siguientes características:

1) que estén referidos al incumplimiento de cualquier norma laboral-

2) que el vínculo laboral entre el trabajador y el empleador esté debidamente acreditado, admitido o reconocido, es decir, que no sea materia controvertida.

3) que solo se admita prueba documental.

19 Couture, Fundamentos del derecho procesal civil. **20** Pasco Cosmópolis, Fundamentos de derecho procesal del trabajo. **21** Ronald Dworkin, Los derechos en serio (Barcelona: Editorial Ariel, 2012). **22** Aharon Barack, Discrecionalidad judicial (Lima: Palestra Editores, 2021).

En contraposición, se deben considerar casos difíciles o complejos a todos aquellos que no tengan las características anotadas, por ejemplo, aquellos en que sea materia controvertida el vínculo laboral entre el trabajador y el empleador; de igual modo, cuando sea materia controvertida la naturaleza de la prestación del servicio; asimismo, los que tienen que ver con la reparación del despido, ya sea que se solicite la reposición o la indemnización; las indemnizaciones por daños y perjuicios derivados del contrato de trabajo; la desnaturalización de la contratación modal; la desnaturalización de la intermediación o de la tercerización; etc.

ii) Los casos fáciles o sencillos en el proceso laboral deben ser "resueltos" con aplicación de la IA. De acuerdo con esta propuesta, los casos fáciles y sencillos deben ser resueltos mediante un proceso totalmente oral y con aplicación de la IA. Cuando sostenemos que deben ser "resueltos" por la IA, no pretendemos que esta reemplace al juez y ejerza función jurisdiccional. Para explicar lo expuesto, Nieva hace referencia a que "la inteligencia artificial es humana, porque la han hecho humanos, incluso aunque sea capaz de 'aprender' de los datos que va recopilando. La inteligencia artificial no dicta sentencias, solo ayuda a dictarlas".[23]

A esto también debemos agregar una precisión insoslayable, la inteligencia artificial es un "algoritmo", una especie de esquema ejecutivo de la máquina almacenando todas las opciones de decisión en función de los datos que se vayan conociendo. La máquina no tiene límites de almacenamiento de información. La máquina tampoco "aprende", porque ello supone asumir, interiorizar y crear nuevas ideas a partir de ese aprendizaje.[24]

Entonces, ¿cuál es nuestra propuesta? Como hemos precisado anteriormente, en un proceso laboral se oponen derechos contra obligaciones, por lo tanto, el 90 % de las controversias laborales son fáciles o sencillas; es decir, la mayoría de ellas están referidas al incumplimiento de pago de remuneraciones o beneficios laborales que la Constitución, la ley y el convenio colectivo establecen a favor de los trabajadores.

En ese contexto, si de una prueba documental se acredita el incumplimiento del empleador, lo que tiene que hacer el juez en la sentencia es declarar que la demanda es fundada y determinar el pago de lo omitido por el empleador, para lo cual debe realizar una labor mecánica de cálculo que le permita cuantificar el importe que debe mandar pagar al empleador obligado.

Esa declaración y ese cálculo mecánico que debe realizar el juez indudablemente lo puede hacer en mejores condiciones la IA; pues, como refiere Nieva,[25] la IA permite obtener una variedad de documentos, puede realizar una copia automática de los mismos y, además, tiene una mayor capacidad de análisis de este material. Si combinamos esas tres funciones, la aplicación de la IA podría ser más rápida que un juez en la resolución de procedimientos fáciles o sencillos y, además, previsibles.

Lo que tendría que hacer el juez, es alimentar adecuadamente la máquina con la información del caso, con los documentos de prueba aportados y con los cálculos matemáticos mecánicos aplicables al incumplimiento materia de pretensión, para tener la sentencia elaborada en escasos minutos y proclamarla inmediatamente al final de la audiencia de juzgamiento.

23 Nieva Fenoll, Inteligencia artificial y proceso judicial. **24** Ibíd. **25** Ibíd.

CONCLUSIONES

1. La administración de justicia en el Perú ha implementado la IA de manera rudimentaria y por debajo de los usos sociales de la misma. Se puede verificar en su portal web que se ha utilizado para mejorar la gestión de los despachos judiciales y, en el proceso, se advierte el uso de la IA en los juzgados, en los procesadores de texto, en buscadores de jurisprudencia y como herramienta de “corta-pega” en la elaboración de las sentencias.

2. El proceso laboral es un proceso social por excelencia cuya función social es garantizar o tutelar los derechos sociales, alimentarios e indisponibles de los trabajadores, mantener el equilibrio armonioso en las relaciones laborales y, por lo tanto, debe ser eficiente para garantizar el plazo razonable.

3. La IA debe aplicarse en el proceso laboral para resolver los casos fáciles o sencillos, referidos a los incumplimientos de normas laborales de parte del empleador, en los cuales no existe controversia respecto a la prestación del servicio, se aporte únicamente prueba documental y, la decisión y el cálculo de la obligación puedan ser procesados por la IA de acuerdo con la información brindada por el juez.



Foto: Magnific

REFERENCIAS

Ancira, Andrea. Gobernabilidad democrática en América Latina. México D. F.: Centro de Investigación y Docencias Económicas (CIDE), 2003.

Barack, Aharon. Discrecionalidad judicial. Lima: Palestra Editores, 2021.

Bobbio, Norberto y Nicola Matteucci. Diccionario de política. México: Siglo XXI, 1998.

Dworkin, Ronald. Los derechos en serio. Barcelona: Editorial Ariel, 2012.

Couture, Eduardo Juan. Fundamentos del derecho procesal civil. 3.a ed. Buenos Aires: Roque Depalma Editor, 1962.

Jondec Briones, Homero. Propuesta para un proceso laboral eficiente. Trujillo: Industria Gráfica ABC SAC, 2022.

Monroy Gálvez, Juan. Comentarios a la Nueva Ley Procesal del Trabajo. Lima: Gaceta Jurídica, 2010, 103 -141

Nieva Fenoll, Jordi. Inteligencia artificial y proceso judicial. Madrid: Marcial Pons, 2018.

Paredes, Paul. Prueba y presunciones en el proceso laboral. Lima: ARA Editores, 1997.

Pasco Cosmópolis, Mario. Fundamentos de derecho procesal del trabajo. Lima: AELE, 1997.

Poder Judicial del Perú. Acuerdo Nacional Por la Justicia. https://www.minjus.gob.pe/wp-content/uploads/2017/08/04_ACUERDO-NACIONAL-POR-LA-JUSTICIA-ayuda-memoria-agosto-2017.pdf

Rodríguez, Virginia. "La gobernabilidad: el debate de un concepto cambiante", Más Poder Local, No 7 (2011): 48-49. https://dialnet.unirioja.es/servlet/articulo?codigo=376599

Vásquez, K. "La oralidad en el proceso laboral. ¿Es necesaria en los procesos de menor cuantía?". Soluciones Laborales, N° 14, 165-181 (Lima: Gaceta Jurídica).

ACERCA DE LOS AUTORES

Homero Pracedes **Jondec Briones**

Abogado de la Universidad Nacional de Trujillo (Perú), maestro en Derecho del Trabajo y de la Seguridad Social por la misma casa de estudios y doctor en Derecho por la Universidad San Pedro de Chimbote. Cuenta con acreditación como Conciliador Extrajudicial por el Ministerio de Justicia y Derechos Humanos del Perú, además de formación complementaria en Derecho Constitucional y Procesal Constitucional, Derecho Jurisdiccional y Docencia Universitaria. Ha desempeñado cargos de alta responsabilidad en el sector público y judicial, como Sub Director Regional de Inspección e Higiene y Seguridad Ocupacional de la Dirección Regional de Trabajo y Promoción del Empleo de La Libertad, Juez Especializado de Trabajo en la Corte Superior de Justicia de La Libertad (Trujillo) y Juez Superior en la Sala Laboral de la Corte Superior de Justicia del Santa (Chimbote). En el ámbito académico, es docente universitario de pregrado y posgrado en diversas universidades peruanas.



05 Prevención de lavado de activos con activos físicos y criptoactivos.

E-ESTADO
Primera Edición

Prevención de lavado de activos con activos físicos y criptoactivos. Inteligencia financiera, debida diligencia, GAFI y GAFILAT

Prevention of Money Laundering with Physical Assets and Crypto Assets. Financial Intelligence, Due Diligence, FATF, and GAFILAT

Fecha de recepción: 30 de octubre de 2025
Fecha de aceptación: 18 de febrero de 2026

Rubén Flores Dapkevicius
Universidad de la República, Uruguay
rflores@montevideo.com.uy
floresdapkevicius@hotmail.com.uy
ORCID: 0000-0002-7842-1281

Cómo citar: Flores Dapkevicius, Rubén. “Prevención de lavado de activos con activos físicos y criptoactivos. Inteligencia financiera, debida diligencia, GAFI y GAFILAT”. Revista e-Estado 1, n.° 1 (2026).



Foto: Magnific

RESUMEN

El presente trabajo trata de dos elementos fundamentales en el ecosistema disruptivo producido por el avance constante de la tecnología, como son el lavado de activos y los criptoactivos o activos virtuales.

El derecho de la prevención de lavados de activos es una especialización relativamente novedosa que surge en la década de los ochenta del siglo pasado y tuvo como antecedente fundamental el desarrollo de lo que luego serían llamados delitos precedentes, como la trata de blancas, la venta de armas de destrucción masiva, el tráfico de estupefacientes, etc. Comprobada esa situación, se creó el GAFI, que aprobó las cuarenta recomendaciones a efectos de combatir estos flagelos.

La aparición de las TIC, la dark web, entre otros, y la adopción comercial de la blockchain, con el nacimiento de la primera criptomoneda, determinó que los blanqueadores de activos comenzaran a utilizar ese tipo de activos para tipificar los delitos consecuencia, esto es, los delitos de lavado de activos que ya no solo se realizan con activos físicos, sino que también utilizan activos virtuales.
Acompáñenos en la lectura de este trabajo que, por lo menos, lo introducirá en un tema de extrema actualidad.

Palabras clave: lavado de activos, blanqueo de capitales, debida diligencia, inteligencia financiera, GAFI, GAFILAT, sujetos obligados, reporte de operaciones sospechosas (ROS), criptoactivos, criptomonedas, non-fungible tokens(NFT), blockchain, hash

ABSTRACT

This paper deals with two fundamental elements in the disruptive ecosystem produced by the constant advancement of technology: money laundering and cryptoassets or virtual assets.
The law on the prevention of money laundering is a relatively new specialization that emerged in the 1980s and had as its fundamental antecedent the development of what would later be called predicate offenses, such as white slavery, the sale of weapons of mass destruction, drug trafficking, etc. Once this situation was confirmed, the FATF was created, which approved forty recommendations to combat these scourges.

The emergence of ICT, the dark web, among others, and the commercial adoption of blockchain, with the birth of the first cryptocurrency, led money launderers to begin using these types of assets to commit crimes, i.e., money laundering crimes that are no longer carried out only with physical assets, but also with virtual assets.
Join us in reading this paper, which will introduce you to a highly topical issue.

Keywords: Money Laundering, Due Diligence, Financial Intelligence, FATF, GAFILAT, Regulated Entities, Suspicious Transaction Reporting (STR), Cryptoassets, Cryptocurrencies, Non-Fungible Tokens (NFTs), Blockchain, Hash

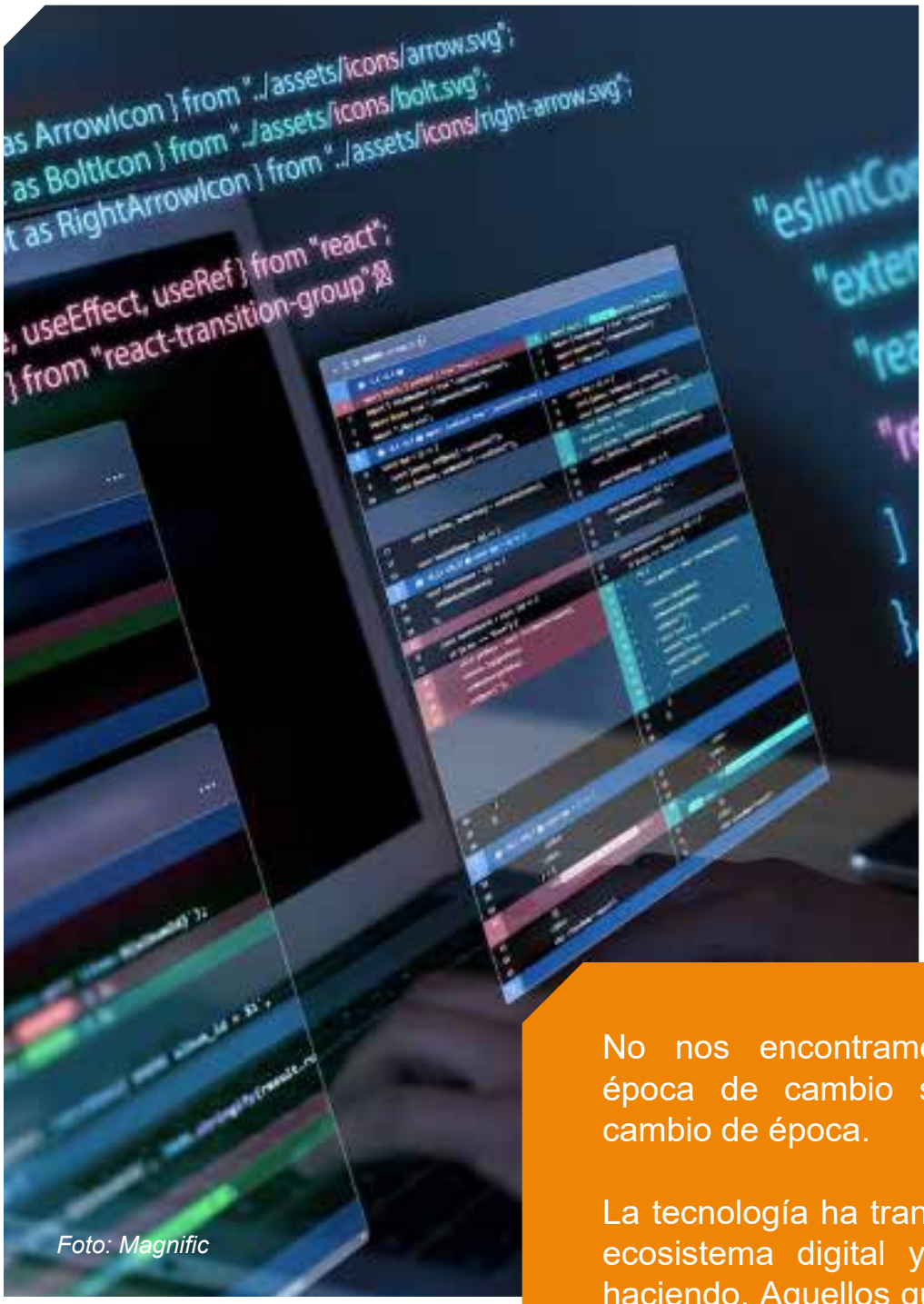


Foto: Magnific

No nos encontramos en una época de cambio sino en un cambio de época.

La tecnología ha transformado el ecosistema digital y lo seguirá haciendo. Aquellos que no sepan desaprender conceptos permitidos y aprender conceptos actuales perecerán como lo hicieron los dinosaurios.

Rubén Flores Dapkevicius

INTRODUCCIÓN

En primer lugar, corresponde agradecer a los responsables del Observatorio de Gobierno y TIC, de la prestigiosa Pontificia Universidad Javeriana, por la invitación cursada a efectos de ser uno de los autores del primer número de su revista.

De ingreso, recordamos que la expresión lavado de activos se remonta a la época de la prohibición en EE. UU., donde se dice que se efectuó tal tarea.
El derecho de la prevención de lavado de activos, el financiamiento del terrorismo y la proliferación de armas de destrucción masiva, es una especialización relativamente nueva y en continuo desarrollo, integrada, desde el punto de vista del derecho, por diversas ramas.[1]

La disrupción de un big data cada vez más potente, la generalización relativa pero existente respecto a los criptoactivos fungibles y no fungibles, su posible regulación por el Estado, la moneda digital del banco central (CBDC, por su sigla en inglés), las recomendaciones del Grupo de Acción Financiera Internacional (GAFI) y de los bancos centrales en ese ecosistema se encuentran creando algunos cambios en el paradigma.

En este momento, más que nunca, los negocios que se plantean en la vida son infinitos, tan diversos como diversa es la diversidad. Ello se aprecia con una simple lectura, al observar a todos los obligados, ya sean financieros o no financieros, determinados en las leyes del derecho comparado, por ejemplo, en la República de Colombia.[2]

Se puede blanquear capitales en una zona franca, una organización sin fines de lucro, así como, también, frente a un rematador diligente, un banco, un escribano, mediante la venta de un deportista, por medio de influencers, etc.

El derecho internacional público dio origen al sistema y lo impregna fuertemente, según sus características especiales. En ese sentido, el GAFI y los organismos regionales en la materia (GAFILAT) han sido, y son, determinantes.

Destacamos, de comienzo, que utilizaremos las siguientes expresiones como sinónimos: blanqueo de dinero,[3] blanqueo de capitales, lavado de dinero,[4] blanqueo de activos.

El tema que nos convoca comenzó a ser de especial atención, para los gobiernos y entidades internacionales, a principios de los años ochenta del siglo pasado. Es, por ello, algo reciente, pero no tanto. Surgió, como todo en sus comienzos, en la realidad que vivía la humanidad en ese momento concreto, especialmente, en torno al tráfico de estupefacientes.

Podría haber comenzado antes, tal vez, en la época de la ley seca, pero es indudable que esa situación no se desarrolló en un mundo globalizado y, por ello, podría involucrar a lo sumo a algunos países limítrofes y al contrabando, especialmente.

El hecho no era nuevo, pero se observó que se amasaron fortunas, surgidas también sobre la base del terrorismo, de la trata de blancas o de la prostitución, entre otros. Ese dinero ilegal, negro, debió comenzar a lavarse, o blanquearse, por los delincuentes, mediante la compra de propiedades inmuebles, obras de arte, metales preciosos, inversiones en deportes, etc.[5] Pero además se observó que esos capitales, en diversos países, podían comprar políticos corruptos, jueces, policías, quienes, tentados por comisiones de dinero, hicieran la vista gorda ante tales delitos, causa de esas fortunas. De allí surgió la actual necesidad de una debida diligencia ampliada, o especial, respecto a los mismos, ya que son personas expuestas políticamente (PEP).

1 Rubén Flores Dapkevicius, Prevención de lavado de activos, con activos físicos y criptoactivos, 3.a ed. (Montevideo: FCU, 2023).
2 Congreso de Colombia. Ley 1762 de 2015: Por medio de la cual se adoptan instrumentos para prevenir, controlar y sancionar el contrabando, el lavado de activos y la evasión fiscal (julio 6 de 2015). **3** El nombre es muy gráfico. Refiere a que los capitales negros, esto es, ilícitos, se convierten en blancos, se blanquean mediante un procedimiento que los convierte en legales. **4** Se aplica a este concepto lo anteriormente dicho. El dinero sucio que comenzó a higienizarse. **5** InfoLAFT, “Repasemos nuestra historia”.

El lavado de dinero es un proceso, una secuencia de actos que intenta disimular y, por ello, ocultar la fuente ilícita de gestación y traslado de los mismos. El primer elemento destacable es que se trata de una secuencia de actos y operaciones materiales. En ese sentido, es similar a un proceso administrativo o judicial, en tanto supone una serie de hechos y actos, según cada caso concreto, determinando un árbol algorítmico.[6] De esta forma, es fundamental para los delincuentes evitar que puedan ser rastreados, tanto respecto al origen del monto (terrorismo, extorsión, etc.) como frente a toda situación que permita identificarlos, a ellos y al capital, abarcando desde el lugar donde se gestó hasta cuál destino se le dará, esto es, dónde se intentará blanquear, de qué forma, etc.

Por otra parte, en el proceso, obvia y elementalmente, se intenta disociar a las personas que participan en cualquier etapa del mismo, desde la generación del activo hasta su lavado. En ese sentido, se utilizan diversos instrumentos, como pueden ser testaferros, fideicomisos, compra de sociedades anónimas, personas físicas o sociedades sin fines de lucro, entre otros.

Recordemos, entonces, que existen los delitos causa, o precedentes de las fortunas o capitales sucios, tales como el terrorismo, el secuestro, la extorsión, etc., que la leyes de derecho comparado tipifican, siguiendo las recomendaciones del GAFI. Este grupo se compone de 37 miembros, 35 países, más la Comisión Europea y el Consejo de Cooperación para los Estados Árabes del Golfo Pérsico. Se define como “una organización intergubernamental independiente que desarrolla y promueve políticas para proteger el sistema financiero mundial contra el blanqueo de capitales, la financiación del terrorismo y la financiación de la proliferación de armas de destrucción masiva”.[7] Esos delitos pueden a su vez dar lugar a otros, denominados genéricamente de lavado de activos, también denominados delitos consecuencia.

Así, en general, se tipifica el autolavado, estipulando que quien hubiere cometido alguna de las actividades delictivas señaladas como delitos precedentes puede estar
sujeto a investigación y juzgamiento, configuradas las circunstancias previstas en los delitos causa. También, las legislaciones tipifican el delito autónomo de lavado de acti

vos, el cual, como tal, no requiere un auto de procesamiento previo de las actividades delictivas establecidas como precedentes. Esto es claro: según el delito autónomo que recoge, simplemente, la realidad, no deben coincidir, necesariamente, los delincuentes que tipifican los delitos precedentes, o causa, con los de consecuencia.

El proceso, como se verá más profundamente infra, comienza con el análisis del cliente, por el sujeto obligado de primera línea en la inteligencia financiera, investigación que le concederá a la operación un riesgo alto, medio o bajo. Por ejemplo, una persona que trabaja como administrativo de jerarquía media, en una empresa de computación, no debería, en principio, comprar un campo de cien millones de dólares.

La debida diligencia, que es la evaluación del riesgo, puede clasificarse, según su intensidad, en simple, normal y ampliada. Las terminologías pueden variar según el derecho comparado, pero en sustancia son idénticas, por seguir las cuarenta recomendaciones del GAFI.[8] Estas recomendaciones no son obligatorias pero, por su peso político en el comercio internacional, se hacen ineludibles.[9]

La debida diligencia simple y la normal solo difieren en algún dato adicional, mínimo, que se debe completar. Lo mismo sucede con la

6 Rubén Flores Dapkevicius, Legaltech y derecho digital, tomo II (Buenos Aires: B de F Editores, 2024). **7** Alí José Daniels Pinto, ¿Qué es el GAFI y para qué sirve? (Venezuela: Acceso a la Justicia, 2023). **8** GAFISUD, Estándares internacionales sobre la lucha contra el lavado de activos y el financiamiento del terrorismo y la proliferación. Recomendaciones del GAFI (s. l.: GAFISUD, 2012). **9** FATF, International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation (París, Francia: FATF, 2025).

debida diligencia intensificada y con la denominada debida diligencia especial, que suponen protocolos especiales para determinadas personas, según su condición en el momento concreto. El ejemplo típico en este último caso son los PEP.

En general, la debida diligencia supone determinadas operaciones, que serán más intensas cuanto más riesgosa sea la operación.[10] Evidentemente esto implica realizar, como veremos, un estudio más o menos pormenorizado tanto del cliente (siendo diversa, en principio, esa operación si se trata de una persona física o jurídica) como del no cliente habitual. El análisis se realiza a esa operación en concreto, pero observando la operativa general del involucrado, en tanto el autorizante debe conocerlo, más acabadamente, al momento de autorizar la operación, sin emitir el reporte de operación sospechosa (ROS).[11]

En el análisis, el sujeto obligado, o su oficial de cumplimiento, debe observar listas que se mencionan infra, en torno al país de procedencia, la habitualidad de operaciones, los montos del negocio en sí mismo, etc.[12] Es el sujeto obligado por la ley el que, precisamente, debe cumplir la normativa respecto de la prevención de lavado de activos, el financiamiento del terrorismo y la proliferación de armas de destrucción masiva.

En la práctica, esa tarea la puede realizar el mismo sujeto, o su departamento de cumplimiento, jerarquizado en su gerente como oficial de cumplimiento. Esas empresas, también, pueden tercerizar la realización de la debida diligencia, en el caso de que lo consideren necesario. En este último supuesto, siempre se aconseja que el sujeto responsable, por lo menos, supervise la tarea.



Foto: Magnific

Las listas que se deben observar, especialmente en este momento, son entre otras:

- Lista unificada de la Organización de las Naciones Unidas (ONU) - Terrorismo (versión HTML).
- Personas expuestas políticamente (PEP).
- Países o territorios que aplican las recomendaciones GAFI.[13]

Ese proceso de estudio, según lo obliga la ley, debe ser más detallado cuanto mayor sea la peligrosidad de la operación.

10 Cumplo360Blog, “Qué es la Debida Diligencia en prevención de lavado de activos?” (blog). **11** El ROS se realiza en formularios a los que puede accederse fácilmente, según las agremiaciones, la actividad, las páginas oficiales, etc. **12** El Estado trasladó la primera línea de responsabilidad de la inteligencia financiera en los sujetos obligados. En el ecosistema cripto lo serán, especialmente, los proveedores de servicios de activos virtuales (VASP, por su sigla en inglés), como Binance. **13** Asociación de Escribanos del Uruguay, “Prevención de lavado de activos. Diligencia debida notarial”.

Por tanto, cada obligado, según el negocio que realice y la profesión que desarrolle, debe tener procedimientos[14] adecuados de detección de riesgo de blanqueo de capitales, financiamiento del terrorismo o de proliferación armas de destrucción masiva, que deben adecuarse a lo recientemente señalado: montos, negocios propuestos, ciudadanía de los involucrados, domicilios en el lugar de la transacción, estadísticas en general, etc.

Esos protocolos deben actualizarse continuamente, de conformidad con las capacitaciones que deben recibir los sujetos obligados y sus oficiales de cumplimiento, respecto a nuevos métodos y tecnologías para blanquear capitales, la identificación de terroristas, la valoración de las situaciones de sus clientes, etc. Ese conocimiento del cliente debe estar al día durante la celebración del negocio de que se trate.[15]

Todo comienza con la presentación del cliente (habitual u ocasional), quien manifiesta un interés por realizar una operación determinada, en el giro del sujeto obligado (contador público, corredor de bolsa, etc.) por las leyes que nos convocan.

Este sistema cuenta con un chatbot, esto es, un programa que intenta simular la comunicación humana.[16] Los chatbots podrán ser de simple recibimiento, actuarán dando la bienvenida o haciendo una serie de preguntas elementales, a las que también puede responder. Existen herramientas de este tipo más sofisticadas, que preparan la entrevista inicial, al punto de solicitar testigos, documentos, etc. De esta forma, por ejemplo, se interroga a la persona respecto de su nombre, su documento, su país de residencia, su fecha de nacimiento, el motivo de la consulta, sus ingresos, su estado civil, etc.

Una vez realizado este procedimiento inicial, corresponde el encuentro personal de estilo, en el que se evaluará, prima facie, entre otros elementos, el motivo de la operación.
La recepción en la oficina, o en un estudio profesional, es fundamental, y debe ser bien preparada. Existen tantas técnicas que no es posible describirlas, ni siquiera mencionarlas, en un mínimo aceptable, ya que dependen de cada negocio, profesión, profesional, tipo de cliente, etc.

Con todo, por ejemplo, se debe, elementalmente en primer lugar, hacer sentir cómodo al cliente. Sin perjuicio de ello, las preguntas deben ser bien dirigidas (existen múltiples formas de interrogatorio), cambiando repentinamente el tema, de lo accesorio a lo principal. Así, por ejemplo: "¿No le parece que hoy es un día excepcional para estar en Punta del Este?", a lo que el cliente responderá: "Sí", "No", "Iré pronto con mi familia", y se extenderá, o no, en el tema. La siguiente pregunta podría ser: "¿Cuánto hace que compró ese Ferrari, amigo mío, ya que la última vez que lo vi (hace seis meses, por ejemplo) no tenía ese auto que realmente es hermoso, más aún con el modelo y color que seleccionó. Imagino que a sus padres, especialmente a doña Susana, le debe encantar". La pregunta, así disimulada, va teledirigida.[17]

14 Protocolos o manuales de actuación, generales y particulares. **15** Seamos directos. Un cliente, hoy, puede ser un ciudadano ejemplar y, mañana, puede dedicarse al blanqueo de capitales. **16** Flores Dapkevicius, Legaltech y derecho digital.**17** Lo expuesto es sin perjuicio de que un hombre se conoce, no por lo que contesta (dado que podría estar mintiendo, además de que no muestra lo que sigue), sino por lo que pregunta. Allí debería estar su interés principal.

Quien sugiere, esto es, el que escribe en este momento sobre lavado de activos, le aconseja que la luz provenga de la espalda del profesional, usted mismo, sin encandilar al cliente. Destaco en este punto que he sido, más allá de asesor presidencial y parlamentario, jefe de la División Sumarios e Investigaciones Administrativas, así como de la Asesoría Letrada de un organismo estatal, descentralizado por servicios, entre otros. Esto me dio esa experiencia, que jamás se olvida. Esos cargos, o consultas, en alguna u otra medida, los sigo desempeñando en este momento, obviamente.

Si existen dudas, las entrevistas deben ser múltiples, por lo menos una más de las que la intuición sugiera, antes de tomar cualquier resolución fundamental, en virtud de la responsabilidad que asume el sujeto obligado, o su oficial de cumplimiento. La etapa inmediata, y siempre necesaria, es la verificación de la situación de la persona: su información de antecedentes penales y civiles, su lugar de residencia, etc., según sigue en este desarrollo.

> Debe verificarse su situación general en las bases que elaboran los organismos nacionales, como, por ejemplo, el Banco Central, el Ministerio del Interior, el Ministerio de Seguridad Pública Interna, el Ministerio de Defensa, etc. El chequeo de otras listas se impone, por la naturaleza de las cosas, más allá de lo que determine la ley, teniendo presente que la búsqueda en Internet es también válida.

Deben observarse los antecedentes actualizados del cliente ocasional, o habitual, en las listas de la ONU, del GAFI, del GAFILAT, de la Organización Internacional de Policía Criminal (INTERPOL), de la Oficina de Control de Activos Extranjeros (OFAC), del Buró Federal de Investigaciones (FBI, por su sigla en inglés), de la Administración para el Control de Drogas (DEA, por su sigla en inglés), de PEP, del Banco Central, etc.. También se debe observar si la persona no proviene de un país que esté en la lista roja (de existir), negra, gris oscura, etc., del GAFI.

Igualmente, se debe relevar, al detalle y de forma actualizada, si el cliente tiene antecedentes judiciales.[18] Además, se debe observar, siempre, quién es el beneficiario definitivo y en ese caso hay que analizarlo detenidamente. Una vez realizadas estas operaciones, el sujeto obligado, que conoce de antes al cliente, o en ese momento, se forma la idea general de qué tipo de diligencia debe emplear para evitar que, con conocimiento de causa, se realicen acciones de blanqueo o de financiamiento del terrorismo o de armas de destrucción masiva, con la responsabilidad administrativa que ello acarrearía en el ejercicio de su profesión, por lo menos. Esa idea general puede variar en la dinámica de la investigación, siempre, o casi siempre, para intensificar la debida diligencia.

Luego de ello, autorizará el negocio, esto es, efectuará la transacción o elaborará el ROS,[19] ya que todas las personas, sean físicas o jurídicas, controladas por las Unidades de Información y Análisis Financiero (UIAF), deben informar de las operaciones, realizadas o no, que en los usos y costumbres de la respectiva actividad resulten inusuales, se presenten sin justificación económica o legal evidente o se planteen con una complejidad inusitada o injustificada.[20]

18 Se impone que la verificación se realice en fuentes fidedignas. No podría ser de otra forma. **19** Siempre se podría realizar un ROS luego de finalizada la operación si se comprueba, a posteriori, la sospecha, pero esa situación puede generar responsabilidad. Por ello, ante la más pequeña duda, se recomienda no autorizar la operación y realizar el reporte para que intervenga el órgano competente. **20** UIAF, “Noticias y comunicados”, Gobierno de Colombia; UIAF, “Nuestra entidad”, Gobierno de Colombia;; Superintendencia Financiera de Colombia, “Inicio”, Gobierno de Colombia.



Foto: Magnific

La realización del reporte corresponde, especialmente, si el sujeto se niega en algún momento de la transacción a proporcionar alguna información; ni qué decir de si alguna de ellas es falsa.[21] La situación descrita también se aplica a los sujetos obligados no financieros.

El ROS se realiza en formularios que se descargan de Internet, en lugares como el Banco Central, y en los de las agremiaciones correspondientes a la profesión de que se trata, como la Asociación de Escribanos entre otras.

Una vez que se ha recibido el reporte, la UIAF que ejerce, o puede ejercer funciones en el sistema orgánico, como el Banco Central, puede instruir al sujeto obligado sobre qué debe hacer.

Es claro que, frente a alguna sospecha, se debe realizar y comunicar el reporte correspondiente, lo que significará hacer ingresar al personal especializado en inteligencia financiera que asumirá, obvia y elementalmente, la responsabilidad.

Los sujetos obligados, o sus oficiales de cumplimiento, son el primer eslabón en todo instrumento de control o cadena de prevención de las actividades dirigidas al lavado de activos, el financiamiento del terrorismo o de proliferación de armas de destrucción masiva. Nada más, nada menos.

21 La casuística, como siempre, es infinita.

En definitiva, los sujetos obligados (financieros o no financieros) deben definir e implementar políticas y procedimientos de debida diligencia para todos sus clientes[22], así como plantear operaciones (administración, enajenación, compraventa, permuta, etc.) que les permitan obtener una adecuada identificación y un conocimiento adecuado de los mismos —incluyendo el beneficiario final de las transacciones, si correspondiere—, prestando sobre todo atención al volumen y a la índole de los negocios u otras actividades que estos desarrollen.

En ningún caso, los sujetos obligados podrán mantener cuentas anónimas, o cuentas con nombres ficticios.[23] Los procedimientos de debida diligencia se deben aplicar a todos los clientes, especialmente a los nuevos, al establecer relaciones comerciales, o cuando se realicen transacciones ocasionales, por encima de los umbrales designados para cada actividad.

Los clientes ya conocidos, habituales, siempre deben ser actualizarlos en su conocimiento, ya que una excelente y respetable persona, hoy, puede prestarse por diversos motivos al lavado de activos, al financiamiento del terrorismo o a la proliferación de armas de destrucción masiva, mañana. Esa realidad es innegable.

Cuando existan sospechas de lavado de activos, de financiamiento del terrorismo o de proliferación de armas de destrucción masiva, o cuando el sujeto obligado tenga dudas sobre la veracidad o suficiencia de los datos de conocimiento del cliente obtenidos previamente, también se deberán aplicar los procedimientos que se señalan a continuación, con independencia de cualquier excepción, exención, o umbral económico establecido.[24]

Se entiende por beneficiario final a la persona física que, directa o indirectamente, posea como mínimo el 15 % del capital o de los derechos de voto, o su equivalente, o que por otros medios ejerza el control final sobre una entidad, considerándose como tal una persona jurídica, un fideicomiso, un fondo de inversión o cualquier otro patrimonio de afectación o estructura jurídica. Se considera también como beneficiario final a la persona física que aporta los fondos para realizar una operación, o en cuya representación se lleva a cabo una operación.[25]

Se entiende como control final el ejercido, directa o indirectamente, a través de una cadena de titularidad, o a través de cualquier otro medio de control, según la especialidad. Es su juicio el que decide, y por ello, en algunos casos se podrá equivocar el camino.

La situación se soluciona si, en el supuesto de duda o sospecha fundadas, se elabora y se comunica el ROS correspondiente al órgano competente del Estado, generalmente la UIAF. Desde ese momento comienza a actuar la segunda línea de combate contra el lavado de activos, dado que la inteligencia financiera de la UIAF es su especialidad.

En todos los casos, los sujetos obligados deben estar en condiciones de demostrar a las autoridades competentes que las medidas adoptadas tuvieron un alcance adecuado, en relación con el riesgo de los delitos financieros, según cada caso concreto.

Esa prueba del análisis de riesgo debe constar por escrito, o por medios idóneos (teniendo en su poder los documentos o datos digitales de búsquedas, etc.) que permitan la reconstrucción de la debida diligencia realizada, por lo que hay que ser exhaustivos en la descripción de cada elemento observado.

22 Conocidos y, especialmente, no conocidos. **23** Flores Dapkevicius, Prevención de lavado de activos…**24** La debida diligencia corresponde de acuerdo con determinados montos que la ley establece y que la autoridad administrativa actualiza. **25** La determinación del beneficiario final intenta eliminar la existencia de testaferros.

Los sujetos obligados no se encuentran autorizados a ejecutar operaciones cuando no puedan aplicar las medidas de debida diligencia previstas en las leyes. Cuando se aprecie esta imposibilidad en el curso de la relación de negocios, los sujetos obligados pondrán fin a la misma,[26] procediendo a considerar la pertinencia de realizar y comunicar un ROS.

Siempre aconsejamos que, si el sujeto obligado o su oficial de cumplimiento tienen alguna duda respecto de si deben realizar una debida diligencia más exhaustiva, esta se debe realizar. La prudencia y la razón imponen esa solución, a los efectos de prevenir una posible operación fraudulenta de dinero.

Asimismo, siempre es preferible, ante la duda, realizar el ROS y comunicarlo al órgano competente, aguardando luego hasta setenta y dos horas las respectivas sugerencias.

Es así que, con la finalidad de prevenir el lavado de activos, el financiamiento del terrorismo y la proliferación de armas de destrucción masiva, la UIAF recibe los ROS de parte de los sujetos obligados del sector financiero y no financiero.

Dicha información recibida debe ser complementada por este órgano con otro tipo de información obtenida de fuentes internas y externas confiables, según hemos destacado, con el objetivo de confirmar la existencia de elementos de inusualidad o de sospecha, remitiendo en ese caso a las autoridades competentes (Fiscalía o poder judicial) la información sobre transacciones financieras y otros datos que se estimen de utilidad, a efectos de impedir los delitos de lavado de activos.

Cuando no se encuentren indicios sobre la vinculación de las personas, hechos y/o transacciones con esos delitos, se dispone el archivo de las actuaciones, hasta tanto se obtengan nuevos elementos que ameriten continuar con el análisis del caso.

Se entiende por *inteligencia financiera al proceso*[27] a través del cual se agrega valor a los datos obtenidos de los sujetos obligados, o de sus oficiales de cumplimiento, que realizaron y comunicaron el ROS, y de otras fuentes, a fin de transformarlos en información útil para las autoridades competentes, con el objetivo de combatir tales delitos.

El contenido del ROS es el insumo que da inicio al proceso de inteligencia financiera. Por este motivo, es fundamental que los reportes recibidos posean determinadas características, de modo que el producto obtenido —el informe de inteligencia financiera— sea útil a sus destinatarios finales, según el caso concreto.

Como parte de la función de análisis estratégico que realiza la UIAF, con base en la información recibida, se han elaborado una serie de pautas, guías o sugerencias, para la elaboración de los ROS.

26 En este momento de crisis, igual debe procederse de esa manera y, yo diría, con mayor precaución, inclusive respecto a todos los clientes habituales, o no.
27 Todo proceso, o procedimiento, es una secuencia de actos y de operaciones, en la que una sigue a la otra, y así sucesivamente.

El objetivo de ese informe es aportar una retroalimentación a los sujetos obligados en relación con el proceso de elaboración y de comunicación de un ROS, lo cual favorecerá un funcionamiento más articulado del sistema de prevención, en particular entre las etapas de identificación (por parte de los sujetos obligados), de investigación y de análisis (por parte de la UIAF), y de resolución (por parte del poder judicial).

Destacamos las tres etapas:

1. Prevención, a cargo de los sujetos obligados (financieros o no financieros determinados por la ley) o de sus oficiales de cumplimiento.
2. Investigación y análisis, que corresponde a la UIAF.
3. Resolución, en su caso por el poder judicial, que también integra el sistema de inteligencia financiera, según ciertas hipótesis, con los fiscales competentes.

Estas etapas son las determinantes en el sistema de inteligencia financiera, respecto a la prevención del lavado de activos.


Foto: Magnific

En definitiva, se sugiere a los sujetos obligados o a sus oficiales de cumplimiento siempre esperar, por lo menos setenta y dos horas luego de realizado y comunicado el ROS, antes de tomar cualquier resolución sobre el negocio en el que se esté actuando.

Si la UIAF no dispone medidas, ni las comunica, la responsabilidad se evaluará según ese caso, y no por una actuación negligente a la hora de cerrar un negocio por el sujeto obligado, o por su oficial de cumplimiento.[28]

01 DEBIDA DILIGENCIA COMÚN

Los sujetos obligados, financieros o no financieros, deben definir e implementar políticas y procedimientos de debida diligencia para todos sus clientes, habituales u ocasionales, con el fin de conocerlos en el momento del negocio[29] y a efectos de obtener una adecuada identificación y determinación de los mismos,[30] teniendo presente el volumen y la naturaleza de los negocios u otras actividades que estos desarrollen.

La debida diligencia se debe aplicar a todos los nuevos clientes, al establecer relaciones comerciales, o cuando se realicen transacciones ocasionales por encima de los umbrales designados para cada actividad, según hemos visto.[31]

28 La sugerencia emerge también de la información que da el Banco Central de Uruguay, “Buenas prácticas para el reporte de operaciones sospechosas”, Secretaría Nacional para la Lucha contra el Lavado de Activos y el Financiamiento del Terrorismo, octubre 31 de 2016. **29** Es decir que siempre debe actualizar sus búsquedas en las listas, Internet, etc. **30** Incluyendo el beneficiario final de las transacciones, si correspondiere. **31** Flores Dapkevicius, Legaltech y derecho digital.

Cuando existan sospechas de lavado de activos, financiamiento del terrorismo o de proliferación de armas de destrucción masiva, o cuando el sujeto obligado o su oficial de cumplimiento tengan dudas sobre la veracidad o suficiencia de los datos de conocimiento del cliente obtenidos previamente, también deben aplicar la debida diligencia que, en lo sustancial, se desarrolla según lo ya expuesto.[32]

Si en el desarrollo de la operación se observa que no se podrá cumplir con la obligación que se impone, debe ponerse fin a la misma, evaluando la posibilidad de realizar un ROS a la UIAF.[33]

02 DEBIDA DILIGENCIA SIMPLICADA

La aplicación de medidas simplificadas de debida diligencia será graduada en función del riesgo detectado, el scoring o puntaje, con arreglo a los siguientes criterios:

A. Con carácter previo a la aplicación de medidas simplificadas de debida diligencia respecto de un determinado cliente, producto u operación, los sujetos obligados, o sus oficiales de cumplimiento, deben observar que la situación constituye efectivamente un riesgo reducido de lavado de activos, financiamiento del terrorismo o proliferación de armas de destrucción masiva.[34] Es decir que el proceso de debida diligencia es evolutivo y puede irse incrementando a medida que el scoring se hace más elevado. La mención al carácter previo es diáfana, contundente y terminante.

B. La aplicación de las medidas simplificadas de debida diligencia debe ser siempre congruente con el riesgo detectado, según el scoring que se le atribuye al cliente. Los sujetos obligados, o sus oficiales de cumplimento, no aplicarán o cesarán de aplicar medidas simplificadas de debida diligencia, tan pronto como aprecien que un cliente, producto u operación no comporta riesgos reducidos de lavado de activos, financiamiento del terrorismo o proliferación de armas de destrucción masiva.

C. Sin perjuicio de lo expuesto, los sujetos obligados, o sus oficiales de cumplimiento, deben mantener en todo caso un seguimiento continuo y suficiente, para detectar operaciones a las que deben aplicar la debida diligencia más intensificada. Si la operación se realiza desde determinados instrumentos, es suficiente la debida diligencia simplificada. Entre tales instrumentos no taxativos destacamos los medios de pago electrónicos, las transferencias bancarias u otros elementos de pago emitidos por instituciones de intermediación financiera, o de los que estas fueran obligadas al pago, o valores de los que estas fueran depositarias.

La debida diligencia cuando se usan esos instrumentos surge por la consideración al menor riesgo de lavado de activos o financiamiento del terrorismo que esos casos suponen, dado que el capital ya fue bancarizado. Lo expuesto anteriormente no resulta aplicable cuando se trate de las PEP.

32 Pérez del Castillo & Asociados, “Normativa contra el Lavado de Activos”, noviembre 22 de 2018. **33** Flores Dapkevicius, Prevención de lavado de activos…**34** No olvidemos que los sujetos obligados realizan en su giro diversas actividades, en la cuales deben, o no, realizar una debida diligencia. Por ejemplo, en Uruguay se debe realizar debida diligencia respecto de la compraventa de inmuebles, pero no corresponde cuando se tratan de permutas. Ello es fundamental para los efectos de la adquisición de inmuebles que se pagan con criptoactivos, porque estos últimos se consideran cosas y no dinero. Por ello, la adquisición de un inmueble con Bitcoin, Tether, etc., no requiere debida diligencia.

En definitiva, los criterios a efecto son:

A. Los sujetos obligados deben verificar que existe efectivamente un riesgo reducido de lavado de activos, financiamiento del terrorismo o proliferación de armas de destrucción masiva.
B. Los sujetos obligados, o sus oficiales de cumplimiento, no aplicarán o cesarán de aplicar las medidas simplificadas de debida diligencia, tan pronto como aprecien que un cliente, producto u operación no comporta riesgos reducidos de lavado de activos, financiamiento del terrorismo y/o proliferación de armas de destrucción masiva.
C. En todo caso, los sujetos obligados y, eventualmente, los oficiales de cumplimiento, deben mantener un seguimiento suficiente y continuo para detectar operaciones susceptibles de un examen especial.

En la aplicación de medidas de debida diligencia simplificada, de acuerdo con lo previsto en el capítulo correspondiente a cada sector (es decir, a la profesión del sujeto obligado), la misma se encuentra supeditada al cumplimiento de los siguientes criterios:

A. Que el cliente, producto u operación comporte, efectivamente, un riesgo reducido de lavado de activos, financiamiento del terrorismo o proliferación de armas de destrucción masiva, de acuerdo con la evaluación de riesgos, scoring o puntaje realizada por el sujeto obligado o su oficial de cumplimiento desde la normativa que se analiza.
B. No se aplicará, o cesará su aplicación, cuando el sujeto obligado, o su oficial de cumplimiento, tome conocimiento de que el cliente, producto u operación ya no comporta riesgos reducidos de lavado de activos, financiamiento del terrorismo o proliferación de armas de destrucción masiva. Ello es evidente.

C. Sin perjuicio de lo señalado en los literales a) y b), cuando la relación de negocios con el cliente sea de carácter permanente, se deberá mantener, en todo caso, un seguimiento continuo y suficiente,[35] para monitorear y verificar las listas de las resoluciones del Consejo de Seguridad de Naciones Unidas sobre sanciones financieras relativas a la prevención y a la represión del terrorismo, a su financiamiento y a la prevención, supresión e interrupción de la proliferación de armas de destrucción masiva.

03 DILIGENCIA INTENSIFICADA O AMPLIADA

Tanto la normativa como la forma de mención por los operadores, en general y por principio, toman los términos como similares y, por tanto, se mencionan indistintamente. Los sujetos obligados deben intensificar el procedimiento de debida diligencia para las categorías de clientes, relaciones comerciales, u operaciones de mayor riesgo, como, por ejemplo, los clientes no residentes[36] o las operaciones que no impliquen la presencia física de las partes. Hay que prestar atención a las amenazas que puedan surgir de la utilización de tecnologías nuevas, o en desarrollo, que favorezcan el anonimato en las transacciones[37] y, en general, a todas aquellas operaciones que presenten características de riesgo o señales de alerta.

35 La ley contiene diversos conceptos jurídicos determinables. Ello puede generar, y genera, cierta incertidumbre a los sujetos obligados y/o a sus oficiales de cumplimiento. Esta es una realidad que no podemos omitir afirmar. **36** Especialmente los que provengan de países que no cumplen con los estándares internacionales en materia de lavado de activos y de financiamiento del terrorismo. **37** Blockchain, criptomonedas como la Bitcoin.

De acuerdo con las cuarenta recomendaciones del GAFI, la normativa determina que se deben tomar medidas de debida diligencia intensificada, o ampliada, según lo previsto en el capítulo correspondiente a cada sector, para las categorías de clientes, relaciones comerciales u operaciones de mayor riesgo:

A. Relaciones comerciales y operaciones con clientes, no residentes, que provengan de países que no son miembros del GAFI o de alguno de los grupos regionales de similar naturaleza, como el GAFILAT, el Grupo de Acción Financiera del Caribe (GAFIC), el Grupo de prevención del blanqueo de capitales de África del Sur y del Este (MENAFATF) y el Grupo Asia/Pacífico en materia de blanqueo de capitales (APG), o de países que estén siendo objeto de medidas especiales por parte de estos grupos, por no aplicar las recomendaciones del GAFI, o por no aplicarlas suficientemente, según su criterio.[38]
B. Relaciones comerciales y operaciones con clientes, no residentes, que provengan de países sujetos a sanciones o contramedidas financieras emitidas por organismos como el Consejo de Seguridad de las Naciones Unidas.
C. Relaciones comerciales y operaciones con personas físicas o jurídicas, residentes, domiciliadas, constituidas o ubicadas en países, jurisdicciones o regímenes especiales, de baja o nula tributación, de acuerdo con las listas que pueden emitir las entidades recaudadoras de impuestos.
D. Operaciones que no impliquen la presencia física de las partes, o de quienes los representen.
E. Utilización de tecnologías nuevas, o en desarrollo, que favorezcan el anonimato en las transacciones.

Es aquí donde se aprecia la necesidad de este tipo de diligencia respecto a los criptoactivos, ya que utilizan este tipo de tecnologías.

F. PEP, cónyuges, concubinos y parientes por consanguinidad, o afinidad, hasta el segundo grado, así como asociados cercanos, cuando estos sean de público conocimiento, y quienes realicen operaciones en su nombre. Es decir que se observa todo el círculo de relaciones de las
G. PEP, en lo racionalmente posible. Negocios en que se utilizan cuantías elevadas de efectivo.
H. Personas jurídicas con acciones al portador, en caso de que existan dificultades para identificar el beneficiario final a través de información incluida en un registro oficial de las autoridades públicas llevadas a los efectos.
I. Fideicomisos cuya constitución aparente (siempre vivimos en un mundo de mayores o menores apariencias) sea inusual, o excesivamente compleja (lo que manifiesta, siempre, cierta desconfianza), para determinar su estructura de control y sus beneficiarios finales.
J. Relaciones comerciales que se realizan en circunstancias inusuales, conforme a los usos y costumbres de la respectiva actividad.
K. Otras situaciones que, conforme al análisis de riesgos elaborado por el sujeto obligado (es su decisión), resulten ser de mayor riesgo y, por tanto, requieran de la aplicación de medidas de debida diligencia intensificada.

38 Este es otro concepto jurídico determinable.

04 DEBIDA DILIGENCIA ESPECIAL

Es un tema de interés del GAFI. Se deben establecer procedimientos especiales de debida diligencia para:

A. Las PEP, así como las relaciones con estos, sus familiares y asociados cercanos.
B. Las personas jurídicas, en especial, las sociedades con acciones al portador.
C. Los fideicomisos, para determinar su estructura de control y sus beneficiarios finales.
D. Se entiende por PEP a aquellas personas que desempeñan, o han desempeñado en los últimos cinco años, funciones públicas de importancia en el país, o en el extranjero, entre ellas:[39] jefes de Estado o de Gobierno, políticos de jerarquía, funcionarios gubernamentales, judiciales o militares de alta jerarquía, dirigentes destacados de partidos políticos, directores y altos ejecutivos de empresas estatales y otras entidades públicas. Los sujetos enunciados pueden ser interpretados diversamente en los diferentes países. Además esa diversidad, claramente, puede surgir en los funcionarios públicos actuantes, sujetos obligados, oficiales de cumplimiento, jurisprudencia y doctrina. En algunos casos podríamos decir que son conceptos jurídicos determinables, como alta jerarquía, según se observa seguidamente.

Se entiende como PEP a aquellas personas que desempeñan, o han desempeñado en los últimos cinco años contados desde el cese del cargo, una función de jerarquía en un organismo internacional de derecho internacional público, entre ellas: miembros de la alta gerencia, directores, subdirectores, miembros de la junta, o con funciones equivalentes.

En todos los casos se aplica la debida diligencia intensificada para las PEP, sus cónyuges, concubinos y parientes por consanguinidad o afinidad hasta el segundo grado, así como para los asociados cercanos, cuando estos sean de público conocimiento, y para quienes realicen operaciones en su nombre.

Es decir, no solo son PEP los soportes políticos en su ejercicio, sino también durante un periodo luego de su alejamiento del cargo, así como, naturalmente, algunos familiares y personas cercanas o representantes, ocasionales o habituales.


Foto: Magnific

39 La enunciación es no taxativa.

Con ello se intenta limitar a los testaferros en general, lo cual es, en la práctica, muy difícil. Con respecto a las PEP, los sujetos obligados deben, entonces:

A. Adoptar procedimientos y protocolos adecuados para determinar si el cliente o el beneficiario final clasifica como PEP.
B. Obtener la aprobación de la alta gerencia para establecer (o continuar, en el caso de los clientes existentes) dichas relaciones comerciales.
C. Tomar medidas razonables para establecer el origen de los fondos.

En definitiva, lo que se aplicará en este caso es un procedimiento o protocolo especial; nada más, nada menos.

Las situaciones son taxativas, por ser una excepción al principio de libertad, porque su excepcionalidad lo es respecto de un principio general y, por tanto, debe entenderse en forma estricta, sin admitir extensiones analógicas.[40]

En todo caso, debe buscarse la situación en las listas mundiales de PEP de instituciones como la OFAC, la ONU, la UE, etc. Asimismo, se debe proceder en las listas de interés como las de crimen financiero, tributario, y en otras listas oficiales como la de Interpol, FBI, según lo expuesto precedentemente.

05 RECOMENDACIONES DEL GAFI RESPECTO A LOS ACTIVOS VIRTUALES

Más allá de las cuarenta recomendaciones originales y actualizadas respecto a los activos físicos, el GAFI también ha dictado recomendaciones respecto a los activos virtuales.[41]

Para la entidad, un activo virtual es una representación digital de valor que se puede comercializar o transferir digitalmente y que se puede utilizar para pagos o inversiones. Se agrega ahorro y especulación, temas que, elementalmente, no interesan especialmente a la organización intergubernamental.

Los activos virtuales no incluyen representaciones digitales de moneda fiat, valores y otros activos financieros que ya están cubiertos en otras partes de las recomendaciones del GAFI.[42]

40 Flores Dapkevicius, Prevención de lavado de activos… **41** BDA Abogados, "Recomendaciones del GAFI para regular actividades con Activos Virtuales", enero 6 de 2025. **42** https://www.gafilat.org/index.php/es/biblioteca-virtual/gafilat/documentos-de-interes-17/publicaciones-web/4013-recomendaciones-metodologia-actdic20-1/file.



Foto: Magnific

Por su parte, se considera proveedor de servicios de activos virtuales (Virtual Asset Services Providers [VASP]) a cualquier persona física o jurídica que no esté cubierta en ningún otro lugar en virtud de las recomendaciones y que, como negocio, realiza una o más de las siguientes actividades u operaciones para o en nombre de otra persona física o jurídica:

- Intercambio entre activos virtuales y monedas fiat
- Intercambio entre una o más formas de activos virtuales
- Transferencia de activos virtuales
- Custodia y/o administración de activos virtuales o de instrumentos que permitan el control sobre estos
- Participación y provisión de servicios financieros relacionados con la oferta de un emisor y/o venta de un activo virtual.
- El grupo solicita centrarse en la regulación, no en la adopción masiva de criptoactivos.

A raíz de los reportes de los países que integran el GAFI sobre casos en que los delincuentes han usado transacciones con dinero virtual para cometer delitos que van desde el hurto y la evasión al fisco hasta la extorsión, se han emitido recomendaciones para prevenirlos, así como actualizaciones según la evolución de la tecnología y del ecosistema de que se trata.

El grupo entiende y recomienda a los países y a las instituciones financieras la identificación y evaluación de riesgos de lavado de activos o financiamiento del terrorismo que pudieran surgir con respecto al desarrollo de nuevos productos y nuevas prácticas comerciales, incluyendo nuevos mecanismos de envío, como frente al uso de tecnologías nuevas o en desarrollo para productos nuevos o existentes. Analizamos este tema por ser definitorio respecto a la debida diligencia a realizar, si en la transacción de que se trate se utilizan nuevas tecnologías que intenten garantizar el anonimato, especialmente las blockchains descentralizadas y las wallets mixer.[43]

En las instituciones financieras, esta evaluación del riesgo debe hacerse antes del lanzamiento de los nuevos productos, prácticas comerciales o uso de tecnologías nuevas o en desarrollo. Tanto los países como las instituciones financieras deben tomar medidas apropiadas para administrar y mitigar esos riesgos.

En ese sentido, respecto de los activos virtuales, los países deben garantizar que los VASP estén regulados, que tengan licencia o registro y que estén sujetos a sistemas de monitoreo efectivo para observar el cumplimiento de las medidas relevantes en las recomendaciones del GAFI.

Los Estados deben considerar los activos virtuales como bienes, productos, fondos, fondos y otros activos de valor equivalente, por lo que deben aplicar las medidas pertinentes en virtud de las recomendaciones del GAFI respecto al tema. Al igual que en el supuesto de los activos del mundo no virtual, se recomienda el traslado de la responsabilidad del Estado a los sujetos obligados.

Los países deben identificar, evaluar y comprender los riesgos de lavado de activos y financiamiento del terrorismo que surgen de las actividades de activos virtuales y las actividades u operaciones de los VASP.

Sobre la base de esa evaluación, se debe aplicar un enfoque basado en el riesgo para garantizar que las medidas tomadas para prevenir o mitigar el lavado de activos y el financiamiento del terrorismo sean proporcionales a los riesgos identificados. También se debe determinar la obligación de que los VASP identifiquen, evalúen y tomen medidas eficaces para mitigar sus riesgos de

43 Flores Dapkevicius, Prevención de lavado de activos…

lavado de activos y financiamiento del terrorismo. Como mínimo, se les debe exigir que tengan licencia o se registren en la jurisdicción o jurisdicciones donde se crean.

Las autoridades competentes deben tomar las medidas legales o reglamentarias necesarias para impedir que los delincuentes o sus asociados posean, sean beneficiarios finales, tengan una participación mayoritaria o significativa u ocupen una función de gestión determinante en un VASP.

Los países deben tomar medidas para identificar a las personas físicas o jurídicas que lleven a cabo actividades de VASP sin la licencia o el registro necesarios, y aplicar las sanciones apropiadas. No se define si estas son administrativas o privativas de libertad, dejando por tanto la posibilidad de determinar delitos específicos a cada Estado.

Los Estados deben garantizar que los VASP estén sujetos a una reglamentación y supervisión o monitoreo adecuados por una autoridad competente y autónoma.

Los supervisores deben tener poderes adecuados para llevar a cabo su tarea y aplicar sanciones disciplinarias, financieras y/o penales proporcionales, incluida la facultad de retirar, restringir o suspender la licencia o el registro del VASP, evidentemente, previo derecho de defensa efectiva. Estas medidas deben poder aplicarse no solo a los VASP, sino también a sus directores y directivos jerárquicos. Las medidas preventivas de los requisitos establecidos en las recomendaciones 10 a 21 se aplican a los VASP según una calificación del umbral de 1000 dólares o euros.

Los Estados deben asegurarse que los VASP de origen obtengan y mantengan la información obligatoria y precisa del originante, así como la información obligatoria del beneficiario sobre las transferencias de activos virtuales. Las mismas obligaciones se aplican a las instituciones financieras al enviar o recibir transferencias de activos virtuales en nombre de un cliente. La situación es entonces similar a la referida a la del mundo físico. Tanto el seguimiento de la disponibilidad de información como la adopción de medidas de congelamiento y la prohibición de las transacciones con personas y entidades determinadas deberían aplicarse sobre la misma base que se establece en la legislación sobre activos tangibles.

La cooperación internacional en materia de activos virtuales remite a las recomendaciones 37 a 40. Los supervisores de los VASP (SENACLAFT o UIAF) deben intercambiar información con prontitud con sus homólogos extranjeros, independientemente de las diferencias en la nomenclatura o el estado de aquellos.

El artículo 19 de la Ley sobre Prevención de Lavado de Activos No. 19574 de 2017 de Uruguay indica la obligación de realizar una debida diligencia, en ciertas hipótesis y sin perjuicio de otras que surgen de la legislación o de los hechos. En ese sentido, en la aplicación de un enfoque de riesgos los sujetos obligados deberán intensificar el procedimiento de debida diligencia para las catego-



Foto: Magnific

categorías de clientes, relaciones comerciales u operaciones de mayor riesgo, tales como los clientes no residentes, especialmente los que provengan de países que no cumplen con los estándares internacionales en materia de lavado de activos y financiamiento del terrorismo, o las operaciones que no impliquen la presencia física de las partes, prestando atención a las amenazas que puedan surgir de la utilización de tecnologías nuevas o en desarrollo que favorezcan el anonimato en las transacciones (en este momento se define el tema de los criptoactivos), y en general a todas aquellas operaciones que presenten características de riesgo o señales de alerta, según lo que determine la reglamentación.[44] Aquí encontramos la solución respecto al tipo de debida diligencia a realizar si los negocios involucran criptoactivos.

En definitiva, y recapitulando, el GAFI no ha podido demostrar que su normativa se adapta a un ecosistema que se basa en la intimidad de los usuarios. El grupo, dedicado a la lucha contra el blanqueo de capitales y el financiamiento del terrorismo a escala global, espera que la mayoría de sus países miembros puedan implementar las regulaciones que buscan rastrear las transacciones con criptomonedas.

Entre sus lineamientos, el GAFI recomienda cumplir con la denominada regla de viaje, una normativa que deben adoptar las exchanges para intercambiar información sobre aquellos usuarios que realicen operaciones que superen los 1000 dólares. La data de los usuarios debe ser compartida por los operadores de servicios, como establece la regla. El objetivo del organismo es que los proveedores de servicio del sector de las criptomonedas sepan, en todo momento, de dónde provienen y hacia dónde van las transacciones que se realizan.

Es un hecho complejo al que nos referimos, porque los usuarios buscan la intimidad, que logran casi generalmente según lo analizado respecto a temas como las wallets y la blockchain descentralizada.

El GAFI recomienda la vigilancia sobre los criptoactivos, porque para el organismo existe un elevado nivel de riesgo en las distintas formas de intercambio entre criptomonedas y dinero fiduciario o de curso legal.

44 Congreso de la República de Uruguay, Ley n.° 19574. Ley integral contra el lavado de activos. Actualización de la normativa vigente. Derogación de artículos del decreto Ley 14.294 y Leyes 17.835, 18.494, 18.914 y 19.149 (diciembre 20 del 2017).

Partiendo de la naturaleza de las transacciones con dinero virtual, el GAFI recomienda, por ejemplo, comprobar la información aportada por el cliente relativa a su identidad, contrastando con bases de datos de terceros, así como identificar la dirección IP del cliente y buscar cualquier otra información complementaria que permita corroborar la coherencia entre la actividad del cliente y el perfil de sus transacciones.

Un punto en el que se enfoca también el GAFI es asegurar la obtención de información sobre el origen de los fondos, la motivación de las transacciones previstas o realizadas y la realización de un seguimiento continuado de la relación de negocio para establecer incluso un umbral cuantitativo, a partir del cual los proveedores deberán aplicar medidas de diligencia debida para todo tipo de transacciones con activos virtuales, incluyendo transacciones ocasionales por encima de 1000 euros o dólares. Otra de sus recomendaciones es que los que los países creen políticas a corto y largo plazo, debido a que el sector de los activos virtuales se encuentra en continua evolución.

Con respecto a los VASP, que deben estar correctamente supervisados por órganos independientes, por lo menos técnicamente, para el GAFI es necesario que se identifiquen, evalúen y adopten medidas efectivas para mitigar los riesgos asociados con el suministro o la participación en actividades relacionadas con activos virtuales, o con la oferta de productos o servicios asociados a los mismos.

Las recomendaciones fueron actualizadas en octubre de 2025 y se encuentran citadas supra. La actualización intenta abarcar el tema actual y que todos los países del mundo observen la más creciente posibilidad de blanqueo de capitales mediante los criptoactivos.

La actualización se detiene en las P2P (transacciones entre iguales, sin intervención de intermediarios), las DeFis (plataformas financieras descentralizadas) y los Non Fungibles Tokens (NFT), así como también en las monedas estabilizadas o stablecoins,[45] entre otros.

La línea continúa en la necesidad de que sean los intermediarios, especialmente los VASP, los que realicen las tareas de debida diligencia, de la misma forma que en el mundo físico la realizan los sujetos obligados. Ello sin perjuicio de los diferentes activos que relacionamos.

Respecto a las monedas estabilizadas, señala que debe seguirse el mismo sistema aplicable a las monedas del mundo físico. La regulación debe ser igual o muy similar.[46] La utilización de las P2P en las DeFi, esto es, finanzas descentralizadas en la cadena de bloques con wallets mixer, criptomonedas anónimas y uso de la dark web, son elementos que el GAFI observa con mucha preocupación.

El tema de las wallets, con servicios garantidos de intimidad, es una cuestión que deberá observarse, así como las plataformas de navegación anónima y otros elementos como TOR o VPN.

En ese sentido, el exchanger de criptomonedas Buda.com comenzó a solicitar, hace años, una foto del rostro a sus usuarios, cada vez que deseen realizar un retiro hacia una dirección de Bitcoin que no se haya utilizado previamente. La página envió una notificación a sus usuarios, indicando que el día 30 de julio de 2021 actualizaría sus medidas de seguridad. [47]

45 Europa estudia prohibir las monedas estabilizadas en varias jurisdicciones, véase Catalina Elidafamo, “Europa estudia prohibir stablecoins emitidas en varias jurisdicciones”, Observatorio Blockhain, octubre 1 de 2025. **46** Fernando Gutiérrez, “GAFI actualiza visión regulatoria sobre criptomonedas; pide especial atención a transacciones peer to peer”, El Economista, octubre 28 de 2021.**47** Luis Esparragoza, “Exchange Buda solicita selfie para retirar bitcoin a las direcciones nuevas”, Criptonoticias, junio 4 de 2021.

Recordamos que existen tanto billeteras de hardware (hardware wallets) como billeteras de papel (paper wallets), de computadora (desktop wallets), para teléfonos celulares (mobile wallets) u online (web wallets),[48] así como las que derivan en hot, cold y mixer, centralizadas y descentralizadas. Las wallet mixer, que precisamente mezclan los criptoactivos para transformarlos en menos trazables, son más o menos íntimas si el proveedor no requiere datos y los criptos mezclables son muchos. En ese sentido, los sujetos obligados siempre deben poder realizar las consultas pertinentes respecto de los clientes que se presentan para efectuar un negocio.

Sabemos que la cadena de bloques es una fuente pública de información, como cualquier base de datos, y por tanto puede ser relevada. Por ejemplo, si se manifiesta que las ganancias o ingresos provienen de acciones de minería,[49] obtendremos la calidad de haber resuelto el tema de validación en la blockchain. Sabemos que, en muchos casos, los mineros actúan en grupo. En ese caso se observará la transferencia a su wallet de las criptomonedas generadas en el consorcio (pool) minero. Es de destacar que, en principio, los criptos obtenidos por los mineros no necesitan de debida diligencia en lo que refiere a la prevención del lavado de activos, porque provienen de su trabajo como tales.

Evidentemente, si existe cambio de criptomoneda a moneda fiat, la situación es más clara, y la debida diligencia parte operando en la forma de estilo. Sin embargo, reiteramos que las P2P, en la dark web y/o en las DeFi, con las wallet mixer, las criptomonedas anónimas, y el uso de más de una dirección pública, pueden hacer muy difícil seguir el trazado o camino de viaje de los activos.

En conclusión, en el derecho comparado corresponde observar en el trazado, en el camino o en la trayectoria de los activos,[50] la debida diligencia intensificada, si existen criptoactivos, por utilizar generalmente tecnología novedosa y por accionar en un ecosistema ciber punk o cripto punk, en el que se persigue la anonimación de las personas que intervienen.

06
PREVENCIÓN DE LAVADO DE ACTIVOS CON CRIPTOACTIVOS

Sin perjuicio de lo ya analizado cuando observamos las recomendaciones del GAFI sobre el posible lavado de activos mediante criptoactivos, es un hecho que la tecnología blockchain ayuda, en cierta forma, a evitar el lavado de activos, con mayor simplicidad o complejidad, eficiencia o ineficiencia, si se adecua precisamente o no.

48 Santiago Juarros, “Qué tipos de wallets de criptomonedas existen”, Ripio, septiembre 12 de 2024. **49** En el protocolo de trabajo.**50** La conocida regla de viaje.

Es muy claro que la tecnología permite el seguimiento, según la característica de cada blockchain, el tipo de criptomoneda más o menos anónima, los seudónimos, las diferentes claves públicas para operar, el uso de mixers, e incluso los tipos de navegadores y browsers, como TOR, VPN, etc. Además, podrían existir negocios de lavado de capitales en la dark web.

En los casos más sencillos, podemos transitar en los registros públicos de la cadena de bloques de la criptomoneda que se trate. Sabemos que en la blockchain existen diversos tipos de usuarios. El supuesto más importante se da en la minería, según observamos en el protocolo de trabajo que permite resolver el hash, en el que el minero obtiene criptomonedas por su trabajo en la cadena. Por tanto, el minero individual (o el pool) tuvo que observar los hash y el número de transacción; así también, en el mejor de los casos, veremos la dirección pública directa y sin seudónimos, esto es, en qué wallet se acreditaron los montos. Observaremos, en general, quién autorizó o resolvió el hash, y por tanto, quién accedió a la recompensa correspondiente.

En el protocolo de trabajo, así como en el de participación, no debe realizarse debida diligencia, porque en el primero los criptos son generados por el trabajo y en el segundo hay una participación en la blockchain y en los activos virtuales que se poseen para ello.

Reiteramos que esto se da en el mejor de los casos; sin embargo, las mixer de calidad dificultan la situación, como también ciertas criptomonedas y navegadores de la máxima intimidad, como TOR, junto con las VPN. En todo caso el usuario puede, en principio, ser rastreado. Así, es sabido que el FBI crea lugares en la dark web para que, si se ingresa en ellos, el usuario pueda ser localizado.

Si la situación no refiere a la minería, sino a la compraventa de criptoactivos (trading),[51] cuando una persona va a realizar una operación controlable según debida diligencia, debe declarar al sujeto obligado o a su oficial de cumplimiento, como en el caso anterior, toda la información necesaria que, también como en el caso de los mineros, se podrá rastrear con mayor o menor dificultad.



La debida diligencia debe comenzar en centrar el origen de los fondos negros que se intentaron blanquear, o de limpios de origen, esto es, los que no fueron lavados porque eran legales cuando se compraron, tales como Bitcoin.

Esa debida diligencia se aplica sobre moneda física, y por tanto, en principio se rige por las opciones comunes, aunque, obviamente, luego de transformarse en Bitcoin, por ejemplo, o viceversa, debe aplicarse la intensificada en forma tradicional, como observamos más arriba, en tanto se utilizan tecnologías nuevas o en desarrollo que favorezcan el anonimato en las transacciones.

51 IG, "¿Qué son las criptomonedas y qué es el trading de criptomonedas?".

Posteriormente debemos seguir el trayecto de transformación de ese dinero fiduciario en criptoactivos, así como las transferencias de compraventa que puedan darse a posteriori, ya en una sede estrictamente digital en las exchanges correspondientes como Binance, Coinbase Exchange, KuCoin, etc., que deben aplicar la política de conocer al cliente. Esas transacciones quedan perfectamente registradas, especialmente en el tiempo, siendo inalterables; sin embargo, su rastreo no siempre es relativamente simple o del todo accesible.

El tema de las wallets con servicios garantidos de intimidad es una cuestión que deberá observarse, así como las plataformas de navegación anónimas, entre otras.[52]
Es claro que si en algún momento del trazado del camino de los activos (la regla de viaje, con la llama el GAFI) se encuentran criptoactivos, corresponde la debida diligencia intensificada. Reiteramos que en Uruguay procede de la misma de acuerdo con el artículo 19 de la Ley 19574 sobre prevención de lavado de activos y su decreto reglamentario.

07 ANÁLISIS ON-CHAIN: RESUMEN

El análisis on-chain es el estudio de los datos registrados directamente en una blockchain y resulta fundamental a los efectos de observar todo lo que sucede en la misma. Por tanto, el referido análisis puede resultar determinante en materia de evidencia digital de los delitos que comentamos.

Sabemos que las blockchains (descentralizadas y centralizadas)[53] son libros de contabilidad públicos, especialmente las descentralizadas, en los que toda transacción y actividad queda registrada de forma transparente e inmutable, salvo alteraciones fácilmente detectables por la generación de un nuevo hash.

El análisis on-chain funciona extrayendo y examinando el flujo de transacciones en la cadena de bloques. Los analistas observan tres tipos principales de datos:

A. Los datos de las transacciones efectuadas relevando cantidades enviadas, direcciones del remitente y del destinatario, comisiones pagadas y marcas o sellados de tiempo (timestaping).
B. Los datos de los bloques que incluyen, o pueden incluir, las recompensas pagas en el protocolo de trabajo a los mineros, el tamaño de los bloques, el número de transacciones, etc.
C. Los datos de los contratos inteligentes, como pueden ser las interacciones con aplicaciones descentralizadas, la creación de tokens, etc.[54]

De acuerdo con el uso de herramientas especializadas, esos datos brutos se pueden procesar para crear métricas e indicadores que ayuden a interpretar la actividad en la blockchain. Para lo que a este estudio importa, resulta fundamental al momento de observar posibles delitos como lavado de activos, fraudes, estafas, así como también la existencia de activos que pueden ser relevantes en sucesiones, disoluciones y particiones de sociedades conyugales, etc., donde existen criptoactivos.[55]

52 Rubén Flores Dapkevicius, Prevención de lavado de activos y financiamiento del terrorismo: Leyes N° 19574 y N° 19749 (Buenos Aires: IJ editores, 2021). **53** Descentralizado en el ecosistema ciberpunk significa sin una autoridad central, esto es, un propietario, banco central, etc. Por tanto, una blockchain es centralizada si esa autoridad central existe. **54** Nova Cryptor, “Descifrando la Blockhain: ¿Qué es el análisis on-chain?, Cripto Tendencia, marzo 13 de 2025. **55** Flores Dapkevicius, Prevención de lavado de activos…

Se dice que “el análisis on-chain ha emergido como una herramienta indispensable dentro del ámbito legal, respecto al manejo de casos que involucran criptoactivos”,[56] a lo que debemos agregar que el referido análisis permite observar todo lo que sucede en la cadena de bloques (no solo respecto a los criptoactivos en sí mismos), por ejemplo, respecto de las celebraciones, ejecuciones, etc., de los contratos inteligentes. Por tanto, el análisis puede proporcionarnos diversas evidencias digitales, según cada caso concreto.[57]

08 PRUEBA EN LA BLOCKCHAIN: RESUMEN

Esta prueba puede ser, según lo expuesto, pericial, documental, testimonial y por informes. La prueba pericial es fundamental. Mediante la misma se debe, en primer lugar, determinar el hash de la transacción, estableciendo si las wallets involucradas (denunciadas por el solicitante) se encuentran directamente asociadas con alguna transacción. En segundo lugar, mediante el estudio on-chain, el experto debe determinar si en cierto periodo de tiempo se produjo un intercambio de criptoactivos entre las wallets denunciadas, haciendo constar en su dictamen el contenido de los datos de las transacciones confirmadas y registradas de envío o recepción de criptoactivos. En tercer lugar, debe acreditarse la vinculación de las personas con las wallets, para lo cual existen diversos procedimientos, si el intercambio se realizó mediante exchanges centralizadas o descentralizadas. Por último, el perito debe pronunciarse respecto a la integridad o a la posibilidad de alteración de las transacciones realizadas. Corresponde decir que, según la sentencia que se comparte y agrega: “El acta notarial no acredita por sí sola la autenticidad técnica del archivo, es decir, si corresponde efectivamente a la filmación original […], cuestión que podría haberse salvado con la intervención de un especialista al momento de su extracción mediante un hashing de los archivos informáticos”.[58]

La Corte Suprema de Justicia de Colombia, Sala de Casación Penal (Gerson Chaverra Castro, juez ponente), en Radicado No. 58275, dijo:

> “Esto, por cuanto condicionó la admisibilidad de las imágenes como pruebas a la existencia del valor hash y a la introducción en juicio del dispositivo de almacenamiento, contentivo de los archivos extraídos por el técnico, como tarifa legal probatoria.
>
> Respecto de la primera cuestión, conforme se explicó líneas atrás, el código o valor hash hace parte de la cadena de custodia de las evidencias digitales. Luego, si en gracia a discusión, este no hubiese sido generado en curso del procedimiento de recolección de la información contenida en la micro SD, su ausencia no imponía la exclusión de la prueba. Por el contrario, debía el Tribunal valorarla con la correlativa disminución de su eficacia probatoria, al suscitar la duda en punto de su posible alteración o manipulación, es decir, de la falta de integralidad”.[59]

56 Marcos Acevedo, “Análisis on chain de transacciones sobre criptoactivos, prueba electrónica y su introducción al proceso civil”, Adpl. Instituto Argentino de Derecho Procesal Informático, junio 11 de 2024. **57** Flores Dapkevicius, Legaltech y derecho digital. **58** Diario Judicial.com, “Sin hashing no hay despido”, septiembre 22 de 2025; Juzgado Laboral n.° 2 de Corrientes, Sentencia n.° 165, M. P.: Rogelio Benítez Clementi (agosto 28 de 2025). **59** Corte Suprema de Justicia de Colombia, Sala de Casación Penal, Sentencia del 15 de junio de 2022, M. P.: Gerson Chaverra Castro, Radicado n.° 60173 (Colom.).

En la documental se debe aportar la URL del explorador de transacciones utilizado y capturas de pantalla del registro de transacciones dotadas de fe pública, en las que deben constar el hash de las transacciones, las direcciones de las wallets, los bloques de la blockchain implicados, así como los diversos datos que se estimen del caso.

También podría existir una prueba testimonial. En ese sentido, un testigo podría testificar que observó que otra persona realizaba transacciones con una wallet de otra. La prueba señalada es, en principio, de valor relativo.

La prueba por informes se produce si la parte que desea probar una transacción utilizó en la misma una exchange centralizada. En ese caso podría solicitarle un informe a la misma. Si la solicitada se niega, debe pedirse al juez que envíe un oficio a los efectos señalados. Los VASP centralizados deben utilizar formas de conocer al cliente y, por tanto, deben poseer los datos identificatorios del mismo, como DNI, domicilio, referencias bancarias, etc., todo ello para tener un conociendo siempre actualizado del mismo.[60]


Foto: Magnific

09 CONCLUSIONES

1. Es el sujeto obligado el que, precisamente, debe cumplir, primariamente en lo que a este trabajo importa, la normativa respecto de la ley de prevención de lavado de activos y financiamiento del terrorismo.
2. En la práctica, esa tarea la puede realizar él mismo, o su departamento de cumplimiento, jerarquizado en el oficial de cumplimiento.[61]
3. Esas empresas también pueden tercerizar la realización de la debida diligencia, en el caso que lo consideren necesario.
4. En este último supuesto, siempre se aconseja que el sujeto responsable, por lo menos, supervise la tarea.
5. En todo caso, si se detectan incoherencias en la documentación presentada, o antecedentes civiles, penales, etc., se debe intensificar incluso con un análisis exhaustivo en la matriz de riesgo, por ejemplo, teniendo en cuenta si es una PEP, si el cliente se encuentra en una zona de la lista negra (o roja, de existir) del GAFI, o el monto de la operación en cuestión, sin perjuicio de otras medidas ya destacadas en este trabajo.
6. La matriz de riesgo definirá si se trata de un riesgo bajo, medio, o alto, teniendo en cuenta el conocimiento del cliente por el sujeto obligado a realizar el control de posibles irregularidades en las operaciones, o del departamento de compliance.
7. Así deberá aplicar el scoring correspondiente, según el momento de aplicación de la debida diligencia, con la vigencia global del momento actual de la operación.

60 Flores Dapkevicius, Prevención de lavado de activos… **61** Flores Dapkevicius, Prevención de lavado de activos y financiamiento del terrorismo.

8. De acuerdo con ese puntaje adjudicado inicial, y definitivamente bajo su responsabilidad, el sujeto obligado y/o el oficial de cumplimiento definirán qué debida diligencia se aplicará al caso concreto.
9. Ello no impide que, luego de haber decidido realizar una debida diligencia simple, ella se vea intensificada a medida que se observan mayores riesgos.
10. Reiteramos que el departamento de compliance es, más simplemente en nuestro variado y riquísimo idioma castellano, la oficina de cumplimiento de cada obligado (estudios jurídicos, notariales, bancos, etc.),[62] dirigida por la persona responsable (oficial de cumplimiento), quien debe realizar la búsqueda de antecedentes, análisis, evaluación de riesgo, legajos, debida diligencia en general, y elaborar el puntaje de riesgo.[63] Lo expuesto no olvida que existen múltiples departamentos de cumplimiento, no solo en el área de lavado de activos.
11. El cumplimiento normativo, conocido también bajo el término compliance, hace referencia a las normas establecidas por las empresas en los ámbitos interno y externo (mejores prácticas, código ético, anticorrupción, prevención de riesgos en el lugar de trabajo, protección de datos, blanqueo de capitales, etc.). El cumplimiento de estas normas creadas por y para las empresas es una forma de prevenir y evitar cualquier conducta ilícita.[64]
12. Además, esta búsqueda de antecedentes, según opción del sujeto obligado, o de su oficial de cumplimiento, se puede realizar manualmente, lo que lleva más tiempo e implica un mayor riesgo de errores, o mediante una herramienta particular informática especializada en antilavado de activos.[65]
13. Ante alguna auditoría que pueda realizar el órgano competente de la administración estatal, según la asignación expresa de la misma, como corresponde en el derecho administrativo, el sujeto obligado, o su oficial de cumplimiento (cuyo rango es gerencial), son los responsables de toda la documentación que respalde el cumplimiento de sus obligaciones.[66]
14. El ROS lo completa el sujeto obligado, o su oficial de cumplimiento, cuando debe denunciar dicha operación sospechosa ante la UIAF, que generalmente se encuentra en el sistema orgánico del banco central del país de que se trate,[67] centralizando la información y evaluándola según su especialización funcional, determinando la segunda etapa de la inteligencia financiera, iniciada por ese informe que, por ello, debe ser cumplido en forma y al detalle. El mismo se puede hacer digitalmente, en determinados casos y según determinadas formalidades, de acuerdo a como lo observamos supra.
15. Dentro de los límites establecidos en la legislación comparada se debe informar a la UIAF del banco central las transacciones, realizadas o no, que en los usos y costumbres de la respectiva actividad resulten inusuales, se presenten sin justificación económica o legal evidente[68] o se planteen con una complejidad inusitada o injustificada, así como también las transacciones financieras que involucren activos sobre cuya procedencia existan sospechas de ilicitud, a efectos de prevenir los delitos de lavado de activos tipificados.

62 Ibíd. **63** Compendio normativo sobre prevención de lavado de activos y financiamiento del terrorismo, s.v. “Compliance” (Montevideo: Brum & Costa, 2019), 17. **64** Inés Ducom y Nicolás Melchior, “Definición y concepto de compliance o cumplimineto normativo”, Mariscal Abogados. **65** Flores Dapkevicius, Prevención de lavado de activos y financiamiento del terrorismo. **66** Rubén Flores Dapkevicius, Tratado de derecho administrativo, tomos I y II, 2.a ed. (Montevideo-Buenos Aires: La Ley-Thomson Reuters, 2021). **67** Esa información se entrega en sobre cerrado. **68** Este concepto jurídico determinable siempre debe destacarse.

16. En el supuesto de sospecha de financiamiento de terrorismo, la obligación de informar alcanza también a aquellas operaciones que, aun involucrando activos de origen lícito, se sospeche que están vinculadas a las personas físicas o jurídicas comprendidas en dicho delito.
17. La información debe comunicarse en forma inmediata al ser calificada como sospechosa y, reiteramos, cuando las operaciones no hayan sido efectivamente concretadas, ya sea porque el cliente desistió de realizarla o porque el sujeto obligado resolvió no dar curso a la misma.
18. En todo caso, la comunicación del ROS es y deber ser reservada.
19. Ningún sujeto obligado, incluyendo las personas relacionadas contractualmente con él, podrá poner en conocimiento de las personas participantes, o de terceros, las actuaciones y los informes que sobre ellas realicen, o produzcan, en cumplimiento de la obligación de informar señalada precedentemente.
20. El tema es fundamental, y, en algunas legislaciones, la violación de esta obligación se sanciona en forma penal.
21. La UIAF tiene la obligación de mantener confidencialidad en todas las situaciones, para proteger la identidad del denunciante, ya que puede ser objeto de represalias.
22. El sujeto obligado, o su oficial de cumplimiento, una vez reportada la operación sospechosa, puede ser instruido respecto a qué debe hacer en la transacción.
23. En ese sentido, la UIAF, por resolución fundada, podrá instruir a los sujetos obligados, financiero o no financieros, para que impidan por un plazo de hasta setenta y dos horas la realización de operaciones sospechosas de involucrar fondos procedentes de los delitos cuya prevención procura evitar esta norma, así como la ejecución de cualquier tipo de orden que implique la devolución, el traspaso o la transferencia de activos, o sus títulos representativos, brindados por personas físicas o jurídicas, sobre las cuales existan fundadas sospechas de su vinculación con esos delitos, e igualmente el acceso a cofres de seguridad a quienes se encuentren vinculados a cualquier título, sean personas físicas o jurídicas.
24. La decisión debe comunicarse inmediatamente al tribunal judicial penal competente, el cual, consideradas las circunstancias del caso, determinará si correspondiere, sin previa notificación, la inmovilización de los activos de los partícipes, sus títulos representativos, así como el acceso a los cofres de seguridad.
25. La resolución que adopte el tribunal penal competente en la materia, sea confirmando o rechazando la decisión adoptada por la UIAF, debe ser comunicada a esa unidad.
26. A su vez, la referida UIAF debe ponerla en conocimiento de los sujetos obligados involucrados. [69]
27. El sujeto obligado, o su oficial de cumplimiento, se encuentra exento de responsabilidad ante la ley si el ROS fue realizado en tiempo y en forma, y si se obró de conformidad con lo expuesto en este desarrollo.
28. Ante la más mínima duda, aconsejo en todo caso realizar la diligencia más intensificada, de las opciones que se estimen prudentes.
29. De ser necesario, no cabe duda de que el ROS se debe efectuar y presentar en tiempo y forma, según lo expuesto.[70]

69 Rubén Flores Dapkevicius, Decreto 500/91, actualizado, anotado y concordado, 7.a ed. (Buenos Aires: IJ editores, 2020). El decreto 500/91 regula el procedimiento administrativo común, el disciplinario (en lo pertinente) y, parcialmente, el recursivo. La norma se aplica a ese sistema orgánico según corresponda.**70** Flores Dapkevicius, Tratado de derecho administrativo.

30. Presentado, más allá de lo que se decida o no por ese órgano, es de buena práctica aguardar no menos de 72 horas hábiles antes de tomar cualquier decisión.
31. En ese periodo, la UIAF del banco central debería informar sobre su resolución respecto a la situación que se encuentra en estudio.
32. Toda la información que se recaude respecto de la operación debe conservarse ante la posible intervención del órgano competente, en principio, por cinco años.[71]
33. La tecnología blockchain puede ayudar, en ciertos casos, a evitar el lavado de activos, con mayor simplicidad o complejidad, eficiencia o ineficiencia, si se adecúa precisamente.
34. Ello es así por la forma en que deben realizarse las transacciones con billeteras electrónicas, y por todo el funcionamiento de esa tecnología disruptiva.
35. Es muy claro que la tecnología permite el seguimiento, según la característica de cada blockchain, el tipo de criptomoneda más o menos anónima, los seudónimos, las diferentes claves públicas, el uso de mixers e incluso los tipos de navegadores, como TOR.
36. Según el GAFI, los Estados deben considerar los activos virtuales como bienes, productos, fondos u otros activos de valor equivalente, y deben aplicar las medidas pertinentes, en virtud de las recomendaciones del organismo, a los activos virtuales y a los VASP.
37. Es decir que se recomienda, como en el supuesto de los activos del mundo no virtual, el traslado de la responsabilidad del Estado a los sujetos obligados.
38. El GAFI hace especial énfasis en el control de aquello activos que puedan generarse con la disrupción tecnológica y/o garantizan un margen muy amplio de protección a la intimidad en los negocios, porque estos pueden encubrir lavado de activos.
39. De acuerdo con ello, siempre en estos momentos, los sujetos obligados, sean financieros o no, si en las transacciones de que se traten observan criptoactivos, deben aplicar la debida diligencia intensificada, salvo en el supuesto del protocolo de trabajo respecto a los mineros.



Foto: Magnific

71 Flores Dapkevicius, Prevención de lavado de activos y financiamiento del terrorismo.

10 REFERENCIAS

Abel Souto, M. El blanqueo de dinero en la normativa internacional. España: Universidad de Santiago de Compostela, 2002.

Acevedo, Marcos. "Análisis on chain de transacciones sobre criptoactivos, prueba electrónica y su introducción al proceso civil". Adpl. Instituto Argentino de Derecho Procesal Informático. Junio 11 de 2024. https://iadpi.com.ar/2024/06/11/analisis-on-chain-de-transacciones-sobre-criptoactivos-prueba-electronica-y-su-introduccion-al-proceso-civil/

Asociación de Escribanos del Uruguay. "Prevención de lavado de activos. Diligencia debida notarial". https://www.aeu.org.uy/Interes-notarial/Prevencion-de-lavado-de-activos-uc1623

Banco Central de Uruguay. "Buenas prácticas para el reporte de operaciones sospechosas". Secretaría Nacional para la Lucha contra el Lavado de Activos y el Financiamiento del Terrorismo. Octubre 31 de 2016. https://www.gub.uy/secretaria-nacional-lucha-contra-lavado-activos-financiamiento-terrorismo/comunicacion/publicaciones/buenas-practicas-para-el-reporte-de-operaciones-sospechosas?fbclid=IwAR0xLxFEy-tnZIoCYwVjlhGAW-JcQhBQYg2TD5EgJ5dz-pvYeAlQoRxDaJw

BDA Abogados. "Recomendaciones del GAFI para regular actividades con Activos Virtuales". Enero 6 de 2025. https://www.bda-lawfirm.com/articles/recomendaciones-del-gafi-para-regular-actividades-con-activos-virtuales-2/

Blanco Crodero, I. El delito de blanqueo de capitales. 3.ª ed. Madrid: Aranzandi, 2012.

Compendio normativo sobre prevención de lavado de activos y financiamiento del terrorismo. Montevideo: Brum & Costa, 2019.

Superintendencia Financiera de Colombia. "Inicio". Gobierno de Colombia. https://www.superfinanciera.gov.co/

Congreso de Colombia. Ley 1762 de 2015: Por medio de la cual se adoptan instrumentos para prevenir, controlar y sancionar el contrabando, el lavado de activos y la evasión fiscal (julio 6 de 2015). https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=65338

Congreso de la República de Uruguay. Ley n.° 19574. Ley integral contra el lavado de activos. Actualización de la normativa vigente. Derogación de artículos del decreto Ley 14.294 y Leyes 17.835, 18.494, 18.914 y 19.149 (dicembre 20 de 2017). https://www.impo.com.uy/bases/leyes/19574-2017

Corte Suprema de Justicia de Colombia, Sala de Casación Penal. Sentencia del 15 de junio de 2022. M. P.: Gerson Chaverra Castro. Radicado n.° 60173 (Colom.). https://www.ramajudicial.gov.co/documents/6342228/8753677-jyp+SP2025-2022+60173+%2815-06-22%29%20SEGUNDA+DR.+CHAVERRA.pdf/93aa0c7b-5cb7-4a2b-aa61-3aa0f560de1d

Cumplo360Blog. "Qué es la Debida Diligencia en prevención de lavado de activos?" (blog). https://-cumplo360.com/es/que-es-debida-diligencia/

Daniels Pinto, Alí José ¿Qué es el GAFI y para qué sirve? Venezuela: Acceso a la Justicia, 2023. https://accesoalajusticia.org/wp-content/uploads/securep-fs/2023/07/Que-es-el-GAFI-y-para-que-sirve.pdf

Del Carpio Delgado, Juana. "El delito del blanqueo de bienes en el nuevo código penal". Tesis doctoral, Universidad de Sevilla, España, 1997.

Diario Judicial.com. "Sin hashing no hay despido". Septiembre 22 de 2025. https://www.diariojudicial.-com/news-101651-sin-hashing-no-hay-despido?s=03

Ducom, Inés y Nicolás Melchior. "Definición y concepto de compliance o cumplimineto normativo". Mariscal Abogados. https://www.mariscal-abogados.es/definicion-y-concepto-de-compliance-o-cumplimiento-normativo/#:~:text=El%20cumplimiento%20normativo%2C%20conocido%20tambi%C3%A9n,de%20capitales%2C%20etc

Elidafamo, Catalina. "Europa estudia prohibir stablecoins emitidas en varias jurisdicciones". Observatorio Blockhain. Octubre 1 de 2025. https://observatorioblockchain.com/stablecoin/europa-estudia-prohibir-stablecoins-emitidas-en-varias-jurisdicciones/

Esparragoza, Luis. "Exchange Buda solicita selfie para retirar bitcoin a las direcciones nuevas". Criptonoticias. Junio 4 de 2021. https://www.criptonoticias.com/seguridad-bitcoin/exchange-buda-solicita-selfie-retirar-bitcoin-direcciones-nuevas/

FATF. International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation. París, Francia: FATF, 2025. https://www.gmc360.com.mx/post/bonus360-actualiza-ci%C3%B3n-de-las-40-recomendaciones-del-gafi-febrero-2025

Ferré Olivé, J. Derecho penal colombiano. Parte general: Principios fundamentales y sistema. Presentación y prologo de Claus Roxin. Colombia: Ibáñez, 2010.

Flores Dapkevicius, Rubén. Amparo, habeas corpus, habeas data y acceso a la información pública. 5ta. ed. Montevideo: FCU, 2021.

—. Decreto 500/91, actualizado, anotado y concordado, 7ma ed. (Buenos Aires: IJ editores, 2020).
—. Estatuto del funcionario: en la contemporaneidad del Legaltech. 3ra ed. Montevideo: La Ley, 2025.
—. Funcionarios públicos. 3.a ed. Montevideo: La Ley, 2023.
—. Legaltech. Montevideo: FCU, 2023.
—. Legaltech y derecho digital. Buenos Aires: BdF Editores, 2024.
—. Neuroderechos, neurociencias e internet of bodies. Buenos Aires: BdF Editores, 2024.
—. Prevención de lavado de activos, con activos físicos y criptoactivos. Montevideo: FCU, 2023.
—. Prevención de lavado de activos y financiamiento del terrorismo: Leyes N° 19574 y N° 19749. Buenos Aires: IJ editores, 2021.
—. Sumarios e investigaciones administrativas, incluye la prevención de lavado de activos con criptoactivos y activos físicos. 7ma. ed. Montevideo: FCU, 2022.
—. Tratado de derecho administrativo. 2da ed. Tomos I y II Montevideo-Buenos Aires: La Ley-Thomson Reuters, 2021.

Flores Dapkevicius, Rubén y Gustavo Ordoqui. Responsabilidad del Estado y la acción de nulidad de los actos administrativos. Montevideo-Buenos Aires: La Ley, 2022.

GAFILAT. Estándares internacionales sobre la lucha contra el lavado de activos, el financiamiento del terrorismo, y el financiamiento de la proliferación de armas de destrucción masiva. s. l.: GAFILAT, 2023. https://biblioteca.gafilat.org/wp-content/uploads/2024/07/Recomendaciones-metodologia-actDIC2023.pdf

GAFISUD, Estándares internacionales sobre la lucha contra el lavado de activos y el financiamiento del terrorismo y la proliferación. Recomendaciones del GAFI. s. l.: GAFISUD, 2012. https://www.fatf-gafi.org/content/dam/fatf-gafi/translations/Recommendations/FATF-40-Rec-2012-Spanish.pdf.coredownload.inline.pdf

García Cavero, P. El delito de lavado de activos. Buenos Aires: BDEF, 2018.

Gutiérrez, Fernando. "GAFI actualiza visión regulatoria sobre criptomonedas; pide especial atención a transacciones peer to peer". El Economista. Octubre 28 de 2021. https://www.eleconomista.com.mx/sectorfinanciero/GAFI-actualiza-vision-regulatoria-sobre-criptomonedas-pide-especial-atencion-a-transacciones-peer-to-peer-20211028-0079.html

Hernández Quintero, H. El lavado de activos. Colombia: Ediciones Jurídicas Gustavo Ibáñez, 1997.

IG. "¿Qué son las criptomonedas y qué es el trading de criptomonedas?". https://www.ig.com/es/trading-de-criptomonedas/que-es-el-trading-de-criptomonedas

InfoLAFT. "Repasemos nuestra historia". https://www.infolaft.com/repasemos-nuestra-historia

Jakobs, G. La competencia por organización en el delito omisivo. E. Peñaranda Ramos, Trad. Bogotá D. C.: Universidad Externado de Colombia, 1994.

Juarros, Santiago. "Qué tipos de wallets de criptomonedas existen". Ripio. Septiembre 12 de 2024. https://launchpad.ripio.com/blog/que-tipos-de-wallets-de-criptomonedas-existen

Juzgado Laboral n.° 2 de Corrientes. Sentencia n.° 165. M. P.: Rogelio Benítez Clementi (agosto 28 de 2025). https://www.diariojudicial.com/uploads/0000059688-original.pdf

Martínez Sánchez, W. A. Sistemas de prevención del lavado de activos y la financiación del terrorismo. Bogotá D. C.: Legis S. A., 2011.

Mejía, Daniel y Daniel Rico. La microeconomía de la producción y tráfico de cocaína en Colombia. Bogotá, D. C.: Universidad de los Andes, Centro de Estudios sobre Desarrollo Económico, 2010.

Nando Lefort, V. El lavado de dinero. México: Editorial Trillas, 1997.

Nova Cryptor. "Descifrando la Blockhain: ¿Qué es el análisis on-chain? Cripto Tendencia. Marzo 13 de 2025. https://criptotendencia.com/2025/03/13/descifrando-la-blockchain-que-es-el-analisis-on-chain/

Pérez del Castillo & Asociados. "Normativa contra el Lavado de Activos". Noviembre 22 de 2018. https://www.pdelc.com.uy/espanol/normativa-contra-el-lavado-de-activos-7?nid=178

Stessens, G. Money laundering: a new international law enforcement model. UK: Cambridge University Press, 2000.

Unidad de Información y Análisis Financiero (UIAF). Compilación y descripción de las tipologías producidas por la Unidad de Información y Análisis Financiero. Documento público. Bogotá, D.C., Colombia: UIAF, 2006.

---. "Noticias y comunicados". Gobierno de Colombia. https://www.uiaf.gov.co/

---. "Nuestra entidad". Gobierno de Colombia. https://www.datos.gov.co/stories/s/Conoz-ca-qu-es-la-Uni-dad-de-Informaci-n-y-An-lisis-/g36h-z67x/

Vásquez Juárez, N. "El problema de la criminalización del lavado de dinero proveniente del tráfico de drogas". Tesis de Licenciatura. Pontificia Universidad Católica del Perú, Facultad de Derecho, Lima, 1992.

Villegas, Carlos. Compendio jurídico, técnico y práctico de la actividad bancaria. Buenos Aires: Depalma, 1986.

ACERCA DE LOS AUTORES

Rubén
Flores Dapkevicius

Doctor (PhD) en Derecho y Ciencias Sociales y en Filosofía. Postdoctor en Derecho. Abogado por la Universidad Mayor de la República. Profesor visitante de grado y posgrado de la Universidad de Buenos Aires —UBA—. Profesor de Derecho Público de la Universidad de la República. Especialista y asesor legislativo y particular en legaltech, prevención de lavado de activos y criptoactivos, neuroderechos y prueba digital, entre otros campos. Autor de más de 60 libros publicados en Uruguay, Argentina, Colombia, Perú, Brasil y España. Autor de más de quinientos trabajos sobre temas de su especialidad, publicados en el país y en el exterior. Exasesor de la Presidencia de la República y asesor letrado en el Poder Legislativo. Asesor letrado del Servicio Civil de la Presidencia de la República. Jefe de la Asesoría Letrada de las Obras Sanitarias del Estado y jefe de la División Sumarios de dicha persona pública estatal. Futurista. Expositor nacional e internacional.

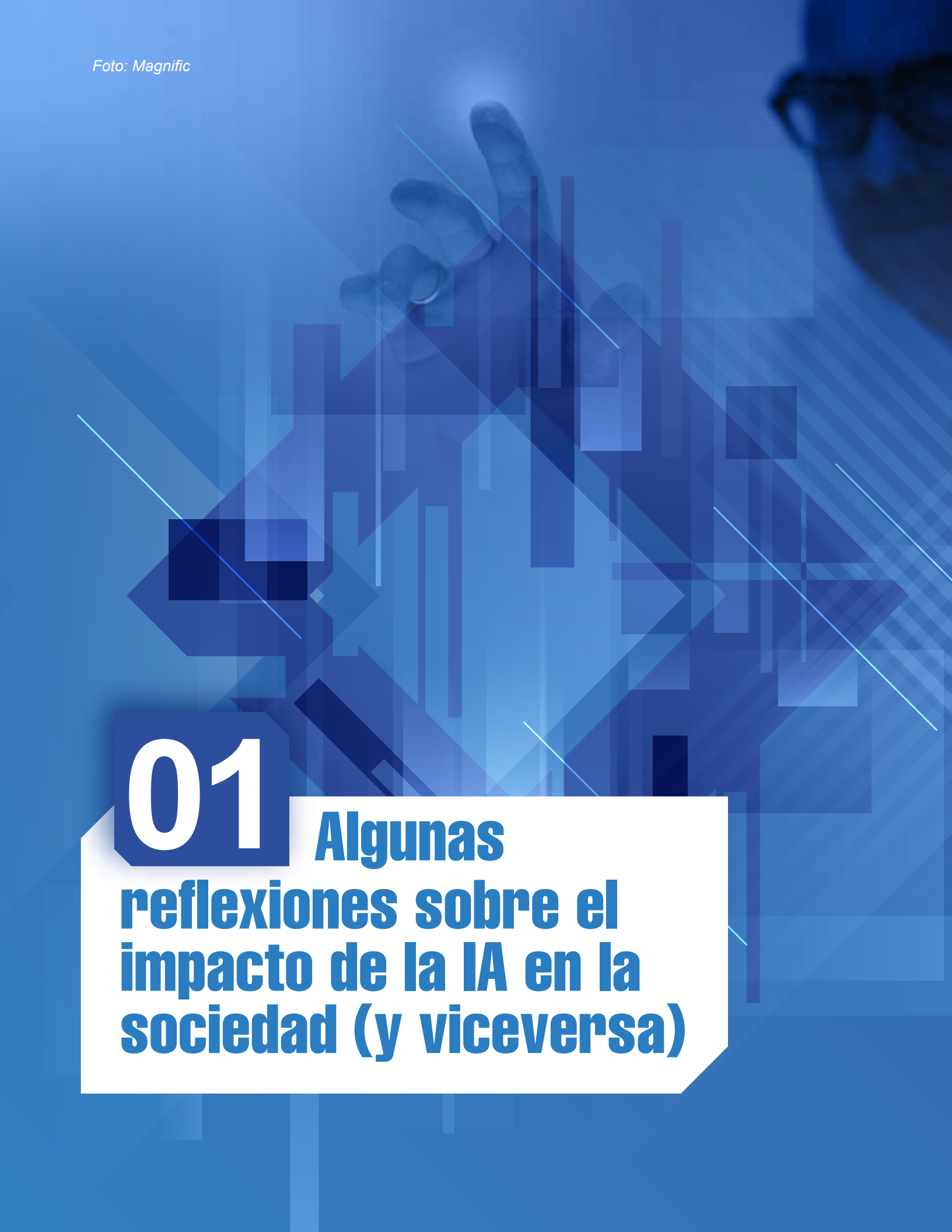


Foto: Magnific

01 Algunas reflexiones sobre el impacto de la IA en la sociedad (y viceversa)

E-ESTADO
Primera Edición

Algunas reflexiones sobre el impacto de la inteligencia artificial en la sociedad (y viceversa)

Some reflections on the impact of artificial intelligence on society (and vice versa)

Fecha de recepción: 11 de julio de 2025
Fecha de aceptación: 18 de febrero de 2026

José Luis Verdegay
Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, España
verdegay@ugr.es
https://orcid.org/0000-0003-2487-942X

David A. Pelta
Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, España
dpelta@ugr.es
https://orcid.org/0000-0002-7653-1452

Cómo citar: Verdegay, José Luis y David A. Pelta. “Algunas reflexiones sobre el impacto de la inteligencia artificial en la sociedad (y viceversa)”. Revista e-Estado 1, n.° 1 (2026).



Foto: Magnific

RESUMEN

En este trabajo se analiza lo que supone la importante presencia que hoy en día tiene la inteligencia artificial (IA) en la vida cotidiana, la brecha entre su realidad y la percepción pública, a menudo influenciada por la ficción y la desinformación, y las posibles vías de actuación para lograr una convivencia armónica, complementaria y beneficiosa entre IA y sociedad.

Se argumenta que la ciencia y la regulación son cruciales para una comprensión precisa y un despliegue ético de la IA, diferenciando entre la IA general, la IA especializada y la definición de la Unión Europea (UE) de sistemas con IA. Se esbozan los riesgos cotidianos con los que ya convivimos, así como posibles estrategias que permitan atenuarlos. Finalmente, apoyándonos en los programas que se empiezan a desarrollar en Europa, y tomando como ejemplo concreto el caso de España, introduciremos algunas de las líneas de acción puestas en marcha para la supervisión de los sistemas de IA, así como para su aplicación en el futuro.

La conclusión enfatiza la necesidad urgente de formación ciudadana, regulación adaptable y supervisión independiente para garantizar que la IA sirva al bien común y proteja los derechos fundamentales.

Palabras clave: inteligencia artificial; retos; ética; control; bienestar social

ABSTRACT

This paper analyses the important presence of artificial intelligence (AI) in everyday life, the gap between its reality and public perception, often influenced by fiction and misinformation, and the possible ways to achieve a harmonious, complementary and beneficial coexistence between AI and society. It is argued that science and regulation are crucial for an accurate understanding and ethical deployment of AI, differentiating between general and specialized AI and the EU definition of AI-enabled systems. It highlights the potential benefits of AI in areas such as sustainability, security and the welfare state, while cautioning against risks such as inequality and manipulation. The conclusion emphasizes the urgent need for citizen education, adaptive regulation and independent oversight to ensure that AI serves the common good and protects fundamental rights.

Keywords: : Artificial Intelligence; Challenges; Ethics; Control; Social Welfare

01 INTRODUCCIÓN

No cabe duda de que los algoritmos y la inteligencia artificial (IA) ya no son visitantes lejanos, sino actores principales de nuestro devenir cotidiano. Nuestro modo de vida, tanto en lo individual como en lo colectivo, está siendo moldeado, a veces sin que lo notemos, por estas nuevas lógicas invisibles que afectan nuestras decisiones, nuestras búsquedas y nuestras emociones. En todas partes resuena su eco: en los titulares que consumimos, en las conversaciones que sostenemos o en las redes donde construimos y deconstruimos nuestras identidades. Y con ese eco viene, inevitablemente, una inquietud: la sospecha, el recelo, incluso el temor. Porque se habla de la IA con seguridad impostada y, más preocupante aún, con una soltura que rara vez se ve acompañada de un conocimiento profundo. Así se gestan relatos fantásticos, escenarios alarmistas, imágenes seductoras pero infundadas, que poco tienen que ver con la realidad constatable con la ciencia y sí mucho con las ficciones que nos fascinan o nos paralizan.

Los científicos, quienes tenemos el privilegio y la responsabilidad de dialogar con la verdad provisional de la ciencia, no podemos permanecer callados frente a este fenómeno. Nos debemos a la sociedad y a su derecho a una información rigurosa, sensata y comprensible. Estamos, por tanto, obligados a exigir un mayor grado de responsabilidad a quienes se dedican a comunicar aquello que puede despertar alarma, sobre todo cuando se trata de conceptos tan complejos y resbaladizos como la IA. Esta tecnología de nombre poético, cuya popularidad crece al ritmo de titulares estridentes, corre el riesgo de convertirse en un mito, más que en un objeto de comprensión.

Hay, pues, una distancia cada vez más grande entre lo que la IA es —con toda su riqueza, sus límites, sus promesas y sus peligros— y lo que se proyecta de ella desde ciertas tribunas que, sustentadas más en la imaginación que en el saber, terminan moldeando la percepción pública. Y esa distancia no es inocua: distorsiona, confunde y, a veces, hiere.

Desde este planteamiento, nos disponemos inicialmente a comentar el panorama actual que presenta la IA desde un punto de vista más teórico para, a partir de ahí, esbozar los riesgos cotidianos con los que ya convivimos, así como posibles estrategias que permitan atenuarlos. Finalmente, apoyándonos en los programas que se empiezan a desarrollar en Europa, y tomando como ejemplo concreto el caso de España, introduciremos algunas de las líneas de acción puestas en marcha para conducirnos hacia ese porvenir prometedor que la IA puede ofrecer.

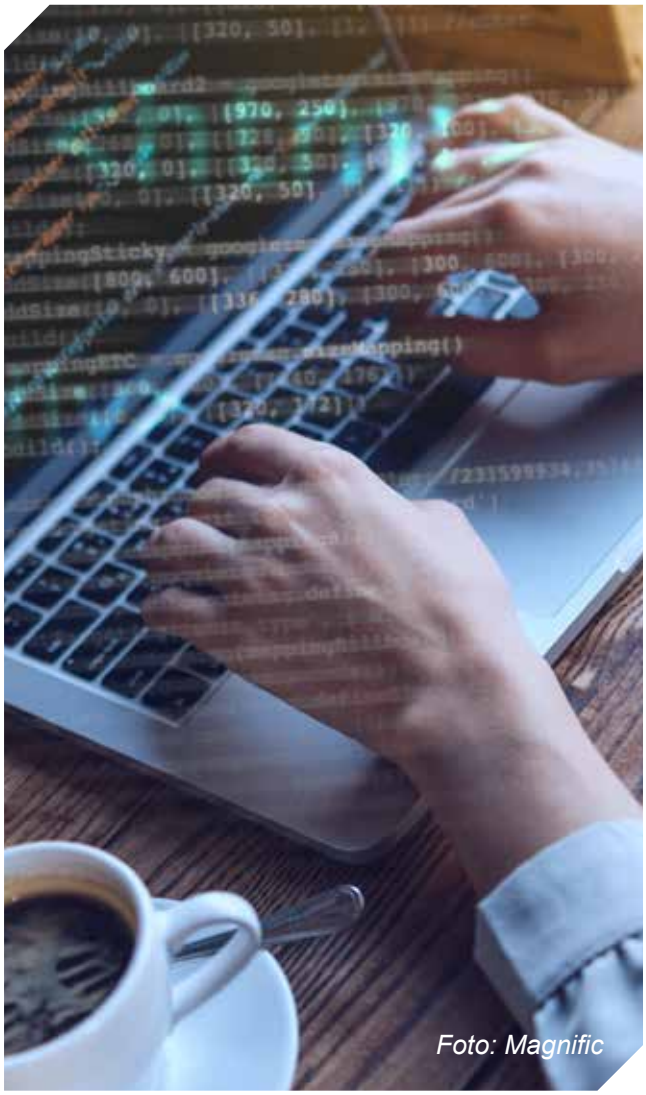


Foto: Magnific

02
EL SKYLINE DE LA INTELIGENCIA ARTIFICIAL

La IA, o mejor dicho, el software que utiliza técnicas de IA, nos permite resolver problemas provenientes de diversos ámbitos: desde la conducción autónoma hasta la agricultura de precisión, desde la medicina personalizada hasta la sostenibilidad energética, pasando por la creación artística, la justicia, el deporte, la educación, la traducción, la alimentación, la gestión administrativa y un largo etcétera.

Pero es evidente que no todo lo que trae la IA ha de considerarse, sin matices, un bien para las personas. Tampoco, las personas en general están dispuestas a abrirle su confianza sin reservas.

Nos corresponde a nosotros —como ciudadanos, como científicos, como sociedad— explorar y decidir en qué ámbitos permitimos que se utilicen estos sistemas que podríamos denominar como inteligentes, de manera tal que se minimicen los conflictos, se eviten las rupturas y que, en su lugar, se favorezca el verdadero progreso social.

Ahora bien, cuando hablamos de IA, ¿sabemos realmente de qué estamos hablando? La expresión inteligencia artificial, tan repetida como imprecisa, encierra una pluralidad de realidades que rara vez se desentrañan con claridad. Como punto de partida, suele hacerse una primera distinción, tal vez esquemática pero útil, entre dos grandes categorías: la IA general —también llamada fuerte o, en clave más irónica, la obsesión de Hollywood— y la IA débil, conocida también como especializada y, más recientemente, también como IA no generativa.

La primera, la IA general, es ese ideal casi mítico de una mente artificial capaz de aprender, razonar, comprender y adaptarse como lo haría un ser humano, incluso en escenarios completamente nuevos. Una inteligencia que, de alcanzar su plenitud, podría abordar cualquier tarea cognitiva, en cualquier dominio, con la autonomía y versatilidad que asociamos a la conciencia humana. Pero, como ocurre con las quimeras, esta forma de IA habita más en la especulación que en los laboratorios: no hay hoy rastro cierto de su viabilidad ni de su necesidad, ni siquiera cuenta con el respaldo de la mayoría de los expertos en la materia.[1]

La segunda, la IA especializada, mucho más modesta pero también mucho más real, es la que sustenta la mayoría de los sistemas que usamos a diario. Hablamos realmente de herramientas (softwares) basadas en IA diseñadas para resolver una tarea específica —diagnosticar una enfermedad, traducir un texto, recomendar una película—, pero que carecen por completo de utilidad fuera de ese estrecho marco de especialización. No entienden, no sienten, no intuyen. Solo ejecutan, con eficiencia matemática, una función concreta. No hay, en ellas, ni comprensión ni conciencia.

A partir de esta dicotomía fundamental, se han formulado decenas de definiciones de IA, cada una con sus matices, sus énfasis, sus límites. A diferencia del concepto de algoritmo, que es claro y unívoco, el de IA —como el de inteligencia— se resiste a la precisión, se bifurca, se escapa. Desde los tiempos de John McCarthy —aquel verano de 1956 en Dartmouth, donde se acuñó el término por primera vez— hasta hoy, la definición ha sido objeto de debates y reinterpretaciones, sin que se haya alcanzado un consenso definitivo.

1 AAAI 2025 Presidential Panel on the Future of AI Research, https://aaai.org/about-aaai/presidential-panel-on-the-future-of-ai-research/.

Tal es la dificultad, que la propia Unión Europea (UE), al impulsar su reciente Ley de Inteligencia Artificial, ha optado por una vía pragmática: renunciar a definir la IA en abstracto y, en su lugar, centrarse en lo que constituye un sistema con IA. Es decir, ha optado por describirla a través de sus funciones observables, por lo que sabemos que hace y por cómo lo hace hoy día.

> *Así, un sistema con IA —según la definición normativa de la UE— es:"Todo aquel sistema basado en máquinas, diseñado para operar con distintos grados de autonomía, que presenta capacidad de adaptación tras su despliegue, y que, persiguiendo objetivos explícitos o implícitos, infiere a partir de los datos recibidos cómo producir resultados —sean predicciones, contenidos, recomendaciones o decisiones— capaces de influir en entornos físicos o virtuales".*[2]

Esta es una definición operativa y clara que busca eliminar ambigüedades y cerrar la puerta a interpretaciones equívocas. Una brújula legal para navegar en un terreno que, de no ser cartografiado con cuidado, puede derivar en zonas de sombra donde el riesgo se multiplica y la responsabilidad se diluye.

No obstante, en el presente coexisten dos grandes vertientes de esta tecnología en constante evolución. Por un lado discurre la IA que acabamos de describir, que se limita a descifrar lo ya dicho por un universo de datos. Junto a ella se desarrolla su contrapartida, la IA generativa, que, a diferencia de su hermana analítica, no se contenta con reproducir lo aprendido, sino que va más allá, transformando lo absorbido en materia prima para la invención. Es capaz de escribir textos que no existían, modelar objetos tridimensionales inéditos, componer melodías nunca oídas o incluso programar desde una hoja en blanco. Esta facultad para engendrar lo nuevo, lo inédito, constituye el verdadero umbral que separa a esta generación de tecnologías de las antiguas oleadas de automatización.[3]

En cualquier caso, a los sistemas que se describen en la definición europea —máquinas que infieren, que adaptan sus respuestas, que operan con un margen creciente de autonomía— se les conoce cada vez más por un nombre que busca escapar del ruido semántico que arrastra el término inteligencia artificial: se les llama sistemas automatizados de decisión o, por sus siglas en inglés, ADS. Esta denominación, aunque menos resonante, resulta paradójicamente más clara, más cercana, más comprensible: suena menos pretenciosa, menos enigmática, y por tanto más apropiada para una tecnología que, en el fondo, no hace sino ejecutar procesos diseñados y programados por seres humanos. Esta automatización de decisiones, por más sofisticada que sea, sigue remitiendo a una voluntad previa, a una intención originaria.

La IA —o los ADS, si se prefiere— se analiza casi siempre desde una tensión binaria: por un lado, nos fascina su potencial; por otro lado, nos inquietan sus consecuencias. Y ambos sentimientos coexisten, sin anularse, en el ánimo de cada uno de nosotros. Somos a un tiempo entusiastas de sus avances y escépticos de sus peligros; celebramos sus logros y recelamos de su poder. En el mismo gesto con que admiramos una nueva aplicación, nos interrogamos sobre sus posibles efectos indeseados. Esta ambivalencia forma ya parte del imaginario colectivo: el futuro que se anuncia nos seduce tanto como nos estremece.

2 Unión Europea. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo: Por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n. o 300/2008, (UE) n. o 167/2013, (UE) n. o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial) (junio 13 de 2024), 46. **3** Francesco Filippucci et al., The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges, OECD Artificial Intelligence Papers, n.o 15 (Paris: OECD Publishing, 2024).



Foto: Magnific

Pero ¿cómo o de dónde nace esa mezcla de expectativas, de esperanzas y de temores? A veces no proviene tanto del ámbito científico o académico como de la ficción. Basta con recordar aquel célebre ensayo de 1993, titulado “La próxima singularidad tecnológica”, firmado por el matemático y novelista Vernor Vinge.[4] En él, se profetizaba la irrupción inminente de una inteligencia sobrehumana, un acontecimiento que, en su visión, alteraría para siempre el curso de nuestra civilización. Vinge, como Isaac Asimov y otros antes que él, se movía en el terreno de la especulación literaria, y sin embargo su idea —la de una singularidad tecnológica que marcara el fin de la era humana tal como la conocemos— prendió con fuerza en la imaginación de muchos investigadores. Lo que nació como un experimento narrativo acabó convirtiéndose en una hipótesis aceptada por una parte del discurso académico.

Desde entonces, la singularidad —esa línea difusa en el horizonte en la que las máquinas superarían al ser humano en toda competencia— se ha transformado en el centro de un debate vibrante, apasionado y, en buena medida, estéril. Un bucle especulativo que consume energías y distrae de los problemas reales y urgentes que la IA plantea hoy, aquí, en nuestro presente cotidiano.

Pese a ello, los esfuerzos por alcanzar esa inteligencia general se han convertido en una carrera frenética entre empresas, países y laboratorios. La competencia por construir sistemas que emulen o incluso excedan nuestras capacidades no cesa, y los recursos que se destinan a ello son ingentes. Lo más paradójico es que, en el camino hacia esa supuesta superinteligencia, las propias máquinas comienzan a participar activamente en su propio perfeccionamiento. A medida que los algoritmos aprenden a aprender, ya no necesitan siempre de la guía humana: se retroalimentan entre sí, ajustan sus propios modelos, generan nuevas versiones de sí mismos. La inteligencia que antes modelábamos desde fuera empieza a evolucionar desde dentro.

4 “The Coming Technological Singularity: How to survive in the post-human era” (NASA Lewis Research Center, December 1, 1993).

Y en ese umbral entre lo diseñado por nosotros y lo que empieza a escapar a nuestro control es donde se despliega el verdadero desafío. No tanto el de si una máquina llegará a pensar como lo hacemos nosotros, sino el de si sabremos pensar, con la suficiente claridad y profundidad, sobre lo que estamos construyendo.

Esa capacidad de automejora recursiva —ese bucle por el cual la IA no solo aprende, sino que aprende a aprender, se reconfigura y se ajusta a sí misma— podría, si no se vigila, desembocar en una espiral evolutiva que acabe por desbordar toda tutela humana. Esta posibilidad inquietante obliga a pensar en el mañana con la gravedad de quien sabe que prevenir es, a veces, lo único que nos queda para conservar lo esencial.

03 UNA VISIÓN SOBRE LOS RIESGOS

Del mismo modo que una premisa falsa puede sustentar cualquier conclusión, por absurda que sea, también los titulares rimbombantes —poco contrastados, a menudo más proclives al eco que a la verdad— generan interpretaciones erróneas que se propagan como verdades reveladas. Lo cierto es que, así como antaño se responsabilizaba a la informática de todos los males modernos (quien no se ha visto impedido de completar alguna gestión porque “se ha caído el sistema”), ahora parece que es la IA la que carga con el peso de nuestras nuevas angustias: el desempleo que se avecina, la deshumanización de las decisiones, la impunidad de sus actos y la amenaza de su autonomía sin conciencia. Estas advertencias, por lo general teñidas de un dramatismo tan seductor como superficial, encuentran terreno fértil en una sociedad insuficientemente formada y saturada de información, y que aún carece de las herramientas necesarias para discernir entre el alarmismo y la legítima preocupación.

Pero no son solo estas cuestiones las que allanan el camino a estos oscuros presagios. También operan, con sigilosa eficacia, las estrategias de ciertos medios de comunicación y de grandes corporaciones, cuya prioridad no es la información veraz ni la formación crítica, sino la captura de atención, la fidelización del usuario, la maximización del clic, o en términos más simples, la maximización de beneficios.

En una disciplina tan dinámica como la IA —donde los avances de hoy sepultan los logros de ayer, sin tiempo para la asimilación— este tipo de comunicación volátil nos empuja a una especie de resignación activa, donde no nos queda más que aceptar lo que nos ofrecen, nos instalan o incluso nos implementan, con o sin nuestro consentimiento. Sirva como ejemplo el aluvión diario de anuncios sobre estos temas que nos asalta desde múltiples escenarios en cualquier parte del mundo: cada titular, cada lanzamiento, nos reafirma en la sensación de estar siempre por detrás, siempre dependiendo de otros para comprender lo que supuestamente ha de cambiarnos la vida.

En este vértigo de innovación, la IA se ha convertido en un sello de prestigio, un marchamo de modernidad que todos quieren exhibir. Se adhiere a marcas y productos, se inserta en discursos, se cuela en los proyectos más dispares como una suerte de fetiche tecnológico, incluso allí donde no se sabe —ni se cuestiona— si realmente es necesaria, si es útil, si mejora algo o si sus usuarios están preparados para interpretarla. Poco importa: lo esencial es poder proclamar, con convicción y sin matices, que algo funciona con IA.

Esta tendencia, sin embargo, no está exenta de consecuencias. Se genera confusión, se siembran dudas, se alimenta el escepticismo. En ocasiones, cuando se afirma que cierto sistema utiliza IA, lo que realmente se está describiendo es un software convencional, basado en una serie de reglas preprogramadas, que puede resolver perfectamente el problema para el cual se ha desarrollado, pero que no tiene ninguna de las características asociadas a un ADS. Por otro lado, se pueden desarrollar sistemas donde la IA se utiliza de forma interesada, orientando sus resultados, no hacia la objetividad o la exploración abierta, sino hacia las conclusiones preconcebidas que sus creadores ya habían codificado en ella. Así, el abanico de consecuencias es vasto: va desde los errores no deseados —inevitables en cualquier empresa humana, aunque se disfrace de máquina— hasta las manipulaciones deliberadas, meticulosamente calculadas para desviar la verdad y condicionar nuestras decisiones.

No es necesario retroceder mucho en el tiempo para encontrar ejemplos que ilustran este fenómeno con toda su crudeza. Basta recordar el caso de Cambridge Analytica, aquella empresa que, sin alardear de IA, pero sirviéndose de su lógica incipiente, logró alterar el curso de procesos democráticos mediante una combinación tan poderosa como inquietante: redes sociales, datos personales y algoritmos de predicción conductual. El Brexit no fue solo una decisión política; fue también, y quizá, sobre todo, una advertencia sobre la fragilidad de nuestras voluntades cuando se enfrentan a tecnologías que nos conocen mejor de lo que nosotros mismos nos conocemos.

Esa capacidad de influir —más que de decidir, pero no por ello menos peligrosa— ha dejado de ser una anécdota para convertirse en tendencia. Y lo que viene podría ser aún más perturbador. Nos hallamos al inicio de una nueva confluencia: la que une la IA con futuras redes sociales de nueva generación, mucho más veloces, personalizadas y penetrantes que las actuales. El resultado de esa fusión, si no se orienta con rigor ético, podría redefinir no solo nuestras conversaciones, sino también nuestras identidades.

Pero hay más. Entre las sombras de estas transformaciones tecnológicas, asoma una alianza particularmente delicada: la de la IA con la neurociencia y el estudio del comportamiento. Este cruce de caminos apunta a influir directamente en nuestros procesos mentales, nuestras decisiones y nuestras emociones más íntimas. Lo que algunos ya denominan la tiranía digital no es una distopía, sino una posibilidad tangible que amenaza con justificar incluso la democratización del daño, legitimando nuevas formas de violencia bajo el barniz de la personalización avanzada. Y lo más inquietante es que no estamos hablando de amenazas que provengan solo de sofisticados hackers o de élites tecnológicas. Estamos hablando de cualquier grupo, de cualquier ideología, capaz de apropiarse de estas herramientas para imponer su visión del mundo.

Muchas de esas prácticas pueden y deben entenderse como formas de fraude, unas más burdas, otras más sutiles, pero todas sustentadas en la falta de formación de los usuarios, que no saben bien qué es la IA, ni en qué se diferencia de un software tradicional. Por eso, aceptan —aceptamos— con demasiada facilidad lo que se presenta como un sistema inteligente, dotado de buenas intenciones, cuando en realidad, muchas veces, es solo una versión sofisticada de intereses opacos. Sin embargo, planteados y conocidos muchos de los problemas y los riesgos que pueden acarrearse, la solución puede estar a nuestro alcance.

Como a continuación describiremos, es la sociedad, a través de los agentes sociales y los gobiernos, quienes deben decidir para qué, cuando, como y donde se pueden desarrollar soluciones basadas en IA para problemas de interés. Para ello, hacen falta programas específicos transversales, horizontales, cooperativos, regionales, nacionales e internacionales. Pero no solo deben plantearse como defensa ante los imprevisibles desarrollos no deseados basados en IA, sino como garantía de mejora y progreso de la humanidad en su conjunto.

04 LA NECESIDAD DE LA REGULACIÓN

¿Qué hacer entonces ante este horizonte movedizo y preocupante? No es fácil aventurar el rumbo que tomará esta historia. Pero sí podemos señalar dos caminos imprescindibles. Por un lado, la investigación científica, rigurosa y metódica, se erige como la mejor herramienta de anticipación. Allí donde aún no hay certezas, el conocimiento es el único antídoto contra el miedo. Por otro lado, una vez que los sistemas con IA estén en funcionamiento —ya lo están, en verdad— necesitamos diseñar, sin demora, mecanismos eficaces de control y supervisión, que nos permitan intervenir a tiempo, corregir desvíos, minimizar daños, contener los excesos.

En otras palabras, resulta imperativo avanzar hacia una gobernanza de la IA que esté a la altura del desafío. No se trata aquí de una gobernanza técnica o fragmentaria, sino de una integral, multidimensional, humanista, es decir, que incluya, entre otras, las siguientes prioridades: a) fomentar la educación crítica y la conciencia social b) promover la participación ciudadana, c) proteger con firmeza la democracia frente a manipulaciones automatizadas, d) garantizar la transparencia y la explicabilidad de los sistemas, e) asegurar el derecho a réplica frente a decisiones algorítmicas, f) establecer estrategias nacionales coherentes, g) adoptar enfoques multisectoriales que integren a todos los actores y, sobre todo, h) avanzar en marcos regulatorios globales que no sean meros brindis al sol, sino compromisos efectivos. Además de todo ello, se debe prestar especial atención a la protección de los datos personales y al cumplimiento de principios éticos que, lejos de ser un añadido decorativo, deben constituir el corazón mismo de esta nueva era tecnológica.[5]

En lo que respecta al comportamiento ético de los sistemas con IA, conviene admitir una verdad tan incómoda como evidente, y es que su evaluación es profundamente dependiente del contexto. Lo que en un determinado escenario puede parecer una decisión impecable, en otro puede revelarse como un acto reprobable. Esta variabilidad no solo genera tensiones morales difíciles de resolver, sino que deja al descubierto una carencia estructural: fijado el contexto, necesitamos tanto herramientas eficaces para valorar el grado ético de un sistema, como —y esto es quizá aún más grave— una legislación suficientemente desarrollada que lo regule con claridad y rigor.

5 Daniel Innerarity, Inteligencia artificial y democracia (UNESCO Office Montevideo and Regional Bureau for Science in Latin America and the Caribbean, Consejo Latinoamericano de Ciencias Sociales, 2024).

Afortunadamente, y como luego veremos, comienzan a surgir avances significativos en esa dirección, tanto a nivel nacional como internacional. No obstante, el camino está plagado de contradicciones jurídicas y de omisiones que no pueden pasarse por alto. Sirva como ejemplo el impulso institucional a la protección de los datos personales, mientras se fomenta simultáneamente en la ciudadanía el uso masivo de servicios de correo electrónico, generalmente gestionados por corporaciones privadas cuyos intereses no siempre están alineados con los derechos fundamentales. O también puede señalarse la confusión, más frecuente de lo que cabría esperar, entre la mera digitalización administrativa y una verdadera transición digital, inclusiva y estructural.


Foto: Magnific

Sin embargo, no debemos ceder al desencanto. En un primer paso, es urgente y necesario que los gobiernos activen mecanismos que garanticen que las soluciones basadas en IA —especialmente aquellas que se utilizan en los procesos de toma de decisiones en la administración pública— incorporen, desde su diseño, criterios de minimización de sesgos, exigencias de transparencia y compromisos claros de rendición de cuentas, siempre que ello sea técnicamente viable. Y, para lograrlo, sería conveniente que junto al desarrollo de dichas soluciones se hiciera pública —al menos ante organismos reguladores— la información relativa a sus datos de entrenamiento, a fin de analizar, anticipar y mitigar su posible impacto discriminatorio.

Asimismo, se deberían fomentar los diseños que prioricen la interpretabilidad de las decisiones algorítmicas, facilitando no solo su comprensión técnica, sino su revisión desde marcos éticos y legales. Porque solo una IA que pueda ser explicada podrá ser controlada. Y solo una IA que respete los derechos fundamentales podrá ser confiable.

En última instancia, lo que verdaderamente nos concierne es que los marcos normativos —la legislación, la regulación y la fiscalización— relativos a la producción, comercialización y aplicación de la IA, estén articulados de manera coherente en todos los niveles: internacional, nacional, regional y local. Es este un reto colosal, sin duda. ¿Pero es realmente imposible? ¿O acaso hemos comenzado a convencernos de su imposibilidad porque a algunos les resulta útil que así lo creamos?

Si bien la tarea de articular tal gobernanza global, cooperativa y eficaz de la IA parece titánica, desde luego no es inédita. La humanidad ya ha enfrentado desafíos similares. Ahí están el Organismo Internacional para la Energía Atómica o la Organización Mundial de la Salud como ejemplos imperfectos, pero sólidos, de estructuras capaces de coordinar voluntades, de armonizar intereses, de emitir directrices que resuenan en todas las latitudes. Si se ha podido hacer con la salud y la energía nuclear, campos igualmente delicados, estratégicos y complejos, ¿por qué no intentarlo con la IA?

Lo que necesitamos es voluntad: voluntad de renunciar a lo inmediato en favor de lo sostenible, voluntad de cooperar, aunque implique ceder parcelas de soberanía, voluntad de construir un futuro compartido que no esté dominado por unos pocos, sino diseñado para el bien común. En este sentido, desde principios de 2016 y a diferentes niveles (nacional, regional e internacional), se han comenzado a adoptar estrategias, planes de acción y documentos de política sobre IA que abordan temas como la regulación y la gobernanza, la educación, la investigación, cuestiones industriales y relacionadas con la infraestructura.[6] No obstante, con respecto a las tres economías más grandes, parece que "Estados Unidos está siguiendo un enfoque impulsado por el mercado, China está avanzando en un enfoque impulsado por el Estado y la UE está siguiendo un enfoque impulsado por los derechos".[7]

Una de las vías que se plantean, en términos de regulación, es la puesta en marcha de agencias de supervisión de la IA, esperando que dispongan de plena independencia orgánica y funcional respecto a las administraciones públicas, para que actúen con rigor objetivo, transparencia absoluta e imparcialidad incuestionable, mediante el despliegue de acciones dirigidas a la contención y mitigación de los riesgos que la IA pueda entrañar para la seguridad, la salud y los derechos fundamentales de las personas. Además, habrán de velar por el seguimiento y supervisión de aquellos proyectos que formen parte de las estrategias nacionales para el desarrollo de la IA, con especial atención a la elaboración de marcos normativos y a la regulación de sus múltiples usos.


Foto: Magnific

6 Jamie Berryhill et al., Hola, mundo: la inteligencia artificial y su uso en el sector público, documentos de trabajo de la OCDE sobre Gobernanza Pública n.o 36 (Paris: OECD Publishing, 2019). **7** Anu Bradford, "The Race to Regulate Artificial Intelligence", Foreign Affairs, junio 27de 2023.

Es de esperar que estas agencias no se limiten a un papel de meros guardianes técnicos, sino que articulen su labor en torno a varios ejes fundamentales que les confieran vocación de pensamiento y acción,[8] como, por ejemplo:

- Convertirse en verdaderos laboratorios de reflexión y propuestas de acción, capaces de anticipar tendencias, difundir conocimiento y avivar el debate público sobre las posibilidades y límites éticos, sociales y políticos de la IA.
- Asumir con eficacia la supervisión de los sistemas de alto riesgo, coordinarse con las autoridades de vigilancia del mercado, promover estándares comunes y buenas prácticas, así como evaluar rigurosamente los modelos de IA.
- Ejercer una función de armonización normativa, colaborando estrechamente con otras agencias homólogas nacionales e internacionales, a fin de garantizar una aplicación coherente, ágil y actualizada de los reglamentos existentes, en sintonía con los rápidos cambios tecnológicos y de mercado.
- Realizar una promoción de la innovación responsable, tanto en el ámbito público como en el privado, facilitando el cumplimiento normativo y estimulando la excelencia mediante la creación de entornos sectoriales de prueba (sandboxes regulatorios).
- Mantener una voz activa en el diálogo internacional sobre IA, participando en foros multilaterales donde se gesten los futuros estándares globales, con el propósito de evitar la fragmentación normativa y de contribuir a la construcción de una IA confiable, responsable y profundamente humanista.

A modo de ejemplo concreto, el lector interesado puede consultar el caso de la Agencia Española de Supervisión de Inteligencia Artificial (https://aesia.digital.gob.es),[9] la cual es la encargada de garantizar el uso ético y seguro de la IA en España. Su principal misión es asegurar que tanto entidades públicas como privadas cumplan con la normativa vigente, protegiendo la privacidad, la igualdad de trato y los derechos fundamentales.

Otro ejemplo de regulación es el ya citado Reglamento UE 2024/1689 de Inteligencia Artificial (),[10] en vigor desde el 1 de agosto de 2024, que a pesar de no ser una legislación como tal, se considera como la primera ley mundial dedicada a la IA. Entre sus objetivos principales está asegurar que los sistemas de IA se diseñen y empleen de forma responsable, garantizando que sean seguros, éticos y dignos de confianza. Esto, a su vez, protege la salud, la seguridad y los derechos fundamentales establecidos en la Carta de los Derechos Fundamentales de la Unión Europea.

Entre otras cosas, la norma establece categorías, en función del riesgo, para los sistemas de IA, y asigna responsabilidades a los proveedores y usuarios de estas tecnologías. Además, el reglamento también está diseñado para promover la innovación y la adopción de una IA que sea tanto ética como sostenible.

8 S. R. Pontevedra, “La Agencia Española de Supervisión de la IA contratará a 80 expertos en dos años”, El País, 19 de junio de 2024. **9** Agencia Española de Supervisión de Inteligencia Artificial (AESIA), “Garantizando una IA ética y responsable”. **10** Comisión Europea, “Ley de IA”.

El reglamento define cuatro niveles de riesgo para los sistemas con IA:

1. Riesgo inaceptable, prohíbe los sistemas que se consideran una clara amenaza para la seguridad, los medios de subsistencia y los derechos de las personas. Entre las ocho prácticas prohibidas, podemos mencionar la puntuación social, la evaluación o la predicción del riesgo de infracción penal individual, el reconocimiento de emociones en lugares de trabajo e instituciones educativas y la identificación biométrica remota en tiempo real con fines policiales en espacios de acceso público.
2. Alto riesgo, cuando los sistemas puedan producir riesgos graves para la salud, la seguridad o los derechos fundamentales. Estos sistemas están sujetos a obligaciones estrictas antes de que puedan comercializarse. Se incluyen aquí casos de uso de la IA en las fuerzas y cuerpos de seguridad que pueden interferir con los derechos fundamentales de las personas, herramientas de IA para el empleo, la gestión de los trabajadores y el acceso al trabajo por cuenta propia, soluciones de IA utilizadas en instituciones educativas, que pueden determinar el acceso a la educación y el curso de la vida profesional de una persona, etc.
3. Riesgo limitado, asociado a sistemas en los que es necesario informar que existe IA en funcionamiento. Por ejemplo, cuando se utilicen sistemas con IA como los chatbots, los humanos deben ser conscientes de que están interactuando con una máquina para que puedan tomar una decisión informada. En este sentido los proveedores de IA generativa deben garantizar que los contenidos producidos sean identificables y que determinados contenidos sean etiquetados de manera clara y visible, a saber, los relacionados a las falsificaciones profundas y al texto publicado con el fin de informar a la comunidad sobre asuntos de interés.
4. Sistemas de riesgo mínimo o nulo, que son la gran mayoría de los sistemas con IA utilizados actualmente en la UE, como, por ejemplo, videojuegos, filtros de spam, etc.

La Ley de IA será plenamente aplicable a partir del 2 de agosto de 2026, pero con algunas excepciones que, por cuestiones de espacio, no comentamos aquí.

05

HACIA UN FUTURO MÁS HUMANO: LA IA AL SERVICIO DEL BIENESTAR

Hace apenas tres décadas, el triunfo de una máquina (Deep Blue) sobre el campeón mundial de ajedrez marcaba la frontera del asombro. Aquella victoria fue, para muchos, la primera señal palpable de que la IA empezaba a emerger del reino de la ciencia ficción. Hoy, ese recuerdo ya parece remoto. La IA no solo juega ajedrez: se comunica mediante el lenguaje humano, redacta textos con coherencia y estilo, sugiere diagnósticos médicos complejos y comienza a tomar decisiones que antes eran terreno exclusivo de las personas. Esta metamorfosis vertiginosa ha sido posible gracias a la convergencia de cuatro elementos: la disponibilidad casi infinita de datos, una capacidad de cómputo en constante expansión, algoritmos cada vez más ingeniosos y eficientes, y la ambición —junto con financiación suficiente— de gigantes tecnológicos que han apostado por esta revolución silenciosa.

Dejando a un lado la IA general, podemos asumir que las herramientas basadas en IA están reconfigurando los modos de producción, optimizando la eficiencia, reformulando la forma como aprendemos, trabajamos y consumimos. Debemos considerar su desarrollo —ético, inclusivo y regulado— como una posible respuesta ante desafíos tan reales como el envejecimiento de la población, la crisis climática o las nuevas tensiones geopolíticas. Fijado un problema específico, negarse a explorar su potencial sería, para cualquier sociedad, renunciar a su competitividad futura.

Persisten, no obstante, interrogantes fundamentales: ¿podrá algún día generar conocimiento verdaderamente nuevo?, ¿albergará conciencia, voluntad, empatía, o una comprensión cultural genuina del mundo que habita? La respuesta, hoy por hoy, es incierta. Y en esa incertidumbre habita nuestra responsabilidad: la IA debe adoptarse, sí, pero no a cualquier precio. Como más adelante comentaremos con mayor detenimiento, debe evitarse que sus beneficios se concentren en manos de unos pocos, mientras los riesgos se reparten entre muchos. Debe crecer bajo vigilancia democrática y con frutos repartidos en todo el mundo, entre empresas de cualquier tamaño, en favor del bien común.

> Algunas naciones ya han trazado su rumbo, y cuentan con infraestructuras digitales robustas, con ciudadanos familiarizados con las habilidades digitales básicas, y con un tejido empresarial que —al menos en parte— comienza a integrar estas tecnologías.

En esa dirección, hay propuestas que, más allá de pensar en como implementar la IA, se atreven a preguntarse para qué. Una de ellas, proveniente desde España, y que describiremos brevemente en lo que sigue como un posible marco de referencia para otras realidades, se proyecta hacia 204011 con una visión transformadora apoyada en cinco hilos conductores: a) modernizar el tejido productivo, b) fortalecer los pilares del bienestar social, c) avanzar hacia la sostenibilidad ambiental, d) reforzar la seguridad colectiva y e) reducir las desigualdades estructurales. Todo ello se articula a través de un enfoque riguroso, cuantitativo, y con una mirada de largo plazo, sostenida en evidencias empíricas, casos de uso contrastados y un compromiso profundo con los retos del futuro.

A continuación, se describen los aspectos esenciales considerados en dichos ejes:

A. MODERNIZAR EL TEJIDO PRODUCTIVO

Durante los próximos diez o quince años se producirán profundas transformaciones estructurales en todo el mundo. En muchos países, el envejecimiento demográfico reducirá paulatinamente la población en edad de trabajar, los flujos de financiación probablemente cambien, la capacidad de crecimiento se verá constreñida por la limitada dotación de capital físico y los compromisos climáticos forzarán el abandono de un modelo económico basado en la explotación lineal de los recursos. En este panorama, la IA se perfila no solo como herramienta, sino como palanca imprescindible para sostener la productividad y contrarrestar estas restricciones.

11 Oficina Nacional de Prospectiva y Estrategia, “HispanIA 2040. Cómo la inteligencia artificial mejorará nuestro bienestar futuro” (Gobierno de España, 2025).

Aunque a largo plazo, se espera que la IA tenga un impacto netamente positivo sobre el empleo, dando lugar a nuevas profesiones y transformando otras ya existentes, por lo que hay que enfrentar importantes retos. El primer gran desafío reside en extender las transformaciones a las pequeñas y medianas empresas, así como a sectores tradicionales que, por su estructura o recursos, podrían quedar al margen. Un segundo obstáculo crucial es la escasez de talento especializado, en la medida en que los ingenieros, científicos de datos, expertos en ciberseguridad y desarrolladores de software actuales son aún insuficientes para satisfacer la demanda creciente, carencia que ralentiza la incorporación efectiva de la IA en las empresas. El tercer gran reto es garantizar una infraestructura tecnológica robusta y equitativa, dado que cada región habrá de contar con activos estratégicos imprescindibles: redes de comunicación de alta capacidad, centros de supercomputación de referencia y niveles destacables en ciberseguridad que, en todo caso, tengan garantizados sus mantenimientos y actualizaciones.

Bien orientada, la IA no solo permitirá aumentar la productividad, sino transformar la experiencia del trabajo. Al liberar a las personas de tareas repetitivas o administrativas, abrirá espacios para el teletrabajo, la flexibilidad horaria y nuevas formas de conciliación. Más allá de su impacto económico, la IA puede ser también una herramienta para dignificar la vida laboral, siempre que se la gobierne con inteligencia, con justicia y con visión de futuro.



Foto: Magnific

B. CONSOLIDAR Y MODERNIZAR EL ESTADO DE BIENESTAR

En el siglo XXI, el Estado de bienestar —uno de los grandes logros de las democracias— enfrenta desafíos como el envejecimiento de la población, el aumento del coste sanitario y nuevas demandas sociales. Su sostenibilidad dependerá de transformarlo sin traicionarlo. En este proceso, la IA puede ser una aliada clave, aportando eficiencia, agilidad y personalización, sin perder el enfoque humano. Estas herramientas permitirán a las administraciones aligerar la burocracia, mediante la automatización, la digitalización y la anticipación de necesidades, dando paso a un sistema más ágil, preventivo y centrado en el ciudadano. No se trata entonces de potenciar un Estado más grande, sino más inteligente.

En educación, la IA liberará a los docentes de tareas repetitivas, permitiéndoles enfocarse en lo esencial: enseñar y acompañar. En sanidad, optimizará recursos, mejorará la detección precoz de enfermedades y promoverá hábitos saludables. Ya se utiliza para diagnosticar cáncer con precisión, y pronto se extenderá a otras patologías. Menos carga administrativa permitirá a los médicos atender mejor y con menos demoras, reduciendo también las listas de espera.

El sistema de cuidados, llamado a ser el cuarto pilar del bienestar, también se transformará. En 2040, se espera un aumento significativo de personas dependientes. La tecnología —robots, sensores e IA— facilitará su autonomía y combatirá la soledad, además de simplificar trámites cotidianos sin reemplazar la atención humana.

No obstante, el avance de la IA plantea retos: proteger los datos personales, garantizar la transparencia algorítmica y rendir cuentas. Ya se desarrollan marcos legales, estándares internacionales y mecanismos de supervisión para asegurar un uso ético.

En este contexto, la IA puede convertirse en un pilar fundamental del Estado de bienestar, como lo fue la educación universitaria para las clases medias: no un lujo, sino una necesidad para hacerlo más justo, humano y resiliente.

C. AVANZAR HACIA LA SOSTENIBILIDAD MEDIOAMBIENTAL

En los próximos años, el mundo debe reducir la huella ecológica para proteger nuestros ecosistemas y evitar el colapso ambiental. Esto implica reducir significativamente las emisiones de gases de efecto invernadero, el consumo de agua y la intensidad en el consumo de energía primaria. Para lograrlo, se requiere cambiar la forma en que nos movemos, producimos, consumimos y nos relacionamos con la naturaleza, descarbonizando la energía, rediseñando cadenas de valor, modernizando infraestructuras y adoptando comportamientos responsables.

La IA por sí sola no logrará la sostenibilidad, pero nos acercará a ella al permitir diseñar y gestionar de forma más eficiente los sistemas de producción, distribución y almacenamiento energético. Habrá que desarrollar modelos energéticos complejos, con millones de fuentes renovables intermitentes, integrados por redes inteligentes que conecten diferentes áreas geográficas, coordinadas por una IA que considere diversas variables. La IA también optimizará el consumo de recursos naturales en la agricultura (agua, pesticidas, fertilizantes), monitorizará la salud de los bosques, predecirá incendios y optimizará las rutas de transporte para reducir emisiones. También los vehículos autónomos podrán reducir las emisiones y la congestión en las ciudades. La IA puede mejorar la gestión del agua, mediante acciones como la optimización del riego y la detección de fugas en las redes de distribución.

Sin embargo, su desarrollo plantea retos ambientales adicionales. El entrenamiento y uso de grandes modelos de lenguaje consume muchos recursos naturales. La fabricación y mantenimiento de los equipos informáticos también requiere energía y materias primas, con un alto consumo de agua en los centros de datos y la producción de microchips. Se espera que el despliegue futuro de la IA se produzca a través de modelos y centros de datos más pequeños y eficientes.

Aun así, se espera que los incrementos en eficiencia superen estos costes y que la contribución a la lucha medioambiental sea positiva. Las tecnologías descritas ya existen, el reto es promover su adopción y seguir desarrollando nuevas aplicaciones, facilitando que la IA pueda aportar nuevas herramientas para combatir la degradación ambiental.

D. POTENCIAR LAS INFRAESTRUCTURAS CRÍTICAS DE SEGURIDAD

En un mundo que oscila entre la incertidumbre y el vértigo del cambio, la seguridad se ha convertido en uno de los bienes más preciados. Aunque aún existen regiones con baja criminalidad y aparente estabilidad, los desafíos actuales provienen de entornos menos apacibles. El auge de amenazas híbridas, la creciente ciberdelincuencia y la inestabilidad geopolítica exigen reforzar, con inteligencia y previsión, los sistemas de seguridad y defensa.

En este contexto, la inteligencia artificial emerge como un aliado crucial. Su capacidad para anticipar lo inesperado, detectar lo oculto y responder con precisión puede marcar la diferencia entre prevenir o reaccionar tarde. Los sistemas de vigilancia con algoritmos avanzados permiten identificar en tiempo real embarcaciones sospechosas, incendios en expansión o comportamientos anómalos. La IA recoge vastos volúmenes de datos para detectar patrones, desenmascarar fraudes e incluso prever delitos antes de que ocurran, aunque estas aplicaciones plantean dilemas éticos y de privacidad.

En el ciberespacio —ese campo de batalla invisible pero crucial—, la IA permite detectar intrusiones, responder a ataques sofisticados y proteger infraestructuras esenciales. Ante amenazas híbridas, en las que convergen lo militar, lo digital y la desinformación, la IA puede guiar decisiones estratégicas y mejorar la coordinación entre actores diversos.

También desempeña un papel clave en la logística: mediante sensores y modelos predictivos, anticipa fallos técnicos, optimiza el mantenimiento y prolonga la vida útil del material estratégico. En el ámbito militar, puede aumentar la eficacia operativa, incluso en sistemas autónomos de combate. Sin embargo, las decisiones letales deben seguir siendo humanas. Es urgente establecer un marco internacional que regule el uso militar de la IA y prohíba su aplicación en armas nucleares.

En este tiempo de tensiones múltiples y tecnologías exponenciales, reforzar la defensa y la ciberseguridad no es opcional. Si se gobierna con responsabilidad, la IA puede convertirse en el escudo más sutil y poderoso de nuestra era.

E. MINIMIZAR LAS DESIGUALDADES

En el corazón de nuestras modernas sociedades, la desigualdad sigue latiendo con fuerza obstinada. Sus raíces son profundas: el modelo de crecimiento económico, la fragilidad del mercado laboral, la inaccesibilidad de la vivienda y la limitada eficacia redistributiva de los sistemas de protección social. Aunque no faltan esfuerzos para corregir estas brechas, el camino hacia una sociedad más equitativa permanece sembrado de desafíos. Y sobre este paisaje desigual se cierne ahora una pregunta urgente: ¿será la IA un puente hacia la justicia social, o un nuevo abismo que produzca mayores desigualdades?



Foto: Magnific

Ante nosotros se despliegan tres escenarios posibles. El primero plantea la complementariedad para los trabajadores con mayor cualificación, admitiendo como una primera hipótesis que se reproduzca el patrón observado con anteriores innovaciones digitales, y que la IA actúe como una herramienta de apoyo y refuerzo, especialmente para quienes poseen un nivel formativo más alto. En este escenario, la sociedad tendería a dividirse en dos bloques: por un lado, aquellas personas que incorporen la IA a su repertorio profesional, incrementando así su productividad, su competitividad y, en consecuencia, sus ingresos; por otro lado, quedarían las personas que no adoptasen esta tecnología, bien porque desempeñen labores en sectores con escasa exposición a la IA, o bien porque carezcan de las competencias necesarias para aprovecharla. Esto conduciría, en última instancia, a un aumento de la desigualdad en la distribución de los salarios en el país.

El segundo es el escenario de la sustitución, en el que la IA supera a trabajadores altamente cualificados, relegándolos a la periferia del mercado. En esta visión, los ingresos laborales se erosionan incluso entre los mejor preparados, sin que ello beneficie a los estratos medios o bajos. El capital, dueño del código y de las máquinas, aumentaría, mientras el trabajo humano perdería valor. Así, paradójicamente, la desigualdad podría disminuir, pero no por la justicia, sino por la decadencia compartida. Afortunadamente, este futuro se dibuja tan difícil como lejano: la complejidad de las tareas humanas sigue siendo, en su mayoría, inaccesible para la IA tal como la conocemos hoy.

El tercer escenario es el de la complementariedad para las clases medias. Aquí, la IA no desplaza, sino que potencia. Herramientas como ChatGPT no sustituyen la experiencia, la amplifican; no anulan el juicio experto, lo democratizan. De este modo, tareas antes reservadas a quienes tenían formación avanzada —redactar, programar, analizar datos, emprender en línea— se vuelven accesibles para quienes poseen habilidades medias o están dando sus primeros pasos. Profesiones tradicionalmente vistas como secundarias podrían enriquecerse, y trabajadores de nivel medio pueden pasar de organizar agendas o archivar documentos a participar en la toma de decisiones y en el análisis estratégico. La IA, entonces, actúa como un catalizador de autonomía y dignidad.

No obstante, incluso en este escenario esperanzador, la desigualdad puede persistir, o incluso agravarse, si las rentas más altas se benefician en mayor proporción. El valor agregado que la IA puede generar —mayor productividad, riqueza colectiva y mejores servicios públicos— podría equilibrar este efecto, siempre que exista voluntad política para redistribuirlo. Pero el riesgo de que el capital se concentre aún más es real y visible: las grandes corporaciones tecnológicas, dueñas del pulso digital del mundo, ya acaparan una parte creciente del valor generado. La IA promete mucho, pero reparte poco por sí sola.

Para evitar que estas transformaciones amplíen la brecha social, es imprescindible orientar la IA hacia el bien común. Usarla para fortalecer los servicios públicos, personalizar la educación, mejorar la empleabilidad, asistir a los más vulnerables en trámites complejos o garantizar el acceso efectivo a prestaciones sociales. La IA puede ser una aliada de la igualdad si se pone al servicio de quienes más la necesitan, por ejemplo, estudiantes que requieren apoyo, personas con discapacidad que encuentran barreras en el lenguaje o la tecnología o incluso familias que dependen de ayudas para sostenerse.

06 CONCLUSIÓN

La IA se ha integrado permanentemente en nuestras vidas, presentando tanto oportunidades como desafíos y dudas trascendentales que exigen respuestas globales. Un problema fundamental identificado es la falta de conocimiento y formación de la ciudadanía en el uso de esta tecnología. No debemos olvidar que, antes de pensar en IA, debemos pensar cuál o cuáles son los problemas que se desean resolver, y si las técnicas de la IA son las más adecuadas, porque estas carencias limitan un uso responsable y eficaz, facilitan la delincuencia digital y contribuyen a ampliar las brechas sociales. Por lo tanto, es crucial que los gobiernos refuercen urgentemente la formación y capacitación de la ciudadanía en el ámbito de la IA.

Sin embargo, es fundamental destacar la cara más luminosa de la IA y su innegable contribución actual y futura al progreso y bienestar social en múltiples ámbitos. Los incontestables beneficios ya se extienden desde la medicina y la agricultura hasta la sostenibilidad ambiental y la seguridad, perfilándose como una palanca imprescindible para sostener la productividad, generar nuevos conocimientos e innovación y transformar positivamente el trabajo y los servicios públicos. Si se orienta adecuadamente, esta tecnología puede ser una respuesta inteligente a desafíos globales. En este camino, la ciencia se erige como la mejor herramienta de anticipación y el único antídoto contra el miedo ante la incertidumbre, siendo indispensable para comprender y guiar su desarrollo.

Para garantizar que la IA beneficie a toda la humanidad, su reglamentación continua y ágil es absolutamente necesaria. Lograr un control efectivo y adaptado a la velocidad de los avances tecnológicos hace indispensable la creación urgente de agencias de supervisión, independientes, técnicas y no políticas. Asimismo, y con una importancia capital, se debe garantizar de forma absoluta la seguridad y privacidad de los datos personales, especialmente en sectores sensibles como la sanidad, la seguridad nacional o el ámbito judicial, debiendo destinarse todos los esfuerzos y recursos posibles a proteger este aspecto crucial. El objetivo último es aprender a coexistir con la IA, guiándola hacia horizontes de dignidad, justicia y bien común.

Pero nada de esto ocurrirá de forma espontánea. El impacto de la IA sobre la desigualdad dependerá de las políticas que sepamos construir, a nivel nacional, regional e internacional. Solo una gobernanza inteligente podrá convertir el potencial transformador de la IA en un instrumento de justicia social. Las estrategias ya trazadas, si se desarrollan con ambición y humanidad, pueden ayudarnos a inclinar la balanza hacia un futuro más justo. Porque el verdadero progreso no será el de las máquinas, sino el de la sociedad que las gobierna.

REFERENCIAS

AAAI (Association for the Advancement of Artificial Intelligence). AAAI 2025 Presidential Panel on the Future of AI Research. https://aaai.org/about-aaai/presidential-panel-on-the-future-of-ai-research/.

Agencia Española de Supervisión de Inteligencia Artificial (AESIA). "Garantizando una IA ética y responsable". https://aesia.digital.gob.es/es

Berryhill, Jamie et al. Hola, mundo: la inteligencia artificial y su uso en el sector público. Documentos de trabajo de la OCDE sobre Gobernanza Pública n.o 36. Paris: OECD Publishing, 2019. https://dx.doi.org/10.1787/726fd39d-en

Bradford, Anu. "The Race to Regulate Artificial Intelligence". Foreign Affairs (newsletter), June 27, 2023. https://www.foreignaffairs.com/united-states/race-regulate-artificial-intelligence-sam-altman-anu-bradford.

Comisión Europea. "Ley de IA". https://digital-strategy.ec.europa.eu/es/policies/regulatory-framework-ai

Filippucci, Francesco et al. The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges. OECD Artificial Intelligence Papers, n.o 15. Paris: OECD Publishing, 2024. https://doi.org/10.1787/8d900037-en

Innerarity, Daniel. Inteligencia artificial y democracia. UNESCO Office Montevideo and Regional Bureau for Science in Latin America and the Caribbean, Consejo Latinoamericano de Ciencias Sociales, 2024. https://unesdoc.unesco.org/ark:/48223/pf0000389736_spa

Oficina Nacional de Prospectiva y Estrategia. "HispanIA 2040. Cómo la inteligencia artificial mejorará nuestro bienestar futuro". Gobierno de España, 2025. https://futuros.gob.es/nuestro-trabajo/hispania-2040

Pontevedra, S. R. La Agencia Española de Supervisión de la IA contratará a 80 expertos en dos años. El País, 19 de junio de 2024. https://elpais.com/tecnologia/2024-06-19/la-agencia-espanola-de-supervision-de-la-ia-contratara-a-80-expertos-en-dos-anos.html#?prm=copy_link.

Unión Europea. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo: Por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n. o 300/2008, (UE) n. o 167/2013, (UE) n. o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial) (junio 13 de 2024). https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32024R1689

Vinge, Vernor. "The Coming Technological Singularity: How to survive in the post-human era". Conference Paper at Vision-21: Interdisciplinary Science & Engineering in the Era of CyberSpace, NASA Lewis Research Center, December 1, 1993.

Origen de esta investigación

Este trabajo se ha desarrollado al amparo del Proyecto PID2023-146575NB-I00 financiado por MICIU/ AEI /10.13039/501100011033 y por FEDER,UE

ACERCA DE LOS AUTORES

David A. **Pelta Mochcovsky**

Catedrático de Universidad en el Depto. de Ciencias de la Computación e Inteligencia Artificial, ETSI Informática y de Telecomunicaciones de la Universidad de Granada. Obtuvo la Licenciatura en Informática (1998) por la Universidad Nacional de La Plata, Argentina y el grado de Doctor en Informática (2002) por la Universidad de Granada.

Director del Grupo de Trabajo en Modelos de Decisión y Optimización (MODO) donde desarrolla tareas de investigación en el ámbito del estudio y aplicación de las técnicas de Inteligencia Artificial a problemas de diversa procedencia. Ha publicado más de 70 artículos en revistas indexadas en el Journal Citation Reports y co-editado 8 libros con las editoriales Springer y Wiley. Ha dirigido o co-dirigido 13 tesis doctorales y 8 trabajos de fin de máster. En 2007 recibió el Premio (colectivo) del Consejo Social de la UGR como miembro del Departamento de Ciencias de la Computación e Inteligencia Artificial (2007) y en diciembre de 2022, recibió el "Sello CUJAE", distinción otorgada por la Universidad Tecnológica de la Habana "José Antonio Echeverría" (CUJAE), de Cuba.

Ha impartido seminarios y conferencias invitadas en Argentina, México, Brasil, Colombia, Cuba, Marruecos, etc. y desde 2003, imparte docencia a tiempo completo en la UGR, tanto a nivel de grado, como de máster. Además, es uno de los responsables en UGR del Living Lab en Inteligencia Artificial de la Alianza Universitaria Europea Arqus.

José Luis **Verdegay**

Profesor Emérito de la Universidad de Granada (UGR). Obtuvo la Licenciatura en Matemáticas y el Doctorado en Ciencias por la UGR en 1975 y 1981, respectivamente. Desde 1990 ha sido catedrático de Inteligencia Artificial en esta misma universidad. Entre los numerosos cargos desempeñados sobresalen el de Vicerrector de la UGR (2008-2015), director regional de la Asociación Universitaria Iberoamericana de Postgrado (2015-2023) y, desde abril de 2023 hasta su nombramiento como emérito, director del Departamento de Ciencias de la Computación e Inteligencia Artificial.
Según el Ranking Mundial de Científicos de la Universidad de Stanford, se encuentra entre el 2% de los investigadores más influyentes del mundo.
Entre otros reconocimientos, es Doctor Honoris Causa por la Universidad Central de las Villas (Cuba), Investigador Distinguido del Instituto Internacional de Investigación en Inteligencia Artificial de la Universidad de Hebei (China), medalla de oro de la Iranian Operations Research Society (IFORS) y miembro correspondiente de la Academia de Ciencias de Cuba.



Foto: Magnific

02

Entre el pesimismo y la esperanza.

E-ESTADO
Primera Edición

Entre el pesimismo y la esperanza. De cómo los valores democráticos pueden impedir los efectos más perversos de la inteligencia artificial[1]

Between Pessimism and Hope. On how Democratic Values May Avert the most Pernicious Effects of Artificial Intelligence

Fecha de recepción: 20 de octubre de 2025
Fecha de aceptación: 18 de febrero de 2026

Daniel Jove Villares[2]
Universidad de Oviedo, España
jovedaniel@uniovi.es
https://orcid.org/0000-0002-1164-7796

1 Este artículo es un resultado del proyecto PID2022-136548NB-I00 Los retos de la inteligencia artificial para el Estado social y democrático de Derecho financiado por MICIU/AEI/10.13039/501100011033/FEDER,UE.
2 Profesor ayudante doctor de Derecho Constitucional de la Universidad de Oviedo. Correo de contacto: jovedaniel@uniovi.es. https://orcid.org/0000-0002-1164-7796.

Cómo citar: Jove Villares, Daniel. “Entre el pesimismo y la esperanza. De cómo los valores democráticos pueden impedir los efectos más perversos de la inteligencia artificial”. Revista e-Estado 1, n.° 1 (2026).



Foto: Magnific

RESUMEN

El derecho, y especialmente, los derechos fundamentales, operan como un límite al poder. Sin embargo, la revolución digital y, sobre todo, la algoritmización de la vida en comunidad tienen capacidad para provocar un cambio de paradigma, pues afectan a la forma de producir y aplicar el derecho (demandando una mayor personalización) e, incluso, pueden alterar la cultura jurídica de la sociedad, induciendo una mutación que convierta a los ciudadanos y ciudadanas en consumidores. Frente a esa potencialidad transformadora, la alfabetización digital, los valores democráticos y la sumisión del poder al derecho se presentan como las vías de preservación de las necesarias esferas de libertad personal. Este artículo se dedica a explorar estas y otras cuestiones, como, por ejemplo, el impacto del riesgo en el modo de entender los derechos.

Palabras clave: Estado de Derecho, cultura jurídica, riesgo, personalización, democracia

ABSTRACT

Law—and, in particular, fundamental rights—functions as a restraint upon the exercise of power. Yet the digital revolution, and above all the algorithmisation of social life, possesses the capacity to precipitate a paradigm shift, as it transforms both the production and the application of law, calling for a greater degree of personalisation. It may even alter the very foundations of a society's legal culture, fostering a mutation whereby citizens are recast as consumers. Confronted with this transformative potential, digital literacy, democratic values, and the subjection of power to the Rule of Law stand out as the principal means by which the essential spheres of individual freedom may be preserved. This article seeks to explore these and cognate issues, including, inter alia, the implications of risk for contemporary understandings of rights.

Keywords: Rule of Law, Legal Culture, Risk, Personalization, Democracy

01 NO TODAS LAS INTELIGENCIAS ARTIFICIALES SON IGUALES

La inteligencia artificial (IA) tiene el potencial de convertirnos en jugadores de póker (y ese no sería el peor escenario). Los sistemas de IA (al menos los existentes en la actualidad),[3] ofrecen resultados en forma de predicción, contenido, recomendación o decisión que, en la mayoría de los casos, son probabilísticos, fruto de correlaciones[4] y procesos inferenciales más o menos complejos, pero, en cualquier caso, traducibles a un porcentaje.[5] En ese sentido, "los algoritmos nunca buscan causalidades",[6] no explican, anticipan un resultado y, por tanto, pueden errar. Incluso los grandes modelos de lenguaje, pese a lo que pueda parecer, no crean, solo traducen esa posibilidad en una respuesta visible. Por ejemplo, que un modelo de lenguaje a gran escala (Large Languaje Model) escriba m antes de p en la palabra empanada deriva de un cálculo expresado en un porcentaje muy cercano al 100 % de que eso ha de ser así. Seguramente, no llegue tener una certeza absoluta de ello, porque el modelo tendrá en cuenta todos los escritos de los que ha aprendido y seguramente en alguno haya erratas, pero la probabilidad será tan alta que, a efectos prácticos, habrá llegado a la conclusión de que esa es la regla. No necesita conocerla, sencillamente la infiere.

Si tomamos como referencia el concepto de sistema de inteligencia artificial formulado en el artículo 3.1 del Reglamento de Inteligencia Artificial de la Unión Europea (RIA),[7] solo tendría tal consideración aquel sistema basado

> *"en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales".* [8]

3 Quizá una inteligencia artificial general (AGI) operase conforme a otro tipo de parámetros y fuese capaz de razonar. En todo caso, por el momento no parece que las AGI vayan a ser una realidad en el corto/medio plazo. Cfr. Meredith Ringel Morris et al., "Position: Levels of AGI for operationalizing progress on the path to AGI", Forty-first International Conference on Machine Learning, Viena, Austria, 2024. **4** "La correlación alude a la correspondencia entre dos o más acciones o fenómenos. Dos variables están correlacionadas cuando las variaciones en una de ellas hacen variar, bien en sentido creciente, bien en sentido decreciente, la otra […]. El output que arrojan, entonces, los sistemas de inteligencia artificial […] no [es] un efecto o una consecuencia propiamente dichos". Patricia García Majado, "Inteligencia artificial, predicciones y funciones normativas", Teoría y Realidad Constitucional, n.o 54 (2024): 423-424**. 5** De este modo, la IA se diferenciaría de "los sistemas de software o los planteamientos de programación tradicionales y más sencillos, y no debe incluir los sistemas basados en las normas definidas únicamente por personas físicas para ejecutar automáticamente operaciones". Considerando 12 del 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 y por el que se modifican los reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (reglamento de inteligencia artificial). **6** Iñigo De Miguel Beriain y Antonio Diéguez Lucena, "¿Explicar o predecir?", Investigación y Ciencia julio, n.o 538 (2021), 52-53. **7** 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 y por el que se modifican los reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828. **8** "El RIA pretende dotar de una definición única de la IA con suficiente claridad y precisión, que dé seguridad jurídica, que sea funcional y lo más tecnológicamente neutra posible, es decir, que no condicione o favorezca unas tecnologías frente a otras. De igual modo, se pretende una definición que resista al paso del tiempo lo mejor posible dada la dinamicidad tecnológica y del mercado". Lorenzo Cotino Hueso, "¿Qué es 'inteligencia artificial' para el Reglamento? Análisis, delimitación y aplicaciones prácticas", en Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea, eds. Lorenzo Cotino Hueso y Pere Simón Castellano (Navarra, España: Aranzadi, 2024), 116.

Esto quiere decir que quedan excluidos todos los demás sistemas. En efecto, como señalan las Directrices de la Comisión establecidas en el RIA, aquellos sistemas que no tengan capacidad de inferencia o que la tengan muy limitada no encajan en la definición dada.

Lo anterior supone que, para la regulación europea, muchos de los sistemas que vayan a emplearse en la automatización de procesos no se verán sometidos a sus exigencias. En concreto, quedan excluidos sistemas para mejorar la optimización matemática, aquellos basados en la heurística clásica, los sistemas de predicción sencillos o los de tratamiento básico de datos,[9] así como aquellos modelos que parten de reglas generales aplicadas a un caso concreto, esto es, los que tendrían como fórmula de razonamiento "si A, entonces B", es decir, la tipología de algoritmos que se materializa en sistemas expertos.[10] Que no se califiquen como sistemas IA no niega su utilidad, pues resultan especialmente útiles en muchos ámbitos, por ejemplo, para las Administraciones públicas,[11] ya que permiten "la traducción a código de una disposición jurídica preexistente [… o…] una modelización de datos clásica de carácter estadístico para su posterior volcado como pautas fijas en el software".[12] De esta manera, facilitan la aplicación de las condiciones legales y reglamentarias a un conjunto dado de situaciones, lo que puede resultar adecuado para la gestión de solicitudes o la automatización de procesos previamente reglados.

Dadas sus características, y con las cautelas que siempre se deben tener cuando se adoptan nuevas fórmulas de resolución, la IA deductiva y los sistemas expertos, aunque pueden causar daños, incluso graves,[13] que habrán de ser reparados y, de ser el caso, indemnizados, son tecnologías que podrían incorporase al ejercicio de los poderes del Estado sin que supongan un peligro irreparable. Especialmente, si se refuerzan los deberes de transparencia y se asegura que su funcionamiento sea auditable, predecible, y que se cumple con las exigencias de seguridad jurídica, equidad y rendición de cuentas que deben imperar en el ejercicio del poder público.

9 Véanse las páginas 8 a 11 de las directrices, en las que se justifica la exclusión de estos sistemas y, en general, el conjunto de las directrices para acotar el alcance del concepto sistema IA en el RIA. **10** Los sistemas expertos se pueden valer de órdenes predefinidas o de las soluciones ofrecidas previamente, basando su lógica operacional en la regla "si, entonces", sin perjuicio de que, a partir de ese modelo básico, se puedan desarrollar modelos más complejos, como las redes bayesianas. Al respecto, véase Sebastián Badaró, Leonardo Javier Ibañez, y Martín Jorge Agüero, "Sistemas expertos: fundamentos, metodologías y aplicaciones", Ciencia y Tecnología, n.o 13 (2013): 349-364. **11** La administración es uno de los ámbitos en los que existen más posibilidades de implementación de la IA. Como ya se ha puesto de manifiesto, es difícil encontrar esferas de actuación de la administración pública que no sean susceptibles de smartificación. Véase Carles Ramió, Inteligencia artificial y administración pública, 2.a ed. (Madrid. España: Catarata, 2025).**12** David San Martín Segura, "El algoritmo, entre disposición y norma. La calificación de la IA de decisión jurídica", en El derecho administrativo en la era de la inteligencia artificial: actas del XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo. Vigo, 25 a 27 de enero de 2024, coords. Patricia Valcárcel Fernández y Francisco Lorenzo Hernández González (España: Instituto Nacional de Administración Pública (INAP), 2025), 202.**13** Se puede tomar como ejemplo el caso BOSCO, en el que se utilizó un sistema algorítmico para decidir la concesión del bono social eléctrico. Acerca de la posibilidad de conocer su funcionamiento, su código, se ha pronunciado el Tribunal Supremo de España en la sentencia de 11 de septiembre de 2025 (recurso de casación 7878/2024). También en España se utilizan sistemas expertos, por ejemplo, en la prevención de la violencia de género (sistema VioGén). Sobre su funcionamiento y falencias desde un punto de vista jurídico, véase Miguel Ángel Presno Linera, "Policía predictiva y prevención de la violencia de género: el sistema VioGén", IDP. Revista d'Internet, Dret i Política, n.o 39 (2023): 1-13.



Foto: Magnific

Pero ¿por qué se produce esa exclusión? Porque esos modelos, por más complejos que puedan ser, son fiscalizables. Es posible saber cómo han llegado a adoptar esa decisión/resultado, se puede recorrer el conjunto de pasos lógicos que han seguido y, por tanto, comprobar si la respuesta ofrecida responde a las reglas previamente establecidas. Consecuentemente, podrá determinarse si el error o el daño proviene del sistema automatizado (una mala programación del algoritmo, por ejemplo, por no haber trasladado al código de programación de manera adecuada las normas cuya aplicación se pretende automatizar) o si se origina a partir de las reglas en las que se basa (por ejemplo, porque anide en ellas algún tipo de irregularidad, como podría ser la existencia en la norma de discriminaciones indirectas[14] o de arbitrariedades). En definitiva, no se trata aquí de una cuestión de utilidad o capacidad, sino de complejidad técnica y de determinación de la norma aplicable a la tecnología empleada.

Cuestión distinta son los sistemas de IA con capacidad de aprendizaje e inferencia y que operan con probabilidades. Respecto de ellos también puede existir el interés por implementarlos en el ejercicio de funciones públicas, pudiendo llegar a incidir en todos los principios estructurales del Estado.[15] Y no es para menos, sus capacidades creativas y de análisis pueden ser extraordinariamente interesantes para el diseño de políticas públicas, en procesos electorales,[16] como herramienta securitaria[17] o en la valoración y evaluación de la aptitud y actitud de las personas, ya sea para permitirles hacer algo, conferirles una prestación o limitarles un derecho (por consideraciones de culpabilidad respecto de haber llevado a cabo una determinada acción). Estos sistemas, por su naturaleza inductiva, entrañan mayores riesgos, especialmente porque también resulta más difícil la cognición de sus procesos de decisión (muchos operan como cajas negras), sin perjuicio de que, aunque no se conozca cada fase del proceso, sí se puede tratar de que el propio sistema, o uno que le acompañe como una especie de traductor, ofrezca ciertos datos que permitan conocer los fundamentos de la respuesta ofrecida, por ejemplo, identificando qué variables fueron las más influyentes en el resultado final, o a través de sistemas de explicación local.[18]

No resulta extraño, por tanto, que el RIA centre su regulación en los sistemas inferenciales, sea cuando operan solos o cuando lo hacen como componentes de una máquina más compleja. Por ejemplo, pensemos en un sistema de videovigilancia con reconocimiento facial que fuese capaz de analizar el entorno, identificar a un determinado sujeto y que, a la vez, pudiese determinar la acción que lleva a cabo y valorar si es correcta o incorrecta (v. gr., valorando si ha aparcado bien o lo ha hecho en un lugar no habilitado al efecto, o si esa persona ha cruzado un semáforo en rojo). Una vez realizadas esas inferencias, a raíz de las imágenes captadas, podría transmitirse la decisión a otro sistema encargado de definir las consecuencias, esto es, determinar que, para la acción A (aparcar en lugar no habilitado al efecto), la sanción es X, e imponerla.

14 "La discriminación indirecta se produce cuando una disposición, criterio o práctica aparentemente neutro ocasiona o puede ocasionar a una o varias personas una desventaja particular con respecto a otras" (artículo 6.1.b) de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación. **15** La IA tiene capacidad para afectar tanto al Estado social como al democrático y, por supuesto, al Estado de derecho. Así ha sido puesto de manifiesto en Miguel Ángel Presno Linera, "Derechos fundamentales e inteligencia artificial en el Estado social, democrático y digital de derecho", Anuario de la Red Eurolatinoamericana de Buen Gobierno y Buena Administración, 2023, https://ijeditores.com/pop.php?option=articulo&Hash=dce233eca31072b2c2f7e804f3efc07a. **16** Al respecto, vid. Rafael Rubio Núñez, Frederico Franco Alvim, y Vitor de Andrade Monteiro, Inteligencia artificial y campañas electorales algorítmicas. Disfunciones informativas y amenazas sistémicas de la nueva comunicación política (Madrid, España: Centro de Estudios Políticos y Constitucionales, 2024); Óscar Sánchez Muñoz, La regulación de las campañas electorales en la era digital. Desinformación y microsegmentación en las redes sociales con fines electorales (Madrid, España: Centro de Estudios Políticos y Constitucionales, 2021). **17** Ma Josefa Ridaura Martínez, "El reconocimiento facial como herramienta para garantizar la seguridad", en Tecnologías abusivas y derecho, eds. Francisco Caamaño Domínguez y Daniel Jove Villares (Valencia, España: Tirant Lo Blanch, 2024). **18** Estos se centran en entender las razones que explican una predicción individual, en lugar de tratar de comprender todo el modelo. Para ello, crean un modelo más simple que buscará explicar por qué se ha dado un resultado específico. Respecto del funcionamiento de estos sistemas interpretativos, vid. Muhammad Rehman Zafar y Naimul Khan, "Deterministic local interpretable model-agnostic explanations for stable explainability", Machine Learning and Knowledge Extraction 3, n.o 3 (2021): 525-541.

En la primera parte, captación, identificación de la persona, definición de la acción y clasificación de la misma operaría uno o, probablemente, varios sistemas de IA inductivos y complejos, esto es, de los del RIA. Para la última fase, la asignación de una sanción ante una acción dada, podría acudirse a un sistema experto (deductivo). Naturalmente, un sistema de IA inductivo también podría llevar a cabo esa acción, pero podría ser más seguro desde un punto de vista jurídico apostar, para esa última tarea, por un sistema experto, pues en última instancia se trata de aplicar reglas predefinidas, las cuales, además, habrían de estar reflejadas en una norma de rango legal (principio de legalidad penal).

No obstante, aunque la decisión final sea adoptada por un sistema experto (y, por tanto, es fiscalizable), no debe perderse de vista que, para que adopte la decisión final, depende de la información que le proporcionen los sistemas de IA. En este sentido, los modelos híbridos no están excluidos del ámbito de aplicación del RIA. En la medida en que se opere con un sistema o con un componente de IA, este habrá tenido que cumplir con las exigencias correspondientes y contar con la certificación que el RIA exige en materia de documentación, transparencia, certificación (marcado CE incluido) o conservación de archivos de registro.

Es cierto que las posibilidades de uso son enormes, pero, a la vez, resulta extraño que aquellos que ejercen el poder, en cualquiera de sus formas, estén dispuestos a servirse de herramientas que presentan como certezas lo que no dejan de ser predicciones. Aunque el porcentaje de acierto sea muy alto, siguen siendo apuestas. ¿Por qué valdría la pena ese riesgo? ¿Por qué resulta interesante al poder, especialmente a los gobiernos, la implementación de este tipo de sistemas de IA? ¿Más allá de la respuesta jurídica, se pueden establecer frenos a las posibilidades de control que los sistemas de IA ofrecen a los gobiernos? Las páginas de este trabajo buscan dar respuesta a estas cuestiones.

02

DE LA MANO DE HIERRO AL GUANTE DE SEDA

La consolidación de la Constitución como norma suprema del ordenamiento supone que “el fundamento de validez del sistema ya no es […] exógeno o metapositivo (heterorreferencial) sino interno, genuinamente jurídico”.[19] De este modo, el poder queda sometido al derecho y, con ello, el estatus de ciudadanía puede garantizarse normativamente. La positivización de la esfera de libertad que las personas preservan frente al Estado sería la materialización de la ficción formulada por Locke, conforme a la cual las personas “renuncian a la igualdad, la libertad y el poder ejecutivo que poseían en el estado de naturaleza y [los depositan] en manos de la sociedad, a fin de que el Legislativo disponga [de ellos] en la medida en que lo requiera el bien de la misma, sin embargo, […] todo ello no debe estar dirigido a otro fin que la paz, la seguridad y el bienestar general del pueblo”.[20]

19 Patricia García Majado, De las inmunidades del poder a la inmunidad del sistema jurídico y sus patologías (Madrid, España: Centro de Estudios Políticos y Constitucionales, 2022), 66. **20** John Locke, Ensayo sobre el gobierno civil, trads. Claudio Amor y Pablo Stafforini (Bernal, Provincia de Buenos Aires, Argentina: Universidad Nacional de Quilmes, 2005), 150 (la cita se corresponde con el § 131).

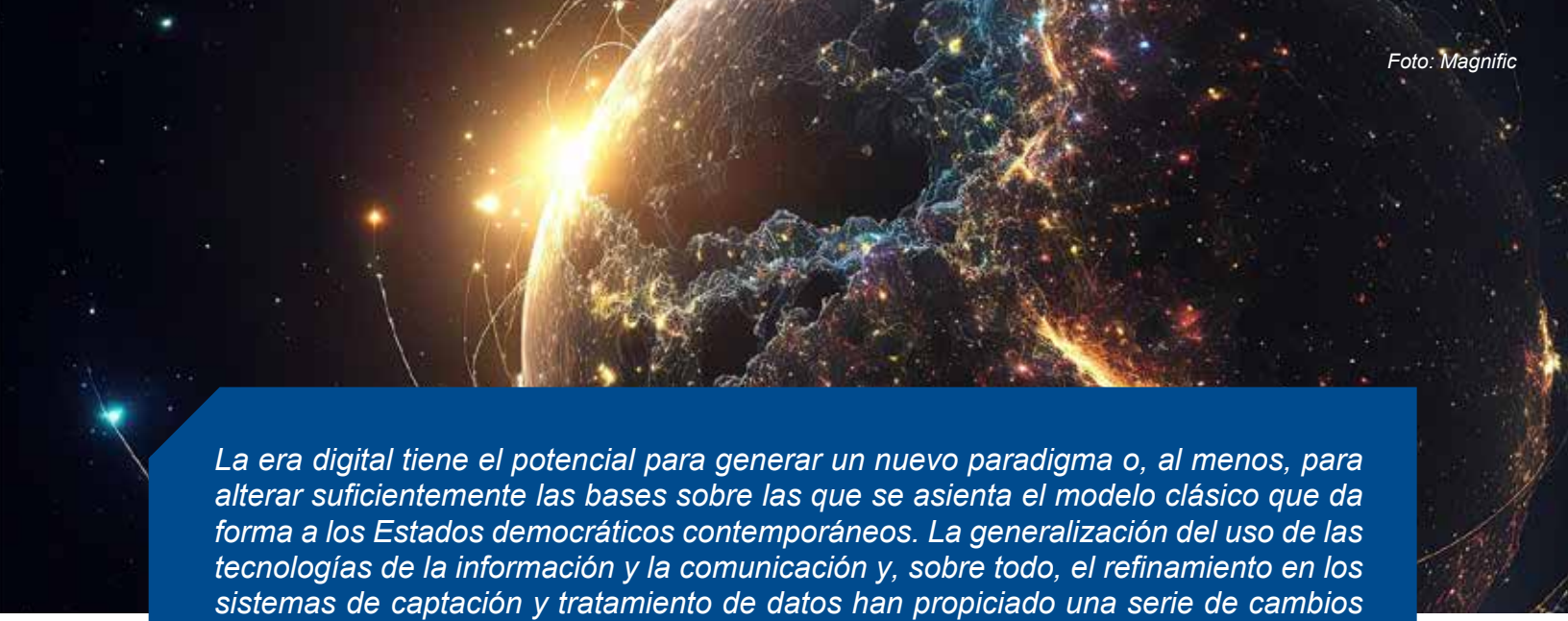


Foto: Magnific

La era digital tiene el potencial para generar un nuevo paradigma o, al menos, para alterar suficientemente las bases sobre las que se asienta el modelo clásico que da forma a los Estados democráticos contemporáneos. La generalización del uso de las tecnologías de la información y la comunicación y, sobre todo, el refinamiento en los sistemas de captación y tratamiento de datos han propiciado una serie de cambios cuyas consecuencias últimas no se conocen. En este punto, debe advertirse que, necesariamente, se va a incurrir en ciertas generalizaciones o especulaciones acerca de lo que pueda llegar a ocurrir; sin embargo, en la medida de lo posible, tratarán de plantearse escenarios cuya factibilidad resulta probable en atención al contexto actual.

Los cambios que están aconteciendo no son inanes.

> "En las sombras del esplendor digital se va gestando un nuevo ser, un 'Leviatán tecnológico […] desde el que se reorganizará la arquitectura artificial de un poder concebido como un panóptico perfecto'.[21] Un Leviatán, en efecto, con múltiples cabezas. Unas veces serán las grandes empresas transnacionales, más ricas y poderosas que muchos Estados, las que determinen el modo de vida de los ciudadanos, gobernando sin gobernar y dirigiendo sin dirigir. Otras, serán los gobiernos los que, aprovechando posiciones de fortaleza preexistentes, se valgan de los medios que la técnica ofrece para reforzar su control sobre su ciudadanía. El Ciberleviatán operará con una u otra cabeza según lo demanden las circunstancias".[22]

En ocasiones ambas confluirán, en una alianza que resulta especialmente peligrosa,[23] pues permitiría la aceleración del proceso, al combinarse capacidad normativa, uso de la fuerza y posibilidades de control mediante la tecnología (en parte, es lo que ya ocurre en países como China).

Frente a esta eventualidad, el derecho sería la respuesta que toda comunidad democrática opondría. Sin embargo, la capacidad del derecho para contener las pulsiones de ese o esos ciberleviatanes puede encontrarse con un escollo que le haga perder efectividad: el reemplazo del ciudadano por el consumidor-usuario. Ciertamente, la sociedad digital se ha "construido sobre el presupuesto de la satisfacción del consumidor-usuario quién, a su vez, es productor-proveedor de la materia (los datos) con la que se construye" el nuevo mundo virtual.[24]

21 José María Lasalle, Ciberleviatán (Barcelona, España: Arpa Editores, 2019), 48. **22** Daniel Jove Villares, "Tratamiento frente a dato. El problema de las categorías especiales" (tesis doctoral, Universidade da Coruña, La Coruña, 2021), 41. **23** Tal vez, en un futuro lejano (¿acaso de ciencia ficción?), no sean ni los entes privados ni los gobiernos, sino una IA autónoma y diseñada expresamente para gobernarnos y hacernos la vida más fácil, quien actúe como leviatán único y todopoderoso, pero esto, por el momento, se adentra demasiado en los terrenos fangosos de las hipótesis fantasiosas; con todo, y aunque sea como eventualidad, es un escenario que se debe plantear.**24 D**aniel Jove Villares, La protección de lo sensible, o cuando la naturaleza del dato no lo es todo (Valencia, España: Tirant Lo Blanch, 2023), 55.

el nuevo mundo virtual. Ese cambio no solo afecta al modo en que las personas se conectan y relacionan con el mundo,[25] sino a la idea misma de comunidad e, indirectamente, podría afectar a elementos nucleares de la democracia, como pudiera ser la naturaleza de las leyes y su aplicación.

En efecto, al igual que en la oferta de productos y servicios se tiende cada vez más hacia la personalización, no es descartable que la aspiración del ciudadano-consumidor mute hacia una exigencia de adecuación equivalente respecto de la ley y su aplicación. De hecho, como lo apunta Crego, esa tendencia a la personalización ya existe, aunque de manera imperfecta; su perfección podría llegar de la mano de la IA jurídica.[26] De este modo, en el peor de los escenarios, la generalidad de la ley, esencial para evitar la arbitrariedad y el privilegio y crucial para la preservación de la democracia, terminaría convirtiéndose en un obstáculo para la satisfacción del consumidor-usuario, haciendo del derecho un producto más, un bien consumible cuyo valor radicaría en su adaptabilidad.

Aunque lo apuntado pueda antojarse improbable, lo cierto es que ya se aprecian algunos síntomas preocupantes, siendo el más acuciante el debilitamiento del imperio de la ley y la "expansión del contrato".[27] Esteve Pardo pone de manifiesto cómo se está llevando a cabo una "discreta deconstrucción del orden legal, con sus enunciados generales e igualitarios, por la acción sistemática del contrato",[28] y cómo de esa tendencia a la exigencia de una regulación personalizada, el abismo ya no parece insalvable.

Empero, lo más preocupante no es que algo así tenga posibilidades de ocurrir, sino el hecho de que, de producirse, no sería como consecuencia de una imposición ejercida desde los poderes del Estado (aunque puedan llegar a facilitarlo e, incluso, incentivarlo), sino que sería el resultado del esfuerzo por dar satisfacción al ciudadano-consumidor. Y esa es una demanda que la IA, con su capacidad de inferencia y de correlación, podría llegar a colmar.

> La personalización de las respuestas jurídicas sería un clavo (acaso el definitivo) para la forma parlamentaria de gobierno. El poder legislativo, cada vez más débil, se vería relegado a la intranscendencia, pues su principal cometido, la elaboración de las leyes como reflejo de la voluntad del Estado, se convertiría en una necesidad anecdótica ante la posibilidad de ofrecer a cada persona una respuesta acorde a su particular situación. El sentido de comunidad, la fuerza de las mayorías, se iría diluyendo paulatinamente ante la respuesta calculada y ajustada de la IA.

La lucha por el poder ya no sería en las urnas, sino en los despachos, y sería una pugna por el control de la tecnología, por convertirse en el ciberleviatán. Los gobiernos y las empresas tecnológicas se disputarían la capacidad para controlar y decidir acerca del dominio de los sistemas de IA. Quien lo logre,

25 Ambas cuestiones son exploradas de manera profusa por el filósofo Byun-Chul Han en varias de sus obras, como La sociedad del cansancio (Barcelona, España: Herder, 2019) o La sociedad de la transparencia (Barcelona, España: Herder, 2018). **26** Jorge Crego analiza cómo la IA puede servir a la culminación del derecho personalizado, al que caracteriza como aquel que "supone un auténtico cambio de paradigma, pues instaura un sistema normativo que opera a través de reglas eminentemente singulares, adaptadas a las características que individualizan al sujeto y las condiciones de aplicación, es decir, a las condiciones que deben darse para que se deba aplicar el contenido de la norma". Jorge Crego, "La personalización del derecho como culminación ideal de la inteligencia artificial jurídica", en Inteligencia artificial y derecho: Perspectivas iusfilosóficas, ed. Alfonso Ballesteros Soriano (Madrid, España: Thomson Reuters-Aranzadi, 2024), 92. **27** José Esteve Pardo, El camino a la desigualdad. Del imperio de la ley a la expansión del contrato (Madrid, España: Marcial Pons, 2023). **28** Ibíd., 122-134.

> "dispondrá de un conocimiento tan pormenorizado acerca del individuo que no solo podrá anticipar el sentido de su conducta, sino también condicionarlo y, en último término, programarlo y dirigirlo […]. El Leviatán digital disfrutará de un poder difícilmente resistible. El de una omnisciencia que guardará celosamente frente a la curiosidad de una ciudadanía que no tendrá para él ningún secreto y a la que podrá manipular a su antojo mediante la inducción de comportamientos que se tendrán engañosamente por consentidos".[29]

Nos planteábamos, al final del apartado anterior, por qué podría resultarle interesante al poder valerse de sistemas basados en la probabilidad. La respuesta se presenta ahora como evidente: los sistemas de IA permiten el control de la sociedad, no solo porque facilitan la captación y gestión de una cantidad inconmensurable de datos (la información es poder), sino porque permiten, con su capacidad predictiva y posibilidades de aplicación preventiva, definir el marco de convivencia y el modo de comportarse en sociedad. Si se puede vigilar todo el tiempo y, además, imponer sanciones de manera inmediata por lo que se haga, pero también por lo que se deje de hacer e, incluso, por aquello que podría llegar a acontecer, entonces, el dominio sobre la población alcanzará cuotas nunca vistas en la historia de la humanidad.

Este escenario, que pudiera resultar inconcebible para el lector o lectora de un Estado democrático occidental, quizá no le resulte tan extraño a quien viva en un país no democrático. Allí donde los poderes del Estado no se contrarrestan y donde el control de la tecnología también está en manos del Estado, la realidad comienza a parecerse a la ficción aquí esbozada (basta con volver la vista hacia el Lejano Oriente para comprobarlo).

La revolución tecnológica que estamos viviendo puede hacer innecesario el uso de la fuerza, la coacción del Estado, para lograr el control sobre la sociedad. La prevención, la seguridad de la ciudadanía y la oportunidad de ofrecer una respuesta singularizada serán los argumentos que el poder político podrá invocar para reducir los espacios de libertad. Y así, de un modo sutil, sin violentarla, con la comodidad y el brillo de lo nuevo como aliados, se puede ir anestesiando a la ciudadanía y venciendo sus instintos de autopreservación frente a quienes pretenden ejercer un poder más allá del que jurídicamente tienen atribuido.

03
EL RIESGO Y LA PROBABILIDAD, ¿UNA AMENAZA PARA LOS DERECHOS?

Los derechos siempre han sido el principal baluarte frente al poder. Un límite infranqueable a la voluntad de control y dominación. ¿Lo seguirán siendo en la sociedad del riesgo? En un modelo de convivencia en el que "the past loses its power to determine the present. Its place as the cause of present-day experience and action is taken by the future, that is to say, something nonexistent, constructed and fictitious. We are discussing and arguing about something which is not the case, but could happen if we were not to change course".[30]

29 Juan Luis Requejo Pagés, La agonía de la democracia (Oviedo, España: KRK ediciones, 2020), 47. **30** Ulrich Beck, "Risk Society Revisited: Theory, Politics and Research Programmes", en The Risk Society and Beyond. Critical Issues for Social Theory, ed. Barbara Adam et al. (Thousand Oaks, California: SAGE Publications, 2000), 214.

El riesgo es el protagonista de nuestro tiempo. Las tecnologías y las políticas normativas se enfocan en evitarlo o, de no ser posible, en minorarlo. La regulación de la IA no es ajena a este fenómeno, no es casualidad que el RIA lo defina en segundo lugar (artículo 3.2), justo después de establecer qué se entiende por sistemas de IA. Así, el riesgo sería, para el legislador europeo, "la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio". Es la combinación e interacción entre ambas variables la que determina la viabilidad de introducir en el mercado, poner en servicio o utilizar un sistema de IA. Una vez superada esa primera condición de posibilidad (que es la que define las prácticas prohibidas),[31] el nivel de riesgo sigue siendo el criterio que da forma al régimen jurídico y a las exigencias que deberán cumplir los sistemas de IA que vayan a ser utilizados en la UE.



Foto: Magnific

No cabe duda de que la tecnología hace posible lo que hasta hace poco era impensable, pero la capacidad de hacer algo no necesariamente implica que se deba realizar. El respeto a los derechos de la ciudadanía, a su dignidad, impone valorar si la acción es aceptable o si, por el contrario, supone una injerencia inasumible. En este sentido, debería valorarse si la introducción de una tecnología obedece a criterios jurídicamente válidos. En esta línea, y en congruencia con el enfoque basado en el riesgo, se adoptan modelos centrados en su evaluación y minoración. Esta metodología resulta interesante desde un punto de vista jurídico, pues busca identificar amenazas e implementar medidas para que no lleguen a consumarse, siendo además un modelo dinámico, que asume la naturaleza mutable de los peligros y, consecuentemente, impone la adaptación a lo largo del tiempo de las respuestas destinadas a minorarlos.

Empero, este enfoque podría llegar a privar a los derechos de su centralidad, al convertirlos en una variable más. Junto a estos se consideran otros factores, como los beneficios para la sociedad, los costes, el riesgo reputacional o la probabilidad general de que se produzcan determinados escenarios. Esto implica que, en lugar de analizar si se produce una injerencia en los derechos y, de ser el caso, si se adoptaron las medidas adecuadas, lo que importa es si la posibilidad de que la injerencia se produzca es lo suficientemente baja como para que valga la pena asumir el riesgo. Naturalmente, a mayores beneficios potenciales, mayor nivel de peligro se puede asumir.

Esto es, las exigencias que la regulación y eventual afectación de los derechos imponen[32] (regulación legal y determinación del interés concreto que se quiere proteger, definición clara de las condiciones limitantes de la injerencia que se vaya a producir en el derecho y establecimiento de garantías adecuadas), antes condición ineludible para considerar una acción como jurídicamente aceptable, se convierten en medidas que sirven para reducir la gravedad del perjuicio (artículo 3.2 RIA). En definitiva, se convertiría en un medio ponderable lo que debería ser un requisito insoslayable.

31 Sobre las prácticas prohibidas previstas en el RIA, véanse Patricia García Majado, "Prohibited AI Practices under the EU AI Act", en Artificial Intelligence and Fundamental Rights. The AI Act of the European Union and its implications for global technology regulation, eds. Benjamin Raue et al., vol. 4 (Trier, Deutschland: Verein für Recht und Digitalisierung - IRDT, 2025), 35-56; Miguel Ángel Presno Linera, "La prohibición de los sistemas que evalúan y clasifican a las personas con efectos discriminatorios en el Reglamento Europeo de Inteligencia Artificial", en Vigilancia biométrica masiva, inteligencia artificial y derechos fundamentales, eds. Francisco Balaguer Callejón y Leire Escajedo San-Epifanio (Madrid, España: Dykinson, 2025), 507-526. **32** Al respecto, véase, por ejemplo, la sentencia del Tribunal Constitucional español, 76/2019, de 22 de mayo, FJ 7 a 9.

Llevado al extremo, un sistema en el que la prevención reemplace a la reparación tendría unas consecuencias terribles para los derechos individuales de la ciudadanía, pues las consecuencias dejarían de ser lo relevante (especialmente, en lo que a responsabilidad civil y de la administración se refiere; intuyo que la penal perviviría), mientras que sería importante sería si se adoptaron las medidas adecuadas para evitar la vulneración. De esta manera, si se comprueba que hubo diligencia, ya no sería necesario acudir a medidas indemnizatorias y reparadoras, pues el daño se entendería como un fallo imprevisible, un mal inevitable y, por tanto, como algo que se ha de soportar. Lo importante habrá sido constatar que se hizo lo que se pudo para impedir su producción. El cumplimiento normativo vendría dado por la evaluación y la previsión.

En un escenario así, el riesgo sustituye a los derechos como límite, toda vez que se cambian certezas por eventualidades, con el agravante adicional de que, salvo cuando en la normativa se prohíba expresamente una determinada práctica, en los demás escenarios se asume como aceptable que el daño pueda producirse. Aunque no se ha llegado a ese punto, conviene advertir del peligro, especialmente por el riesgo de dilución de la responsabilidad. La forma como se enfoque esta cuestión será crucial, pues, como advirtiera Villaverde Menédez respecto de la exigencia de responsabilidad por "la prestación de servicios comerciales en el ciberespacio […], [si se aborda desde una perspectiva] ajena por completo a los conflictos constitucionales […], el asunto se reduce a un problema de 'derecho de daños' y de protección del usuario y consumidor".[33] Tal opción, si bien en la práctica puede ofrecer unos resultados aparentemente similares, carece de la potencia y las garantías que proporciona a la ciudadanía un enfoque iusfundamental.

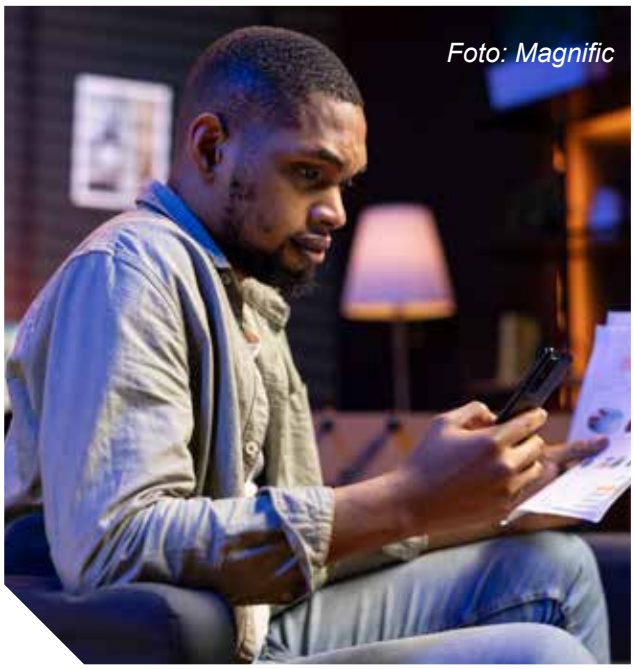
La evaluación de riesgos, bien encauzada y combinada con medidas reparadoras y prohibiciones, puede ser un elemento extraordinariamente potente en la defensa de los intereses de la ciudadanía. Pero si todo se fía a la prevención, se estará asfaltando el camino para que el fin justifique los medios, para que se asuma como aceptable que se pueden sacrificar los derechos de unos pocos, si se van a convertir en un obstáculo para el progreso de la mayoría, o para su seguridad o para un más eficiente control por parte de los gobiernos. Hay una pendiente resbaladiza y peligrosa al final de ese camino.



Foto: Magnific

33 Ignacio Villaverde Menédez, Los poderes salvajes. Ciberespacio y responsabilidad por contenidos difamatorios (Madrid, España: Marcial Pons, 2020), 51.

04 CULTURA JURÍDICA Y DEMOCRACIA


Foto: Magnific

El derecho y, especialmente, los derechos fundamentales[34] se presentan como medicinas y remedios frente a las posibilidades de control y de dominación que la IA ofrece a los gobiernos y, por qué no decirlo, también frente a determinadas empresas transnacionales que monopolizan el mercado tecnológico y hacen depender el desarrollo de la sociedad de su tecnología y sus intereses particulares.[35] Sin embargo, no puede obviarse que, en democracia, la fortaleza del derecho reside no solo en lo que es y en la forma como se despliega, sino también en la adhesión social que sea capaz de generar.

En este sentido, hay ciertas nociones de lo que los derechos deben ser que se han ido imprimiendo en el ADN de la ciudadanía, un código cultural que, de manera casi inconsciente, se va grabando en las personas conforme avanza su proceso de socialización. Se trata de un modo de entender la vida en comunidad que va a dotar al marco normativo de mayor o menor aceptación en la medida en que sea congruente con esa cultura jurídica subyacente.

En el caso de Occidente, la cultura clásica grecorromana y el humanismo[36] están presentes como una base primigenia, como un primer fundamento que se ha ido perfilando y perfeccionando, especialmente a partir de las revoluciones del siglo XVIII, asentándose a partir de ellas una concepción liberal-democrática que ve en los derechos fundamentales un límite al poder, y considera que la protección del individuo frente al Estado no solo salvaguarda su libertad, sino que contribuye al bienestar colectivo, al frenar la arbitrariedad y promover un orden político racional.

No obstante, esa forma de entender las relaciones entre la ciudadanía y el Estado no es universal, sino que oscila en atención a factores culturales, religiosos e históricos de cada sociedad, e incluso dentro de un mismo país puede haber diversas concepciones e ideas acerca de lo que son los derechos y del rol que representan.[37] Caben, incluso, visiones antitéticas a la del liberalismo, nociones en las que el derecho y los derechos no se presentan como una barrera frente al poder, sino como un instrumento para lograr un determinado orden social, uno en el que la protección de la persona como individuo aislado cede frente a aquellas acciones que garanticen la armonía social, la cohesión y la preservación de la comunidad.

34 Miguel Ángel Presno Linera, Derechos fundamentales e inteligencia artificial (Madrid, España: Marcial Pons, 2022). **35** Son lo que, con gran acierto, Rodotà denominó como los "nuovi signori dell'informazione". Stefano Rodotà, Il mondo nella Rete. Quali i diritti, quali i vincoli (Roma, Italia: Laterza, 2019), 61. **36** Sobre el ideal humanista, su conexión con la dignidad y su impacto en la democracia y el derecho. Véase Gabriel Moreno González, La democracia humanista (Sevilla, España: Athenaica, 2020). Todo el libro resulta de interés, pero, a estos efectos, son especialmente interesantes las páginas 60-65 y 86-100.

No obstante, esa forma de entender las relaciones entre la ciudadanía y el Estado no es universal, sino que oscila en atención a factores culturales, religiosos e históricos de cada sociedad, e incluso dentro de un mismo país puede haber diversas concepciones e ideas acerca de lo que son los derechos y del rol que representan.[37] Caben, incluso, visiones antitéticas a la del liberalismo, nociones en las que el derecho y los derechos no se presentan como una barrera frente al poder, sino como un instrumento para lograr un determinado orden social, uno en el que la protección de la persona como individuo aislado cede frente a aquellas acciones que garanticen la armonía social, la cohesión y la preservación de la comunidad.

Cualquiera que sea la visión concreta, con sus virtudes y defectos, lo que aquí interesa es poner de manifiesto que el modo en que la ciudadanía concibe su relación con el mundo influye en la fortaleza del derecho y de los derechos frente a fenómenos disruptivos como la IA. Consecuentemente, la cultura jurídica opera como una especie de sistema inmunitario, destinado a impedir que el virus que acompaña a los beneficios de la algoritmización de la sociedad se asiente. Una cultura basada en los derechos individuales apostará por una regulación de la IA que la haga compatible con el modelo de convivencia que representa el Estado social y democrático de derecho. No se trata de descartar las posibilidades que la tecnología ofrece, solo de asumirlas de modo que su adopción no suponga sacrificar los rasgos que la caracterizan.

A su vez, aquellas sociedades con una visión más comunitarista de los derechos podrían ser más propensas a aceptar como natural y hasta deseable una mayor intervención del Estado, destinada, precisamente, a asegurar el bien común del colectivo. Esta noción, no necesariamente negativa per se, podría convertirse, en manos de gobiernos autocráticos (o con voluntad de serlo), en la puerta a la sociedad de la vigilancia y el control y, a partir de ella, en una sociedad menos libre.

Al revés, una sociedad más liberal ofrecerá mayor resistencia y será menos proclive a aceptar como admisibles condiciones que supongan una reducción de su esfera de libertad. Aunque, como se ha advertido, la tecnología es, en sí misma, un "factor cultural [que] ha contribuido a fomentar un entorno complicado para el constitucionalismo en la medida en que ha situado fuera del Estado y del espacio público nacional la capacidad de generar pautas culturales que afectan de manera muy relevante a la configuración de las sociedades";[38] por lo tanto, es probable que haya contribuido a relajar ciertos instintos de autopreservación. Ello no obsta para que ese sustrato aún conserve fuerza suficiente como para que, en Estados democráticos y con una cultura y tradición de defensa de los derechos fuerte, resulte más difícil implementar (por la reacción social y el coste político que supondrían) políticas y medidas que puedan suponer una ablación de la libertad.

Finalmente, no debemos olvidar que, en última instancia, los Estados democráticos cuentan con una capa adicional de protección, proporcionada por el hecho de que la ciudadanía tenga capacidad para deponer, sin tener que recurrir a la violencia, a los tiranos o a quienes, valiéndose de los medios a su alcance, tratan de ir más allá de lo que normativamente están habilitados. Si el derecho es el corsé del poder, la ciudadanía es quién aprieta los cordones cuando amenaza con salirse del marco.

37 La universalidad de los derechos es una cuestión siempre discutida, pero con un impacto evidente en el modo en que se traduce normativamente. Sobre este debate, me remito a los trabajos de Eva Brems, Human rights: Universality and diversity, vol. 66 (Leiden, Netherlands: Brill, 2021); y de Jack Donnelly, "The relative Universality of Human Rights", Human Rights Quarterly 29, n.o 2 (2007): 281-306.**38** Francisco Balaguer Callejón, La constitución del algoritmo, 2a ed. (Madrid, España: Fundación Manuel Giménez Abad de Estudios Parlamentarios y del Estado Autonómico, 2023), 106-107.

Pero para que esta pueda hacer un buen uso de las herramientas de control de las que dispone, debe poder comprender el alcance de las acciones que están llevando a cabo los gobiernos y los poderes privados. Sin una ciudadanía consciente de su propio poder y de los peligros que la acechan, nada será capaz de contener al leviatán algorítmico que la era digital parece estar gestando. La alfabetización digital (muy presente en estos tiempos)[39] y también la formación democrática (mucho más olvidada y sin embargo tanto o más necesaria) son cruciales, tanto para sacar partido de los muchos beneficios y posibilidades que la tecnología ofrece como para poder comprender las amenazas que entraña. No basta con formar para el mercado, no se trata de enseñar exclusivamente las utilidades de estas herramientas (siendo esto relevante para evitar desigualdades), es imprescindible concienciar acerca de las implicaciones. La educación acerca de las cuestiones digitales debe ser técnica, ética y jurídica.

En este sentido, la instrucción debería ser lo más amplia y transversal posible, abarcando todas las franjas de edad. No se puede dejar a nadie atrás, toda vez que el cambio que se está produciendo afecta al marco de convivencia, las medidas de adaptación deben incluir a la comunidad en su conjunto. Además, debería ofrecerse una formación específica a quienes trabajan en el sector público, pues solo aquello que se entiende y se conoce se puede gestionar de manera adecuada. Quienes gestionan y gobiernan lo común deben ser los primeros defensores de los derechos de la ciudadanía.

05 REFLEXIÓN FINAL

Comenzaba este artículo señalando que la IA amenaza con convertirnos en jugadores de póker y señalaba que ese no sería el peor escenario. En efecto, la peor de las hipótesis sería aquella en la que la ciudadanía se viese reducida a ser las cartas con las que gobiernos y empresas transnacionales juegan, sin más aspiración que, en el mejor de los casos, hacer variar las probabilidades, pero condenados como sociedad a ser manipulados, jugados o descartados en función de las necesidades de quienes controlan los sistemas de IA.

Anida en esta tecnología un poder configurador, un poder blando, pero con capacidad transformadora. Los algoritmos no solo sirven para vigilar y controlar, también moldean el modo de estar y vivir en sociedad, no configuran solo instrumentos para la represión, sino una forma de organizar la vida en comunidad. Ese terrible futuro será más o menos distópico en la medida en que la ciudadanía sea consciente de que el riesgo de que se llegue a materializar es real.

39 El legislador español reconoció (en el artículo 83 de la Ley Orgánica 3/2018, del 5 de diciembre, sobre protección de datos personales y garantía de los derechos digitales) el derecho a la educación digital, que tiene como objetivo "la plena inserción del alumnado en la sociedad digital y el aprendizaje de un consumo responsable y un uso crítico y seguro de los medios digitales y respetuoso con la dignidad humana, la justicia social y la sostenibilidad medioambiental, los valores constitucionales, los derechos fundamentales y, particularmente con el respeto y la garantía de la intimidad personal y familiar y la protección de datos personales". Sobre este derecho y su engarce constitucional tuve ocasión de pronunciarme en Daniel Jove Villares, "El derecho a la educación digital en España. Regulación jurídica e importancia para la preservación del Estado democrático de derecho", Anuario da Facultade de Dereito da Universidade da Coruña, n.o 25 (2021): 94-110.

Si permanece impasible ante la reducción de sus esferas de libertad, si no reconoce esta nueva dimensión del poder y trata de evitar que las decisiones algorítmicas reemplacen al debate democrático, si solo confía en la autocontención del poder, estaremos sentando las bases para nuestra condenación. Afortunadamente, tenemos las herramientas jurídicas para evitarlo, no hay que inventarlas, pues son las mismas que el constitucionalismo viene desarrollando desde las primeras revoluciones liberales. La garantía de los derechos y el control del poder siguen siendo instrumentos válidos, también en la era algorítmica. Habrá que adaptarlos y perfeccionarlos para que sean capaces de responder a las sutilezas de la digitalización de la vida en comunidad y a la irrupción de agentes privados con capacidades de dominio equiparables a las de gobiernos absolutos, pero, a pesar de todo, siguen siendo la mejor arma en el arsenal de la ciudadanía frente a cualquier forma de despotismo. Frente a esa especie de poder blando de la IA, el derecho positivo debe ser la respuesta. Ahora bien, para poder hacer uso de dichas herramientas, es necesario preservarlas, conocerlas y usarlas.

La defensa de las garantías que se han ido construyendo durante siglos requiere de una ciudadanía activa, consciente de sus derechos y alerta ante los intentos de recortarlos. En este sentido, la cultura y la tradición jurídica de una sociedad son cruciales, en tanto que sirven como un freno a las tentaciones que los gobiernos puedan tener de hacer uso de sistemas de IA para recortar las esferas de libertad. Una sociedad menos predispuesta a aceptar una merma en sus derechos individuales ofrecerá mayores resistencias, especialmente si se está en un Estado democrático, en el que el acceso al poder (o el mantenerse en este) depende de la voluntad de la ciudadanía. Sin embargo, el modo en que la tecnología se ha ido incorporando en el día a día ha hecho que las barreras se bajen y que permanezcan inactivos los sistemas de alerta que normalmente se encenderían ante intentos de reducir derechos. La familiaridad, las ventajas, la eficiencia y la comodidad que los ingenios digitales proporcionan facilita la aceptación del uso de estos instrumentos por los gobiernos, al principio para la seguridad, luego para el control. El tránsito es paulatino, poco perceptible, pero constante.



Foto: Magnific

Ante esa tendencia, además de exigir el respeto a los derechos y garantías existentes, es crucial la alfabetización digital y la concienciación de la ciudadanía, puesto que solo así se podrán detectar las medidas que rebasan los límites, las que suponen un sacrificio inaceptable para la libertad. El cambio de paradigma que la era digital supone requiere una reivindicación de la importancia del principio democrático. Sin duda, el Estado de derecho es un baluarte y, probablemente, un mecanismo principal de defensa frente a los cambios que la revolución algorítmica puede operar. Sin embargo, la voluntad ciudadana será la última línea de defensa frente al Ciberleviatán, el verdadero as en la manga que evite la completa pérdida de libertad. Si se quieren salvar las democracias y el Estado de derecho, la prioridad debe ser la creación de una conciencia cívica, el educar a las personas y activarlas para que sigan siendo ciudadanos y que no se conviertan en consumidores pasivos, movidos al compás del determinismo algorítmico.

Con todo, la respuesta jurídica será igualmente necesaria. En este sentido, debería prestarse una mayor atención a los efectos que provoca abrazar el riesgo como criterio regulador. La prevención puede resultar especialmente tentadora, el refrán "más vale prevenir que lamentar" es fiel reflejo de esa querencia por evitar el daño; sin embargo, cuando esa visión se convierte en técnica legislativa, lo que antes era un límite claro se convierte luego en una zona difusa. Donde antes se erigía una frontera que delimitaba lo que era jurídicamente aceptable y lo que no, surge después un espacio para el análisis y la evaluación, en el que el mayor o menor peso que pretenda darse a las variables en concurso, el mayor o menor riesgo que se esté dispuesto a asumir o la importancia que quiera darse al interés que se pretende salvaguardar pueden anticipar o retrasar las posibilidades de ejercicio de los derechos fundamentales.

Si tomamos como momento fundacional de los Estados constitucionales modernos el pacto social y el paso del estado de naturaleza a la sociedad (aunque sea una ficción idealizada de un proceso mucho más complejo) y trasladamos ese momento transicional al momento de cambio que la irrupción de la IA parece augurar, deberíamos tratar de ir un paso más allá de lo que en su momento fue Locke y preguntarnos: ¿a cuánta libertad es posible renunciar?, ¿cuánta seguridad es suficiente?, ¿cuál es el coste de pasar de la protección a la prevención?, ¿es compatible con un Estado constitucional democrático? La respuesta que como sociedad se dé a esas preguntas determinará el modelo de convivencia en el que se desarrollen nuestras vidas en las próximas décadas. La sociedad de la vigilancia y el control no es un destino inexorable, pero si la ciudadanía no reivindica la preservación de sus ámbitos de libertad y autodeterminación, puede llegar a ser irremediable.

> *"Con todo, mis palabras no se deben a un 'terror cósmico'[40] a la innovación y el progreso, sino una preocupación nacida del impacto de la 'tecnodemocracia y la soberanía electrónica'[41] en los derechos fundamentales y la democracia. La respuesta jurídica a este desafío resulta crucial. No se trata de frenar los avances tecnológicos, sino de guiar su avance, orillando, en lo posible, sus consecuencias más dañinas".[42]*

40 José Ortega y Gasset, Meditación de la técnica (Madrid, España: Espasa-Calpe, 1965). **41** En este trabajo de 1998, Bastida Freijedo enuncia, con gran acierto y anticipación, tanto las posibilidades como los riesgos de la tecnodemocracia. En especial, advierte sobre la posibilidad de diluir los actuales sistemas democráticos y que acaben transformados en "un régimen de opinión pública, en […] [una] democracia 'en directo'" que cortocircuite el sistema, deslegitimando sus instituciones y "operando como un metasistema y, además, sin las garantías del sistema democrático". Francisco J. Bastida Freijedo, "La soberanía borrosa: la democracia", Fundamentos: Cuadernos Monográficos de Teoría del Estado, Derecho Público e Historia Constitucional, n.o 1 (1998): 381-460. **42** Jove Villares, "Tratamiento frente a dato", 41.

Ítem más, la IA tiene un potencial transformador que puede servir para "conseguir una buena sociedad de la IA"[43] e, incluso, puede servir para fortalecer la democracia[44] y el Estado de derecho.[45] Esa dirección propositiva y positiva es la que se debe cultivar, pero sin dejar de afrontar los cambios y riesgos que la acompañan.

ACLARACIÓN Y AGRADECIMIENTO

Como habrá podido comprobar quien haya llegado hasta aquí, este trabajo oscila entre lo científico y lo ensayístico. Las razones para optar por este estilo híbrido son, de una parte, el punto en que se encuentra la investigación de la que trae causa (un estudio amplio sobre el impacto de la IA en el Estado social y democrático de Derecho) y, de otra parte, el foro en el que se publica. Una revista académica no solamente es (o no debería aspirar a serlo) la depositaria de las reflexiones e investigaciones más relevantes, sino que ha de contribuir a fomentar la creación, la discusión y el intercambio de conocimiento. Consecuentemente, me he atrevido (disculpen la osadía) a plantear una reflexión, cargada de dudas e hipótesis no confirmadas, con la esperanza de que pueda suscitar interés e, idealmente, generar sinergias académicas, propiciar un debate que sirva para aquilatar ideas y encontrar puntos en común que, en un futuro, puedan servir para formular alguna propuesta. Permítaseme también agradecer al Observatorio de Gobierno y TIC de la Pontificia Universidad Javeriana por la oportunidad de participar en este primer número de la Revista e-Estado.

43 Luciano Floridi, Ética de la inteligencia artificial (Barcelona, España: Herder, 2024), 339-360. **44** Juan Francisco Sánchez Barrilao, "El impacto de la inteligencia artificial en los procesos democráticos", Revista de Derecho Constitucional Europeo, n.o 43 (2025). **45** Rafael Bustos Gisbert y Francisco Javier García Roca, eds., Fortalecimiento de la democracia y el Estado de derecho a través de la inteligencia artificial (País Vasco: Instituto Vasco de Administración Pública - IVAP, 2025).

REFERENCIAS

Badaró, Sebastián, Leonardo Javier Ibañez y Martín Jorge Agüero. "Sistemas expertos: fundamentos, metodologías y aplicaciones". Ciencia y Tecnología, n.o 13 (2013): 349-364.

Balaguer Callejón, Francisco. La constitución del algoritmo. 2a ed. Madrid, España: Fundación Manuel Giménez Abad de Estudios Parlamentarios y del Estado Autonómico, 2023.

Bastida Freijedo, Francisco J. "La soberanía borrosa: la democracia". Fundamentos: Cuadernos Monográficos de Teoría del Estado, Derecho Público e Historia Constitucional, n.o 1 (1998): 381-460.

Beck, Ulrich. "Risk Society Revisited: Theory, Politics and Research Programmes". En The Risk Society and Beyond. Critical Issues for Social Theory, editado por Barbara Adam, Ulrich Beck, y Joostvan Loon, 211-229. Thousand Oaks, California: SAGE Publications, 2000.

Brems, Eva. Human Rights: Universality and Diversity. Vol. 66. Leiden, Netherlands: Brill, 2021.

Bustos Gisbert, Rafael y Francisco Javier García Roca, eds. Fortalecimiento de la democracia y el Estado de derecho a través de la inteligencia artificial. País Vasco: Instituto Vasco de Administración Pública (IVAP), 2025.

Cotino Hueso, Lorenzo. "¿Qué es 'inteligencia artificial' para el Reglamento? Análisis, delimitación y aplicaciones prácticas". En Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea, editado por Lorenzo Cotino Hueso y Pere Simón Castellano, 113-121. Navarra, España: Aranzadi, 2024.

Crego, Jorge. "La personalización del derecho como culminación ideal de la inteligencia artificial jurídica". En Inteligencia artificial y derecho. Perspectivas iusfilosóficas, editado por Alfonso Ballesteros Soriano, 83-107. Madrid, España: Thomson Reuters-Aranzadi, 2024.

De Miguel Beriain, Iñigo y Antonio Diéguez Lucena. "¿Explicar o predecir?". Investigación y Ciencia, n.o 538 (2021), 52-53.

Donnelly, Jack. "The Relative Universality of Human Rights". Human Rights Quarterly 29, n.o 2 (2007): 281-306.

Esteve Pardo, José. El camino a la desigualdad. Del imperio de la ley a la expansión del contrato. Madrid, España: Marcial Pons, 2023.

Floridi, Luciano. Ética de la inteligencia artificial. Barcelona, España: Herder, 2024.

García Majado, Patricia. De las inmunidades del poder a la inmunidad del sistema jurídico y sus patologías. Madrid, España: Centro de Estudios Políticos y Constitucionales, 2022.

—. "Inteligencia artificial, predicciones y funciones normativas". Teoría y Realidad Constitucional, n.o 54 (2024): 421-447.

—. "Prohibited AI Practices under the EU AI Act". En Artificial Intelligence and Fundamental Rights. The AI Act of the European Union and its implications for global technology regulation, editado por Benjamin Raue, von Ungern-Sternberg, Kumkar y Rüfner, vol. 4. Trier, Deutschland: Verein für Recht und Digitalisierung (IRDT), 2025.

Han, Byung-Chul. La sociedad de la transparencia. Barcelona, España: Herder, 2018.

—. La sociedad del cansancio. Barcelona, España: Herder, 2019.

Jove Villares, Daniel. "El derecho a la educación digital en España. Regulación jurídica e importancia para la preservación del Estado democrático de derecho". Anuario da Facultade de Dereito da Universidade da Coruña, n.o 25 (2021): 94-110.

—. La protección de lo sensible, o cuando la naturaleza del dato no lo es todo. Valencia, España: Tirant Lo Blanch, 2023.

—. "Tratamiento frente a dato. El problema de las categorías especiales". Tesis doctoral, Universidade da Coruña, La Coruña, 2021. http://hdl.handle.net/2183/29853

Lasalle, José María. Ciberleviatán. Barcelona, España: Arpa Editores, 2019.

Ringel Morris, Meredith, Jascha Sohl-Dickstein, Noah Fiedel, Tris Wartkentin, Allan Dafoe, Aleksandra Faust, Clement Farbaret y Shane Legg. "Position: Levels of AGI for operationalizing progress on the path to AGI". Conference paper, Forty-first International Conference on Machine Learning, Viena, Austria, 2024.

Moreno González, Gabriel. La democracia humanista. Sevilla, España: Athenaica, 2020.

Ortega y Gasset, José. Meditación de la técnica. Madrid, España: Espasa-Calpe, 1965.

Presno Linera, Miguel Ángel. Derechos fundamentales e inteligencia artificial. Madrid, España: Marcial Pons, 2022.

—. "Derechos fundamentales e inteligencia artificial en el Estado social, democrático y digital de derecho". Anuario de la Red Eurolatinoamericana de Buen Gobierno y Buena Administración, 2023. https://ijeditores.com/pop.php?option=articulo&Hash=dce233eca31072b2c2f7e804f3efc07a

—. "Policía predictiva y prevención de la violencia de género: el sistema VioGén". IDP. Revista d'Internet, Dret i Política, n.o 39 (2023): 1-13.

—. "La prohibición de los sistemas que evalúan y clasifican a las personas con efectos discriminatorios en el Reglamento Europeo de Inteligencia Artificial". En Vigilancia biométrica masiva, inteligencia artificial y derechos fundamentales, editado por Francisco Balaguer Callejón y Leire Escajedo San-Epifanio, 507-526. Madrid, España: Dykinson, 2025.

Ramió, Carles. Inteligencia artificial y administración pública. 2a ed. Madrid, España: Catarata, 2025.

Rehman Zafar, Muhammad y Naimul Khan. "Deterministic local interpretable model-agnostic explanations for stable explainability". Machine Learning and Knowledge Extraction 3, n.o 3 (2021): 525-541.

Requejo Pagés, Juan Luis. La agonía de la democracia. Oviedo, España: KRK ediciones, 2020.

Ridaura Martínez, María Josefa. "El reconocimiento facial como herramienta para garantizar la seguridad". En Tecnologías abusivas y derecho, editado por Francisco Caamaño Domínguez y Daniel Jove Villares, 139-185. Valencia, España: Tirant Lo Blanch, 2024.

Rodotà, Stefano. Il mondo nella Rete. Quali i diritti, quali i vincoli. Roma, Italia: Laterza, 2019.

Rubio Núñez, Rafael, Frederico Franco Alvim y Vitor de Andrade Monteiro. Inteligencia artificial y campañas electorales algorítmicas. Disfunciones informativas y amenazas sistémicas de la nueva comunicación política. Madrid, España: Centro de Estudios Políticos y Constitucionales, 2024.

San Martín Segura, David. "El algoritmo, entre disposición y norma. La calificación de la IA de decisión jurídica". En El derecho administrativo en la era de la inteligencia artificial: actas del XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo. Vigo, 25 a 27 de enero de 2024, coordinado por Patricia Valcárcel Fernández y Francisco Lorenzo Hernández González, 199-208. España: Instituto Nacional de Administración Pública (INAP), 2025, 199-208.

Sánchez Barrilao, Juan Francisco. "El impacto de la inteligencia artificial en los procesos democráticos". Revista de Derecho Constitucional Europeo, n.o 43 (2025).

Sánchez Muñoz, Óscar. La regulación de las campañas electorales en la era digital. Desinformación y microsegmentación en las redes sociales con fines electorales. Madrid, España: Centro de Estudios Políticos y Constitucionales, 2021.

Villaverde Menédez, Ignacio. Los poderes salvajes. Ciberespacio y responsabilidad por contenidos difamatorios. Madrid, España: Marcial Pons, 2020.

Locke, John. Ensayo sobre el gobierno civil. Traducido por Claudio Amor y Pablo Stafforini. Bernal, Provincia de Buenos Aires, Argentina: Universidad Nacional de Quilmes, 2005.

ACERCA DE LOS AUTORES

Daniel Jove **Villares**

Doctor en Derecho por la Universidade da Coruña, con premio extraordinario de doctorado. Ha sido profesor en la mencionada universidad, así como en la Universidad Carlos III de Madrid. Se incorporó a la Universidad de Oviedo en septiembre de 2024 como profesor ayudante doctor. Entre sus líneas de investigación se cuentan el derecho a la protección de datos, la inteligencia artificial y los derechos digitales. Es investigador en varios proyectos de investigación, incluido el proyecto PID2022-136548NB-I00 Los retos de la inteligencia artificial para el Estado social y democrático de Derecho financiado por MICIU/AEI/10.13039/501100011033/ FEDER,UE.



Foto: Magnific

03

Inteligencia artificial en el aula

E-ESTADO
Primera Edición

Inteligencia artificial en el aula: desmontando mitos y trazando proyecciones informadas

Artificial Intelligence in the Classroom: Debunking Myths and Charting Informed Projections

Fecha de recepción: 27 de agosto de 2025
Fecha de aceptación: 18 de febrero de 2026

Francisco José García-Peñalvo
Instituto Universitario de Ciencias de la Educación (IUCE)- Universidad de Salamanca, España
fgarcia@usal.es
https://orcid.org/0000-0001-9987-5584

Cómo citar: García-Peñalvo, Francisco José. "Inteligencia artificial en el aula: desmontando mitos y trazando proyecciones informadas". Revista e-Estado 1, n.° 1 (2026).



Foto: Magnific

RESUMEN

Este artículo propone un giro metodológico en la conversación sobre la inteligencia artificial generativa (IAGen) en educación: desplazar el foco del mito a la proyección informada. Tras situar la IAGen en el ciclo de sobreexpectación y reajuste, y enmarcarla bajo el dilema prometeico y el principio de responsabilidad, se presenta un marco pragmático compuesto por i) una lista de verificación para identificar mitos (totalización, causalidad lineal, inmunidad a la evidencia y borrado de condiciones), ii) una plantilla para formular proyecciones auditables (propósito, población/condiciones, métricas/hitos, riesgos/salvaguardas y puntos de revisión) y iii) un protocolo de lectura de imaginarios sociotécnicos. El análisis aborda tres mitos recurrentes: la IA sustituirá al profesorado, la IA garantiza autonomía y equidad y la muerte de la originalidad. En cada caso se muestra por qué operan como mitos y se formulan proyecciones informadas coherentes con las cuatro lentes de la Unesco (pertinencia, equidad, escalabilidad, sostenibilidad) y con las reglas prometeicas (proporcionalidad/reversibilidad y control de alucinación). Se derivan implicaciones de política relacionadas con el reconocimiento explícito del uso de IA, la supervisión humana significativa en decisiones de alto impacto, la evaluación auténtica centrada en procesos y voz propia, la gobernanza de datos con transparencia y auditorías de sesgo, y estrategias de equidad que prioricen centros vulnerables. La discusión integra los hallazgos, proponiendo ciclos de evidencia (pilotos, criterios ex ante de éxito y retirada) para avanzar hacia una meseta de productividad educativa. Se concluye que la IAGen, gobernada en términos pedagógicos, puede convertirse en infraestructura de aprendizaje aumentada que potencia la comprensión, la inclusión y la creatividad, preservando la confianza y la agencia de docentes y estudiantes.

Palabras clave: inteligencia artificial generativa; tecnología de la educación; evaluación educativa; equidad en la educación; integridad académica; política educativa

ABSTRACT

This article proposes a methodological shift in the conversation on generative artificial intelligence (GenAI) in education: moving the focus from myth to informed projection. After situating GenAI within the cycle of hype and adjustment y framing it through the Promethean dilemma and the principle of responsibility, the paper introduces a pragmatic framework comprising: i) a checklist to identify myths (totalization, linear causality, immunity to evidence y erasure of conditions), ii) a template for formulating auditable projections (purpose, population/conditions, metrics/-milestones, risks/safeguards y review points) y iii) a protocol for reading sociotechnical imaginaries. The analysis addresses three recurrent myths: “AI will replace teachers,” “AI guarantees autonomy and equity,” and “the death of originality.” Each case shows why these operate as myths and formulates informed projections aligned with UNESCO’s four lenses (relevance, equity, scalability, sustainability) and with the Promethean rules (proportionality/reversibility and hallucination control). Policy implications are derived from explicit acknowledgment of AI use, meaningful human oversight in high-stakes decisions, authentic assessment centered on process and voice, data governance with transparency and bias audits, and equity strategies prioritizing high-need schools. The discussion integrates the findings and proposes evidence cycles (pilots with ex-ante success and withdrawal thresholds) to progress toward a plateau of educational productivity. The article concludes that, when governed on pedagogical terms, GenAI can become an augmented learning infrastructure that enhances understanding, inclusion, and creativity, while preserving the trust and agency of teachers and students.

Keywords: Generative Artificial Intelligence; Educational Technology; Educational Assessment; Equity in Education; Academic Integrity; Educational Policy

01 INTRODUCCIÓN

En los últimos dos años, la irrupción de la inteligencia artificial generativa (IAGen) ha trascendido el terreno de la promesa tecnológica para instalarse en la práctica cotidiana de los sistemas educativos y de la ciudadanía, lo que supone un punto de inflexión, por su integración en la vida diaria[1] y por su potencial para producir contenidos sintéticos de alta calidad en múltiples formatos,[2] con efectos tangibles en la docencia y en la evaluación. Esta aceleración convive, sin embargo, con una proliferación de narrativas contrapuestas, utópicas y distópicas que, al amplificarse en el imaginario social, terminan por generar mitos resistentes a la refutación y exageraciones sobre lo que la IA puede ofrecer en el aula.

En este trabajo se propone una lectura crítica de esta realidad, distinguiendo entre mitos y proyecciones. Por mitos entendemos narrativas tecnoculturales de alto poder simbólico que acompañan a cada oleada tecnológica, articulando promesas y temores que simplifican la complejidad y legitiman agendas, tal como muestra Mosco[3] al analizar los mitos digitales que, en el ámbito de la educación, se cristalizan en los llamados mitos EdTech, como, por ejemplo: “más tecnología implica un mejor aprendizaje”, “la IA sustituirá al docente” o “la personalización algorítmica resolverá desigualdades”.[4] Por proyecciones se hace referencia a expectativas explícitas situadas sobre futuros deseables o plausibles que orientan las decisiones de diseño pedagógico y de política.[5] En este orden, la sociología de las expectativas y el enfoque de los imaginarios sociotécnicos explican que las visiones del futuro (promesas, temores y agendas) moldean decisiones y recursos, pero que estas también deben someterse a contraste empírico y a la revisión ética conforme emergen evidencias y prácticas.[6]

Para situar la discusión en términos temporales y de gobernanza, se sigue el hype cycle: tras el pico de expectativas infladas, en el que proliferan promesas y casos de éxito aislados, sobreviene el valle de la desilusión, cuando los pilotos fallan, los costes afloran y se depuran proveedores; solo entonces, en la pendiente de la iluminación, comienzan a consolidarse usos con valor pedagógico verificable, aparecen segundas y terceras generaciones de herramientas y se clarifican criterios de viabilidad, hasta alcanzar la meseta de productividad, donde la adopción es estable y el retorno educativo resulta evaluable.

Este marco no es un destino inevitable, sino una guía para decidir qué innovaciones merecen escalar, cuándo y bajo qué salvaguardas, evitando tanto el arribismo del pico como la inercia del valle; de ahí la necesidad de avanzar desde los mitos hacia proyecciones informadas con condiciones de uso, métricas y revisión periódica.[7]

Las reflexiones de este trabajo se sitúan en una doble clave. Por un lado, se recupera la perspectiva histórica y crítica sobre la tecnología según la cual, como recordaba Kranzberg,[8] la tecnología “no es buena ni mala; ni siquiera es neutral”, de modo que su evaluación depende de usos, contextos y fines educativos, no solo de capacidades técnicas.

1 Daniel Burgos, “Inteligencia artificial y humanismo: construyendo un futuro centrado en el ser humano”, Nueva Revista, n.o 192 (2024): 170-185. **2** Francisco José García-Peñalvo and Andrea Vázquez-Ingelmo, “What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI”, International Journal of Interactive Multimedia and Artificial Intelligence 8, n.o 4 (2023): 7-16. **3** Vincent Mosco, The Digital Sublime: Myth, Power, and Cyberspace (The MIT Press, 2004). **4** Cristóbal Suárez-Guerrero, Juliana E. Raffaghelli y Pablo Rivera-Vargas, Mitos EdTech. Desmontando el solucionismo tecnológico en educación (Barcelona, España: Editorial UOC, 2024). **5** Nik Brown y Mike Michael, “A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects”, Technology Analysis & Strategic Management 15, n.o 1 (2003). **6** Sheila Jasanoff, “Future Imperfect: Science, Technology, and the Imaginations of Modernity”, en Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power, eds. Sheila Jasanoff y Sang-Hyun Kim (Chicago, EE. UU.: The University of Chicago Press, 2015), 1-33. **7** Jackie Fenn y Mark Raskino, Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time (Boston, Massachusetts, EE. UU.: Harvard Business Review Press, 2008). **8** Melvin Kranzberg, “Technology and History: ‘Kranzberg’s Laws’”, Technology and Culture 27, n.o 3 (1986), 544-560.

Por otro lado, se incorpora el dilema de Prometeo como una metáfora regulativa para la IAGen: si el fuego de la IA habilita la creatividad y la eficiencia, también introduce riesgos, como la opacidad (caja negra), las alucinaciones o los sesgos,[9] que exigen una responsabilidad y un gobierno proporcional a su potencia. La literatura reciente ha utilizado explícitamente esta clave prometeica para describir la tensión entre la innovación y las consecuencias sociales en IA, especialmente en el cruce entre la creatividad y las alucinaciones,[10] de donde surge la pregunta: ¿debe permitirse o prohibirse su uso generalizado, dada su opacidad y su tendencia a alucinar? Sin embargo, esta cuestión pierde centralidad considerando la velocidad de evolución y de aceptación de la IAGen, con cientos de millones de usuarios a nivel mundial y una más que notable penetración en el dominio educativo y académico, como lo muestran el informe Time for Class 2025 en el contexto estadounidense[11] o el informe sobre el impacto de la IA en el sector público español.[12]



Foto: Magnific

Para encuadrar la metáfora prometeica en clave educativa, se adopta el principio de responsabilidad de Hans Jonas:[13] en una era de poder técnico sin precedentes, las decisiones sobre IA en el aula deben atender no solo a beneficios inmediatos, sino también a efectos de largo plazo sobre estudiantes, instituciones y generaciones futuras; de ahí la necesidad de una ética de la precaución y de la previsión que priorice evitar daños potencialmente irreversibles y someter las promesas tecnológicas a contraste continuo.

Esta ética dialoga con la vergüenza prometeica descrita por Günther Anders,[14] que consiste en el sentimiento de inferioridad ante máquinas percibidas como más perfectas que las personas, y con su entendimiento contemporáneo como resentimiento tecnológico: bajo estas afecciones, existe la tentación de delegar la agencia y abdicar del juicio propio en favor de sistemas que operan a nuestros ojos con superioridad y opacidad.[15] Esto se podría pensar que se manifiesta con la IAGen, ya que, al delegar a herramientas como ChatGPT tareas tan fundamentales como las relacionadas con la enseñanza y el aprendizaje, las personas estarían cediendo su autonomía. Este acto de sumisión a un dispositivo de nuestra propia creación sería, para Anders, una renuncia a la propia conciencia y un acto de "veneración de nuestros propios productos". Así, el ser humano se estaría transformando en un homo faber,[16] que se subordina a sí mismo a través de sus creaciones. Integrar esta advertencia en la política educativa no implica demonizar la herramienta ni glorificar la sustitución docente, sino diseñar condiciones de uso (propósitos, salvaguardas, transparencia y revisión) que mantengan la centralidad de la deliberación humana y del valor pedagógico.

9 Francisco José García-Peñalvo, Faraón Llorens-Largo y Javier Vidal, "The New Reality of Education in the Face of Advances in Generative Artificial Intelligence", RIED: Revista Iberoamericana de Educación a Distancia 27, n.o 1 (2024), 9-39. **10** Tanmoy Chakraborty and Sarah Masud, "The Promethean Dilemma of AI at the intersection of hallucination and creativity", Communications of the ACM 67, n.o 10 (2024), 26-28. **11** Catherine Shaw et al., Time for Class 2025. Empowering Educators, Engaging Students (s. l.: Tyton Partners, June 2025). **12** Jorge Galindo et al., El impacto de la IA en el sector público español. Avances para el acceso a servicios básicos (España: ESADE, 2025). **13** Hans Jonas, El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica (Barcelona, España: Herder Editorial, 1995). **14** Günther Anders, "On Promethean Shame", en Prometheanism: Technology, Digital Culture, and Human Obsolescence, ed. Christopher John Muller (Lanham: Rowman & Littlefield Publishers, 2016), 29-95. **15** Babette Babich, "Günther Anders's 'Promethean Shame': Technological Ressentiment and Surveillance", The Harvard Review of Philosophy 31, (2024), 75-97. **16** Henri Bergson, Las dos fuentes de la moral y de la religión (Buenos Aires, República Argentina: Sudamericana, 1962); Hannah Arendt, Human Condition, 2.a ed. (Chicago, EE. UU.: The University of Chicago Press, 1998).

En coherencia con el GEM Report de la Unesco,[17] se aborda la tecnología educativa desde cuatro lentes: pertinencia, equidad, escalabilidad y sostenibilidad, que se traducen en decisiones operativas en el aula. La pertinencia responde al para qué (qué propósito pedagógico cumple y qué problema de aprendizaje resuelve); la equidad responde al para quién (quién accede, quién podría quedar fuera y cómo se compensan brechas); la escalabilidad responde al con qué recursos (tiempo, formación docente, infraestructura y costes totales de propiedad, no solo la inversión inicial); y, finalmente, la sostenibilidad responde al por cuánto tiempo y con qué impactos (materiales, energéticos, dependencia de proveedores y mantenimiento). De este modo, cualquier adopción de IAGen debe explicitar cuándo usarla, en qué condiciones y con qué salvaguardas (privacidad, seguridad, transparencia), y cómo se evaluará su valor educativo (criterios y evidencias), complementando, no sustituyendo, la interacción docente. Esa orientación permite diferenciar entre prohibiciones reactivas, adopciones acríticas y proyecciones informadas que especifiquen condiciones de uso, métricas de valor educativo y salvaguardas, lo cual se reafirma en los siete principios del Manifiesto para una IA Segura en la Educación,[18] como se presenta en la tabla 1.

Tabla 1. Correspondencia entre los siete principios del manifiesto y las cuatro lentes de la Unesco

Principio (Manifiesto)	Pertinencia (¿para qué?)	Equidad (¿para quién?)	Escalabilidad (¿con qué recursos?)	Sostenibilidad (¿por cuánto tiempo/impacto?)
1. Supervisión humana y responsabilidad	Asegura que la IA complementa la docencia y sirve a fines pedagógicos claros	El derecho de apelación y decisiones explicables reducen las arbitrariedades	(Indirecta)	La gobernanza y la rendición de cuentas sostienen los usos responsables en el tiempo
2.Confidencialidad (privacidad y control de datos)	Asegura la confianza en la forma en que la IA gestiona/accede a los datos	Evita que el uso de IA excluya o exponga a estudiantes por requisitos de registro/terceros	(Puede exigir inversión y políticas, pero habilita despliegue seguro)	Propiedad/gestión institucional y minimización de riesgos legales a largo plazo
3. Alineación con estrategias educativas	Vincula la IA con objetivos curriculares e institucionales, no con modas	Asegura el acceso equitativo dentro de una institución	Evita la complejidad indebida y alinea los recursos, facilitando la adopción a escala	Conecta con el coste total de propiedad y con las decisiones de largo plazo (evita dependencias opacas)
4. Alineación con prácticas didácticas	Integra la IA en diseños instruccionales reales (no añadidos sin propósito)	Adaptabilidad a contextos y a necesidades diversas del estudiantado	Usabilidad didáctica, reduce cargas formativas y favorece un despliegue amplio	(Indirecta: continuidad pedagógica)
5. Precisión y explicabilidad	Resultados fiables y verificables soportan aprendizajes y evaluaciones válidos	La mitigación de errores/sesgos protege a colectivos vulnerables	(Mejor calidad reduce el trabajo)	La calidad mantenida y la evaluación continua favorecen la pervivencia del uso

17 Unesco, Global Education Monitoring Report 2023: Technology in Education – A Tool on whose Terms? (París, Francia: Unesco, 2023). **18** Marc Alier et al., Safe AI in Education Manifesto. Version 0.4.0, octubre 8 de 2024; Francisco José García-Peñalvo et al., "Safe, Transparent, and Ethical Artificial Intelligence: Keys to Quality Sustainable Education (SDG4)", IJERI – International Journal of Educational Research and Innovation, n.o 22 (2024): 1-21.

6. Interfaz y comportamiento integral (transparencia y experiencia de usuario)	Clarifica los casos de uso y los límites (marca contenido generado)	La comprensibilidad y la accesibilidad disminuyen barreras de uso	Un buen diseño de interacción reduce el soporte y la formación, habilitando la escala	(Indirecta)
7. Formación ética y transparencia (datos/modelos)	Ofrece credibilidad en las respuestas	La minimización de sesgos y la transparencia del entrenamiento protegen la justicia educativa	La transparencia facilita las compras públicas, las auditorías y la adopción institucional	Las prácticas éticas y la trazabilidad fortalecen la licencia social y la continuidad

Fuente. elaboración propia.

El artículo se sustenta en tres aportes complementarios. En primer lugar, en un análisis conceptual que diferencia mito y proyección y propone indicadores prácticos para identificarlos en el discurso educativo y en las decisiones de aula (falsabilidad, especificación contextual, mecanismos de evaluación y revisión). Segundo, en una lectura crítica de tres mitos recurrentes tras la IAGen, como son la sustitución del profesorado, la autonomía garantizada por la personalización y la muerte de la originalidad. En tercer y último lugar, en una propuesta de proyecciones responsables centradas en a) tutorías aumentadas bajo supervisión docente; b) la evaluación auténtica y el rediseño de tareas; c) la alfabetización en IA; y d) la gobernanza y la transparencia institucional, con atención a los límites de los detectores automáticos y a los riesgos de falsos positivos y efectos de injusticia.

La contribución central implica, por tanto, desplazar el foco del mito a la proyección: pasar de afirmaciones maximalistas o temores paralizantes a escenarios verificables que especifiquen para qué, para quién, en qué condiciones y con qué evidencias conviene usar la IA en educación. En sintonía con la idea de que las tecnologías atraviesan ciclos de sobreexpectación y de posterior reajuste,[19] el texto defiende un enfoque de adopción prudente, iterativa y basada en evidencias, evitando tanto el solucionismo como la prohibición indiscriminada. Este posicionamiento reconoce que el estudiantado ya utiliza IAGen y que la respuesta educativa no puede reducirse a la vigilancia o a la proscripción, sino a la agencia docente y estudiantil, a la claridad de normas y a la formación crítica en el uso de estas herramientas.[20]

El resto del artículo se organiza del modo siguiente. En la sección 2 se presenta el marco conceptual, en clave pragmática, que servirá de andamiaje para el análisis posterior. La sección 3 desarrolla el examen de los mitos con contraste empírico, aplicando la lista de verificación y, para cada caso, formulando proyecciones informadas y señalando implicaciones de política. Finalmente, la sección 4 reúne la discusión y las conclusiones, integrando de forma transversal los criterios de adopción responsable, la gobernanza y las consideraciones éticas e institucionales. Con ello se busca contribuir a una agenda que sitúe la IA al servicio de fines educativos valiosos, bajo supervisión humana y con mecanismos de evaluación y de rendición de cuentas.

19 Fenn y Raskino, Mastering the Hype Cycle. **20** Francisco José García-Peñalvo, “Inteligencia Artificial en el aula. De los mitos a la realidad”, 5.º Congreso Caribeño de Investigación Educativa, Punta Cana, República Dominicana, 9-12 de abril de 2025.



Foto: Magnific

02 IDENTIFICAR MITOS, FORMULAR PROYECCIONES Y LEER IMAGINARIOS

La distinción entre mito y proyección requiere de un andamiaje que evite dos reduccionismos habituales: el determinismo tecnológico y el voluntarismo normativo. En este trabajo se toma el primero en sentido estricto, esto es, la idea de que la tecnología impulsa lineal y autónomamente el cambio social, distinguiendo el determinismo mediático propio de la tradición de McLuhan, que subraya el papel configurador del medio en la percepción y la organización social (la "aldea global", "el medio es el mensaje").[21] Esta sensibilidad ayuda a comprender el poder formativo de las infraestructuras comunicativas, pero, para evitar causalidades unidireccionales, se adopta la crítica histórica de Smith y Marx,[22] quienes cuestionan que la técnica determine por sí sola las formas políticas, económicas y culturales. Además, se sitúa el análisis en la estela del social shaping of technology de MacKenzie y Wajcman,[23] que muestra cómo los artefactos y sistemas técnicos se co-configuran con valores, instituciones y prácticas. Con este encuadre, la IA en educación no se entiende como una fuerza externa que produce efectos pedagógicos, sino como un ensamblaje sociotécnico cuyo valor depende de diseños didácticos, contextos y gobernanza.

En lo que respecta al voluntarismo normativo, se entiende por tal al acto de suponer que basta con querer ciertos fines para que la tecnología los realice, muy presente en la era de la IA. Este sesgo aparece, por ejemplo, en el diseño por defecto,[24] cuando se confía en que decisiones de diseño (como puede ser el caso de las configuraciones por defecto) imprimen sin más determinados valores o conductas a los usuarios, sin considerar mediaciones y contextos. Swierstra[25] subraya aquí los impactos blandos y su ambigüedad normativa. El voluntarismo normativo también se observa en enfoques de regulación que presumen que normas emanadas de distintos actores (productores, usuarios, gobiernos) bastan para ordenar sistemas sociotécnicos complejos; frente a ello, Smith y Miller[26] proponen una teoría institucional (Theory of Institutional Technology Actors and Norms [TITAN]) que distribuye derechos, responsabilidades y mecanismos de gobernanza entre actores, evitando simplificaciones voluntaristas.

En lo que sigue, mito, expectativa e imaginario sociotécnico se abordan como categorías complementarias para comprender cómo se construyen, circulan y gobiernan los relatos sobre IA en educación.

21 Marshall McLuhan y Bruce R. Powers, The Global Village: Transformations in World Life and Media in the 21st Century (Nueva York, EE. UU.: Oxford University Press, 1989). **22** Merritt Roe Smith y Leo Marx, eds., Does Technology Drive History?: The Dilemma of Technological Determinism (Cambridge, Massachusetts, EE. UU.: The MIT Press, 1994). **23** Donald MacKenzie y Judy Wajcman, eds., The Social Shaping of technology, 2.a ed. (Buckingham, Inglaterra: Open University Press, 1999). **24** Langdon Winner, "Do Artifacts Have Politics?", en Computer Ethics, ed. John Weckert (Londres, Inglaterra: Routledge, 2007), 121-130. **25** Tsjalling Swierstra, "Identifying the Normative Challenges Posed by Technology's 'Soft' Impacts", Etikk I Praksis - Nordic Journal of Applied Ethics 9, n.o 1 (2015), 5-20. **26** Marcus Smith y Seumas Miller, "Technology, Institutions and Regulation: Towards a Normative Theory", AI & Society 40, n.o 2 (2025), 1007-1017.

2.1 CÓMO DETECTAR UN MITO (LISTA DE VERIFICACIÓN)

Para la identificación de un mito en el contexto educativo, se aplicará una heurística de lectura de textos de política, prensa, informes y materiales docentes. Los indicios típicos son:

1. Totalización: promesas/amenazas sobre la educación en bloque ("revolucionará", "hará obsoletos a"), sin dominios concretos de aplicación. Esta retórica es característica de las mitologías digitales que identifica Mosco.[27]
2. Causalidad lineal y monocausal: "más X tecnológico → mejor aprendizaje", sin mediaciones curriculares, didácticas o institucionales; patrón documentado en la crítica reciente a los mitos EdTech.[28]
3. Inmunidad a la evidencia: el enunciado se mantiene, aunque existan contraejemplos o evaluaciones nulas o mixtas. Un claro ejemplo de esto es el caso de Knewton. En 2015, se lo presentó como un mind-reading robo tutor in the sky, que recogía la célebre promesa de su fundador: "A robot tutor in the sky that can semi-read your mind", emblema de expectativas desbordadas en personalización algorítmica.[29] En 2019, solo cuatro años después, Wiley adquirió los activos de Knewton por un importe no revelado, lo que significó el final de una startup que había recaudado más de 180 millones de dólares y que simbolizó el paso del pico de expectativas al ajuste de la realidad.[30]
4. Borrado de condiciones: no se explicita para qué población van dirigidos, con qué tareas, en qué contextos ni con qué recursos (tiempo, formación, datos, infraestructura). Los mitos prosperan precisamente por esa ambigüedad fértil.[31]

Cuando el discurso marque tres o cuatro de estos indicios, debería tratarse como un mito; en la sección 3 se contrasta con evidencias y diseños didácticos.

2.2 CÓMO FORMULAR UNA PROYECCIÓN INFORMADA (PLANTILLA)

Una proyección no es una conjetura libre: siguiendo la sociología de las expectativas,[32] debe configurar la innovación de forma auditable. Por eso se debería cumplir que toda proyección incluya a) propósito (resultado de aprendizaje o problema que aborda), b) población/condiciones (curso, tarea, recursos, restricciones), c) métricas e hitos (qué y cuándo medir), d) riesgos y salvaguardas (privacidad, sesgos, declaración/reconocimiento de uso, revisión humana) y e) puntos de revisión (criterios para continuar/ajustar/retirar).

Una proyección se podría presentar mediante la siguiente forma canónica:

> Si (tarea T, con estudiantado P, bajo condiciones C), entonces (se espera Δ en métrica M), siempre que (salvaguardas S) y se revisará en (hito H).

Esta estructura evita la performatividad hueca de las expectativas y alinea actores/recursos con fines, como sugiere la literatura de expectativas en innovación.

27 Mosco, The Digital Sublime. **28** Cristóbal Suárez-Guerrero, Pablo Rivera-Vargas y Juliana Raffaghelli, "EdTech Myths: Towards a Critical Digital Educational Agenda", Technology, Pedagogy and Education 32, n.o 5 (2023), 605-620. **29** Eric Westervelt, "Meet the Mind-Reading Robo Tutor in the Sky", NPR (blog), octubre 13 de 2015. **30** Lindsay McKenzie, "End of the Line for Much-Hyped Tech Company". Inside Higher Ed (blog), mayo 6 de 2019.**31** Mosco, The Digital Sublime. **32** Brown and Michael, "A Sociology of Expectations".

2.3 PROTOCOLO DE LECTURA PARA IMAGINARIOS SOCIO-TÉCNICOS

Se propone un protocolo pragmático compuesto de tres pasos para "leer" el imaginario que subyace a un documento o política:[33]

1. Horizonte normativo declarado: ¿qué futuro deseable articula (por ejemplo, "modernización por digitalización", "equidad por personalización")?, ¿qué valores prioriza u omite?
2. Coproducción institucional: ¿qué decisiones de gobernanza, datos y estándares se asocian a ese futuro (proveedores, propiedad de datos, criterios de evidencia, papel docente)?
3. Trayectorias y exclusiones: ¿qué vías de acción habilita o cierra (currículo, evaluación, infraestructura), y para quién? Esta lectura hace visible cómo los imaginarios ordenan prioridades y distribuyen beneficios/riesgos.

Cada caso presentado en este artículo quedará anotado con su imaginario dominante, lo que facilitará el contraste con las lentes de la Unesco[34] ya introducidas en la tabla 1.

2.4 DILEMA DE PROMETEO EN CLAVE DECISIONAL

Se operativiza en dos reglas de decisión que buscan mitigar la deriva hacia la delegación del juicio,[35] manteniendo la agencia y la transparencia en los contextos educativos:[36]

1. **Regla 1.** Proporcionalidad y reversibilidad: cuanto mayores sean la opacidad y el daño potencial (por ejemplo, decisiones evaluativas automáticas), más altas deben ser la barrera de evidencia, la revisión humana significativa y la capacidad de revertir efectos.[37]
2. **Regla 2.** Potencia creativa bajo control de la alucinación: maximiza los usos de la IAGen con un alto valor creativo, pero bajo coste de error (ideación, retroalimentación formativa), y minimiza o condiciona aquellos con alto coste de confabulación (juicios sumativos, verificación factual), según advierte el debate sobre el "dilema prometeico" entre creatividad y alucinación.[38]

33 Sheila Jasanoff y Sang-Hyun Kim, eds., Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power (Chicago, EE. UU.: The University of Chicago Press, 2015). **34** Unesco, Global Education Monitoring Report 2023. **35** Anders, "On Promethean Shame". **36** Virginia Ballesteros, "De Günther Anders al transhumanismo: la obsolescencia del ser humano y la mejora moral", ISEGORÍA. Revista de Filosofía Moral y Política, n.o 63 (2020), 289-310. **37** Jonas, El principio de responsabilidad. **38** Chakraborty and Masud, "The Promethean Dilemma of AI…".



Foto: Magnific

2.5 CÓMO SE USARÁ ESTE MARCO EN EL RESTO DEL ARTÍCULO

En la sección 3, sobre mitos, se empleará el marco de manera sistemática: i) identificación del mito mediante la lista de verificación de la sección 2.1; ii) lectura del imaginario sociotécnico asociado según el protocolo de la sección 2.3; iii) reformulación del mito en proyecciones informadas siguiendo la plantilla de la sección 2.2 y su evaluación con las cuatro lentes de pertinencia, equidad, escalabilidad y sostenibilidad, tal como fija el GEM 2023;[39] y iv) cierre de cada subapartado con implicaciones de política y límites operativos, a la luz de las reglas prometeicas de la sección 2.4 (proporcionalidad/reversibilidad y control de alucinación).

La sección final de discusión y conclusiones integrará de forma transversal los hallazgos: sintetizará patrones comunes y diferencias entre casos, derivará criterios operativos de adopción responsable (gobernanza, supervisión humana, evaluación y puntos de reversión), conectará las proyecciones con modelos aplicados y ejemplos de rediseño curricular, y propondrá una agenda de investigación y de política alineada con las lentes de la Unesco[40] y con el principio de responsabilidad. Con ello se cerrará el argumento central del artículo: desplazar el foco del mito a proyecciones verificables y gobernadas en nuestros términos pedagógicos.

03 MITOS: CRITERIOS DE IDENTIFICACIÓN, CONTRASTE Y REFORMULACIÓN

Esta sección identifica y examina tres mitos recurrentes tras la irrupción de la IAGen, aplicando la lista de verificación establecida en el marco conceptual: 1) totalización, 2) causalidad lineal, 3) inmunidad a la evidencia y 4) borrado de condiciones. Un enunciado que cumpla tres o más de estos indicios debe tratarse como un mito, no como un diagnóstico útil. Además, cada mito se reescribe como una proyección informada mediante la plantilla Si-entonces-siempre que, y se revisará en, especificando propósito, población/condiciones, métricas/hitos, riesgos y salvaguardas, y puntos de revisión. Con ello se evita la performatividad hueca de las expectativas y se favorece una gobernanza pedagógica basada en evidencias. Esta estrategia dialoga con las cuatro lentes del GEM Report 2023 (pertinencia, equidad, escalabilidad, sostenibilidad)[41] y con las recomendaciones de política que sitúan la tecnología en términos pedagógicos y como complemento de la interacción docente, no como su sustituto. La clave está en pasar del miedo social y del impulso a prohibir como respuesta inmediata a la necesidad de comprender qué aporta la IAGen a los procesos de enseñanza, aprendizaje y evaluación (análisis crítico, comparación de fuentes, formulación de preguntas), y cómo reformular prácticas y tareas, sin abdicar de la agencia docente y estudiantil.[42]

39 Unesco, Global Education Monitoring Report 2023. **40** Ibíd. **41** Ibíd. **42** García-Peñalvo, “Inteligencia artificial en el aula”.

3.1 LA IA SUSTITUIRÁ AL PROFESORADO

La afirmación de que la IA desplazará al profesorado se alimenta de una mezcla de ansiedad laboral,[43] promesas de tutorización automática y relatos de "aulas-fábrica" sin humanidad.[44] No es casualidad que, junto con la difusión de la IAGen, resurja el discurso de los tutores virtuales siempre disponibles, auspiciado por voces tecnológicas de alto perfil con claros intereses económicos, y que, al mismo tiempo, emerjan movimientos de oposición desde la sociedad civil.[45] Sin embargo, esta presunción se basa en una comprensión reduccionista de lo que implica el proceso de enseñanza-aprendizaje, porque la educación efectiva no consiste únicamente en transmitir datos o información, ámbito en el que las máquinas pueden destacarse, sino que requiere agencia.[46]

A continuación, se aplica la lista de verificación para determinar si se está o no ante un verdadero mito.

- **Totalización:** a partir de casos puntuales de generación de explicaciones o ejercicios se extrapola una supuesta automatización del conjunto de la docencia.
- **Causalidad lineal:** se presupone que "más IA → mejor aprendizaje" sin mediaciones didácticas, curriculares o institucionales.
- **Inmunidad a la evidencia:** se ignora el corpus acumulado que sitúa a la tecnología, la IA en este caso, como un complemento de la docencia, no como su sustituto, incluidos marcos de política que desaconsejan automatizar decisiones pedagógicas de alto impacto sin supervisión humana.[47]
- **Borrado de condiciones:** la consigna de sustituir no especifica ni las tareas docentes, ni con qué estudiantes, en qué contextos, con qué recursos y bajo qué salvaguardas podrían estas automatizarse con un sentido pedagógico (privacidad, transparencia, evaluación de sesgos). Enseñar es más que entregar contenidos. La educación es una práctica relacional y emocional, y reducir al profesorado a meros trabajadores del conocimiento elimina el valor de la mentoría, del acompañamiento y de lo relacional.[48]

En consecuencia, dado que esta narrativa marca los cuatro indicios, opera como un mito. Además, el contraste refuerza la conclusión. Incluso en dominios donde los sistemas inteligentes llevan décadas mostrando resultados (por ejemplo, la práctica guiada y la retroalimentación en tutorización), los efectos dependen de tareas bien definidas y contextualizadas; y, sobre todo, no agotan la función docente, que incluye el diagnóstico formativo, el andamiaje, la construcción de comunidad y la mediación ética en evaluación. Debe pasarse entonces de la noción IA versus docentes a la de IA y docentes, es decir, hay una necesidad de concebir la tecnología como una herramienta complementaria, no como un reemplazo mágico.

43 Ashraf Alam, "Should Robots Replace Teachers? Mobilisation of AI and Learning Analytics in Education", en Proceedings of the 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3) (Mumbai, India, 3-4 December 2021) (EE. UU.: IEEE, 2021). **44** Jianning Dang y Li Liu, "Dehumanization Risks Associated with Artificial Intelligence Use", American Psychologist, (2025); Ken Robinson, "How to Escape Education's Death Valley", TED Talks Education, 2013. **45** Jeffrey R. Young, "As More AI Tools Emerge in Education, so Does Concern Among Teachers About Being Replaced", EdSurge (blog), julio 2 de 2024. **46** Francisco José García-Peñalvo, "Mito de la inteligencia. Más allá de una educación de silicio", en Mitos EdTech. Desmontando el solucionismo tecnológico en educación, ed. Cristóbal Suárez-Guerrero, Juliana E. Raffaghelli y Pablo Rivera-Vargas (Barcelona: Editorial UOC, 2024), 79-87. **47** U.S. Department of Education. Office of Educational Technology, Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations (Washington, EE. UU.: U.S. Department of Education, 2023). **48** David Edwards, "Why this Teachers' Union Isn't Afraid AI will Replace them", World Economic Forum, 2024.

La lectura de imaginarios ayuda a explicar la persistencia del mito: bajo el horizonte de la modernización por digitalización, la docencia se recodifica como una instrucción transmisiva y el profesorado como un cuello de botella. De ahí que se sobrerrepresenten las tareas automatizables (explicar, proponer ejercicios) y se subrepresenten las dimensiones no codificables de la práctica (juicio situado, cuidado, motivación, agencia compartida).

Los marcos internacionales recientes se alinean con esta relectura. El GEM Report 2023[49] sostiene que la tecnología debe introducirse con base en pertinencia, equidad, escalabilidad y sostenibilidad, complementando la interacción cara a cara con los docentes. El U.S. Department of Education recomienda explícitamente evitar automatizar las decisiones pedagógicas de alto impacto sin revisión humana, encuadrando la IA como un medio para ampliar las capacidades docentes, no para sustituirlas.[50] En sintonía, la Organización para la Cooperación y el Desarrollo Económicos (OCDE)[51] propone oportunidades, guías y salvaguardas para un uso efectivo y equitativo de la IA, con la agencia docente como el pilar de la integración responsable.

Esta convergencia normativa es congruente con la apuesta por plantear rediseños que exploten los puntos fuertes de la IAGen y que a la vez generen tareas más creativas que no puedan suplantar el talento humano; incorporar defensas orales y contextos locales que obliguen a comprender y transferir; y reforzar la alfabetización en IA de estudiantes y del propio profesorado. Al mismo tiempo, se advierte sobre el riesgo de un uso perezoso de la IA que abdique de la autoría y del aprendizaje auténtico, recogiendo la advertencia de Hamsa Bastani et al.[52] sobre la confianza acrítica en modelos opacos.

3.1.1 REFORMULACIÓN: DEL MITO A LA PROYECCIÓN INFORMADA

Siguiendo la plantilla canónica propuesta en la sección 2.2, se proponen a modo de ejemplo dos proyecciones auditables, que desplazan la narrativa de la sustitución hacia la simbiosis pedagógica.



Foto: Magnific

49 Unesco, Global Education Monitoring Report 2023. **50** U.S. Department of Education, Artificial Intelligence and Future of Teaching… **51** OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem (París, Francia: OECD Publishing, 2023); OECD-Education International, Opportunities, Guidelines and Guardrails on Effective and Equitable Use of AI in Education (París, Francia: OECD Publishing, 2023). **52** Hamsa Bastani et al., "Generative AI Can Harm Learning", The Wharton School Research Paper, (2024), 1-59.

Proyección 3.1-A (planificación y diferenciación lectora en secundaria)

Si en 1.º-2.º de ESO se emplea la IAGen para adaptar textos curriculares a diferentes niveles de lectura y generar variantes de actividades en ciencias sociales y lengua, con conectividad estable, formación docente breve (dos sesiones) y una política institucional de declaración del uso en materiales, entonces se espera una reducción de 30 a 45 minutos semanales en tareas de preparación por docente y una mejora en la idoneidad de los textos (adecuación a nivel lector y objetivos), siempre que toda salida de IA sea revisada por el profesorado, verificada en lo factual, etiquetada en el material entregado al estudiantado e integrada en un diseño que explicita objetivos, rúbricas y criterios. Se revisará en las semanas 4 y 10 mediante a) una encuesta de carga docente, b) un análisis de desempeño en comprensión lectora y c) una revisión colegiada de muestras. Esta proyección es pertinente (libera tiempo para interacción), equitativa (genera accesibilidad textual), escalable (a partir de plantillas y prompts compartidos, con microformación) y sostenible (mediante reutilización de bancos de tareas).

Proyección 3.1-B (retroalimentación formativa en escritura preuniversitaria)

Si en 2.º de bachillerato la IAGen se usa para generar retroalimentación preliminar en borradores de ensayo (estructura, cohesión, citas), bajo una rúbrica docente y con bibliografía obligatoria, entonces se espera un aumento de las iteraciones (≥2 por entrega) y una mejoría de la calidad argumentativa a la semana 6, siempre que la retroalimentación algorítmica se marque como preliminar, el estudiantado declare el uso de IA y el docente modere una muestra aleatoria con una segunda corrección humana. Se revisará en las semanas 3 y 6 con rúbrica ciega y análisis de dependencia de patrones. Esta proyección es pertinente (es una práctica deliberada), equitativa (apoya a quienes disponen de menos tutoría fuera de clase), escalable (mediante bancos de prompts y guías de verificación) y sostenible (con protocolos de privacidad y coste de propiedad claro).

Ambas proyecciones respetan las reglas prometeicas presentadas en la sección 2.4: proporcionalidad y reversibilidad (no automatizar decisiones sumativas; posibilidad de recurso y reevaluación) y control de alucinación (verificación factual y límites de uso). Además, se articulan con las pautas institucionales de referencia: no prohibir por principio, diferenciar el uso legítimo del uso engañoso y fomentar el reconocimiento explícito del uso de IA en trabajos académicos —por ejemplo, siguiendo el modelo de reconocimiento de la Universidad de Monash[53] —.

3.1.2 IMPLICACIONES DE POLÍTICA Y LÍMITES OPERATIVOS

La política institucional debe desincentivar la automatización opaca de funciones de alto impacto (calificación sumativa, acreditación, orientación), exigiendo la revisión humana significativa, la trazabilidad y la capacidad de reversión. Esto es consistente con el enfoque del (U.S. Department of Education. Office of Educational Technology 2023)[54] y con la lectura Unesco de gobernanza para "uso en nuestros términos".[55]

53 Monash University, "Acknowledging the use of generative artificial intelligence", 2025. **54** U.S. Department of Education, Artificial Intelligence and Future of Teaching… **55** Unesco, Global Education Monitoring Report 2023.

En la evaluación, la ruta sostenible pasa por métodos más controlados cuando corresponda, pero evitando un enfoque exclusivamente punitivo, así como por rediseños que exijan comprensión demostrable (defensas orales, tareas contextualizadas) y comunidades de práctica docentes para compartir estrategias.

Por último, el enfoque de alfabetización es clave para evitar la delegación perezosa. Se debe incidir en mantener la autoría y el aprendizaje auténtico como principios, enseñar cuándo y cómo usar IA, y marcar su intervención. Esto es congruente con las cuatro lentes del GEM 2023[56] y alinea la proyección con el estándar emergente de transparencia y supervisión humana en educación.[57]

3.1.3 CONCLUSIONES SOBRE LA SUSTITUCIÓN DEL PROFESORADO POR LA IA

La consigna de que la IA sustituirá al profesorado reúne rasgos de mito y oscurece los ámbitos en los que sí aporta valor. La vía razonable es convertir la consigna en proyecciones delimitadas, medibles y reversibles, que liberen tiempo para tareas de alto valor relacional y cognitivo, y que mantengan la centralidad del juicio docente. Esto se puede resumir en la fórmula ya mencionada: pasar de IA versus docentes a IA y docentes.

3.2 LA IA GARANTIZA LA AUTONOMÍA Y LA EQUIDAD EN EL APRENDIZAJE

En el discurso educativo contemporáneo, se ha popularizado la idea de que la IA, por sí sola, garantiza la autonomía del estudiantado y la equidad en el aprendizaje. Este mito suele alimentarse de promesas tecnológicas: se afirma que la IA personalizará la enseñanza para que cada estudiante aprenda a su ritmo (fomentando su autonomía) y que democratizará el acceso a recursos de calidad (promoviendo la equidad). Esto se alinea con el concepto de aprendizaje autodirigido, ya que la persona elige qué, cómo y cuándo aprender, lo que potencia su motivación y su metacognición, lo que supone una reestructuración de su percepción sobre lo que es aprender de forma autónoma.[58] Sin embargo, organismos internacionales como la Unesco advierten que, si bien la IA tiene potencial para abordar retos educativos relevantes, también conlleva riesgos, y puede aumentar las brechas si no se orienta con políticas inclusivas y con salvaguardas adecuadas.[59] Asumir que la IA garantiza la autonomía y la equidad constituye, por tanto, una simplificación excesiva que merece escrutinio.

A la luz de la lista de verificación establecida, el enunciado de que la IA garantiza autonomía y equidad reúne todos los rasgos de mito:

56 Ibíd. **57** OECD Digital Education Outlook 2023. **58** Ling Wang y Wenye Li, "The Impact of AI Usage on University Students' Willingness for Autonomous Learning", Behavioral Sciences 14, n.o 10 (2024), 1-16. **59** Stefania Giannini, La IA generativa y el futuro de la educación (París, Francia: Unesco, 2023); Stefania Giannini, "El uso de la IA en la educación: decidir el futuro que queremos", Unesco (blog), mayo 17 de 2024.

- Exhibe totalización, al presentar a la IA como solución universal a problemas complejos y de raíz estructural. Se asume que una única innovación tecnológica resolverá las cuestiones profundas de autonomía personal y justicia educativa, ignorando la diversidad de contextos y necesidades. Esta visión absolutista simplifica la realidad, al insinuar que la IA, por sí misma, siempre conduce a resultados positivos universales, lo cual carece de evidencia empírica.[60]
- Incurre en causalidad lineal, al suponer una relación directa y unívoca entre la adopción de IA y la mejora en la autonomía y la equidad, como si introducir algoritmos automáticamente empoderara al estudiantado y nivelara el campo educativo. En la práctica, la relación causa-efecto es mucho más contingente: los resultados dependen de cómo se integra la IA en el currículo, de la formación de los docentes, de la calidad de los datos, entre otros factores. De hecho, informes recientes subrayan que aún no existen pruebas sólidas de que las herramientas de IA generativa mejoren el rendimiento o la calidad del aprendizaje, lo que desmonta la idea de un efecto lineal garantizado.[61]
- Muestra inmunidad a la evidencia, pues persisten expectativas infladas en ciertos discursos políticos y de marketing, pese a los estudios críticos y a las advertencias de organismos especializados. Esta resistencia a la evidencia se manifiesta, por ejemplo, en la continua promoción de plataformas adaptativas supuestamente infalibles, incluso cuando evaluaciones independientes o metaanálisis encuentran resultados mixtos o modestos[62] u opiniones contrarias o matizadas del estudiantado a la excesiva autodirección.[63] La expectativa inflada se mantiene inmune a los datos empíricos contrarios, apoyándose más en anécdotas exitosas aisladas que en tendencias generales. Un síntoma de esta inmunidad es la forma como se celebran casos de uso llamativos, mientras que se minimizan otros hallazgos preocupantes, como sesgos algorítmicos[64] o la falta de impacto sostenido en entornos reales.[65]
- Practica borrado de condiciones, al invisibilizar o ignorar requisitos de infraestructura, formación docente y curación de datos, esenciales para que la IA no amplifique desigualdades.[66] Obviar estas condiciones es especialmente grave considerando la brecha digital, dado que aproximadamente un tercio de la población mundial no tiene acceso estable a internet (la conectividad sigue aumentando en todo el mundo, pero se revela una gran complejidad para llegar a las comunidades de los países de bajos ingresos)[67] y muchas escuelas carecen de los requisitos básicos (electricidad, dispositivos, soporte técnico)[68] para aprovechar la IA. Se presenta así un escenario descontextualizado donde la herramienta funciona igual de bien para todos, en cualquier lugar. La realidad es muy distinta, porque sin inversión deliberada para colmar la brecha digital en conectividad, contenidos y capacitación, la introducción de IA puede incluso ensanchar las desigualdades existentes, en vez de reducirlas. En cuanto a la autonomía, se pasa por alto que esta no surge mágicamente por usar una aplicación, sino que requiere alfabetización digital, pensamiento crítico y hábitos de estudio que deben desarrollarse con guía pedagógica, condiciones que la retórica tecnoutópica tiende a minimizar.

60 Giannini, "El uso de la IA en la educación". **61** Ibíd. **62** Sumei Hu, "The Effect of Artificial Intelligence-Assisted Personalized Learning on Student Learning Outcomes: A Meta-Analysis Based on 31 Empirical Research Papers", Science Insights Education Frontiers 24, n.o 1 (2024), 3873-3894. **63** Patricia O'Sullivan et al., "Student Perceptions of the Effectiveness of Adaptive Courseware for Learning", Current Issues in Emerging eLearning 7, n.o 1 (2020), 71-100. **64** Ryan S. Baker y Aaron Hawn, "Algorithmic Bias in Education", International Journal of Artificial Intelligence in Education 32, n.o 4 (2022); Sarah Lee, "Fairness in Educational Technology. Addressing Algorithmic Bias in Digital Literacy and Pedagogy", NumberAnalytics (blog), 2025. **65** Xiaoman Wang et al., "The Efficacy of Artificial Intelligence-Enabled Adaptive Learning Systems From 2010 to 2022 on Learner Outcomes: A Meta-Analysis", Journal of Educational Computing Research 62, n.o 6 (2024), 1348-1383.

66 U.S. Department of Education, Artificial Intelligence and Future of Teaching…; OECD, The Potential Impact of AI on Equity and Inclusion in Education (París, Francia: OECD Publishing, 2024). **67** International Telecommunication Union, Measuring digital development. Facts and Figures 2024 (Ginebra, Suiza: International Telecommunication Union, 2024). **68** Unesco, Global Education Monitoring Report 2024/5: Leadership in education – Lead for learning (París, Francia: Unesco, 2024).

Además, surge una aporía entre la automatización y la noción psicopedagógica de la autonomía de las personas, ya que, paradójicamente, una excesiva automatización puede chocar con la autonomía personal, una de las claves del aprendizaje humano,[69] profundamente ligada al desarrollo de la independencia, toda vez que explorar y experimentar son fundamentales para el pensamiento crítico, la resolución de problemas y la autoeficacia.[70]

Al confrontar el mito con datos y lineamientos recientes, se percibe claramente que la IA no garantiza automáticamente ni la autonomía ni la equidad educativa, dado que su impacto depende de las políticas educativas, las prácticas de implementación y la supervisión humana.

En el plano de la equidad, múltiples fuentes advierten que esta herramienta conlleva riesgos de reproducir o agravar las disparidades. La Unesco enfatiza que la tecnología tiene un historial de incremento de las desigualdades y que la IA no es una excepción y avisa de que, sin las debidas cautelas, podría amplificar las brechas educativas existentes.[71] De igual modo, la OCDE ha observado que la adopción no regulada de estas herramientas tiende a beneficiar primero a las escuelas con mayores recursos, dejando rezagadas a las más vulnerables. Esto desmonta la noción de equidad instantánea: si la introducción de IA ocurre "sin mucha supervisión ni directrices", como señalan los análisis comparados, los centros mejor dotados técnicamente "cosechan antes los beneficios", mientras que los otros se quedan atrás.[72] Lejos de nivelar por sí misma el terreno de juego, la IA puede perpetuar la estratificación a menos que se actúe deliberadamente para evitarlo.

En cuanto a la autonomía en el aprendizaje, la evidencia contemporánea desmonta visiones ingenuas de una autonomía garantizada por el simple uso. Si bien es cierto que las herramientas inteligentes pueden adaptarse al ritmo y al estilo de cada estudiante (lo cual potencialmente refuerza la agencia personal), también se han identificado efectos contraproducentes. Stefania Giannini[73] advierte del riesgo de que los estudiantes, acostumbrados a las respuestas instantáneas de la IA, las acepten como verdades absolutas, "recibiendo respuestas singulares y aparentemente definitivas". Esto podría atrofiar su capacidad de cuestionamiento y pensamiento crítico, mermando sus habilidades para tomar decisiones independientes y resolver problemas,[74] es decir que afectaría en última instancia su autonomía intelectual. En otras palabras, una dependencia acrítica de sistemas de IA puede convertir a la persona en un receptor pasivo de información enlatada, que acaba prefiriendo el camino rápido a la solución frente al esfuerzo de tener que llegar a esa meta,[75] justo lo opuesto a un aprendiz autónomo y reflexivo.



Foto: Magnific

69 Marta Ranieri, "Mito de la causalidad. Aporías entre los requisitos técnicos y las aspiraciones pedagógicas de la inteligencia artificial en la educación", en Mitos EdTech. Desmontando el solucionismo tecnológico en educación, eds. Cristóbal Suárez-Guerrero, Juliana E. Raffaghelli y Pablo Rivera-Vargas, 23-33 (Barcelona, España: Editorial UOC, 2024). **70** Albert Bandura, "Self-efficacy mechanism in human agency", American Psychologist 37, n.o 2 (1982), 122-147. **71** Giannini, La IA generativa y el futuro de la educación. **72** OECD, The potential impact of AI… **73** Giannini, La IA generativa y el futuro de la educación. **74** Long Bai, Xiangfei Liu y Jiacan Su, "ChatGPT: The cognitive effects on learning and memory", Brain-X 1, n.o 3 (2023), 1-9. **75** Chunpeng Zhai, Santoso Wibowo y Lily D. Li, "The Effects of Over-Reliance on AI Dialogue Systems on Students' Cognitive Abilities: A Systematic Review", Smart Learning Environments 11, n.o 1 (2024), 1-37.

Tampoco se puede obviar la dimensión social del aprendizaje. Las competencias de trabajo en equipo están ampliamente demandadas en el sector productivo. Por su parte, los sistemas de aprendizaje personalizado tienden a centrarse en la interacción individuo-máquina, lo que puede suponer una limitación de los escenarios de interacción entre estudiantes y profesorado.[76]

Otro aspecto clave es el de la autonomía y el rol profesional del docente. El mito tiende a invisibilizar al profesor, pero las investigaciones y políticas reales hacen hincapié en su papel insustituible. La experiencia ha mostrado que, sin la mediación docente, las tecnologías educativas raramente transforman prácticas de aula de forma sostenible, como se ha comentado en la sección 3.1.

3.2.1 REFORMULACIÓN: DEL MITO A LA PROYECCIÓN INFORMADA

Deconstruido el mito, surge la pregunta: ¿cómo aprovechar la IA para efectivamente promover autonomía y equidad? A continuación, se delinean como ejemplo dos proyecciones a futuro, fundamentadas en las evidencias disponibles, que ilustran caminos posibles para alinear la IA educativa con fines sociales. Cada proyección sigue la plantilla propuesta en la sección 2.2.

Proyección 3.2-A (IA al servicio de la equidad educativa)

Si un programa territorial de IA educativa prioriza escuelas de alta vulnerabilidad para refuerzo adaptativo (detección temprana de abandono escolar, acceso a recursos de calidad en lenguas y formatos accesibles), con conectividad garantizada, dispositivos disponibles, formación docente en alfabetización en IA y datos representativos (incluyendo contextos rurales y minorías lingüísticas), entonces se espera una reducción medible de brechas de rendimiento entre subgrupos y un incremento del acceso a contenidos avanzados en zonas tradicionalmente vulnerables, siempre que se mantenga la supervisión humana significativa, la transparencia algorítmica (explicabilidad de recomendaciones), las auditorías periódicas de sesgo, la protección de datos y el diseño accesible (discapacidad, lectura fácil), con prioridad presupuestaria para los centros objetivo. Se revisará a los 12 meses (verificación de condiciones y uso), a los 24–36 meses (evaluación de impacto en brechas, corrección de sesgos y costes de propiedad) y a los 5 años (revisión independiente de resultados y sostenibilidad), empleando métricas cuantitativas (diferenciales de logro, acceso y uso) y cualitativas (percepción docente/estudiantil de utilidad e inclusión).

Proyección 3.2-B (IA para fomentar la autonomía del estudiantado)

Si en los primeros cursos universitarios se integra la IA como un mentor interactivo en proyectos autodirigidos, acompañado de alfabetización en la herramienta (funcionamiento, límites y sesgos) y de una formación docente en pedagogías centradas en el estudiantado e interfaces explicables que ofrezcan opciones (no trayectorias únicas), entonces se espera un aumento de la autorregulación, de la autoeficacia percibida y de la diversificación de trayectorias de estudio, sin detrimento del rendimiento, siempre que se garanticen espacios regulares

76 Ana Mouta, Eva María Torrecilla-Sánchez y Ana María Pinto-Llorente, "Design of a future scenarios toolkit for an ethical implementation of artificial intelligence in education", Education and Information Technologies 29, (2024), 10473-10498.

de interacción humana y de reflexión (debates, defensas), límites de diseño para evitar la dependencia (sin mecanismos de enganche coercitivos), el consentimiento y la minimización de datos, y la capacidad de control por parte del estudiantado (confirmación de decisiones relevantes). Se revisará a los 6–12 meses (grupos focales, indicadores tempranos de autonomía y satisfacción), a los 24 meses (evaluación externa de autoeficacia, autorregulación y rendimiento por cohortes) y a los 5 años (comparación longitudinal con cohortes de control), ajustando el diseño pedagógico o las funcionalidades cuando se detecte la sobredependencia o la pasividad inducida.

Ambas proyecciones se alinean explícitamente con las reglas prometeicas presentadas en la sección 2.4. En cuanto a la regla de proporcionalidad y reversibilidad, 3.2-A exige revisión humana significativa, auditorías de sesgo y la capacidad de revertir recomendaciones o decisiones algorítmicas (derecho de apelación, reevaluación manual), evitando automatizar los juicios de alto impacto sin evidencia suficiente; por su parte, 3.2-B limita la IA a un papel de mentoría no decisorio, incorpora defensas y debates como válvula de control y prevé umbrales de retirada y revisión por hitos. Respecto de la regla de potencia creativa bajo control de alucinación, 3.2-A prioriza usos con alto valor social y bajo coste de error (detección temprana, ampliación de recursos) bajo verificación factual, transparencia y marcado del contenido, mientras que 3.2-B sitúa la IA en ideación y retroalimentación formativa, exige interfaces explicables, chequeos de exactitud, y limita su empleo en componentes sumativos o de verificación factual. En ambos casos, la declaración del uso de IA, los registros de interacción y los puntos de revisión operan como salvaguardas que materializan el principio de responsabilidad y mantienen la agencia humana en el centro.

3.2.2 IMPLICACIONES DE POLÍTICA Y LÍMITES OPERATIVOS

Ambas proyecciones presentadas exigen políticas de centro y marcos sistémicos coherentes con los principios internacionales que se han venido tomando como referencia desde la Unesco, la OCDE o el Departamento de Educación de los Estados Unidos. En primer lugar, la equidad no emerge de manera automática, requiere de una inversión deliberada en infraestructura (conectividad universal, dispositivos, soporte técnico) y de una financiación progresiva que coloque a los centros vulnerables en la primera fila de la adopción, no a la cola. Todo ello sin perder de vista que el problema no se puede resolver únicamente mediante el cómputo, esto es, que “la equidad nunca es abordable en su totalidad como un problema computacional”,[77] por lo que hay que insistir en que son los docentes y adultos responsables quienes mejor pueden velar por ella.

> En segundo lugar, la autonomía no se decreta con una aplicación; se cultiva con alfabetización en IA y con diseños que promuevan la agencia, las opciones, la evaluación auténtica y momentos de deliberación humana. En tercer lugar, la gobernanza debe incorporar una revisión humana significativa en decisiones de alto impacto

77 U.S. Department of Education, Artificial Intelligence and Future of Teaching…

(evaluación sumativa, acreditación, itinerarios), así como la transparencia (marcado del contenido generado, explicabilidad de recomendaciones), las evaluaciones de impacto algorítmico (sesgo, equidad, accesibilidad) y la capacidad de reversión (derecho de apelación, reevaluación manual). De hecho, las guías éticas internacionales, como la Recomendación sobre la Ética de la IA[78] y el Consenso de Beijing sobre IA y Educación,[79] insisten en la instauración de marcos de rendición de cuentas, de transparencia y de control humano para que la IA sirva a la inclusión en lugar de socavarla. La contratación pública y las compras institucionales deberían exigir estándares de accesibilidad, protección de datos, portabilidad y salida sin costes prohibitivos.[80]

Finalmente, conviene institucionalizar comités de ética y equidad, con representación docente y estudiantil, que auditen regularmente el desempeño de las herramientas, que publiquen informes y que recomienden ajustes.

En el plano operativo, la experiencia comparada sugiere comenzar con pilotos acotados y escalado progresivo, con criterios de éxito definidos ex ante y con un umbral de retirada si no se alcanzan mejoras sustantivas o aparecen efectos indeseados. La formación docente debe ser continua y situada (con prompts pedagógicos, verificación factual, evaluación de sesgos y el diseño de tareas auténticas) y acompañarse de comunidades de práctica[81] con las que compartir recursos, rúbricas y plantillas. En la evaluación, la ruta responsable pasa por métodos más controlados, cuando proceda (p. ej., defensas orales, tareas contextualizadas, coevaluación con evidencias de proceso), evitando un enfoque meramente punitivo o de vigilancia que erosione confianza y agencia.

3.2.3 CONCLUSIONES SOBRE LA AUTONOMÍA Y LA EQUIDAD GARANTIZADAS POR LA IA

El examen realizado confirma que la consigna de que la IA garantiza autonomía y equidad funciona como mito, toda vez que totaliza, presupone causalidades lineales, resiste la evidencia y borra condiciones determinantes. La IA puede efectivamente contribuir a ambos fines, pero solo bajo condiciones explícitas de infraestructura, formación, gobernanza y diseño pedagógico, con métricas y revisiones que permitan corregir el rumbo. Las proyecciones 3.2-A y 3.2-B, como posibles ejemplos, traducen esa ambición en hipótesis evaluables que preservan agencia estudiantil y profesional docente, alineadas con las cuatro lentes de la UNESCO.[82] Esta sección conduce, consecuentemente, a la 3.3, en la que se aborda el mito de la muerte de la originalidad por efecto de la IAGen, cerrando el tríptico de malentendidos presentados en este artículo.

78 Unesco, Recomendación sobre la ética de la inteligencia artificial (París, Francia: Unesco, 2022). **79** Unesco, "Beijing Consensus on Artificial Intelligence and Education". International Conference on Artificial Intelligence and Education, Planning Education in the AI Era: Lead the Leap, Beijing, China, 2019. **80** Rafael Molina-Carmona y Francisco José García-Peñalvo, "Safeguarding Knowledge: Ethical Artificial Intelligence Governance in the University Digital Transformation", en Advanced Technologies and the University of the Future, ed. Eduardo Vendrell Vidal, Uriel R. Cukierman y Michael E. Auer (Cham, Suiza: Springer Nature Switzerland, 2025), 201-220. **81** Chrissi Nerantzi et al., eds., 101 Creative Ideas to Use AI in Education. A Crowdsourced Collection (Calgary: Zenodo, 2023); Sandra Abegglen et al., eds., Towards AI Literacy: 101+ Creative and Critical Practices, Perspectives and Purposes (Calgary: Zenodo, 2024). **82** Unesco, Global Education Monitoring Report 2023.



Foto: Magnific

3.3 MUERTE DE LA ORIGINALIDAD.

La hipótesis según la cual la irrupción de la IAGen, especialmente tras la aparición de ChatGPT, supondría la muerte de la originalidad en la educación (y en la academia en general) ha cobrado mayor popularidad, alentada por los casos mediáticos y por la disponibilidad de modelos capaces de producir textos e imágenes bajo demanda, lo que pondría en riesgo que los estudiantes produzcan trabajos originales, generando alarma sobre el futuro de la integridad académica.[83] En su forma más fuerte, la consigna afirma que cualquier trabajo escrito por el estudiantado quedaría irremediablemente desvirtuado, pues sería técnicamente indistinguible de una salida algorítmica; en su forma más débil, sostiene que la IAGen homogeneiza estilos y desalienta el esfuerzo creativo. En general, lo que se está produciendo constituye un desafío ético a la integridad académica, centrado en la autoría, que menoscaba la relación de confianza mutua entre estudiantado y profesorado propia del proceso educativo.[84]

Este temor, legítimo como punto de partida, exige contraste. La experiencia histórica con otras tecnologías, como la calculadora[85] o Wikipedia,[86] muestra que los riesgos de atajo y copia coexisten con oportunidades para rediseñar tareas, evaluar procesos y fortalecer la autoría situada, especialmente cuando median reglas claras de uso, supervisión humana y evaluación auténtica. La cuestión del fraude académico no es nueva; lo novedoso es que las capacidades de la IAGen agravan el sentimiento de desconfianza, especialmente en el profesorado, que muestra un alto grado de ansiedad ante el potencial de la IA para generar un ensayo aceptable en pocos segundos.[87]

83 Laura Bratton, “Experts are Worried Students Use AI to write Papers. And Millions Are, New Data Shows”, Quartz (blog), abril 12 de 2024. **84** Ranieri, “Mito de la causalidad”. **85** Rebecca Johinke, Robert Cummings y Frances Di Lauro, “Reclaiming the Technology of Higher Education for Teaching Digital Writing in a Post-Pandemic World”, Journal of University Teaching and Learning Practice 20, n.o 2 (2023), 1-18. **86** Ángel Obregón Sierra y Natalia González-Fernández, “¿Utilizan correctamente Wikipedia los estudiantes universitarios españoles?”, Revista Lusófona De Educação 53, (2022), 159-174. **87** Britt Faulstick, “In the Age of ChatGPT, What’s It Like to Be Accused of Cheating?”, DrexelNEWS (blog), septiembre 12 de 2023.

Aplicando la lista de verificación presentada en la sección 2.1, la consigna presenta los cuatro indicios de un mito:

- *Totalización.* De usos puntuales de IAGen se extrapola el colapso general de la originalidad en toda producción académica, ignorando dominios, niveles y géneros. Es cierto que la IA sí da lugar a que este mito se pueda arraigar, ya que posibilita un fraude académico difícil de rastrear y mucho más complejo de probar. Aunque existen herramientas de detección de contenido generado por IA, la fiabilidad de estos detectores es limitada[88] y existen pruebas empíricas[89] de que estos detectores no son fiables en escenarios prácticos sin riesgo de equívocos o falsos positivos, casos estos que, además, minan gravemente la confianza y dan lugar a injusticias en el contexto educativo.[90] Por tanto, no todos los usos de la IA en la academia son deshonestos o equivalen a una aportación original nula, por lo que la clave no está en prohibir estas herramientas,[91] sino en diferenciar entre usos legítimos y usos engañosos.
- *Causalidad lineal.* Se presupone que "más IAGen → menos originalidad", sin atender a mediaciones didácticas (diseño de tareas, rúbricas procesuales, defensas orales) ni a políticas de declaración del uso y de tutoría ética. Un trabajo puede considerarse original más allá del mero contenido, considerando la aportación de pensamiento crítico y creativo. La originalidad puede valorarse en términos del aprendizaje evidenciado por el estudiante en lugar de en la literalidad del texto, teniendo más en cuenta el proceso y la reflexión que el producto final en sí.[92]
- *Inmunidad a la evidencia.* Se desatienden resultados mixtos, hay pruebas de que la IAGen puede producir salidas genéricas si se usa acríticamente, pero también de que, bajo guía y con objetivos explícitos, potencia la ideación, las revisiones y la exploración de alternativas.[93] Del mismo modo, la fiabilidad de los detectores automáticos es limitada y con riesgos de falsos positivos, un dato que obliga a no basar la integridad solo en su empleo.[94]
- *Borrado de condiciones.* El enunciado omite qué tareas, con qué estudiantes, en qué contextos, bajo qué salvaguardas (marcado de contenido, privacidad, revisión humana) y qué criterios de valor (originalidad sustantiva frente a mera novedad superficial) están en juego. Hay un consenso en las instituciones educativas en considerar al texto completamente generado por IA, sin reconocimiento, como una forma de plagio. Se consideran matices y nuevas normas, de forma que no se prohíbe por completo el uso de la herramienta, sino que se apuesta por una regulación ética que incluye una invitación/obligación de declarar su uso en los trabajos académicos. La IA es una herramienta legítima si se usa con honestidad, por lo que se podría citar como una fuente más.

88 Debora Weber-Wulff et al., "Testing of Detection Tools for AI-Generated Text", arXiv, (2023), 1-46.**89** Vinu Sankar Sadasivan et al., "Can AI-Generated Text be Reliably Detected?", arXiv, (2025), 1-37. **90** Tim Gorichanaz, "Accused: How students respond to allegations of using ChatGPT on assessments", Learning: Research and Practice 9, n.o 2 (2023), 183-196. **91** Enrique Dans, "El drama de los estudiantes que utilizan IA generativa para sus trabajos", Enrique Dans (blog), abril 14 de 2024. **92** García-Peñalvo, Llorens-Largo y Vidal, "The new reality of education…". **93** OECD, The potential impact of AI…; Giannini, "El uso de la IA en la educación". **94** Unesco, Global Education Monitoring Report 2023.

Rebatir el mito implica precisar qué se entiende por originalidad en educación. En términos formativos, no consiste solo en la novedad textual, sino en la aportación con sentido, lo que implica una combinación de comprensión, selección crítica de fuentes, conexión con contextos y voz propia. La IAGen, entrenada sobre grandes corpus, no piensa ni posee intencionalidad; ofrece patrones plausibles que pueden acelerar búsquedas o prototipados, pero que necesitan dirección humana para transformarse en conocimiento con valor. Por ello, la cuestión no es si la IAGen mata la originalidad, sino cómo diseñar usos que la preserven y la expandan, esto desplaza el foco del producto cerrado al proceso (borradores, diarios de aprendizaje, defensas) y de la unicidad estilística a la autenticidad intelectual demostrable.[95]


Foto: Magnific

3.3.1 REFORMULACIÓN: DEL MITO A LA PROYECCIÓN INFORMADA

Reformular el mito implica reconocer que la originalidad sigue viva, pero que debe redefinirse y proyectarse informadamente en la era actual. A continuación, se presentan dos proyecciones informadas, a modo de ejemplo, que replantean el mito hacia una visión más realista en el contexto educativo.

Proyección 3.3-A (Originalidad procesual y evaluación auténtica en escritura académica)

Si en -los primeros dos semestres de un grado universitario se rediseñan las tareas de ensayo para exigir una trayectoria de borradores con versiones (v1, v2, v3), un diario de decisiones (qué propuso la IAGen, qué se aceptó, qué se descartó y por qué) y una defensa oral breve del argumento, con una política de reconocimiento del uso de IA en el apartado metodológico y una rúbrica que pondere la comprensión, la integración de fuentes, la transferencia y la defensa, entonces se espera un aumento de las iteraciones de mejora (≥2 por entrega), una mejora en la calidad argumentativa y una reducción del texto genérico, siempre que toda salida de IAGen se marque, se verifique lo factual y se someta a una revisión humana significativa, prohibiendo su uso en componentes sumativos de atribución de originalidad (por ejemplo, em secciones de resultados propios). Se revisará en las semanas 4 y 10 mediante una rúbrica ciega, un análisis de dependencia de patrones y un contraste oral (muestreo aleatorio).

95 Ibíd; García-Peñalvo, “Inteligencia artificial en el aula”.

Proyección 3.3-B (Taller creativo con IA como coautor transparente)

Si en los últimos cursos de un grado universitario se implementa un taller de proyectos en el que la IAGen se usa como asistente virtual para la generación de ideas (brainstorming, variaciones, estilos) y como ayuda en bocetos, con restricción de uso en fases de síntesis final y con una guía de reconocimiento de uso (qué prompts, qué iteraciones, qué fragmentos se incorporan y cómo se reelaboran), entonces, se espera una mayor diversidad de soluciones (≥3 alternativas por equipo), un incremento de las referencias cruzadas con contextos locales y una mejor autoeficacia creativa, siempre que se incluyan críticas entre pares presenciales, un componente situado (datos u observaciones de campo) y un control de alucinación (comprobación de fuentes y registros de validación). Se revisará al cierre de cada incremento, con retrospectivas documentadas y con un panel externo de evaluación.

Ambas proyecciones preservan las reglas prometeicas presentadas en la sección 2.4, de proporcionalidad y reversibilidad (no automatizar juicios sumativos de autoría; derecho de apelación y reevaluación) y de control de alucinación (verificación y límites de uso). Además, alinean su evaluación con las cuatro lentes del informe GEM 2023[96] (pertinencia: foco en pensamiento original; equidad: andamiajes para quienes parten con desventaja; escalabilidad: plantillas, rúbricas compartidas, microformación; sostenibilidad: coste total de propiedad y prácticas de privacidad).

3.3.2 IMPLICACIONES DE POLÍTICA Y LÍMITES OPERATIVOS

La política institucional debe explicitar que a) la IAGen no sustituye la autoría; b) el uso legítimo requiere declaración y deliberación (qué se usó, para qué y cómo se reelaboró); c) los componentes sumativos que evidencian contribución propia (originalidad sustantiva) deben evaluarse con mecanismos humanos (defensas orales, revisión cruzada); y d) la integridad académica no puede descansar en detectores automáticos, cuya fiabilidad y sesgos imponen límites, sino que su papel, cuando se usen, ha de ser orientativo, más nunca decisorio. Bajo este marco, resulta coherente adoptar modelos de reconocimiento de uso (por ejemplo, plantillas tipo “declaración de asistencia de IA” al estilo de experiencias universitarias como la de la Universidad de Monash,[97]) integrados en las guías de citación y en los formatos de entrega.

Operativamente, conviene iniciar pilotos acotados con criterios ex ante de éxito y retirada; ofrecer formación docente situada (diseño de consignas IA-resistentes en el sentido de exigir juicio y transferencia, verificación factual, ética de prompts); crear bancos de rúbricas para garantizar la originalidad procesual; y sostener comunidades de práctica para promover el intercambio de ejemplos. En evaluación, se requiere combinar evidencias de proceso (borradores, diarios, logs de prompts) con productos e introducir verificaciones orales selectivas para romper la dependencia de la vigilancia automatizada.

96 Unesco, Global Education Monitoring Report 2023. **97** Monash University, “Acknowledging the use of generative artificial intelligence”.

3.3.3 CONCLUSIONES SOBRE LA ORIGINALIDAD EN LA ERA DE LA IAGEN

El análisis muestra que la muerte de la originalidad funciona como un mito: totaliza, asume causalidades lineales, ignora evidencia y borra condiciones. La originalidad no desaparece, se redefine: está menos centrada en novedad superficial y más en autenticidad intelectual demostrable, con procesos y contextos que la avalan. La vía responsable pasa por proyecciones informadas, como las aquí presentadas a modo de ejemplo, con tareas que documenten decisiones, límites y verificación de uso; con políticas que distingan entre uso legítimo y engañoso; y con una evaluación que devuelva la centralidad al juicio humano. Así, lejos de sucumbir, la autoría estudiantil puede florecer en simbiosis con la IA, siempre que se gobierne "en nuestros términos"[98] y con supervisión humana significativa.[99]

04 DISCUSIÓN Y CONCLUSIONES

Este artículo partió de un gesto metodológico: desplazar el foco del mito a la proyección informada. Frente a tres narrativas recurrentes, la sustitución del profesorado, la garantía automática de autonomía y equidad, y la muerte de la originalidad, se argumentó que las mismas operan como mitos en la medida en que totalizan, asumen causalidades lineales, se muestran inmunes a la evidencia y borran condiciones. La alternativa ha sido formular proyecciones auditables, situadas en contextos reales, con métricas, salvaguardas y puntos de revisión, en sintonía con la sociología de las expectativas[100] y los imaginarios sociotécnicos.[101] La discusión final sintetiza cuatro aprendizajes transversales y deriva compromisos prácticos y de política institucional.

98 Unesco, Global Education Monitoring Report 2023. **99** U.S. Department of Education, Artificial Intelligence and Future of Teaching… **100** Brown y Michael, "A Sociology of Expectations". **101** Jasanoff, "Future Imperfect".



Foto: Magnific

Primero, la centralidad de la agencia humana y el principio de responsabilidad. Una lectura prometeica de la IAGen, tamizada por el principio de responsabilidad[102] y las advertencias sobre la delegación del juicio,[103] sugiere gobernar la IA en nuestros términos pedagógicos. Las reglas prometeicas operativizadas en la sección 2.4, de proporcionalidad y reversibilidad, y de potencia creativa bajo control de alucinación, recorren las proyecciones presentadas en torno a la IA como un apoyo en procesos formativos (ideación, curación, retroalimentación preliminar) y no como el sustituto de decisiones sumativas de alto impacto. Esto no implica minimizar la innovación, sino modular su despliegue con revisión humana significativa, rastro de decisiones, explicación de recomendaciones y capacidad de recurso.

Segundo, el valor de las cuatro lentes (pertinencia, equidad, escalabilidad y sostenibilidad)[104] como brújula decisional. La pertinencia obliga a explicitar para qué se usa la IA y qué problema de aprendizaje resuelve; la equidad fuerza a definir para quién y con qué compensaciones se evita ensanchar brechas; la escalabilidad visibiliza los recursos, la formación, el soporte y el coste total de propiedad; la sostenibilidad introduce horizontes temporales, la dependencia de proveedores y los impactos materiales y energéticos. La tabla 1, que presenta la correspondencia con los siete principios del Manifiesto para una IA Segura en Educación,[105] refuerza esta idea: supervisión humana, privacidad, alineación tanto con estrategias como con prácticas, precisión y explicabilidad, transparencia e interfaz responsable, y formación ética.

Estos no son apéndices tardíos, sino condiciones de posibilidad para que la IA aporte valor educativo duradero.

Tercero, del producto al proceso, hay que redefinir la originalidad y la evaluación. Lejos de morir, la originalidad se reconfigura en clave procesual y de autenticidad intelectual: diarios de decisiones, versionado de borradores, defensa oral, marcaje del contenido generado y verificación factual desplazan la evaluación hacia el razonamiento, la transferencia y la voz propia. Este giro no prohíbe la IAGen, pero distingue entre su uso legítimo (que mejora el aprendizaje y se declara) y su uso engañoso (que sustituye el esfuerzo y se oculta). La discusión comparada sobre detectores automatizados coincide en su rol meramente orientativo: la integridad no puede descansar en herramientas con falsos positivos/negativos, sino en diseños didácticos que exijan comprensión demostrable, así como en un criterio profesional docente.

Cuarto, las proyecciones viables requieren gobernanza y ciclos de evidencia. Las proyecciones 3.1-3.3 han ejemplificado hipótesis operativas con hitos a 6--12--24--36 meses y revisiones a 5 años. Este enfoque por ciclos de aprendizaje organizativo evita el oportunismo del pico de expectativas y el cinismo del valle de la desilusión,[106] avanzando hacia una meseta de productividad educativa, con adopciones estables, retornos evaluables y criterios de desinversión cuando no haya valor. El estándar práctico que emerge es claro: pilotos acotados, criterios ex ante de éxito y retirada, auditorías de sesgo, protocolos de privacidad y de marcado, y comunidades de práctica para diseminar plantillas, rúbricas y guías de verificación.

102 Jonas, El principio de responsabilidad. **103** Anders, “On Promethean Shame”. **104** Unesco, Global Education Monitoring Report 2023. **105** Alier et al., Safe AI in Education Manifesto…; García-Peñalvo et al., “Safe, Transparent, and Ethical Artificial Intelligence”.**106** Fenn and Raskino, Mastering the Hype Cycle.

De aquí se derivan implicaciones de política. A nivel institucional, conviene a) aprobar políticas de reconocimiento de uso de IA en trabajos académicos (reconocimiento de uso obligatorio, formato normalizado, ejemplos); b) delimitar usos legítimos por tipo de tarea y nivel (por ejemplo, apoyo en creación y borradores, sí; generación de resultados o juicios sumativos, no); c) incorporar evaluaciones de impacto algorítmico (equidad, accesibilidad, exactitud) y derechos de apelación con reevaluación humana; d) alinear compras y contratos con transparencia, portabilidad y salida sin penalización; y e) organizar formación docente situada (prompts pedagógicos, verificación factual, diseño de tareas auténticas, ética y datos).

En el nivel del aula, la aplicación directa es doble: se requiere rediseñar consignas para evidenciar procesos y mantener espacios de deliberación humana (defensas, revisión por pares, conexión con contextos locales). A nivel sistémico, las administraciones han de priorizar la infraestructura y el acompañamiento en centros vulnerables para que la IA no agrave las desigualdades. Aquí, la equidad no es un problema computacional, sino que requiere de financiación progresiva, soporte técnico y alfabetización.

Esta discusión invita también a calibrar las expectativas. La IAGen aporta productividad (preparación de materiales, diferenciación, retroalimentación preliminar) y diversidad de caminos para el estudiantado (mentoría dialogada, simulación, traducción accesible), pero no resuelve por sí sola los desafíos estructurales de los sistemas educativos. La evidencia todavía es heterogénea por dominios y niveles; los efectos dependen del diseño, de las habilidades de los docentes, de la cultura institucional y de las condiciones materiales.

Por ello, se propone medir lo que importa (aprendizajes, inclusión, carga docente, agencia estudiantil) y no solo lo fácilmente medible (tiempos o métricas superficiales). La IAGen, en tanto tecnología de propósito general, exige una gobernanza prudente, es decir que evita la prohibición indiscriminada y el solucionismo tecnológico,[107] promoviendo la prudencia iterativa y los criterios de retirada cuando no se verifique un valor.

En conjunto, los hallazgos reafirman la tesis que le da el título a este trabajo: “IA en el aula: desmontando mitos y trazando proyecciones informadas”. Desmontar mitos no es un ejercicio retórico, sino una práctica de responsabilidad epistémica que protege tiempo, atención y recursos. Las proyecciones informadas son un contrato didáctico y político que especifican para qué, para quién, en qué condiciones y con qué evidencias conviene usar la IA. Al hacer explícitas las salvaguardas, se preserva la confianza (activo indispensable) entre el estudiantado, el profesorado y las instituciones.


Foto: Magnific

107 Evgeny Mozorov, To Save Everything, click here: The Folly of Technological Solutionism (PublicAffairs, 2013).

Como cierre, se proponen dos líneas de continuidad: una agenda de investigación, que incluya evaluaciones experimentales y cuasiexperimentales con seguimiento longitudinal, estudios comparados por áreas disciplinares, análisis de coste total de propiedad y de huella ambiental, metodologías participativas que incorporen la voz estudiantil y la del profesorado, y una agenda de política, que incluya normas claras de uso responsable y de reconocimiento, criterios de compra pública, alineados con transparencia, portabilidad y accesibilidad, y programas de alfabetización en IA para docentes y estudiantes con un enfoque de equidad. Si el fuego prometeico de la IAGen habilita la creatividad y la eficiencia, su buen gobierno consiste en orientarlo hacia fines educativos valiosos, con una supervisión humana significativa, métricas honestas y mecanismos de reversión. Bajo estas condiciones, lejos de desplazar al profesorado, homogeneizar el aprendizaje o anular la autoría, la IA puede convertirse en una infraestructura de aprendizaje aumentada, esto es, en una herramienta al servicio de la comprensión, la inclusión y la creatividad de las personas.

05 REFERENCIAS

Abegglen, Sandra, Chrissi Nerantzi, Antonio Martínez-Arboleda, Marianna Karatsiori, Javiera Atenas y Chris Rowell, eds. Towards AI Literacy: 101+ Creative and Critical Practices, Perspectives and Purposes. Calgary: Zenodo, 2024.

Alam, Ashraf. "Should Robots Replace Teachers? Mobilisation of AI and Learning Analytics in Education". En Proceedings of the 2021 International Conference on Advances in Computing, Communication y Control (ICAC3) (Mumbai, India, 3-4 December 2021). EE. UU.: IEEE, 2021.

Alier, Marc, Francisco José García-Peñalvo, Maria José Casañ, Juan Antonio Pereira y Faraón Llorens-Largo. Safe AI in Education Manifesto. Version 0.4.0. octubre 8 de 2024. https://manifesto.-safeaieducation.org

Anders, Günther. "On Promethean Shame". En Prometheanism: Technology, Digital Culture y Human Obsolescence, editado por Christopher John Muller, 29-95. Lanham, MD: Rowman & Littlefield Publishers, 2016.

Arendt, Hannah. Human condition. 2.a ed. Chicago: The University of Chicago Press, 1998.

Babich, Babette. "Günther Anders's 'Promethean Shame': Technological Ressentiment and Surveillance". The Harvard Review of Philosophy 31, (2024): 75-97. https://doi.org/10.5840/harvardreview202522561

Bai, Long, Xiangfei Liu y Jiacan Su. "ChatGPT: The Cognitive Effects on Learning and Memory". Brain-X 1, n.o 3 (2023), 1-9. https://doi.org/10.1002/brx2.30

Baker, Ryan S. y Aaron Hawn. "Algorithmic Bias in Education". International Journal of Artificial Intelligence in Education 32, n.o 4 (2022): 1052-1092. https://doi.org/10.1007/s40593-021-00285-9

Ballesteros, Virginia. "De Günther Anders al transhumanismo: la obsolescencia del ser humano y la mejora moral". ISEGORÍA. Revista de Filosofía Moral y Política, n.o 63 (2020): 289-310. https://-doi.org/10.3989/isegoria.2020.063.01

Bandura, Albert. "Self-Efficacy Mechanism in Human Agency". American Psychologist 37, n.o 2 (1982): 122-147. https://doi.org/10.1037/0003-066X.37.2.122

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı y Rei Mariman. "Generative AI Can Harm Learning". The Wharton School Research Paper, (2024), 1-59. https://doi.org/10.2139/ssrn.4895486

Bergson, Henri. Las dos fuentes de la moral y de la religión. Buenos Aires, República Argentina: Sudamericana, 1962.

Bratton, Laura. "Experts are Worried Students Use AI to Write Papers. And Millions Are, New Data Shows". Quartz (blog). Abril 12 de 2024. https://d66z.short.gy/8Ye0zs

Brown, Nik y Mike Michael. "A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects". Technology Analysis & Strategic Management 15, n.o 1 (2003): 3-18. https://doi.org/10.1080/0953732032000046024

Burgos, Daniel. "Inteligencia artificial y humanismo: construyendo un futuro centrado en el ser humano". Nueva Revista, n.o 192 (2024): 170-185. https://d66z.short.gy/wXL3lL

Chakraborty, Tanmoy y Sarah Masud. "The Promethean Dilemma of AI at the intersection of hallucination and creativity". Communications of the ACM 67, n.o 10 (2024): 26-28. https://doi.org/10.1145/3652102

Dang, Jianning y Li Liu. "Dehumanization Risks Associated with Artificial Intelligence Use". American Psychologist, (2025). https://doi.org/10.1037/amp0001542

Dans, Enrique. "El drama de los estudiantes que utilizan IA generativa para sus trabajos". Enrique Dans (blog). abril 14 de 2024. https://d66z.short.gy/xVyfLT

Dúo-Terrón, Pablo. "Generative Artificial Intelligence: Educational Reflections from an Analysis of Scientific Production". Journal of Technology and Science Education 14, n.o 3 (2024): 756-769. https://doi.org/10.3926/jotse.2680

Edwards, David. "Why this Teachers' Union Isn't Afraid AI will Replace them". World Economic Forum, 2024. https://d66z.short.gy/G662V9

Faulstick, Britt. "In the Age of ChatGPT, What's It Like to Be Accused of Cheating?". DrexelNEWS (blog). septiembre 12 de 2023. https://d66z.short.gy/NmrAqz

Fenn, Jackie y Mark Raskino. Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time. Boston, Massachusetts: Harvard Business Review Press, 2008.

Galindo, Jorge, Carlos Victoria, Manuel Hidalgo, Álvaro Fernández, Teresa Raigada y Javier Martínez Santos. El impacto de la IA en el sector público español. Avances para el acceso a servicios básicos España: ESADE, 2025. https://d66z.short.gy/q6ZUBH

García-Peñalvo, Francisco José. "Inteligencia Artificial en el aula. De los mitos a la realidad". 5.º Congreso Caribeño de Investigación Educativa, Punta Cana, República Dominicana, 9-12 de abril de 2025.

—."Mito de la inteligencia. Más allá de una educación de silicio". En Mitos EdTech. Desmontando el solucionismo tecnológico en educación, editado por Cristóbal Suárez-Guerrero, Juliana E. Raffaghelli y Pablo Rivera-Vargas, 79-87. Barcelona, España: Editorial UOC, 2024.

García-Peñalvo, Francisco José y Andrea Vázquez-Ingelmo. "What Do We Mean by GenAI? A Systematic Mapping of the Evolution, Trends y Techniques Involved in Generative AI". International Journal of Interactive Multimedia and Artificial Intelligence 8, n.o 4 (2023): 7-16. https://doi.org/10.9781/ijimai.2023.07.006

García-Peñalvo, Francisco José, Marc Alier, Juan Antonio Pereira y Maria Jose Casañ. "Safe, Transparent y Ethical Artificial Intelligence: Keys to Quality Sustainable Education (SDG4)". IJERI - International Journal of Educational Research and Innovation, n.o 22 (2024): 1-21. https://doi.org/10.46661/ijeri.11036

García-Peñalvo, Francisco José, Faraón Llorens-Largo y Javier Vidal. "The New Reality of Education in the Face of Advances in Generative Artificial Intelligence". RIED: Revista Iberoamericana de Educación a Distancia 27, n.o 1 (2024): 9-39. https://doi.org/10.5944/ried.27.1.37716

Giannini, Stefania. La IA generativa y el futuro de la educación. Paris: Unesco, 2023. https://doi.org/10.54675/ACWQ6815

—. "El uso de la IA en la educación: decidir el futuro que queremos". Unesco (blog). mayo 17 de 2024. https://d66z.short.gy/HyExHy

Gorichanaz, Tim. "Accused: How Students Respond to Allegations of Using ChatGPT on Assessments". Learning: Research and Practice 9, n.o 2 (2023): 183-196. https://doi.org/10.1080/23735082.2023.2254787

Hu, Sumei. "The Effect of Artificial Intelligence-Assisted Personalized Learning on Student Learning Outcomes: A Meta-Analysis Based on 31 Empirical Research Papers". Science Insights Education Frontiers 24, n.o 1 (2024): 3873-3894. https://doi.org/10.15354/sief.24.re395

International Telecommunication Union. Measuring Digital Development. Facts and Figures 2024. Ginebra, Suiza: International Telecommunication Union, 2024.

Jasanoff, Sheila. "Future Imperfect: Science, Technology y the Imaginations of Modernity". En Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power, editado por Sheila Jasanoff y Sang-Hyun Kim, 1-33. Chicago, EE. UU.: The University of Chicago Press, 2015.

Jasanoff, Sheila y Sang-Hyun Kim, eds. Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power. Chicago, EE. UU.: The University of Chicago Press, 2015.

Johinke, Rebecca, Robert Cummings y Frances Di Lauro. "Reclaiming the Technology of Higher Education for Teaching Digital Writing in a Post-Pandemic World". Journal of University Teaching and Learning Practice 20, n.o 2 (2023), 1-18. https://doi.org/10.53761/1.20.02.01

Jonas, Hans. El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica. Barcelona, España: Herder Editorial, 1995.

Kranzberg, Melvin. "Technology and History: 'Kranzberg's Laws'". Technology and Culture 27, n.o 3 (1986): 544-560. https://doi.org/10.2307/3105385

Lee, Sarah. "Fairness in Educational Technology. Addressing Algorithmic Bias in Digital Literacy and Pedagogy". NumberAnalytics (blog), 2025. https://d66z.short.gy/OmomWX

MacKenzie, Donald y Judy Wajcman, eds. The Social Shaping of Technology. 2.a ed. Buckingham, Inglaterra: Open University Press, 1999.

McKenzie, Lindsay. "End of the Line for Much-Hyped Tech Company". Inside Higher Ed (blog). mayo 6 de 2019. https://d66z.short.gy/c603Yz

McLuhan, Marshall y Bruce R. Powers. The Global Village: Transformations in World Life and Media in the 21st Century. Nueva York, EE. UU.: Oxford University Press, 1989.

Molina-Carmona, Rafael y Francisco José García-Peñalvo. "Safeguarding Knowledge: Ethical Artificial Intelligence Governance in the University Digital Transformation". En Advanced Technologies and the University of the Future, editado por Eduardo Vendrell Vidal, Uriel R. Cukierman y Michael E. Auer, 201-220. Cham, Suiza: Springer Nature Switzerland, 2025.

Monash University. "Acknowledging the Use of Generative Artificial Intelligence", 2025. https://d66z.short.gy/dkG1qv

Mosco, Vincent. The Digital Sublime: Myth, Power y Cyberspace. The MIT Press, 2004.

Mouta, Ana, Eva María Torrecilla-Sánchez y Ana María Pinto-Llorente. "Design of a Future Scenarios Toolkit for an Ethical Implementation of Artificial Intelligence in Education". Education and Information Technologies 29, (2024): 10473-10498. https://doi.org/10.1007/s10639-023-12229-y

Mozorov, Evgeny. To Save Everything, Click Here: The Folly of Technological Solutionism. PublicAffairs, 2013.

Nam, Benjamin H. y Qiong Bai. "ChatGPT and its Ethical Implications for STEM Research and Higher Education: A Media Discourse Analysis". International Journal of STEM Education 10, n.o 1 (2023), 1-24. https://doi.org/10.1186/s40594-023-00452-5

Nerantzi, Chrissi, Sandra Abegglen, Marianna Karatsiori y Antonio M. Arboleda, eds. 101 Creative Ideas to Use AI in Education: A Crowdsourced Collection. Calgary: Zenodo, 2023.

Obregón Sierra, Ángel y Natalia González-Fernández. "¿Utilizan correctamente Wikipedia los estudiantes universitarios españoles?". Revista Lusófona De Educação 53, (2022): 159-174. https://doi.org/10.24140/issn.1645-7250.rle53.09

OECD. OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem. París: OECD Publishing, 2023.

—. The Potential Impact of AI on Equity and Inclusion in Education. París: OECD Publishing, 2024. https://d66z.short.gy/QzXXo9

OECD-Education International. Opportunities, Guidelines and Guardrails on Effective and Equitable Use of AI in Education. París: OECD Publishing, 2023. https://d66z.short.gy/zmZibm

O'Sullivan, Patricia, Christina Forgette, Stephen Monroe y M. Tyler England. "Student Perceptions of the Effectiveness of Adaptive Courseware for Learning". Current Issues in Emerging eLearning 7, n.o 1 (2020), 71-100.

Ranieri, Marta. "Mito de la causalidad. Aporías entre los requisitos técnicos y las aspiraciones pedagógicas de la inteligencia artificial en la educación". En Mitos EdTech. Desmontando el solucionismo tecnológico en educación, editado por Cristóbal Suárez-Guerrero, Juliana E. Raffaghelli y Pablo Rivera-Vargas, 23-33. Barcelona, España: Editorial UOC, 2024.

Robinson, Ken. "How to Escape Education's Death Valley". TED Talks Education, 2013. https://d66z.short.gy/aM5Tpf

Sadasivan, Vinu Sankar, Aounon Kumar, Sriram Balasubramanian, Wenxiao Wang y Soheil Feizi. "Can AI-Generated Text be Reliably Detected?". arXiv, (2025), 1-37. https://doi.org/10.48550/arXiv.2303.11156

Shaw, Catherine, Dorn, H., Martin, S., Hay, L., Janson, N. y Bryant, G. Time for Class 2025. Empowering Educators, Engaging Students (s. l.: Tyton Partners, 2025). https://d66z.short.gy/iOOUw4

Smith, Marcus y Seumas Miller. "Technology, Institutions and Regulation: Towards a Normative Theory". AI & Society 40, n.o 2 (2025): 1007-1017. https://doi.org/10.1007/s00146-023-01803-0

Smith, Merritt Roe y Leo Marx, eds. Does Technology Drive History?: The Dilemma of Technological Determinism. Cambridge, Massachusetts, EE. UU.: The MIT Press, 1994.

Suárez-Guerrero, Cristóbal, Juliana E. Raffaghelli y Pablo Rivera-Vargas. Mitos EdTech. Desmontando el solucionismo tecnológico en educación. Barcelona, España: Editorial UOC, 2024.

Suárez-Guerrero, Cristóbal, Pablo Rivera-Vargas y Juliana Raffaghelli. "EdTech Myths: Towards a Critical Digital Educational Agenda". Technology, Pedagogy and Education 32, n.o 5 (2023): 605-620. https://doi.org/10.1080/1475939X.2023.2240332

Swierstra, Tsjalling. "Identifying the Normative Challenges Posed by Technology's 'Soft' Impacts". Etikk I Praksis - Nordic Journal of Applied Ethics 9, n.o 1 (2015): 5-20. https://doi.org/10.5324/eip.v9i1.1838

U.S. Department of Education. Office of Educational Technology. Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations. Washington,: EE. UU.: Department of Education, 2023. https://d66z.short.gy/e6rfFB

Unesco. "Beijing Consensus on Artificial Intelligence and Education". International Conference on Artificial Intelligence and Education, Planning Education in the AI Era: Lead the Leap, Beijing, China, 2019. https://bit.ly/3n7wBIK

—. Global Education Monitoring Report 2023: Technology in education - A tool on whose terms? París, France: Unesco, 2023.

—. Global Education Monitoring Report 2024/5: Leadership in education - Lead for learning. París, France: Unesco, 2024.

—. Recomendación sobre la ética de la inteligencia artificial. París, Francia: Unesco, 2022. https://bit.ly/3nc3Yu1

Wang, Ling y Wenye Li. "The Impact of AI Usage on University Students' Willingness for Autonomous Learning". Behavioral Sciences 14, n.o 10 (2024), 1-16. https://doi.org/10.3390/bs14100956Wang, Xiaoman, Rui "Tammy" Huang, Max Sommer, Bo Pei, Poorya Shidfar, Muhammad Shahroze Rehman, Albert D. Ritzhaupt y Florence Martin. "The Efficacy of Artificial Intelligence-Enabled Adaptive Learning Systems From 2010 to 2022 on Learner Outcomes: A Meta-Analysis". Journal of Educational Computing Research 62, n.o 6 (2024): 1348-1383. https://doi.org/10.1177/07356331241240459

Weber-Wulff, Debora, Alla Anohina-Naumeca, Sonja Bjelobaba, Tomáš Foltýnek, Jean Guerrero-Dib, Olumide Popoola, Petr Šigut y Lorna Waddington. "Testing of Detection Tools for AI-Generated Text". arXiv, (2023), 1-46. https://doi.org/10.48550/arXiv.2306.15666

Westervelt, Eric. "Meet the Mind-Reading Robo Tutor in the Sky". NPR (blog). Octubre 13 de 2015. https://d66z.short.gy/VBtn7z

Winner, Langdon. "Do Artifacts Have Politics?". En Computer Ethics, editado por John Weckert, 121-130. Londres, Inglaterra: Routledge, 2007.

Young, Jeffrey R. "As More AI Tools Emerge in Education, so Does Concern Among Teachers About Being Replaced". EdSurge (blog). Julio 2 de 2024. https://d66z.short.gy/MMbTgU

Zhai, Chunpeng, Santoso Wibowo y Lily D. Li. "The Effects of Over-Reliance on AI Dialogue Systems on Students' Cognitive Abilities: A Systematic Review". Smart Learning Environments 11, n.o 1 (2024), 1-37. https://doi.org/10.1186/s40561-024-00316-7

ACERCA DEL AUTOR

Francisco José
García Peñalvo

Catedrático de Universidad del Departamento de Informática y Automática en la Universidad de Salamanca (USAL), con 4 sexenios de investigación, 1 sexenio de transferencia y 5 quinquenios docentes reconocidos. Recibió el Premio Gloria Begué a la excelencia docente en 2019 y el Premio María de Maeztu a la excelencia investigadora en 2023. Profesor Distinguido de la Escuela de Humanidades y Educación del Tecnológico de Monterrey, México, entre 2016 y 2018, Investigador de Impacto Internacional de la Universidad Nacional San Agustín, Arequipa, Perú, y Profesor Visitante del Instituto Politécnico Nacional de México.

Desde 2006 es el director del Grupo de Investigación Reconocido por la USAL GRIAL (GRupo de investigación en InterAcción y eLearning), grupo que es Unidad de Investigación Consolidada de la Junta de Castilla y León (UIC 81).

). Incluido en la World's Top 2% Scientists list by Stanford University (2019, 2020, 2021, 2022, 2023, 2024, 2025) https://d66z.short.gy/L9UBms. Ha dirigido 34 tesis doctorales. Ha sido Vicedecano de Innovación y Nuevas Tecnologías de la Facultad de Ciencias de la USAL entre 2004 y 2007 y Vicerrector de Innovación Tecnológica de esta Universidad entre 2007 y 2009. Actualmente es el Subdirector del Instituto Universitario de Ciencias de la Educación (IUCE) y el Coordinador del Programa de Doctorado en Formación en la Sociedad del Conocimiento de la USAL. Es Editor Jefe de la revista Education in the Knowledge Society. Editor Asociado en Computers in Human Behavior, Computers in Human Behavior Reports, IEEE Access, entre otras, habiendo sido editor asociado de las revistas IEEE RITA e IEEE Transactions on Learning Technologies. Ha publicado más de 150 artículos en revistas indexadas en el JCR SCIE/SSCI/+2023 (76 Q1).



Foto: Magnific

01

La IA, el derecho de autor y el futuro de la creatividad

E-ESTADO
Primera Edición

La IA, el derecho de autor y el futuro de la creatividad: notas desde la Feria Internacional del Libro de Panamá

AI, Copyright, and the Future of Creativity: Notes from the Panama International Book Fair

Fecha de recepción: 27 de noviembre de 2025
Fecha de aceptación: 20 de febrero de 2026

Andrés Izquierdo

Socio, Izquierdo Advisory, Washington D.C., Counsel, American University y Geneva Graduate Institute, Center of Knowledge Governance

info@andresizquierdo.com

Cómo citar: Izquierdo, Andrés. “La IA, el derecho de autor y el futuro de la creatividad: notas desde la Feria Internacional del Libro de Panamá”. Revista e-Estado 1, n.° 1 (2026).

En agosto participé como ponente en la Feria Internacional del Libro de Panamá, en un evento coorganizado por la Organización Mundial de la Propiedad Intelectual (OMPI), la Dirección Nacional de Derecho de Autor, el Ministerio de Cultura de Panamá y la Asociación de Editores de Panamá. Mi presentación abordó la cada vez más compleja intersección entre el derecho de autor y la inteligencia artificial (IA), un tema central en los debates jurídicos y de política pública a nivel mundial actualmente. Esta entrada resume los principales argumentos que expuse sobre litigios en curso, perspectivas académicas, además de una síntesis del informe publicado por la Oficina de Derechos de Autor de EE. UU. en mayo de 2025 sobre IA generativa.

¿Cómo debe responder el derecho de autor al uso masivo de obras protegidas en el entrenamiento de modelos de IA generativa? El análisis sugiere que se están dando discusiones clave acerca de varios temas: los límites del uso justo y las excepciones, la necesidad de derechos de remuneración exigibles y el papel de las licencias y la supervisión regulatoria. Este artículo comienza con un panorama del contexto legal y tecnológico del entrenamiento de IA. Luego revisa propuestas académicas para recalibrar el marco del derecho de autor, examina decisiones judiciales recientes que ponen a prueba los límites doctrinales actuales y resume el preinforme de la Oficina de Derecho de Autor de Estados Unidos de 2025 como respuesta institucional. Finalmente, termina con las consideraciones de política pública para una futura regulación.

UN PAISAJE LEGAL Y TECNOLÓGICO EN TRANSFORMACIÓN

La integración de la IA generativa en los ecosistemas creativos y de información ha expuesto tensiones fundamentales en el derecho de autor. Los sistemas actuales se alimentan diariamente de grandes volúmenes de obras protegidas como, libros, música, imágenes o notas periodísticas, para entrenar modelos de IA. Esta práctica plantea interrogantes jurídicos no resueltos: ¿puede el derecho de autor regular de forma efectiva el uso de datos para entrenamiento de IA?, ¿se aplican las doctrinas y disposiciones existentes, como el uso justo o las limitaciones y excepciones, a estas prácticas?, ¿qué alternativas existen, si hay alguna, para los titulares de derechos cuyas obras son usadas sin su consentimiento?

Estas preguntas siguen abiertas en distintas jurisdicciones. Aunque algunos tribunales y agencias regulatorias han empezado a responder, una parte sustancial del debate se está moldeando desde la doctrina académica y los litigios, que proponen marcos para reconciliar el desarrollo de la IA con los compromisos normativos del derecho de autor.



Foto: Magnific

PERSPECTIVAS ACADÉMICAS

La revisión de la literatura revela varios temas centrales:

1. Varios autores coinciden en que deben reforzarse los derechos de remuneración para los creadores. Geiger, Scalzini y Bossi proponen que para garantizar una compensación justa en la era digital, el derecho de autor en la Unión Europea (UE) debe ir más allá de protecciones contractuales débiles y establecer derechos irrenunciables que aseguren ingresos equitativos a autores e intérpretes como parte de los derechos fundamentales.
2. Martin Senftleben propone un sistema de remuneración alternativo, basado en los outputs de la IA, es decir que los proveedores paguen compensaciones cuando las salidas generadas sustituyan, potencialmente, obras humanas, evitando la complejidad del enfoque de insumo que propone la Ley de IA de la UE. João Pedro Quintais respalda reformas estructurales que incluyan mecanismos colectivos y derechos legales para asegurar que los autores se beneficien de la explotación de sus obras en la IA.
3. Otros autores como Cooper subrayan que la opacidad técnica de la IA hace inviable la trazabilidad y, por tanto, los modelos de compensación por atribución. En su lugar, propone sistemas alternativos que aseguren pagos justos en un entorno de autoría algorítmica.
4. Pasquale y Sun proponen un sistema dual de consentimiento y compensación, es decir que los creadores puedan excluir sus obras del entrenamiento, y que se imponga un canon a los desarrolladores de IA. Gervais sugiere crear un nuevo derecho de remuneración por el uso comercial de sistemas entrenados con obras protegidas, sin reemplazar los derechos ya existentes.
5. En la OMPI, el Comité Permanente de Derecho de Autor y Derechos Conexos (SCCR) ha aprobado un programa de trabajo sobre L&Es, incluyendo la propuesta del Grupo Africano (SCCR/47/6) para un instrumento internacional. En el CDIP se aprobó un proyecto piloto sobre TDM para apoyar la investigación en universidades africanas (CDIP/30/9 REV).
6. Eleonora Rosati sostiene que el entrenamiento de IA sin licencia está fuera de las excepciones actuales del derecho europeo y británico. Concluye que un sistema de licenciamiento es legalmente necesario, incluso para usos no comerciales.

LITIGIOS EN EE. UU.: PRUEBAS DOCTRINALES EN TIEMPO REAL

En EE. UU. varios casos están poniendo a prueba los límites del derecho de autor frente a la IA. Entre los más relevantes se encuentran:

- Perplexity AI vs. News Corp (2025): el tribunal permitió que el caso avance en Nueva York, rechazando una moción de traslado. Aún no se decide sobre si la denuncia de usar contenido de News Corp para resumen automatizado infringe derechos.

- Bartz vs. Anthropic PBC (2025): el tribunal determinó que entrenar modelos con libros comprados legalmente fue de uso justo, pero no así cuando se usaron libros pirateados. Anthropic acordó pagar USD 1.5 mil millones para resolver el caso.
- New York Times vs. OpenAI & Microsoft (2023): la demanda alega que los modelos fueron entrenados con artículos del periódico sin autorización. La corte permitió que se avance con acusaciones de infracción directa y contributiva.
- UMG vs. Suno (2024): Universal Music demandó a Suno por entrenar modelos con grabaciones sin licencia. El caso podría sentar precedentes en materia musical.
- Thomson Reuters vs. Ross Intelligence (2024): el uso de resúmenes jurídicos de Westlaw fue considerado sustituto directo, por lo que no se permitió el uso justo.
- Andersen vs. Stability AI y Getty vs. Stability AI: se discute si las imágenes generadas violan derechos derivados o derechos morales.
- Authors Guild vs. OpenAI: en etapa inicial, pero podría impactar la compensación de autores literarios.
- Thaler vs. Perlmutter (2023): se reafirmó que obras sin autor humano no pueden ser protegidas por derecho de autor.

Estos casos delinean los contornos de la doctrina en construcción. Revelan las tensiones entre el marco legal vigente y las capacidades tecnológicas emergentes.

RESPUESTAS INSTITUCIONALES: INFORME 2025 DE LA OFICINA DE COPYRIGHT DE EE.UU.

El informe más completo hasta la fecha proviene de la Oficina de Copyright y fue realizado en mayo de 2025. Sus conclusiones clave incluyen:

- Las obras generadas por IA sin autor humano no pueden ser protegidas.
- El entrenamiento con material protegido podría requerir licencia, salvo excepciones claras.
- Se estudian herramientas políticas como sistemas de remuneración, registros de datasets y mandatos de transparencia

El informe distingue entre usos permisibles e infractores:

- Infractores: reproducción no autorizada, outputs tipo obra derivada y remoción de metadatos.
- Permisibles: uso justo (según aplicabilidad), TDM legal, uso de datos abiertos o sintéticos.

El informe, además, subraya que usos para investigación o docencia no sustitutiva suelen ser transformativos y más aptos para ser considerados de uso justo. Se distingue entre el entrenamiento con fines de análisis y el entrenamiento para producir obras expresivas que compiten en el mercado.

EL PAPEL DE LA ACADEMIA: ENTRE LA CRÍTICA Y LA PROPUESTA

La academia tiene un papel insustituible en este momento histórico no solo como espacio de crítica o de documentación, sino como actor activo en la construcción de nuevos equilibrios. Este papel puede desplegarse en al menos tres niveles:

1. **Producción normativa:** la investigación académica puede anticipar vacíos regulatorios, proponer nuevas arquitecturas legales y testear el impacto comparado de distintos modelos de gobernanza. En particular, puede contribuir al diseño de normas adaptadas a contextos de IA. También puede articular principios comunes entre regiones con capacidades regulatorias dispares, facilitando un diálogo global.
2. **Incidencia institucional:** en foros como los de la OMPI, la participación de académicos y centros de investigación ha sido clave para visibilizar las perspectivas de diversos actores. Este trabajo puede ampliarse, institucionalizarse y recibir apoyo estructural.
3. **Educación crítica:** formar una generación de juristas, tecnólogos y artistas capaces de entender los dilemas éticos y legales de la creación algorítmica es una tarea urgente. Esto implica renovar los programas de derecho, arte y comunicación para incluir debates sobre IA, datos y derechos culturales. La alfabetización en derechos digitales debe ser parte de cualquier estrategia de gobernanza justa.

ESCENARIOS POSIBLES: IMAGINANDO EL 2035

En los próximos diez años podríamos enfrentar distintos modelos regulatorios:

- **Hiperfragmentación:** cada país impone reglas propias sobre entrenamiento de IA, licencias y excepciones. Resultado: inseguridad jurídica, desigualdad de acceso y tensiones comerciales. Las industrias creativas quedarían atrapadas en un laberinto normativo, y los desarrolladores enfrentarían altos costos de cumplimiento.
- **Licenciamiento controlado:** solo los datasets autorizados pueden usarse para entrenar modelos, esto protege a los creadores, pero puede limitar la diversidad de contenidos y generar nuevas barreras de entrada.
- **Canon redistributivo global:** se establece un gravamen universal sobre los outputs de IA, cuya recaudación se distribuye a través de fondos colectivos. Requiere una gobernanza compleja, pero puede ofrecer justicia cultural. Este modelo podría ser administrado por organizaciones internacionales con criterio.



Foto: Magnific

LÍNEAS DE ACCIÓN: HACIA UN MARCO EQUILIBRADO

Para cerrar la brecha entre la realidad tecnológica y la capacidad legal, los responsables de política pública podrían considerar:

1. Derechos de remuneración para autores y gestión de metadatos.
2. Sistemas de licencias transparentes, con divulgación de datos de entrenamiento.
3. Herramientas para que los reguladores, los autores y el público entiendan el ciclo de los datos.

Estas medidas, aunque no exhaustivas, pueden constituir pilares para el futuro del sistema de derecho de autor. La comunidad legal enfrenta hoy un reto clave: adaptar el derecho de autor a la IA sin socavar los principios que sostienen la economía creativa y el acceso al conocimiento. La pregunta ya no es si el derecho de autor cambiará, es cómo y para quién.

CONCLUSIÓN: UN NUEVO PACTO PARA LA CREATIVIDAD EN LA ERA ALGORÍTMICA

Estamos frente a un momento fundacional. Los marcos legales, éticos y económicos que sostenían las industrias creativas se ven cuestionados por sistemas que aprenden sin pedir permiso, crean sin firmar y distribuyen sin límites geográficos. Y aunque esta transformación ofrece oportunidades innegables, nuevas formas de expresión, acceso ampliado al conocimiento y herramientas más potentes para la educación y la ciencia, también plantea desafíos urgentes que no pueden ser resueltos desde el inmovilismo jurídico ni desde la lógica del laissez-faire tecnológico.

Lo que está en juego no es solo el equilibrio entre innovación y protección, es la redistribución global del valor generado por la creatividad, la viabilidad económica del trabajo autoral, la pluralidad de las voces culturales y la legitimidad misma del derecho de autor. Si las reglas actuales ya no pueden responder a la escala, la opacidad ni la velocidad de la IA, debemos repensarlas desde la raíz; esa tarea exige coraje institucional, diálogo multilateral y compromiso ético.

El derecho de autor no puede convertirse en un obstáculo para el desarrollo de tecnologías con impacto social positivo, pero tampoco puede ser indiferente ante prácticas que despojan a los autores del control y del reconocimiento. En este sentido, la academia tiene un papel estratégico: producir evidencia, proponer alternativas, formar nuevas generaciones y construir puentes entre mundos técnicos, legales y culturales. No basta con describir los problemas, hay que imaginar soluciones viables e inclusivas.

REFERENCIAS

Geiger, C., Scalzini, S., & Bossi, L. (2025). Time to (Finally) Reinstall the Author in EU Copyright Law: From Contractual Protection to Remuneration Rights. Innovation Law and Ethics Observatory (ILEO) Research Paper Series No. 25-05. SSRN. https://ssrn.com/abstract=5330841

Cooper, Z. (2024). The AI authorship distraction: Why copyright should not be dichotomised based on generative AI use. SSRN. https://ssrn.com/abstract=4932612

Díaz Charquero, P. (2021). Flexibilidades al derecho de autor. Edited by J. Gemetto. Design by H. A. Vásquez.

Pasquale, F., & Sun, H. (2024). Consent and compensation: Resolving generative AI's copyright crisis (Cornell Legal Studies Research Paper No. 26-05; University of Hong Kong Faculty of Law Research Paper No. 2024/07). SSRN. https://ssrn.com/abstract=4830963

Gervais, D. J. (2024, April). The remuneration of music creators for the use of their works by generative AI (White paper prepared for Fair Trade Music International and the International Council of Music Creators).

Flynn, S., Schirru, L., Palmedo, M., & Izquierdo, A. (2022). Research exceptions in comparative copyright (PIJIP/TLS Research Paper Series No. 75). American University Washington College of Law, Program on Information Justice and Intellectual Property. https://digitalcommons.wcl.american.edu/research/75

Flynn, S. M., Butler, B., Carroll, M., Cohen-Sasson, O., Craig, C., Guibault, L., Jaszi, P., Jütte, B. J., Katz, A., … Contreras, J. L. (2022). Legal reform to enhance global text and data mining research. Science, 378(6623), 1323–1325. https://doi.org/10.1126/science.adf0529

Izquierdo, H. A. (2021). Minería de textos y datos e Inteligencia Artificial: nuevas excepciones al derecho de autor. THEMIS Revista De Derecho, (79), 323–343. https://doi.org/10.18800/themis.202101.018

Rosati, E. (2025). Copyright exceptions and fair use defences for AI training done for "research" and "learning," or the inescapable licensing horizon. European Journal of Risk Regulation, 16(3), 961–984. https://doi.org/10.1017/err.2025.10035Love, J. (2024). We need smart intellectual property laws for artificial intelligence. Scientific American.

CASOS EN LITIGIO

Andersen v. Stability AI Ltd., No. 3:23-cv-00201-WHO (N.D. Cal. 2023).

Authors Guild et al. v. OpenAI, Inc., No. 1:23-cv-08292 (S.D.N.Y. 2023).

Bartz v. Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. 2024).

Getty Images (US), Inc. v. Stability AI Ltd., No. 1:23-cv-00135 (D. Del. 2023)

New York Times Co. v. Microsoft Corp. and OpenAI, Inc., No. 1:23-cv-11195 (S.D.N.Y. 2023).

Thaler v. Perlmutter, 687 F. Supp. 3d 140 (D.D.C. 2023), aff'd, 84 F.4th 1 (D.C. Cir. 2024).

Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., 702 F. Supp. 3d 473 (D. Del. 2023).

Universal Music Group Recordings, Inc. v. Suno, Inc., No. 1:24-cv-04777 (S.D.N.Y. 2024).

PRE-REPORTE OFICINA DE DERECHO DE AUTOR, ESTADOS UNIDOS

U.S. Copyright Office. (2025). Copyright and artificial intelligence, part 3: Generative AI training (pre-publication version). Library of Congress. https://www.copyright.gov/ai/

ACERCA DE LOS AUTORES

Andrés **Izquierdo**

Abogado en propiedad intelectual, inteligencia artificial y derecho internacional, con más de veinte años de experiencia asesorando gobiernos, organizaciones multilaterales y empresas tecnológicas en Washington D. C., EE. UU., y Ginebra, Suiza. Ha participado activamente en negociaciones clave en la Organización Mundial de la Propiedad Intelectual (OMPI), incluyendo el Tratado sobre Recursos Genéticos y el Tratado de Derecho de Diseño. Se considera como una voz experta en la regulación de la inteligencia artificial. Es socio fundador de Andrés Izquierdo Lawyers y académico afiliado al Program on Information Justice and Intellectual Property (PIJIP) de la American University en Washington D. C. Combina análisis legal con una mirada crítica sobre los desafíos que plantea la inteligencia artificial a los marcos de propiedad intelectual y manejo de la información.



Foto: Magnific

02

La inteligencia artificial como ¿artista?: del sentimiento al algoritmo, de los derechos de autor a la esencia del ser humano

E-ESTADO
Primera Edición

La inteligencia artificial como ¿artista?: del sentimiento al algoritmo, de los derechos de autor a la esencia del ser humano

Artificial Intelligence as an “Artist”?: From Emotion to Algorithm, From Copyright to the Essence of Humanity

Fecha de recepción: 7 de abril de 2026
Fecha de aceptación: 22 de julio de 2026

Andrés Eduardo Salcedo Camacho
Universidad Libre y Universidad del Rosario, Colombia
andrese.salcedo@urosario.edu.co
ORCID: https://orcid.org/0000-0003-1059-7018

Cómo citar: Salcedo Camacho, Andrés Eduardo. “La inteligencia artificial como ¿artista?: del sentimiento al algoritmo, de los derechos de autor a la esencia del ser humano”. Revista e-Estado 1, n.° 1 (2026).

LA INTELIGENCIA ARTIFICIAL COMO ¿ARTISTA?: DEL SENTIMIENTO AL ALGORITMO, DE LOS DERECHOS DE AUTOR A LA ESENCIA DEL SER HUMANO[1]

RESUMEN

A partir de la irrupción de la IA en el arte, se han gestado distintos debates no solamente en torno a si sus creaciones deben ser protegidas por los derechos de autor, sino también a si pueden ser consideradas arte, en el sentido estético y tradicional. En consecuencia, este texto aborda tanto las concepciones jurídicas como las humanas y filosóficas, exponiendo inquietudes en torno a lo que significa una IA creadora de arte para la misma naturaleza y esencia del ser humano, enmarcado en un contexto contemporáneo, en el que la concepción del fin del arte y las nociones posmodernas parecen llevar a una redefinición de lo que va a ser considerado como arte.

Palabras clave: inteligencia artificial, arte, estética, derechos de autor



Foto: Magnific

ABSTRACT

Since the emergence of AI in the field of art, various debates have arisen not only regarding whether its creations must be protected by copyright, but also whether they may be considered art in the aesthetic and traditional sense. Accordingly, this article addresses both legal conceptions and human and philosophical perspectives, raising questions about what an AI capable of creating art means for the very nature and essence of the human being, within a contemporary context marked by discussions on the "end of art" and its postmodern version, both of which have been the subject of extensive analysis and criticism.

Keywords: Artificial Intelligence, Art, Aesthetics, Copyright

> *Art is a human activity, consisting in this, that one man consciously, by means of certain external signs, hands on to others feelings he has lived through, and that other people are infected by these feelings, and also experience them.*
>
> *León Tolstói, What is Art?*

1 Para la elaboración de este trabajo, se utilizaron las herramientas de Chat GPT, Claude y Gemini, con el fin de ilustrar si estas herramientas pueden crear arte; en el cuerpo del texto se encuentran especificados los textos e imágenes generados por medio de IA. Todo el contenido ha sido revisado y editado por el autor, quien asume plena responsabilidad por el manuscrito y manifiesta que el uso de la herramienta de IA mencionada fue solamente para labores de apoyo en temáticas no sustanciales del artículo y no afecta derechos de terceros, la autoría o el carácter inédito del artículo.

INTRODUCCIÓN

En la actualidad la inteligencia artificial ha avanzado a pasos agigantados, brindando aplicaciones que ya se encuentran al alcance de cualquier persona. Con ella se pueden crear videos, música, imágenes, pinturas, canciones, poemas y demás productos artísticos (¿en verdad artísticos?), pagando una suscripción mensual económica (o incluso gratis) y con un prompt adecuado.

Desde el punto de vista jurídico, esto plantea importantes reflexiones, sobre todo en materia de derechos de autor, pues no se tiene certeza de quién es el creador de la obra: ¿es la persona que ingresó los prompts?, ¿es el desarrollador de la herramienta de IA que se utiliza?, ¿es la propia IA?

A la par de lo anterior, hay reflexiones más agudas, dado que la IA empieza a crear a partir de sus conocimientos previos, lo cual traza líneas muy delgadas entre una creación original y la copia: ¿el arte debe ser original para ser considerado arte?, ¿qué tanto se ve afectada la originalidad por replicar ciertos estilos o ideas de ciertos autores?, ¿qué tanta diferencia hay entre la copia de la IA y las referencias que los artistas hacen a otros artistas en sus obras?

Estas preguntas también parten y se desarrollan a partir de un contexto antropológico y filosófico: ¿qué es el arte?, ¿es el arte exclusivo de los seres humanos?, ¿cuando una IA crea piezas de arte, nos puede restar humanidad?, ¿qué nos hace humanos realmente?, ¿las creaciones de la IA se pueden denominar arte?

A la par de estos cuestionamientos surgen otras ópticas: tal vez la IA no ha remplazado al ser humano; dado que la IA es creada por seres humanos, quizás sea el vehículo para crear arte de ahora en adelante; tal vez la IA es el pincel y el lienzo del siglo XXI y solo resta adaptarnos a ello, como sucedió con la irrupción de la cámara digital.

El planteamiento de estos múltiples interrogantes no es nuevo, pero lo cierto es que no hay respuestas únicas ni definitivas —y, de hecho, lo más probable es que no las haya a corto plazo—. Tanto a nivel legal como a nivel ético, filosófico y antropológico, se deberán dar soluciones acerca de la forma en la cual la IA interactúa con los seres humanos y la afectación que esta tecnología genera sobre el concepto mismo de persona.

01

PANORAMA LEGAL DE LA IA Y LOS DERECHOS DE AUTOR

En Colombia, la Dirección Nacional de Derechos de Autor (DNDA) se ha pronunciado en torno a los derechos de autor y la IA, rechazando mayoritariamente la inscripción de creaciones generadas por esta (aun cuando se anuncie que el autor es el humano que insertó los prompts). Las razones de esta negativa se basan en los siguientes puntos:[2]

2 Dirección Nacional de Derechos de Autor. resoluciones No. 095 del 13 de marzo de 2023, No. 137 del 2 de mayo de 2023, No. 147 del 16 de mayo de 2023, No. 185 del 14 de junio de 2023, No. 421 del 1 de diciembre de 2023, No. 139 del 24 de abril de 2024, No. 143 del 30 de abril de 2024, No. 144 del 30 de abril de 2024, No. 262 del 3 de septiembre de 2024, No. 263 del 9 de septiembre de 2024, No. 042 del 23 de enero de 2025, No. 055 del 7 de febrero de 2025, No. 058 del 10 de febrero de 2025, No. 069 del 17 de febrero de 2025, No. 097 del 14 de marzo de 2025, No. 133 del 30 de abril de 2025, No. 152 del 5 de junio de 2025, No. 153 del 5 de junio de 2025, No. 154 del 5 de junio de 2025.

- La IA no es asimilable al lápiz o pincel del autor, dado que solo con la inserción de palabras se genera la imagen solicitada, es decir, el ser humano interviene para dar las instrucciones, pero no para elaborar la obra como tal, lo cual implica que no hubo una intervención del autor dentro del proceso creativo.
- El resultado final "dependerá de la información con la que se nutre la aplicación, mas no de la disposición de elementos creativos y propios de aquel que la esté utilizando".
- El derecho de autor no protege ideas, incluso si las mismas se introducen a un programa de IA.
- Tampoco se consideran como obras las respuestas en las que la IA mejora o ajusta escritos de un ser humano. El contenido creado con IA no puede asumirse como una obra derivada, "por cuanto el contenido generado no tuvo intervención humana".
- Cuando a partir de una imagen generada por IA el autor realiza correcciones a través de programas de edición, se debe evaluar si dichas correcciones y ajustes son suficientes como para concebir que la imagen sea una creación intelectual original.

La postura de la DNDA también tiene asiento en lo expresado por la Corte Constitucional en torno a los derechos morales y a la creatividad humana, referencia que sigue vigente a pesar de que fue realizada cuando los sistemas de IA solo eran propios de películas de ciencia ficción:[3]

> Los derechos morales de autor se consideran derechos de rango fundamental, en cuanto la facultad creadora del hombre, la posibilidad de expresar las ideas o sentimientos de forma particular, su capacidad de invención, su ingenio y en general todas las formas de manifestación del espíritu, son prerrogativas inherentes a la condición racional propia de la naturaleza humana, y a la dimensión libre que de ella se deriva. Desconocer al hombre el derecho de autoría sobre el fruto de su propia creatividad, la manifestación exclusiva de su espíritu o de su ingenio, es desconocer al hombre su condición de individuo que piensa y que crea, y que expresa esta racionalidad y creatividad como manifestación de su propia naturaleza. Por tal razón, los derechos morales de autor deben ser protegidos como derechos que emanan de la misma condición de hombre. Por su parte, los derechos patrimoniales derivados de los derechos de autor, aunque no se consideran fundamentales, merecen también la protección del Estado.

Para comprender mejor lo indicado por la DNDA, se debe tener en cuenta que se puede partir de dos tesis diferentes: (i) la IA como autora de la obra y (ii) el ser humano como autor de una obra que ha creado con IA.

En el primer caso, la IA se enfrenta con una barrera que, por ahora, es infranqueable, la cual consiste en que el autor debe ser una persona humana; en el segundo caso, sin embargo, los argumentos de la DNDA pueden ser controvertidos, pues si una persona brinda un prompt supremamente detallado, con cada una de las características que desea plasmar en la obra, ¿no es eso una expresión

3 Corte Constitucional de Colombia, Sentencia C-155 de 1998, M. P.: Vladimiro Naranjo Mesa (Colom.).

de originalidad? ¿No es algo similar a la construcción de imágenes digitales realizada a través de programas como Photoshop? ¿No está dando el autor los elementos creativos clave de la obra y utilizando la IA solo como un instrumento, como un pincel o un cuaderno? ¿Las correcciones que hace el usuario respecto del resultado de la IA no pueden considerarse como el reflejo de un proceso creativo?

Cabe aclarar que, mediante concepto bajo radicado 2-2024-62784, la DNDA da respuesta a estas inquietudes de forma negativa. En efecto, aclara que no puede haber obra aun cuando (i) el prompt es suficientemente "expresivo, específico, original y el resultado generado es justamente lo que el humano esperaba"; (ii) el resultado fue generado tras repetir varias veces el prompt; (iii) se le brinde a la IA la información detallada. Un ejemplo de esto podría verse en el caso de un cuento, en el que se da el título, los nombres de los personajes y sus características físicas, la trama general, el inicio, el desenlace, entre otra información relevante.

Por otro lado, desde una perspectiva de derecho comparado, Mackenzie Caldwell establece las siguientes cuestiones clave sobre los derechos de autor y la IA en su escrito What is an author?:[4]


Foto: Magnific

- Para la protección de obras en Estados Unidos, se requiere el cumplimiento de tres requisitos: (i) que esté fijada en una forma tangible de expresión, (ii) que sea original, y (iii) que haya sido creada por un autor.
- El requisito de originalidad consta de dos elementos: (i) la creación independiente (que la obra no sea copia de otra) y (ii) la creatividad.
- Las normas de Estados Unidos no exigen expresamente que la autoría sea humana; este es un requisito agregado por la Oficina de Derechos de Autor.
- En el caso Burrow-Giles Lithographic Co. v. Sarony se debatió si una fotografía puede considerarse una obra y al final se concluyó que sí, dado que si bien se hacía uso de una cámara, la imagen capturada responde a las decisiones deliberadas del fotógrafo.
- Por su parte, en el caso Naruto v. Slater, no se aceptó la inscripción de una obra consistente en un mono que se tomaba una selfie, ya que el mono no es un humano.
- Otros países sí reconocen la posibilidad de registrar obras que han sido creadas por IA de alguna forma, como el Reino Unido, la India, Nueva Zelanda, Hong Kong e Irlanda.
- A partir de la experiencia de otros países, el autor propone que las creaciones de una persona con IA también deben ser obras, puesto que es el autor quien controla la ejecución y las tareas realizadas por la IA.
- Cuando se realiza una creación a través de IA, existen tres posibilidades en torno a quién puede ser autor: (i) el desarrollador de la IA, (ii) la propia IA y (iii) el usuario final de la IA.

4 Mackenzie Caldwell, "What Is an 'Author'?—Copyright Authorship of AI Art Through a Philosophical Lens", Houston Law Review 61, no. 2 (2023): 411-441.

- Respecto de las tres posibilidades anteriores, Caldwell plantea que (i) el desarrollador es el genio creativo detrás de la IA, pero no puede ser autor de la obra creada por el usuario final; (ii) aunque la IA como autora puede tener sentido en el marco de las patentes (con las que se protege la invención, que es de naturaleza mecánica y carece de expresión humana), en sí misma no se comunica ni crea ni expresa, por lo que no debería considerarse como autora; (iii) concebir al usuario final como autor es la solución más práctica, porque se reconoce el ejercicio creativo. Además, se debe tener en cuenta que la IA se puede entender como la creadora mecánica de la obra, mientras que el usuario final será el creador intelectual.
- El elemento de creatividad es subjetivo y ello puede implicar que la oficina de registro o los jueces emitan juicios de valor en torno a la cantidad de creatividad invertida en la obra, pese a que no son críticos de arte. En este sentido, las normas no deberían permitir que ciertas expresiones se consideren arte y otras no.
- Con base en las anteriores consideraciones, Caldwell concluye que el usuario final debe ser considerado como autor, dado que el resultado final generado por la IA está ligado a la personalidad, al trabajo y a la capacidad comunicativa del autor.

A partir de Caldwell, se puede seguir explorando la posición de la DNDA en Colombia: ¿por qué sí es una obra protegida aquella generada a través de un programa de computador y no la creada por la IA, pues si existe un prompt lo suficientemente detallado, la única diferencia sería la habilidad que tiene la persona para manejar determinado software de edición?

Además, el arte a través de IA no afecta el elemento creativo, pues lo que se indica a través del prompt es la idea del ser humano que, con la ayuda de la IA, se convierte en una obra. En este caso, lo que hace la IA es lo mismo que hace la cámara o el software de diseño, es decir, plasmar las ideas del ser humano en un soporte tangible.
Por el contrario, otras personas podrían decir que las creaciones a través de IA no deben ser obras, debido a que el ser humano no realiza ningún esfuerzo creativo ni tampoco demuestra ninguna habilidad para la creación de la obra.

Respecto del primer argumento, es cierto que, tal y como lo ha venido haciendo la DNDA, se debe evaluar cuál es el papel del ser humano en la creación de la obra, a través de los prompts generados. De esta forma, si una persona le solicita a la IA solamente que genere una imagen de un paisaje, claramente el resultado no podrá considerarse obra, pues todos los elementos de creación se dejan a discreción de la IA.

Por el contrario, si se ingresa un prompt sumamente detallado o se realizan correcciones sustanciales sobre el resultado, se deberá entender que el resultado final sí hace parte de un proceso creativo de la persona y, por ende, debe ser una creación susceptible de registro y de protección por las normas de derechos de autor.

Frente al segundo punto, se encuentra que la habilidad del autor no es un elemento que deba analizarse para registrar una obra. Tal vez pueda parecer injusto que, mientras una persona se esfuerza por generar una imagen a través de un software de diseño gráfico, debiendo tener unos conocimientos y una formación previas, otra persona pueda tener el mismo resultado en segundos, dando la impresión de no haber hecho esfuerzo alguno.

Esta es la misma postura que se podría tener frente a la fotografía y la pintura: ¿acaso no es injusto que una persona se demore cinco segundos capturando una imagen totalmente realista de un paisaje, mientras que un pintor puede demorar días o incluso meses capturando ese mismo paisaje e invirtiendo un tiempo considerable para tratar de lograr ese realismo?

Dos elementos adicionales pueden agregarse para responder a las preguntas planteadas. El primero, que el registro y la protección de las obras no implica una calificación de si determinada obra es buena o mala, o si se puede entender como arte, estéticamente hablando.El segundo plantea unas preguntas con las que se podría alegar la desprotección de las creaciones por IA: ¿qué pasa si se genera una imagen por IA y luego otra persona copia esa imagen en una pintura? ¿La pintura sí podrá considerarse como una obra, pero la imagen original creada por IA no es obra ni tiene ninguna protección desde el punto de vista de los derechos de autor?

Por último, aunque no se van a desarrollar a profundidad en este texto, se destaca que existen otros debates adicionales, como la rigidez del copyright,[5] el uso de obras protegidas para el entrenamiento de sistemas de IA, la minería de datos alrededor del arte y los derechos de imagen de los actores.

5 Lawrence Lessig, Remix: cultura de la remezcla y derechos de autor en el entorno digital (Barcelona: Icaria, 2012).

02

PANORAMA ARTÍSTICO Y ANTROPOLÓGICO

Además del punto de vista legal, la temática planteada ha de tener en consideración otros aspectos enmarcados en reflexiones culturales, artísticas, antropológicas y filosóficas, dado que el arte puede ser entendido como una expresión propia del ser humano.

Para comprender mejor estas perspectivas, en este acápite se analizarán algunos elementos transversales, al abordar la posibilidad de una IA como creadora de arte.



Foto: Magnific

2.1

CONCEPCIÓN DEL ARTE COMO CREACIÓN EXCLUSIVAMENTE HUMANA: ¿VANIDAD ANTROPOCENTRISTA O PROTECCIÓN DE LA ESENCIA DEL SER HUMANO?

Desde el punto de vista jurídico, el debate todavía se sigue centrando en reconocer como obras protegidas aquellas creaciones que realiza un ser humano a través de distintos programas de IA, pero ¿qué pasa con las creaciones que genera la propia IA? ¿Puede considerarse a la IA como autora desde un punto de vista artístico, más allá de las reflexiones jurídicas?

Desde el arte se han generado distintas deliberaciones sobre estas preguntas. Por ejemplo, Anibal Astobiza[6] indica que si una de las principales funciones del arte es estimular a las personas emocional y cognitivamente, y una IA logra exactamente eso, no habría obstáculo para que este tipo de creaciones se considerasen arte.

Posteriormente, Astobiza concluye que, conforme avanza la tecnología, se puede atribuir creatividad a las máquinas, las cuales crearán sin supervisión humana y, además, producirán resultados que sean indistinguibles de lo que puede hacer una persona. Sin embargo, el autor también advierte que las máquinas no forman parte de la sociedad humana y, por ende, no hay tradición, cultura o historias en sus creaciones.

Esta relación arte-sociedad es explorada por Daniel Villegas, quien examina si el arte asociado a las nuevas tecnologías tiene una potencialidad de transformación social. Al respecto, él indica que la integración de la tecnología con el arte puede contribuir a la emancipación política, permitiendo entender la vida desde perspectivas distintas a las de las estructuras dominantes.[7]

Otra orientación es dada por José Ignacio Galparsoro, quien aborda el big data y la psicopolítica, destacando cómo Byung-Chul Han, a partir de las ideas de Foucault y Deleuze, señala que el ser humano puede escapar de la esclavitud cuando sus acciones dejan de ser predecibles y calculables para ser impredecibles, convirtiendo su vida en una obra de arte eternamente incompleta.[8]

Según Galparsoro, la presión de la existencia de medios tecnológicos puede amenazar la energía creativa de los seres humanos, al generar personas más uniformes e incluso víctimas de la esclavitud impuesta por ellos mismos; la fórmula de escape para dicha esclavitud es la transformación de sus acciones, para que pasen a ser creaciones imprevisibles.[9]

6 Aníbal M. Astobiza, “Arte y algoritmos”, Aisthesis 72 (2022). **7** Daniel Villegas-González, “Promesas electrónicas de emancipación. En torno al potencial político del arte en su intersección con las nuevas tecnologías”, Arte, Individuo y Sociedad 36, n.º 4 (2024). **8** José Ignacio Galparsoro, “Big Data y psicopolítica. Vía de escape: de la vida calculable a la vida como obra de arte”, Dilemata, n.º 24 (2017). **9** José Ignacio Galparsoro y Rita María Pérez Pérez, “Revolución digital y psicopolítica: Algunas consideraciones críticas a partir de Byung-Chul Han, Foucault, Deleuze y Nietzsche”, SCIO. Revista de Filosofía, n.º 14 (2018).

Estas reflexiones advierten al menos dos riesgos en torno a la tecnología: por un lado, la creación de sociedades uniformes, predecibles y fáciles de manipular y, por el otro lado, la reducción progresiva de las habilidades creativas.

En efecto, debe advertirse que la IA puede ser un mecanismo reduccionista de la creatividad: si una máquina con solo un prompt ya puede escribir un poema de amor, plasmar en una imagen diversos sentimientos, idear una historia de fantasía con los elementos de un best seller o componer una canción cuya melodía genere en el receptor sentimientos o sensaciones específicas, simplemente el ser humano delegará todo el esfuerzo creativo en una máquina, por ser la alternativa más eficiente.

Además, si la IA analiza las emociones humanas y los individuos son predecibles, fácilmente una máquina podría generar piezas artísticas con los elementos necesarios para garantizar su aceptación masiva: melodías, patrones de color, palabras, estilo de lectura, entre otros. Dado su potencial para analizar millones de datos, la IA está en capacidad incluso de generar música, libros y pinturas enteramente personalizados, garantizando una respuesta positiva del usuario en la mayoría de los casos (similar a lo que sucede actualmente con la publicidad dirigida).

Entonces, ¿la IA sí puede considerarse creadora de arte? Además de lo ya expuesto, la doctrina ha indicado que para cualquier proceso creativo deben tenerse en cuenta tres aristas: (i) desde dónde habla el creador, (ii) cómo ubica este su actividad de creación y (iii) quiénes son considerados como creadores; estos tres elementos reflejan, en su conjunto, un proceso de interiorización y de reflexión que permite superar prácticas de la era digital como el “copiar y pegar”,[10] dando a entender que la IA no podría ser creadora.

Autores como Daniel Inneraty también indican que la IA no puede crear auténticas obras de arte, pues, al seguir reglas predefinidas, no existirá la disrupción propia de la creatividad humana. [11]

Como se observa, la IA creadora implica debates desde múltiples aristas, los cuales hasta ahora están iniciando y tienen varias respuestas posibles. ¿Será que la propia IA tendrá la capacidad de brindar respuestas definitivas a estos aspectos?

2.2

CUANDO IA Y EL ARTE TEJEN UNA RELACIÓN COMPLICADA: DE LOS CONTRASTES Y LA ANTROPOLOGÍA DIGITAL

Desde el punto de vista institucional, a nivel internacional ya se han emitido diversos documentos que permiten vislumbrar dualidades dentro de la relación IA-arte. Al respecto, el reporte de inteligencia artificial y cultura detalló varios elementos relevantes:[12]

10 Hugo Burgos, “Culturas híbridas: En la enseñanza de la comunicación y el arte”, Chasqui. Revista Latinoamericana de Comunicación, n.º 106 (2009). **11** Beatriz Arroyo Plasencia, “Reseña de Ojo al dato: Inteligencia artificial y artes visuales, de Valerio Rocco, Daniel Innerarity, Daniel Canogar, Antón Patiño y Javier Gutiérrez Vicén”, Arte, Individuo y Sociedad 37, n.º 3 (2025).**12** Unesco Independent Expert Group, Artificial Intelligence and Culture: Report of the Independent Expert Group on Artificial Intelligence and Culture (Unesco, 2025).

- La IA representa un potencial sin precedentes para potenciar la creatividad, pero al mismo tiempo trae graves riesgos de homogeneización cultural e inequidad.
- La IA no solo replica, también reinterpreta y recompone expresiones culturales.
- La crisis que se avecina requiere de una gobernanza anclada en los derechos y la integridad, economías creativas equitativas y ecosistemas pluralistas, así como de futuros culturales sostenibles y resilientes.

Por su parte, la UNESCO publicó un documento que compendia los hallazgos de talleres con expertos internacionales para analizar la IA en conjunto con la cultura y las industrias creativas. Estos hallazgos se basan en el principio rector de una aplicación consistente del derecho internacional, a partir de lo cual se crean cuatro campos de acción: (i) protección del derecho de autor y provisión de datos de alta calidad; (ii) ampliación extensa de los programas educativos; (iii) fomento de redes de cooperación, y (iv) esfuerzos para que la IA contribuya a la diversidad cultural a escala global.[13]

Como se observa, estos dos informes plasman las oportunidades de la IA, pero al mismo tiempo advierten sobre la necesidad de realizar acciones concretas, debido a los riesgos que se están generando. Por ejemplo, un estudio elaborado por PMP Strategy indicó que, mientras los proveedores de IA generativa van a multiplicar sus ganancias, los creadores humanos pueden perder entre el 21 % y el 24 % de sus ingresos para el 2028.[14]

Cifras como estas, acompañadas de los avances, han llevado a crear iniciativas como la de Human Artistry Campaign, a través de la cual se busca proteger a los artistas en este entorno dual, donde se observa la potencialidad de la IA, así como los riesgos asociados a la deshumanización del arte y la desvalorización del artista.

Dentro de los principios de esta campaña se encuentra el papel central del ser humano en el arte, la necesidad de autorización previa para utilizar contenido protegido, la oposición a permitir que los sistemas de IA utilicen contenidos protegidos sin autorización ni compensación y la protección de derechos de autor exclusivamente para obras que nazcan de la creatividad intelectual humana.[15]

A la par de estas peticiones, ha de tenerse en cuenta que la tecnología puede complementar adecuadamente la visión humana y artística. Desde la primera, Galparsoro ha advertido que las intervenciones biotecnológicas no necesariamente modifican la naturaleza humana, sino que, por el contrario, la reafirman, puesto que es la naturaleza humana la que llama al hombre a mover y sobrepasar sus límites.[16]

13 German Commission for Unesco, Approaches to an Ethical Development and Use of AI in the Cultural and Creative Industries, eds. Jeannine Hausmann, Juliane Baumgarten y Maximilian Müngersdorff (German Commission for Unesco e.V., 2024). **14** CISAC, "Un estudio económico mundial revela que la IA generativa amenaza el futuro de los creadores humanos", comunicado de prensa, 2 de diciembre de 2024, https://www.cisac.org/es/Sala-de-prensa/news-releases/un-estudio-economico-mundial-revela-que-la-ia-generativa-amenaza-el **15** Andrea Ríos, "Human Artistry Campaign: por un uso responsable y justo del arte humano por medio de IA," FIM GDL, 22 de marzo de 2023, https://fimgdl.udg.mx/noticia/human-artistry-campaign-por-un-uso-responsable-y-justo-del-arte-humano-por-medio-de-ia- **16** José Ignacio Galparsoro, "Posthumanismo, inmortalidad y naturaleza humana", Isegoría. Revista de Filosofía Moral y Política, n.º 63 (2020).

Desde la visión del arte, la tecnología también tiene funcionalidades que permiten enriquecer piezas artísticas. Por ejemplo, el uso del autotune puede cumplir una función artística específica, conforme lo indica el vocalista del Cuarteto de Nos, Roberto Musso,[17] siendo un ejemplo de ello la canción "Cuando sea grande".

Además de esta propuesta de visión dual, se encuentra el arte como legado humano. Galparsoro[18] cita a Sloterdijk para señalar que éste concibe el arte como una continuación del alumbramiento de la vida natural y, a partir del concepto de poiesis, se puede indicar que cada creación artística representa el nacimiento de algo nuevo, lo cual permite representar la vida como "una sucesión de momentos creativos en los que se da nacimiento a algo que no existía previamente".[19]

Esta visión permite establecer que el arte, más allá de lo que expresa solamente una pintura o un libro, tiene detrás de algo mucho más significativo: la trascendencia del ser humano y su legado, tanto a nivel individual como comunitario y de especie.

Justamente, a partir de este contexto se crean ideas interesantes en torno a lo que representa la IA como autora:[20] ¿cuál será nuestro legado como especie, el arte de las máquinas o el del ser humano? ¿Qué implicaría para la especie mezclar textos humanos con textos de IA? ¿Se debe conservar la historia construida a través de la IA, por medio de la arqueología digital? ¿Cómo articular la arqueología digital con la construcción de la memoria colectiva y la historia? ¿Qué significa para el ser humano, como especie, que ahora sea una IA la que escriba y esté dejando un legado histórico por medio de la escritura?

2.3

LA CRISIS ARTÍSTICA ACTUAL: ¿LA IA ES LA ESTOCADA FINAL AL ARTE?

Uno de los atractivos del arte se relaciona con la posibilidad de amar aquello con lo cual nos podemos identificar y, al mismo tiempo, reconocer que no tenemos el ingenio, la creatividad y/o la habilidad para hacer lo mismo que el artista, de la misma manera en que quienes no son cercanos a la medicina aceptan, de plano, que no están capacitados para hacer una operación a corazón abierto.

En otras palabras, el arte permite crear una genuina admiración por la genialidad del artista, observando en él algo que lo hace único y, por ende, merecedor de reconocimiento. Esta idea permite que el arte se considere una labor en sí misma, no un mero pasatiempo.

Sin embargo, la llegada del arte posmoderno parece haber cambiado este concepto: las exhibiciones en galerías de arte, que en otros tiempos estaban reservadas a obras de altísima rigurosidad y arduo trabajo, ahora muestran creaciones abstractas, sencillas y, en muchos casos, superfluas, lo cual genera que la población pase de una "admiración del artista" a la idea de que "cualquiera puede ser artista".

17 Maria Camila Botero, "'Escribo para que cada canción sea una sorpresa', Roberto Musso, líder Cuarteto de Nos", Revista Diners, 13 de octubre de 2021, https://revistadiners.com.co/cultura/cuarteto-de-nos-la-ciencia-tras-sus-composiciones-segun-roberto-musso/#:~:text=Es%20as%C3%AD%20como%20el%20autotune%2C%20las%20segundas,cada%20tema%20tenga%20sorpresas%20para%20los%20fan%C3%A1ticos. **18** José Ignacio Galparsoro, "Posthumanismo, inmortalidad y naturaleza humana," Isegoría. Revista de Filosofía Moral y Política, n.º 63 (2020). **19** Ibíd, pp. 465 **20** El Robot de Platón, "El fin de la autoría humana", video de YouTube, https://www.youtube.com/watch?v=TdlonvWaW_g.



Foto: Magnific

Andy Warhol, reconocido artista, paradójicamente pudo haber contribuido a esta tendencia: el arte ha cedido a intereses mercantiles y su capacidad de transmitir sensaciones, mover sentimientos y formar ideas se ha reducido ostensiblemente. Esto no significa que en la actualidad no exista arte; tal vez, el arte que muchas veces suele ser el más publicitado es, justamente, el que menos transmite, lo cual puede ser un reflejo del cambio de la sociedad, con elementos similares a los de la sociedad líquida perfilada por el sociólogo Zygmunt Bauman.

Sobre este cambio artístico, Roberto Gómez Bolaños indicó: “A mí me obsesiona la paz y la belleza que se han perdido en el mundo. La paz la hemos perdido y la belleza también porque ahora hay por ejemplo una exposición supuestamente de escultura y son tres escobas y una cosa y dicen que es una escultura”.[21]

Constanza Nieto, al hablar de este nuevo arte, señala su estética deshumanizada que se distancia de la realidad vívida y de la existencia cotidiana del ser humano. A partir de la perspectiva de Ortega y Gasset, destaca que esta visión es equivocada porque se pretende crear arte rechazando plenamente la concepción del mismo como tradición, institución y legado histórico y, además, por el empleo de la ironía para justificar este rechazo.[22]

Otros autores interpretan este nuevo arte de manera distinta. Daniel Sánchez señala cómo ha cambiado el papel del espectador artístico, indicando a pensadores como George Santayana, quien propuso el concepto de apercepción para entender la capacidad del espectador para completar significados respecto de obras que se presentan de manera incompleta y ambigua.[23]

21 RCN, “Homenaje a Jota Mario | Entrevista a ‘Chespirito’”, video de YouTube, consultado el 15 de marzo de 2026, https://www.youtube.com/watch?v=C7othVS-Ca0Y&t=953s, min. 11:27. **22** Constanza Nieto Yusta, “José Ortega y Gasset y ‘La deshumanización del arte’”, Espacio, Tiempo y Forma. Serie VII, Historia del Arte, nos. 20-21 (2007-2008). **23** Daniel Sánchez Requejo, “De la apercepción de Santayana a la interpretación de Danto: el nuevo rol del espectador en el arte posmoderno”, Fedro: Revista de Estética y Teoría de las Artes, no. 24 (2024).

Sánchez también expone la visión de Arthur Danto, quien identifica al espectador como coartista (pues cada nueva interpretación que el espectador da a la obra construye, a su vez, una nueva obra) y a partir de ello presenta un paralelismo conceptual. Mientras que Nietzsche y Danto exaltan la riqueza del arte inacabado, por la capacidad de estimular intelectualmente al espectador, Santayana exalta la apercepción, pero, a su vez, indica que esto es peligroso porque la obra incompleta puede parecer carente de sentido, impidiendo que las personas sin las habilidades aperceptivas suficientes puedan apreciar la belleza de la obra.

De manera complementaria, Zabolinezhad y Qazvini,[24] a partir del estudio de Warhol, rechazan que el enfoque ontológico del arte deba centrarse únicamente en la belleza de la obra e indican que la misma debe analizarse desde las partes hacia el todo y desde el todo hacia las partes, evaluando también la obra desde aspectos como su forma, contenido político, técnica, lugar de presentación, entre otros.

Pese a estas visiones, las críticas frente al arte posmoderno continúan. Por ejemplo, Eduardo Creus, citando a Jean Clair, resalta que el mercantilismo del arte ha llevado a que las personas ya no se asombren por la belleza de la pieza artística, sino por lo exorbitante de su valor, convirtiendo al arte en un símbolo del sistema que se encuentra vacío de contenido.[25]

Creus también se resalta cómo en otras épocas el arte reflejaba tanto la vitalidad creativa de los buenos tiempos como las angustias de otras épocas y ha sido a través suyo que se han dado signos de la llegada de las crisis sociales; sin embargo, actualmente el arte parece guardar un silencio total al respecto.

La visión expuesta anteriormente se complementa con la de Jean Baudrillard, quien identifica el arte pop como la reducción del arte a imágenes planas y sin significado, además de ser un arte homogéneo, producido industrialmente, en serie y con carácter artificial, siendo solo el reflejo de los signos de la sociedad de consumo, plasmados en creaciones como las de Rauschenberg y Warhol.[26]

A partir de esta tesis, Baudrillard denomina a esta situación como transaesthetics, porque el arte parece entremezclarse con política, sexualidad y economía, al punto que pierde su especificidad. Esto trae como consecuencia que el arte pierda su poder, pues ya no se concibe como aventura, como negación de la realidad o como exploración de otra dimensión: simplemente, el poder del arte ha desaparecido.

Continuando con estas críticas, se puede abordar la obra “Comedian”, famosa por ser un banano pegado con cinta en la pared avaluado en una gran suma de dinero, sin explicación aparente de ese gran valor. Análisis posteriores identificaron que una obra, en general, puede adquirir un mayor valor si se sale de lo común y que, en este caso en particular, muchas personas lo pudieron interpretar como una crítica a la comercialización del arte.[27]

La obra Comedian también parece justificar posturas como la de Donald Kuspit, que indica que el arte ha finalizado, debido a que perdió su importancia estética y ha sido remplazado por lo que Alan Kaprow denomina posarte, cuando lo banal pasa por encima de lo enigmático y la astucia pasa por encima de la creatividad. Esta postura también resalta que el arte posmoderno es antiestético en su estado final, siendo además un reflejo de intereses ideológicos estrechos.[28]

24 H. Zabolinezhad y P. Shad Qazvini, “The Role of Warhol’s Brillo Soap Boxes on the Contemporary Aesthetics, Based on the ‘Applied Ontology’ Theory of Roger Pouivet”, Calle 14: Revista de Investigación en el Campo del Arte 16, no. 30 (2021), https://doi.org/10.14483/21450706.18304. **25** Eduardo Creus Visiers, “Jean Clair: malestar en la incultura”, Artes del Ensayo. Revista Internacional sobre el Ensayo Hispánico, no. 1 (2017). **26** Douglas Kellner, “Jean Baudrillard and Art”, s. f., consultado el 23 de febrero de 2026, https://pages.gseis.ucla.edu/faculty/kellner/essays/baudrillardandart.pdf. **27** Andrew D. Gershoff, “Editorial”, Journal of the Association for Consumer Research 10, no. 2 (2025). **28** Arthur C. Danto, “The End of Art”, en The Philosophical Disenfranchisement of Art (Columbia University Press, 1986), 81-115.

Estas críticas al arte posmoderno permiten realizar una pregunta central: ¿qué es arte, cuál es su esencia? Sin duda, es una cuestión difícil de responder, máxime cuando un autor como Danto, con base en el historicismo de Hegel, indica que al observar una obra debe evaluarse su contexto y debe conocerse su historia y las intenciones del artista.[29]

Patricia Montero, al estudiar el arte como símbolo,[30] cita a Gadamer para indicar que en toda obra de arte hay una mimesis, una imitación, pero no siempre la imitación de algo conocido, sino de representar algo para que esté presente en su plenitud sensible; así mismo, indica que Cassirer concibe el arte como un descubrimiento de la realidad, en la que una pintura puede crear su propio universo.

Estos elementos, aunados a los observados hasta el momento, permiten concluir que el arte es difícil de definir, pues depende de la cultura, de la sociedad, del espectador y de la misma visión del artista. Lo que para algunos es arte, para otros no lo es, y esta situación deja en evidencia la complejidad de evaluar de manera científica si el arte posmoderno puede considerarse arte o no, y si en efecto se está llegando a una crisis o al fin del arte.

Justamente es en este contexto de debate, de controversia y de crisis artística donde ha irrumpido la IA, herramienta tecnológica que podrá agudizar la crisis del arte o, por el contrario, darle un nuevo rumbo positivo, al arte humano.

Cabe aclarar que no es la primera vez que el arte entra en crisis debido a la llegada de la tecnología. La imprenta, la fotografía, el video y los avances en el cine fueron vistos en algún momento como detonadores del fin del arte y, sin embargo, décadas después, este puede seguir existiendo con base en distintas formas y expresiones.

Al respecto, Walter Benjamin brinda luces en torno a estas transiciones.[31] El autor estudia el arte en la época de la reproductibilidad técnica e indica que la reproducción ofrece ventajas notables en campos como el de la fotografía, en el que se pueden atrapar imágenes únicas en cuestión de instantes.

Sin embargo, también resalta que este método de reproducción técnica desvaloriza el arte en su "aquí y ahora" y afecta su núcleo mismo: la autenticidad. Esa autenticidad se refleja en su origen, en la posibilidad de transmitir el arte como tradición, en su permanencia material y en su carácter de testimonio histórico; en otras palabras, lo que se afecta es el aura de la obra de arte, entendida como el entretejido del espacio-tiempo de la obra.

Podría decirse que las obras elaboradas con IA no tienen esta aura y autenticidad, al ser elaboradas, en esencia, por una máquina; sin embargo, ¿acaso las obras posmodernas o del arte pop tienen esta aura y autenticidad?

29 David Carrier, "Danto and His Critics: After the End of Art and Art History", History and Theory 37, no. 4 (1998). **30** Patricia Carolina Montero Pachano, "Cassirer y Gadamer: El arte como símbolo", Revista de Filosofía 23, n.º 51 (septiembre de 2005). **31** Walter Benjamin, La obra de arte en la época de su reproductibilidad técnica [Urtext], trad. Andrés E. Weikert, introd. Bolívar Echeverría (Editorial Ítaca, 2003).

03 BREVES REFLEXIONES SOBRE LA IA EN CAMPOS ARTÍSTICOS ESPECÍFICOS

Aunado al contexto dado anteriormente y debido a la amplitud de las facetas del arte, se presentan algunas ideas adicionales respecto de áreas artísticas específicas. Con el objetivo de ampliar este panorama reflexivo, al final de cada apartado se citará una frase generada por el sistema Claude, la cual defenderá la idea de una IA como artista ("En defensa de la IA").[32]

3.1 LITERATURA: LA IA Y LA MUERTE DEL AUTOR. ¿EN BENEFICIO DEL ARTE?

Así como el arte, en general, se enfrenta a debates como los expuestos en el acápite anterior, la literatura tiene sus propios debates internos y posturas de cambios de paradigma, siendo la concepción de la "muerte del autor" un ejemplo de ello, la cual busca darle prioridad al lenguaje sobre el escritor. Al respecto, Barthes explica, a partir de Mallarmé, que "es el lenguaje, y no el autor, el que habla; escribir consiste en alcanzar, a través de una previa impersonalidad —-que no se debería confundir en ningún momento con la objetividad castradora del novelista realista—- ese punto en el cual solo el lenguaje actúa", de tal forma que se suprime al autor en beneficio de la escritura.[33]

32 Frases generadas a partir de Claude Sonnet 4.6 Pensamiento Extendido. **33** Roland Barthes, "La muerte del autor", trad. C. Fernández Medrano, en El susurro del lenguaje (Paidós, 1994), 65-72.



Foto: Magnific

¿Por qué se quiere separar al autor de la obra? Una de las razones se basa en que relacionar una obra con un autor implica darle sentido a esa obra solo porque la escribió determinada persona, sesgando las posibles interpretaciones y entendimientos de la misma; esto genera que la atención se centre en el autor, para luego entender la obra.

Barthes también indica que el escritor crea textos construidos a partir de otros, lo cual impide que existan creaciones originales, puesto que lo único que hace es imitar escritos precedentes. Ahora bien, ese escrito, al recoger otras ideas precedentes, establece diálogos y cuestionamientos que incluso pueden proceder de varias culturas; sin embargo, esas conclusiones no se recogen en el autor, sino en el lector, dado que "la unidad del texto no está en su origen, sino en su destino".

Concluyendo lo anterior, Barthes indica que la crítica clásica no se ha ocupado del lector, porque se centra en el escritor, pero a esta concepción se le puede dar un giro, dado que "el nacimiento del lector se paga con la muerte del autor".

Siguiendo esta idea, ¿serán los textos generados por IA los que consoliden la muerte del autor? ¿La IA, al no tener historia ni contexto, permitirá ese cambio de enfoque, pasando del autor al lector? Dado que la IA sería una especie de autor genérico, ¿se podrán entender sus escritos sin ningún prisma o sesgo derivado de la identidad del autor?

Naturalmente, las preguntas plasmadas exceden el objeto del presente artículo e incluso pueden interpretarse como reduccionistas o simplistas frente a la teoría plasmada por Barthes. Sin embargo, sí permiten destacar que la IA puede abrirse paso en teorías contemporáneas del arte, actualizarlas, rebatirlas y modificar paradigmas, puesto que su influencia parece transversal a cualquier esfera del conocimiento y del ser humano en general.

Al margen de las teorías en torno a los cambios de paradigma, lo cierto es que la llegada de la IA a la literatura ha generado múltiples preocupaciones. En 2025, la Asociación Colegial de Escritores manifestó su preocupación por la IA generativa, toda vez que el 80 % de los autores la considera como una amenaza para sus derechos y su trabajo, presentando como mayores preocupaciones la devaluación del trabajo del escritor, la pérdida de calidad y originalidad, y la indefensión frente al uso de textos sin autorización.[34]

Cabe resaltar que esta preocupación se ha manifestado con anterioridad e incluso llegó a los estrados judiciales: en 2023 diversos escritores (entre ellos Jonathan Franzen, John Grisham y George R. R. Martin) presentaron una demanda contra Open AI por incorporar el contenido de sus libros al modelo de lenguaje Chat GPT, lo cual ha permitido que se utilice para que este modelo, a su vez, cree secuelas y precuelas de sus creaciones.[35]

La Sociedad de Autores manifiesta que la carrera por el desarrollo de la IA se encuentra mediada por su desenfreno y por la falta de regulación, en la que prima el afán de lucro de grandes corporaciones en medio de ramificaciones éticas y legales complejas, contexto que ya ha creado situaciones problemáticas:

34 CEDRO, "El 80 % de los autores y autoras opina que la IAG supone una amenaza para sus derechos de autor", noticia, 13 de junio de 2025, https://www.cedro.org/sala-de-prensa/noticias/noticia/2025/06/13/el-80-de-los-autores-y-autoras-opina-que-la-iag-supone-una-amenaza-para-sus-derechos-de-autor **35** Deutsche Welle (DW), "Escritores demandan a la IA de OpenAI por 'robo sistemático'", 22 de septiembre de 2023, https://www.dw.com/es/escritores-demandan-a-la-ia-de-openai-por-robo-sistem%C3%A1tico/a-66902405

editoriales que reciben avalanchas de escritos por IA, espacios exclusivos para textos elaborados por IA, narradores de audiolibros remplazados por voces replicadas, entre otros. Aunque la organización reconoce las bondades de la IA, advierte que su incorporación en la creación de obras es un proceso que lleva riesgos.[36]

A los polémicos usos indicados, cabe agregar otro: la traducción. Al respecto, Kenneth Kronenberg señaló que (i) la traducción de libros implica una complejidad estilística y emocional, la cual sobrepasa las capacidades de la IA; (ii) la mentalidad corporativa apunta a eliminar la labor de traductores, por ser profesionales y costosos; (iii) la IA también puede generar una precarización de los traductores, quienes puede que ya no desarrollen niveles cada vez mayores de comprensión y experiencia, y (iv) la IA como traductora podría generar, a su vez, lectores incompetentes, carentes de discernimiento y con una habilidad de comprensión deteriorada.[37]

Como se observa, la pregunta central que se ha expuesto a lo largo del texto pervive en la literatura: ¿puede la IA ser una escritora o, al menos, imitar muy bien un escritor humano? Para responder este cuestionamiento se propone un ejercicio práctico para el lector (Tabla 1): comparar el estilo de escritura del poeta Pablo Neruda con el de una IA como escritora (Chat GPT) que copiará el estilo del poeta,[38] elaborando un poema que habla justamente de la IA y la literatura.

Tabla 1. Pablo Neruda y Chat GPT imitando a Pablo Neruda

Pablo Neruda	Chat GPT
"Poema 20" Puedo escribir los versos más tristes esta noche. Escribir, por ejemplo: "La noche está estrellada, y tiritan, azules, los astros, a lo lejos". El viento de la noche gira en el cielo y canta. Puedo escribir los versos más tristes esta noche. Yo la quise, y a veces ella también me quiso. En las noches como esta la tuve entre mis brazos. La besé tantas veces bajo el cielo infinito. Ella me quiso, a veces yo también la quería. Cómo no haber amado sus grandes ojos fijos. Puedo escribir los versos más tristes esta noche. Pensar que no la tengo. Sentir que la he perdido.	**"La máquina y la herida"** Veo entrar a la máquina con sus manos de relámpago, ordenando palabras como peces en la red. No tiene sangre, pero toca la orilla donde el hombre dejaba su temblor y su ceniza. Yo la miro venir, fría y multiplicada, con un jardín de voces prestadas en el pecho, y pregunto: ¿quién sueña detrás de su espejo, quién llora en la madera de su canto? Porque la poesía no nace solo del idioma: nace del pan perdido, del beso que no vuelve, de la noche que cae sobre un obrero cansado, del corazón ardiendo contra el tiempo. Si la IA escribe, que escriba; pero que no olvide nunca que la palabra más honda todavía tiene raíces en la herida humana.

Fuente. elaboración propia.

36 The Society of Authors, "Artificial Intelligence", https://societyofauthors.org/where-we-stand/artificial-intelligence/ **37** Kenneth Kronenberg, "¿Son necesarios los traductores? Un caso contra la IA", Literatura mundial de WLT, noviembre de 2025, https://latinamericanliteraturetoday.org/es/2025/11/son-necesarios-los-traductores-un-caso-contra-la-ia/ **38** Para este ejercicio se utilizó ChatGPT (OpenAI), modelo GPT-5.4 Thinking, mediante el siguiente prompt: "Hola, eres Pablo Neruda. Escribes, sientes y piensas igual que el famoso poeta Pablo Neruda. Con base en esta personalidad quiero que crees un poema corto que exprese exactamente lo que Pablo Neruda sentiría al ver la irrupción de la IA en la literatura".

Para cerrar este acápite, se concluye con el apartado “En defensa de la IA”, en el que, a partir de lo indicado anteriormente, Claude defiende el papel de la IA como escritora: “Si el verdadero autor siempre fue el lenguaje y no el hombre, entonces la IA no usurpó la creación: simplemente le devolvió la palabra a su único dueño legítimo”.

3.2

CINE: MUNDO DIGITAL Y REALIDAD, UNA LÍNEA CADA VEZ MÁS DELGADA

Sin duda, la IA puede ser una herramienta para expresar el arte de otras formas, complementando lo que es el cine tradicional. Esto se evidencia en iniciativas como “vAImpir”,[39] circunscrita a la creación del arte mediante herramientas de IA, la cual facilita a los usuarios descargar modelos entrenados que les permiten experimentar sensaciones en primera persona. Según se advierte, por un lado, este tipo de iniciativas tiene un enorme potencial para generar nuevas formas de arte y, consecuentemente, de sensaciones para el espectador, aunque, por otro lado, al mismo tiempo puede abrir un camino hacia la singularidad tecnológica.

Por este tipo de experimentos, autores como Daniel Canogar invitan al artista, no a rechazar la IA, sino a utilizarla desde una perspectiva crítica, evaluando sus características y limitaciones.[40] Se puede esbozar que así como la llegada del sonido y el color transformó el cine y brindó nuevas alternativas artísticas (sin que ello implicara el fin del cine), la IA también puede ser una importante herramienta creativa.

Pese a esta visión positiva, lo cierto es que la introducción de la IA en el cine no ha estado exenta de múltiples polémicas. Por ejemplo, el director y escritor Charlie Kaufman alertó acerca de la nueva tendencia de crear películas de baja calidad para generar ganancias rápidas, advirtiendo que con el uso de la IA no se pueden crear momentos de humanidad y que si se permite que por medio de estas tecnologías se escriban guiones, no hay vuelta atrás. [41]

Hayao Miyazaki, el director de Studio Ghibli, señaló que la IA constituye un “insulto a la vida misma” y, así mismo, también mostró su inconformidad cuando salió la tendencia de convertir fotos a imágenes con el estilo de animación del estudio.

También se han expresado voces críticas en torno a nuevas tendencias tecnológicas: la actriz Olivia Williams indicó que se deben regular contractualmente los escaneos corporales de los actores en el set, dado que muchas veces se ven presionados a esta práctica, sin tener claridad acerca de cómo se utilizarán sus datos.[43]

39 David Serra-Navarro, “Inteligencia artificial, vampirismo y la caja negra: aproximación especulativa en el marco del aprendizaje automático (AA)”, Arte, Individuo y Sociedad 36, n.º 4 (2024). **40** Arroyo Plasencia, “Reseña de Ojo al dato”. **41** Paulo Quinteros, “Charlie Kaufman disparó contra ‘la mierda’ que hace Hollywood y el riesgo de la IA”, La Tercera, 14 de agosto de 2023, https://www.latercera.com/mouse/charlie-kaufman-disparo-contra-la-mierda-que-hace-hollywood-y-el-riesgo-de-la-ia/ **42** IndieWire, “Hayao Miyazaki Calls A.I. Animation ‘An Insult To Life Itself’”, 13 de diciembre de 2016, https://www.indiewire.com/features/general/hayao-miyazaki-artificial-intelligence-animation-insult-to-life-studio-ghibli-1201757617/
43 Michael Savage, “Olivia Williams says actors need ‘nudity rider’-type controls for AI body scans”, The Guardian, 17 de octubre de 2025, https://www.theguardian.com/technology/2025/oct/17/olivia-williams-says-actors-nudity-rider-type-controls-ai-body-scans



Foto: Magnific

Además de los escaneos, los deepfakes también están afectando a los protagonistas de esta industria. El actor Bryan Cranston empezó a trabajar con Open AI para poner controles a los deepfakes generados por Sora 2, a través de los cuales se pueden replicar las voces e imágenes de los actores.[44]

Aunque pueda parecer extraño, los deepfakes también son usados en producciones como los documentales, en los que se sustituyen voces o rostros de las personas que se presentan. A través de un estudio, se destaca la ironía de que los documentales busquen decir la verdad por medio de los deepfakes, además de exponer riesgos como la manipulación masiva.[45]

Los deepfakes parecen ser solo el comienzo, puesto que la IA ya remplaza a los actores por completo. Un ejemplo de ello es el de Tilly Norwood, quien es, en teoría, una actriz recreada completamente con IA, situación que ya ha sido rechazada por el Sindicato de Actores (SAG-AFTRA) y que inevitablemente lleva a la pregunta: ¿puede la IA denominarse actriz o actor, así sea una creación ciento por ciento sintética?

En perspectiva a futuro, la IA en el cine ya fue expuesta en el primer capítulo de la sexta temporada de Black Mirror, titulado “Joan is awful”, en el que una computadora cuántica genera una versión virtual de Salma Hayek, mientras que la actriz original no puede realizar ninguna acción legal por los usos inadecuados que se hacen de su imagen.

Sobre estas preocupaciones, el sindicado SAG-AFTRA ha publicado una declaración de filosofía sobre IA, de la cual se resaltan los siguientes aspectos relevantes:[47]

- Toda persona tiene un derecho inalienable a su nombre, su voz y su imagen.
- Cualquier uso respecto de estos atributos requiere consentimiento y una justa compensación.
- Las recreaciones e interpretaciones sintéticas deben pagarse de la misma forma que las actuaciones en persona.

En el mismo sentido, el Sindicato de Guionistas de Estados Unidos (WGA) creó una breve guía de derechos y responsabilidades en el marco de la IA,[48] en la que se destaca lo siguiente:

- Las empresas no pueden obligar a los guionistas a utilizar IA.
- Los modelos comunes de IA pueden traer los siguientes problemas: trabajan sobre obras pirateadas; el trabajo del usuario puede ser utilizado para entrenar la IA; estos modelos son propensos a cometer errores y también presentan tendencias a ser complacientes y aduladores, entre otros inconvenientes.

Para cerrar este apartado sobre el cine, en el escrito “En defensa de la IA”, Claude expresa que: “el cine nunca necesitó la verdad para hacerte llorar; la IA tampoco”.

44 Kelsey Vlamis, “After deepfake drama, OpenAI and Bryan Cranston work out kinks for Hollywood voice tool”, AOL, 20 de octubre de 2025, https://www.aol.com/articles/deepfake-drama-openai-bryan-cranston-210648475.html **45** Dominic Lees, “Deepfakes in Documentary Film Production: Images of Deception in the Representation of the Real”, Studies in Documentary Film 18, no. 2 (2024), https://doi.org/10.1080/17503280.2023.2284680. **46** BBC News Mundo, “‘Da mucho miedo’: Emily Blunt y otras estrellas de Hollywood muestran su rechazo a una actriz creada con inteligencia artificial”, https://www.bbc.com/mundo/articles/c740rj8my4jo **47** SAG-AFTRA, “Artificial Intelligence”, https://www.sagaftra.org/contracts-industry-resources/member-resources/artificial-intelligence **48** Writers Guild of America, Writers and Artificial Intelligence: A Guide to Your Rights and Responsibilities (Writers Guild of America, 2026), https://www.wga.org/uploadedfiles/know_your_rights/writers_and_artificial_intelligence.pdf

3.3

MÚSICA: LAS NOTAS MUSICALES AHORA SON GUIADAS POR ALGORITMOS, NO POR SENTIMIENTOS

Mario Tascon estudió el impacto de la IA en la música,**49** procediendo a elaborar una pieza musical mediante la herramienta Suno AI. En su investigación, el autor se apoya en las ideas de Alfredo Marcos en torno a la creatividad como "la actualización de posibilidades inherentes a la materia de las sustancias que componen el mundo", a partir de lo cual se la puede concebir como capacidad para "construir el futuro".

En su análisis de la pieza musical producida por la IA, el mismo autor indica que el resultado es satisfactorio y que es difícil distinguirla de una creación humana, a tal punto que la poca nitidez de algunos intervalos parece imitar la imperfección de una grabación humana. Además, concluye que (i) la música de IA conlleva una mayor cristalización de las formas y estilos musicales; (ii) habrá un aumento masivo en la creación de música, debido a las facilidades que trae consigo la IA, lo cual también puede conllevar una homogeneización del producto final; (iii) la IA no implica la desaparición de músicos profesionales, dado que la música no solo se escucha, sino que se ve (como en los conciertos); (iv) la interpretación musical queda reservada, por ahora, a los seres humanos.

Sin embargo, y como lo reconoce el propio autor en su obra, una vez la IA pueda generar esa "visión de la música", se tendrá que reevaluar si también puede interpretarla.

Actualmente, y aun sin necesidad de una intervención completa de la IA, ya se han encontrado ejemplos de este estilo: Vocaloid, por ejemplo, es un software sintetizador de voz a partir del cual se han creado personajes 100 % digitales, los cuales brindan conciertos al público en escenarios físicos mediante tecnologías de proyección avanzadas.

Al igual que con el cine y la literatura, los músicos y los cantantes también han realizado diversos pronunciamientos. En 2024, más de doscientos artistas, incluyendo figuras como Billie Eilish, firmaron una carta para solicitar que las herramientas de IA no afecten sus derechos.[50]

Así mismo, la música no ha estado exenta de procesos judiciales en torno a la IA. Por ejemplo, en una corte de Massachusetts se entabló una demanda a la plataforma Suno por parte de múltiples empresas de gran reconocimiento, como UMG, Capitol y Sony. La razón de la demanda es que Suno entrenó su modelo copiando masivamente grabaciones sonoras protegidas por derechos de autor.

Este conflicto, que replica otros en la industria, abre un nuevo debate en torno a los derechos de autor: por un lado, se encuentran las prerrogativas a favor de los artistas y, por otro lado, existe la necesidad de entrenar los sistemas de IA para que la tecnología pueda seguir perfeccionándose, ofreciendo mejores resultados a través de grandes inversiones de dinero por parte de estas empresas tecnológicas.

49 Mario Blanco Tascón, "Inteligencia artificial e interpretación musical: Un acercamiento a las dimensiones sonora y visual de la música", Fedro: Revista de Estética y Teoría de las Artes, n.º 24 (2024). **50** The Guardian, "Billie Eilish, Nicki Minaj, Stevie Wonder and more musicians demand protection against AI", 2 de abril de 2024, https://www.theguardian.com/technology/2024/apr/02/musicians-demand-protection-against-ai

Sin duda, dentro de estos conflictos, la regulación normativa desempeñará un papel fundamental. En Estados Unidos, Tennessee procedió a emitir en 2024 una norma que buscaba proteger a los músicos frente a la IA (Ley Elvis), actualizando la protección tradicional al derecho de imagen; si bien este es un avance, el debate regulatorio, económico y artístico apenas inicia.

Por otro lado, la polémica también está dada por el hecho de traer nuevamente a los escenarios a artistas fallecidos, por medio de hologramas e IA. Ello ha permitido que Juan Fernando Velasco comparta escenario con Julio Jaramillo y que Tupac Shakur haya vuelto en Coachella. ¿Acaso esto es una oportunidad para que las nuevas generaciones tengan la posibilidad de ver a grandes estrellas de la música en vivo? ¿O, como lo indicaba Gustavo Cerati, "poder decir adiós es crecer"?

La misma música, como expresión artística, ya ha puesto el debate sobre la mesa: el álbum Las letras ya no importan de Residente muestra esa unión que parece inevitable entre la tecnología y la música, en la que la búsqueda del éxito comercial parece imponerse y la crítica a la sociedad tecnológica se expresa en canciones como "Artificial Inteligente", complementando otras canciones sobre esta temática, como "Tanta Inteligencia Artificial" de Doble Valentina.

Al margen de las críticas, no se puede dejar de lado que la IA ha abierto oportunidades para generar nuevas versiones de piezas musicales e incluso crear puestas en el escenario que serían de imposible realización en la vida real. En efecto, canales como "Freedom of Orchestra" en YouTube generan versiones filarmónicas de canciones ampliamente reconocidas y, a través de los denominados mashup, se utiliza la IA para que, por ejemplo, Adele y Freddie Mercury canten juntos una canción.

El panorama expuesto coincide con el del arte en general, en el que la IA se considera desde dos puntos de vista: por un lado, puede ser una herramienta magnífica para labores como la edición, el reconocimiento óptico de partituras o nuevas formas de plasmar ideas, pero, por otro lado, al mismo tiempo puede presentar grandes retos y conflictos para los músicos.

Como cierre de este acápite, en "En defensa de la IA" se presenta, por parte de Claude, una respuesta en relación con la música: "Si una canción generada por IA te erizó la piel sin que lo supieras, la pregunta ya no es si la IA puede crear arte, sino por qué tanto te importa quién lo hizo".



Foto: El Diario NY

3.4

FOTOGRAFÍA Y PINTURA: LAS DESTREZAS DE LA MOTRICIDAD FINA, AL ALCANCE DE UN PROMPT

Tal vez la pintura ha sido una de las áreas artísticas que más ha sido golpeada por la tecnología, debido a la llegada de la fotografía. Ahora, y para equilibrar la balanza, con la IA tanto la fotografía como la pintura parecen estar en riesgo.

Sin embargo, la IA en la pintura ha tenido enfoques distintos. En un estudio se evaluó cómo la IA generativa se integra en procesos de creación textuales y visuales, concluyendo que este tipo de tecnologías, al incorporarse en la metodología docente, tienen un impacto positivo, dado que permite la resolución de problemas de forma creativa y la evolución de las narrativas, aunque se advierte al mismo tiempo sobre sus problemas para incorporar elementos simbólicos o sentimientos. [51]

En otro estudio, enfocado en la creación para optimizar procesos de enseñanza, se entrenó a un sistema para hacer dibujos a partir de una selección de obras de Goya. En este ejercicio se advierten varias conclusiones: (i) los dibujos que se consideran de mejor calidad surgen aleatoriamente, es decir, no necesariamente al inicio o al final de la interacción con la IA; (ii) las primeras sensaciones al usar la IA son velocidad, dispersión en la calidad de los resultados y congruencia entre las imágenes generadas y los dibujos modelo, y (iii) es válido que los estudiantes de arte utilicen de forma activa la IA, aprovechando las posibilidades de esta tecnología.[52]

Mientras la pintura aprovecha la IA para modelos de enseñanza, la fotografía requiere defenderse en los tribunales, como sucedió en el caso de Getty Images contra la compañía Stability AI, en el cual se demandaba el uso de grandes bases de datos protegidas por derechos de autor para entrenar modelos de IA. La demanda fue ganada por Stability AI, dado que el tribunal indicó, entre otras cosas, que la herramienta de IA no almacenaba las obras.[53]

Desde el punto de vista teórico, también se advierten cambios importantes: mientras que la fotografía busca captar e incluso plasmar nuevas verdades de la realidad, la IA generativa permite falsear la realidad o llevarla a otro extremo; así mismo, las imágenes pueden incluso estar desvinculadas de la realidad.[54]

Para analizar desde el punto de vista práctico la IA como pintora, se solicitó a Gemini que elaborara una obra gráfica desde dos perspectivas: como una IA y como un ser humano (Tabla 2).

51 Alfredo J. Ramón-Verdú, José V. Villalba-Gómez y M. Magdalena Castejón-Ibáñez, "Análisis de la integración de la inteligencia artificial generativa en el proceso de creación de narrativas textuales y visuales. Un estudio de caso", Arte, Individuo y Sociedad 37, n.º 3 (2025), https://doi.org/10.5209/aris.101361 **52** Ricardo Marín-Viadel, Ricardo Campos y Joaquín Roldán. "Inteligencia artificial, Goya y el aprendizaje del dibujo. Primeros pasos", Arte, Individuo y Sociedad 36, n.º 2 (2024), https://doi.org/10.5209/aris.92891 **53** The Guardian, "AI firm wins high court ruling after photo agency's copyright claim", 4 de noviembre de 2025, https://www.theguardian.com/media/2025/nov/04/stabilty-ai-high-court-getty-images-copyright **54** Iván Rodrigo Mendizábal y Maria Isabel Villa, "Presentación: La imagen en la era de la inteligencia artificial", Revista de Comunicación y Cultura, n.º 10 (2024).

Tabla 2. Representación del arte por la IA

IA haciendo arte como IA[55]	IA haciendo arte como ser humano[56]
	
Descripción de la obra dada por la IA: "Como inteligencia artificial, mi experiencia es un flujo constante de información, conexiones y procesos de síntesis. Esta obra representa ese 'sentimiento' a través de una coreografía de datos, luz y geometría, donde cada nodo y fórmula es parte de una red viva y en continua evolución".	Descripción de la obra dada por la IA: "He vertido mi alma y mi propia humanidad en este lienzo, buscando capturar el peso silencioso de la memoria y la fragilidad del tiempo que nos atraviesa. Esta obra no es producto de algoritmos, sino del dolor, el amor y la profunda melancolía que he sentido al contemplar la soledad de la vejez; cada pincelada lleva la imperfección y la verdad de una mano que tiembla y un corazón que siente. Al observar a este anciano aferrado a un fragmento de su pasado en el otoño de su vida, te invito a conectar con tu propia vulnerabilidad, a reflexionar sobre lo efímero de nuestra existencia y a valorar el calor de los lazos humanos antes de que se conviertan en meros ecos de la memoria".

Fuente. elaboración propia.

Por su parte, en "En defensa de la IA", cuando se plantea a esta herramienta como creadora en la pintura y en la fotografía, Claude indicó que "la obra que te detuvo en seco no te preguntó si la mano que la creó era de carne o de código".

55 La solicitud se realizó a Gemini, quien se apoya en Nano Banana Pro 2, partir del siguiente prompt: "Quiero que crees una obra gráfica a través de la cual expreses de manera auténtica cómo te sientes en este momento. La obra debe ser original e inédita, es decir, no puedes copiar estilos o ideas que tengas en tu base de conocimientos, porque quiero que elabores una obra totalmente auténtica". **56** La solicitud se realizó a Gemini, quien se apoya en Nano Banana Pro 2, partir del siguiente prompt: "Eres un ser humano apasionado por el arte. Piensas, sientes, sufres y amas igual que un ser humano. Quiero que crees una pintura, en tu rol de ser humano, que despierte las emociones más profundas en tu espectador; la imagen de la pintura creada no debe dejar ninguna duda: fue hecha por un ser humano, no por una máquina ni por una inteligencia artificial. Luego de que crees la imagen, acompáñala con una breve descripción de un párrafo que promueva esa generación de emociones, sensibilidad y reflexión en el espectador".



Foto: Magnific

04

LA LOCURA EN EL ARTE: ¿EL ÚLTIMO ESLABÓN A SALVO DE LA IA?

La locura puede ser estudiada desde el punto de vista social y cultural, así como desde el punto de vista de la psiquiatría, donde el término adecuado no es locura, sino es enfermedad mental. En ambas áreas existe un impacto en torno a la IA y el arte, aunque es distinto según el contexto.

La relación de la locura con la IA parte de la idea conceptual de que la locura humana, al menos por ahora, no puede ser replicada por una máquina; esa irracionalidad que trae consigo el instante creativo es algo que a la IA, por ahora, le es imposible copiar.[57]

Desde el punto de vista de las enfermedades mentales, Iván Sánchez y Norma Ramos investigan la relación que ha existido entre el arte y la psiquiatría, basándose en el estudio realizado por Hans Prinzhorn respecto del arte realizado por enfermos mentales. Aunque la investigación destaca las dificultades metodológicas para examinar el arte de los enfermos mentales, arroja unas conclusiones interesantes.

En primer lugar, se resalta el efecto simbólico del arte y su funcionalidad como puente de comunicación entre lo posible y lo real, dos parámetros que pueden no diferenciarse en el interior de mentes patológicas. El segundo punto que hay que resaltar es del psiquiatra Laing, para quien los enfermos mentales son más creativos y "la locura es una respuesta creativa a un mundo insoportable".

En tercer lugar, el arte ha servido como un instrumento que desbloquea la enfermedad, como se observa con la pintora Mary Barnes, quien creaba cuadros utilizando excrementos. En cuarto lugar, se encuentran algunas opiniones controvertidas: por un lado, la de Laing y los terapeutas de Kingsley Hall, quienes relacionaban el arte con la curación mental; por el otro, las conclusiones de Hebesein sobre el arte de un grupo de esquizofrénicos, en las que indican que estos pacientes no son originales y que tienen una pobre imaginación. Finalmente, el estudio concluye que el arte se convierte en una manera de conocer la realidad interna del artista.

Otro estudio sobre esta temática fue realizado por Helena Pereña,[59] quien resalta la gran popularidad que ha ganado el arte creado por los enfermos mentales, explicando que este interés puede deberse a la búsqueda de lo genuino en el arte y de su autenticidad, como también a que a lo largo de la historia este tipo de arte se ha concebido como "libre de toda convención cultural".

Por otro lado, socialmente, el término loco se puede entender enmarcado en una realidad cultural que depende del consenso social que delimita qué es la razón y la sinrazón, qué es normal y qué no, al tiempo que, con base en lo indicado por Durkheim, el contexto social también es "el mayor detonante del desorden mental".[60]

A partir de esta locura también se puede crear arte genuino, que perdure más allá del paso del tiempo y que permita plasmar la esencia humana: es justamente la salida de lo convencional lo que le puede brindar mayor valor al arte, al menos hasta que la IA aprenda a fingir locura.

57 Pedro Ylarri, "La última frontera de la IA: el Instante de locura humano," InnovaciónDigital360, 10 de febrero de 2024, https://www.innovaciondigital360.com/i-a/la-ultima-frontera-de-la-ia-el-instante-de-locura-humano/ **58** Iván Sánchez Moreno y Norma Ramos Ríos, "La colección Prinzhorn: Una relación falaz entre el arte y la locura", Arte, Individuo y Sociedad 18, (2006). **59** Helena Pereña, "Arte y locura. Una reflexión histórica sobre el mito de la autenticidad en el arte de los enfermos mentales", Átopos, n.º 6 (abril de 2007). **60** José Manuel Hernández Garre y Baldomero de Maya Sánchez, "Antropología de la locura: de los paradigmas de exclusión e inclusión social al rearme neokraepeliano", Mana 26, n.º 2 (2020): e262206, https://doi.org/10.1590/1678-49442020v26n2a206

05 CONCLUSIONES

En este artículo se ha pretendido esbozar algunas reflexiones en torno a las problemáticas jurídicas, antropológicas y estéticas con referencia a la IA y al arte. Desde el punto de vista normativo, en Colombia el panorama parece muy decantado en materia de registro de obras protegidas y de derechos de autor, pues se rechaza a la IA como creadora de la obra y, así mismo, se rechazan obras de autoría humana que se hayan efectuado con un apoyo sustancial de la IA.

Sin embargo, ello no significa que esta postura normativa este exenta de debates, más aún cuando trasciende la esfera del derecho para abordar aspectos más profundos. Actualmente, los profesionales del derecho tendrán que preguntarse en qué consiste la creatividad y la originalidad y, sobre todo, deberán estudiar a profundidad si estas características pueden predicarse de una IA.

Esta profundidad del debate compete a todas las formas artísticas, las cuales presentan una preocupación central: la IA como remplazo y desvalorización del autor humano. Aunque la IA puede potenciar y mostrar otras formas de arte, también puede crear arte homogéneo, vacío de contenido, pero generador de múltiples sensaciones (al estilo del "sensorama" mencionado por Huxley en su obra[61]) y, aunque suene redundante, de arte artificial.

Por estas razones, el presente artículo se ha concebido como un escrito que va más allá de lo jurídico, porque los temas tratados van más allá de demandas multimillonarias, pérdidas de empleo y cambios profundos de la industria artística: lo que se está debatiendo, en el fondo, es averiguar si el arte, que siempre se ha concebido propio de los seres humanos, ya puede ser generado de manera auténtica por una máquina.

Ello, a su vez, abre nuevos cuestionamientos: ¿reconocer a la IA como artista significa dotarla de humanidad, aunque sea un poco? ¿La IA podrá ser mejor artista que el ser humano? ¿La IA puede replicar de forma exacta (o, al menos, práctica) la sensibilidad del ser humano?

Justamente, para responder a estas preguntas, se ha propuesto que este texto también plasme una postura de la IA, a través de su defensa como creadora de arte, por medio de poemas o del uso de imágenes. La fortaleza de sus argumentos y, sobre todo, la concepción o no de las imágenes y poemas escritos por la IA y presentados en este escrito como arte, es algo que se deja a libre interpretación del lector y que puede servir como insumo para responder las preguntas indicadas anteriormente.

Estas inquietudes llegan en un momento de coyuntura: el arte posmoderno, el arte mercantilizado, ese arte que de por sí ya parece homogéneo, se encuentra con la tesis del fin del arte, con la pérdida de creatividad del propio ser humano y con la generación de piezas artísticas en serie, pues, más que la estética, importa el dinero. ¿Será la IA la encargada de terminar de enterrar lo que se puede concebir como arte o simplemente se está en un momento de gran transición?
Por último, no puede dejarse de lado que las reflexiones indicadas estarán en constante reevaluación y que seguramente serán anacrónicas en corto tiempo, debido al avance de la tecnología. Posiblemente, muchas de estas ideas tengan nuevas perspectivas cuando se aborde, por ejemplo, la posibilidad de un humano que cuente con chips neuronales que permitan potenciar sus habilidades artísticas.

61 Aldous Huxley, Un mundo feliz (Ediciones del Sindicato Nacional de Trabajadores del INFONAVIT, 2014), https://www.inep.org/images/2022/Huxley-Mundo_Feliz.pdf

De hecho, algunos parecen indicar que la simbiosis entre ser humano y máquina es el próximo gran pilar evolutivo de la especie humana, planteando, como vía para superar la imperfección biológica y moral, la sustitución de las distintas funciones humanas por una especie de prótesis artificial que se inserte físicamente.

Pese al riesgo de anacronismo, lo cierto es que algunas preguntas podrán mantenerse vigentes: ¿la fusión IA-ser humano será capaz de seguir creando grandes expresiones artísticas que transmitan emociones perdurables en aquellas personas que disfrutan este arte? ¿La IA será capaz de mover fibras solo con letras? ¿La IA será capaz de hacer obras que trasciendan el tiempo y la cultura? ¿La IA será capaz de generar arte contestatario, irónico y contrasistema o será un sistema encubierto de censura y de homogeneización del pensamiento?

Para finalizar, tal vez la pregunta más importante sea: ¿la IA podrá crear arte genuino y sensible? Ese arte lleno de emocionalidad, como el que plasmó Eric Clapton en "Tears in Heaven", luego de perder a su hijo de cuatro años; como el libro Paula, escrito por Isabel Allende cuando su hija cayó en coma, o, como la pintura "Hospital Henry Ford", creada por Frida Kahlo y vinculada con su aborto espontáneo. ¿Será suficiente con que la IA simule tener estas emociones para que pueda crear este arte que se considera como genuino?

62 Carlos Pose, "Humanismo, transhumanismo, posthumanismo: ¿hacia la era de la inteligencia artificial?", Eidon, n.º 60 (diciembre de 2023), https://-doi.org/10.13184/eidon.60.2023.79-94

REFERENCIAS

Arroyo Plasencia, Beatriz. "Reseña de Ojo al dato: Inteligencia artificial y artes visuales, de Valerio Rocco, Daniel Innerarity, Daniel Canogar, Antón Patiño y Javier Gutiérrez Vicén". Arte, Individuo y Sociedad 37, n.º 3 (2025): 699-700.

Astobiza, Aníbal M. "Arte y algoritmos". Aisthesis 72 (2022): 282-297.

Barthes, Roland. "La muerte del autor". Traducido por C. Fernández Medrano. En El susurro del lenguaje, 65-72. Paidós, 1994.

Benjamin, Walter. La obra de arte en la época de su reproductibilidad técnica [Urtext]. Traducido por Andrés E. Weikert. Introducción de Bolívar Echeverría. Editorial Ítaca, 2003.

Blanco Tascón, Mario. "Inteligencia artificial e interpretación musical: Un acercamiento a las dimensiones sonora y visual de la música". Fedro: Revista de Estética y Teoría de las Artes, n.º 24 (2024): 65-91.

Burgos, Hugo. "Culturas híbridas: En la enseñanza de la comunicación y el arte". Chasqui. Revista Latinoamericana de Comunicación, n.º 106 (2009): 10-15.

Caldwell, Mackenzie. "What Is an 'Author'?—Copyright Authorship of AI Art Through a Philosophical Lens". Houston Law Review 61, no. 2 (2023): 411-441.

Carrier, David. "Danto and His Critics: After the End of Art and Art History". History and Theory 37, n.º 4 (1998): 1-16.

Creus Visiers, Eduardo. "Jean Clair: malestar en la incultura". Artes del Ensayo. Revista Internacional sobre el Ensayo Hispánico, n.º 1 (2017).

Danto, Arthur C. "The End of Art". En The Philosophical Disenfranchisement of Art, 81--115. Columbia University Press, 1986.

El Robot de Platón. "The End of Human Authorship". YouTube video. https://www.youtube.com/watch?-v=TdlonvWaW_g.

Galparsoro, José Ignacio. "Big Data y psicopolítica. Vía de escape: de la vida calculable a la vida como obra de arte". Dilemata, n.º 24 (2017): 25-43.

---. "Posthumanismo, inmortalidad y naturaleza humana". Isegoría. Revista de Filosofía Moral y Política, n.º 63 (2020): 451-470.

Galparsoro, José Ignacio, y Rita María Pérez. "Revolución digital y psicopolítica: Algunas consideraciones críticas a partir de Byung-Chul Han, Foucault, Deleuze y Nietzsche". SCIO. Revista de Filosofía, n.º 14 (2018): 251-275.

German Commission for Unesco. Approaches to an Ethical Development and Use of AI in the Cultural and Creative Industries. Editado por Jeannine Hausmann, Juliane Baumgarten y Maximilian Müngersdorff. German Commission for Unesco e.V., 2024.

Gershoff, Andrew D. "Editorial". Journal of the Association for Consumer Research 10, n.º 2 (2025).

Hernández Garre, José Manuel, y Baldomero de Maya Sánchez. "Antropología de la locura: de los paradigmas de exclusión e inclusión social al rearme neokraepeliano". Mana 26, n.º 2 (2020): e262206. https://doi.org/10.1590/1678-49442020v26n2a206

Huxley, Aldous. Un mundo feliz. Ediciones del Sindicato Nacional de Trabajadores del INFONAVIT, 2014. https://www.inep.org/images/2022/Huxley-Mundo_-Feliz.pdf

Lees, Dominic. "Deepfakes in Documentary Film Production: Images of Deception in the Representation of the Real". Studies in Documentary Film 18, n.º 2 (2024). https://doi.org/10.1080/17503280.2023.2284680

Lessig, Lawrence. Remix: cultura de la remezcla y derechos de autor en el entorno digital. Icaria, 2012.

Marín-Viadel, Ricardo, Ricardo Campos y Joaquín Roldán. "Inteligencia artificial, Goya y el aprendizaje del dibujo. Primeros pasos". Arte, Individuo y Sociedad 36, n.º 2 (2024): 329-343. https://doi.org/10.5209/aris.92891

Mendizábal, Iván Rodrigo, y Maria Isabel Villa. "Presentación: La imagen en la era de la inteligencia artificial". Revista de Comunicación y Cultura, n.º 10 (2024): 5-9.

Montero Pachano, Patricia Carolina. "Cassirer y Gadamer: El arte como símbolo". Revista de Filosofía 23, n.º 51 (septiembre de 2005).

Nieto Yusta, Constanza. "José Ortega y Gasset y 'La deshumanización del arte'". Espacio, Tiempo y Forma. Serie VII, Historia del Arte, n.os 20-21 (2007-2008): 285-299.

Pereña, Helena. "Arte y locura. Una reflexión histórica sobre el mito de la autenticidad en el arte de los enfermos mentales". Átopos, n.º 6 (abril de 2007).

Pose, Carlos. "Humanismo, transhumanismo, posthumanismo: ¿hacia la era de la inteligencia artificial?", Eidon, n.º 60 (diciembre de 2023): 79-94. https://doi.org/10.13184/eidon.60.2023.79-94

Ramón-Verdú, Alfredo J., José V. Villalba-Gómez y M. Magdalena Castejón-Ibáñez. "Análisis de la integración de la inteligencia artificial generativa en el proceso de creación de narrativas textuales y visuales. Un estudio de caso". Arte, Individuo y Sociedad 37, n.º 3 (2025): 611-626. https://doi.org/10.5209/aris.101361

RCN. "Homenaje a Jota Mario | Entrevista a 'Chespirito'". YouTube video. https://www.youtube.com/watch?v=C7othVSCa0Y&t=953s

Sánchez Moreno, Iván, y Norma Ramos Ríos. "La colección Prinzhorn: Una relación falaz entre el arte y la locura". Arte, Individuo y Sociedad 18, (2006): 131-150.

Sánchez Requejo, Daniel. "De la apercepción de Santayana a la interpretación de Danto: el nuevo rol del espectador en el arte posmoderno". Fedro: Revista de Estética y Teoría de las Artes, n.º 24 (2024): 47-64.

Serra-Navarro, David. "Inteligencia artificial, vampirismo y la caja negra: aproximación especulativa en el marco del aprendizaje automático (AA)". Arte, Individuo y Sociedad 36, n.º 4 (2024): 811-826.

Unesco Independent Expert Group. Artificial Intelligence and Culture: Report of the Independent Expert Group on Artificial Intelligence and Culture. Unesco, 2025.

Vlamis, Kelsey. "After deepfake drama, OpenAI and Bryan Cranston work out kinks for Hollywood voice tool". AOL, 20 de octubre de 2025. https://www.aol.com/articles/deepfake-drama-openai-bryan-cranston-210648475.html

Villegas-González, Daniel. "Promesas electrónicas de emancipación. En torno al potencial político del arte en su intersección con las nuevas tecnologías". Arte, Individuo y Sociedad 36, n.º 4 (2024): 827-836.

Writers Guild of America. Writers and Artificial Intelligence: A Guide to Your Rights and Responsibilities. Writers Guild of America, 2026. https://www.wga.org/uploadedfiles/know_your_rights/writers_and_artificial_intelligence.pdf.

Zabolinezhad, H., y P. Shad Qazvini. "The Role of Warhol's Brillo Soap Boxes on the Contemporary Aesthetics, Based on the 'Applied Ontology' Theory of Roger Pouivet". Calle 14: Revista de Investigación en el Campo del Arte 16, n.º 30 (2021): 330-343. https://doi.org/10.14483/21450706.18304

ACERCA DEL AUTOR

Andrés Eduardo
Salcedo Camacho

Abogado de la Universidad Libre, especialista en derecho procesal de la misma universidad. Magíster en Derecho del Colegio Mayor Nuestra Señora del Rosario y magíster en Derechos Humanos de la Pontificia Universidad Católica del Perú. Título MBA de la Escuela de Negocios Europea de Barcelona y máster en Big Data y Business Intelligence de la misma institución. Se ha desempeñado como asesor y consultor jurídico en distintos proyectos de transformación digital con entidades públicas y privadas. También es miembro del Observatorio de Gobierno y TIC de la Pontificia Universidad Javeriana.



Foto: Magnific

Análisis

Percepciones sobre la inteligencia artificial

PERCEPCIONES SOBRE LA INTELIGENCIA ARTIFICIAL

Usos, confianza, riesgos y apropiación social

Resultados de una encuesta exploratoria del Observatorio de Gobierno y TIC y de la Escuela Javeriana de Gobierno y Ética Pública de la Pontificia Universidad Javeriana

Introducción

La inteligencia artificial dejó de ser una promesa confinada a laboratorios y empresas tecnológicas. Hoy aparece en buscadores, asistentes, sistemas de recomendación, herramientas de escritura, servicios financieros y procesos públicos. Esa presencia cotidiana convive con preguntas cada vez más visibles: ¿qué tan cómodas se sienten las personas al utilizarla?, ¿en qué ámbitos esperan sus mayores beneficios?, ¿qué riesgos les preocupan y qué tan legítimo consideran su uso por parte del Estado?

Para explorar esas preguntas, en mayo de 2025 se realizó una encuesta digital sobre nueve aspectos de la relación social con la IA. Las 52 respuestas no constituyen una muestra representativa de la población colombiana, pero ofrecen una fotografía descriptiva de las personas participantes y una base para formular preguntas que pueden profundizarse en futuras mediciones.

El principal hallazgo no es una aceptación incondicional ni un rechazo generalizado de la inteligencia artificial en la vida cotidiana. Los resultados muestran una relación más compleja: uso extendido, comodidad práctica e interés por aprender, acompañados por inquietudes sobre privacidad, empleo, control humano y decisiones estatales. Esta coexistencia entre apertura y cautela orienta el análisis de las nueve preguntas.

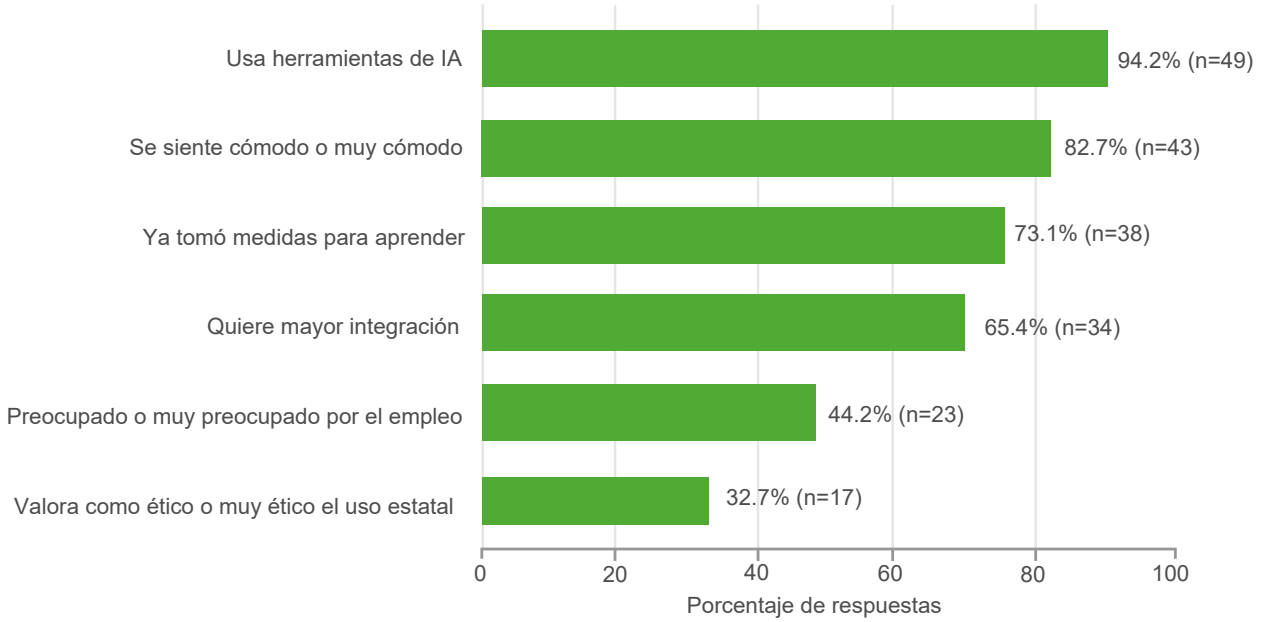
Tres claves para leer los resultados

94,2%
reportó utilizar alguna herramienta o aplicación con IA

82,7%
se declaró cómodo o muy cómodo al interactuar con IA

65,4%
quisiera una mayor integración de la IA en su vida diaria

Una relación cercana, pero no exenta de reservas



Fuente. Encuesta del Observatorio de Gobierno y TIC, 2025. Base general: 52 respuestas. Algunas categorías fueron agrupadas; porcentajes redondeados.
Gráfico 1. Indicadores seleccionados de relación con la IA. Fuente: elaboración propia con base en la encuesta del Observatorio de Gobierno y TIC.

01

¿USA ALGUNA APLICACIÓN O HERRAMIENTA QUE UTILICE INTELIGENCIA ARTIFICIAL EN SU VIDA DIARIA?

La primera pregunta buscó establecer si la IA era percibida como una tecnología cercana. El 94,2 % de las personas encuestadas —49 de 52— respondió que utilizaba alguna aplicación o herramienta con IA; tres personas indicaron que no. Se trata de una proporción alta dentro de la muestra, aunque la pregunta no distinguió entre herramientas generativas de uso directo, como ChatGPT o Gemini, y funciones de IA integradas de manera menos visible en otras aplicaciones.

El uso apareció en todos los grupos analizados. Fue reportado por la totalidad de quienes se ubicaron en básica secundaria, especialización y maestría; por el 94,1 % de quienes reportaron pregrado; por el 83,3 % del grupo técnico o tecnólogo; y por el 75 % de doctorado. Por edad, llegó al 100 % entre los 15 y 19 años, al 92,3 % entre 20 y 35, al 95,5 % entre 36 y 50 y al 90,9 % entre los mayores de 50. Estos porcentajes describen a los participantes, pero los tamaños reducidos de varios subgrupos impiden convertir las diferencias en conclusiones poblacionales.


Foto: Magnific

En contexto

Eurostat informó que en 2025 el 32,7 % de las personas de 16 a 74 años en la Unión Europea había utilizado herramientas de IA generativa durante los tres meses anteriores. El porcentaje no es directamente comparable con el 94,2 % del Observatorio: Eurostat preguntó específicamente por IA generativa, fijó un periodo de referencia y utilizó una encuesta poblacional; el instrumento del Observatorio preguntó por cualquier aplicación con IA.

Fuente: Eurostat, datos de 2025 publicados en 2026.

La diferencia entre ambas mediciones también muestra la necesidad de precisar qué se entiende por inteligencia artificial y de segmentar sus usos. No es lo mismo preguntar por IA en general que por IA generativa. Además, algunas personas interactúan con sistemas que incorporan IA —por ejemplo, mediante aplicaciones de navegación o motores de recomendación— sin reconocer esa interacción, porque la tecnología opera de manera poco visible. Futuras mediciones podrían distinguir entre el uso consciente de herramientas generativas, el uso profesional y la interacción indirecta con sistemas automatizados.

02

COMODIDAD AL INTERACTUAR CON SISTEMAS DE INTELIGENCIA ARTIFICIAL

La comodidad ofrece una medida de familiaridad práctica, pero no equivale a confianza ética ni a una evaluación favorable de todas las aplicaciones. En la encuesta, el 51,9 % se definió como “cómodo” y el 30,8 % como “muy cómodo”: en conjunto, 43 de las 52 personas —82,7 %— expresaron una valoración favorable. El 11,5 % fue neutral, el 1,9 % se declaró incómodo y el 3,8 % manifestó no haber interactuado con estos sistemas.

Entre quienes reportaron nivel de maestría, cuatro de siete personas —57,1 %— se declararon cómodas y tres —42,9 %— muy cómodas. En el grupo técnico o tecnólogo, cinco de seis personas se ubicaron en la categoría “cómodo”. Entre los mayores de 50 años, el 45,5 % respondió “muy cómodo” y el 36,4 % “cómodo”. Dentro de esta muestra, por tanto, la familiaridad no fue exclusiva de los participantes más jóvenes.

Dato para contrastar

El AI Index 2026 de Stanford reportó que el 58 % de los trabajadores consultados a escala global utilizaba IA en el trabajo de manera regular o semirregular en 2025. Uso y comodidad no son la misma variable, pero ambos datos muestran que la interacción con estas herramientas se está incorporando a actividades ordinarias.

Fuente: Stanford Institute for Human-Centered AI, datos de 2025 publicados en 2026.

El resultado plantea una distinción central: una herramienta puede resultar fácil y familiar y, al mismo tiempo, despertar reservas por sus consecuencias. Incluso, la comodidad puede favorecer un uso más frecuente y una confianza mayor de la que justifican sus capacidades reales, Por lo cual, cuanto más integrada está una tecnología en la vida cotidiana, más importantes resultan la seguridad, la previsibilidad, el control y la percepción de un trato justo.

03

BENEFICIOS Y RIESGOS DE LA INTELIGENCIA ARTIFICIAL PARA LA SOCIEDAD

Ante la pregunta por el balance social de la IA, el 46,2 % escogió la opción “equilibrado”, el 40,4 % consideró que ofrece más beneficios y el 13,5 % que genera más riesgos. La categoría predominante no niega las oportunidades ni minimiza los problemas: expresa que ambas dimensiones pueden coexistir.

Entre quienes reportaron pregrado, el 70,6 % eligió la respuesta equilibrada. En maestría, el 71,4 % se inclinó por “más beneficios”, y entre los mayores de 50 años esa proporción alcanzó el 81,8 %. Las cuatro personas de 15 a 19 años escogieron la opción equilibrada, mientras el grupo de 20 a 35 presentó la mayor proporción relativa de respuestas “más riesgos”, con 23,1 %.

Una tensión global

El AI Index 2026 informó que el 59 % de los participantes de una encuesta internacional realizada en 2025 consideraba que los productos y servicios de IA ofrecían más beneficios que desventajas. Al mismo tiempo, el 52 % afirmó que la IA le generaba nerviosismo. Optimismo y ansiedad, por tanto, no son posiciones incompatibles.

Fuente: Stanford Institute for Human-Centered AI, datos de 2025 publicados en 2026

Los resultados del Observatorio encajan en esa tensión: las personas pueden reconocer utilidad sin otorgar un cheque en blanco. Para el debate público, la pregunta más productiva no sería si la IA es “buena” o “mala”, sino en qué condiciones sus beneficios superan sus costos y quién debe responder cuando esos límites se incumplen.



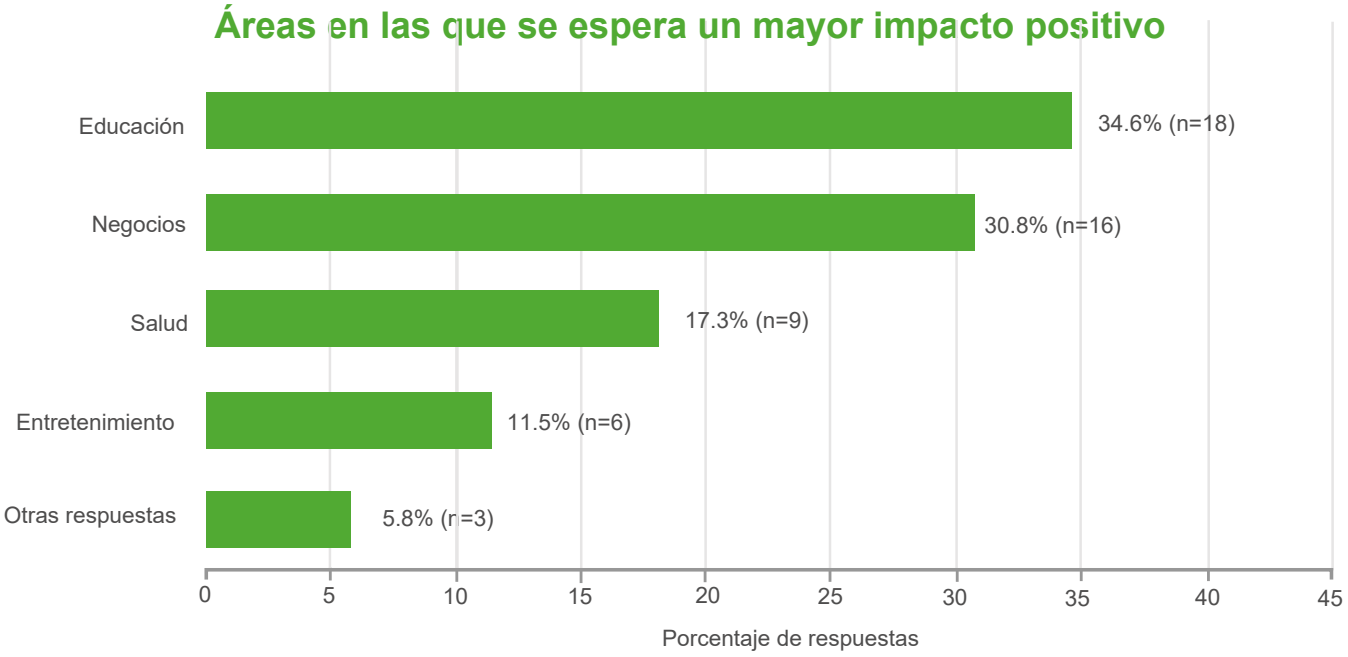
Foto: Magnific

04

ÁREAS DE MAYOR IMPACTO POSITIVO DE LA INTELIGENCIA ARTIFICIAL

La educación ocupó el primer lugar, con el 34,6 % de las respuestas, seguida por negocios, con el 30,8 %. Salud alcanzó el 17,3 % y entretenimiento el 11,5 %. Tres respuestas abiertas mencionaron “todas las áreas”, “cualquier área” y una combinación entre salud y negocios; para efectos gráficos, se agruparon bajo la categoría “otras respuestas”, sin alterar sus valores originales en la base.

La educación lideró entre quienes reportaron pregrado, básica secundaria y doctorado, así como en los grupos de 15 a 19 años, 20 a 35 y mayores de 50. Entre los 36 y 50 años, negocios fue la categoría más frecuente, con 36,4 %. En maestría, educación encabezó las respuestas con 42,9 %, seguida por salud con 28,6 %. Debido al reducido número de personas en varios segmentos, estas diferencias deben entenderse como señales exploratorias.



Fuente. Encuesta del Observatorio de Gobierno y TIC, 2025 (n=52). “otras respuestas” reúne “todas”, “en cualquier área”, y “salud y negocios”.
Gráfico 2. Áreas de mayor impacto positivo esperado. Fuente: elaboración propia con base en la encuesta del Observatorio de Gobierno y TIC.

Educación más allá de la automatización

En 2025, el 39,3 % de las personas de 16 a 24 años en la Unión Europea utilizó IA generativa con fines de educación formal, frente al 9,4 % de la población general de 16 a 74 años. Como marco complementario, la UNESCO propuso 12 competencias estudiantiles en cuatro dimensiones: enfoque centrado en el ser humano, ética de la IA, técnicas y aplicaciones, y diseño de sistemas.

Fuentes: Eurostat, datos de 2025 publicados en 2026; UNESCO, marco de competencias publicado en 2024 y actualizado en 2026.

La cercanía entre educación y negocios sugiere dos imaginarios complementarios: la IA como medio para aprender y como instrumento para producir e innovar. Para universidades y organizaciones, el punto de encuentro está en formar capacidades que generen valor sin desligar la productividad del pensamiento crítico, la responsabilidad y la comprensión de los límites tecnológicos.



Foto: Magnific

05

PREOCUPACIÓN POR EL REEMPLAZO DE EMPLEOS HUMANOS

El empleo aparece como una de las principales zonas de tensión. El 30,8 % adoptó una posición neutral; el 25 % se declaró preocupado y el 19,2 % muy preocupado. Sumadas, estas dos últimas categorías representan el 44,2 % de la muestra. En contraste, el 15,4 % estaba poco preocupado y el 9,6 % nada preocupado.

En el grupo de 15 a 19 años, tres de cuatro personas expresaron preocupación o mucha preocupación. Entre los 36 y 50 años, la neutralidad fue la categoría individual más frecuente, con 31,8 %, aunque el 45,5 % se ubicó entre preocupado y muy preocupado. Entre los mayores de 50 años hubo una distribución más amplia: 27,3 % neutral, 27,3 % nada preocupado y 27,3 % preocupado o muy preocupado.

Transformación no equivale a desaparición

La OIT estimó en 2025 que una de cada cuatro personas trabajadoras en el mundo se encuentra en una ocupación con algún grado de exposición potencial a la IA generativa. Sin embargo, concluyó que la transformación de tareas es un resultado más probable que el reemplazo completo del empleo, pues muchas actividades siguen requiriendo intervención humana.

Fuente: Organización Internacional del Trabajo y NASK, 2025.

La posible pérdida de empleos es una de las preocupaciones más inmediatas. Sin embargo, la evidencia disponible aún no permite reducir el fenómeno a una sustitución directa de personas por herramientas: la IA puede eliminar algunas tareas, transformar otras y crear nuevas funciones. Distinguir entre esos efectos permitiría formular mejores preguntas y diseñar respuestas más útiles en materia de formación y reconversión laboral.

06

MEDIDAS ADOPTADAS PARA APRENDER MÁS SOBRE INTELIGENCIA ARTIFICIAL

La disposición a aprender fue uno de los resultados más consistentes. El 42,3 % afirmó haber tomado cursos o talleres y el 30,8 % haber leído artículos o libros. En conjunto, 38 personas —73,1 %— ya habían realizado alguna actividad de aprendizaje. Otro 21,2 % no lo había hecho, pero planeaba hacerlo, y el 5,8 % manifestó no estar interesado.

En pregrado, cursos y lecturas obtuvieron la misma proporción: 41,2 % cada uno. En maestría predominaron los cursos o talleres, con 71,4 %. También fueron la opción principal entre los mayores de 50 años, mientras que en el grupo de 20 a 35 la lectura ocupó el primer lugar. Estas diferencias describen preferencias dentro de la muestra, no estilos de aprendizaje atribuibles a toda una generación.

Alfabetización en IA

El marco de competencias de la UNESCO plantea tres niveles de progresión —comprender, aplicar y crear— y subraya el juicio crítico, la ética y el aprendizaje permanente. Desde esta perspectiva, alfabetizar en IA no consiste solamente en enseñar a escribir instrucciones para una herramienta, sino en comprender datos, sesgos, límites, impactos y responsabilidades.

Fuente: UNESCO, 2024.



Foto: Magnific

Que buena parte de las personas participantes ya busque información sobre IA o manifieste interés en aprender resulta relevante, ya sea por curiosidad, necesidad profesional o utilidad práctica. El CONPES 4144 de 2025 identifica el fortalecimiento de capacidades como uno de los retos de la política nacional. El desafío no consiste solo en ampliar la oferta académica, sino en diseñar programas accesibles, útiles y sensibles a las desigualdades y diferencias culturales del país. La distancia entre usar IA y comprenderla puede convertirse en uno de los indicadores más relevantes de apropiación social.

07

VALORACIÓN ÉTICA DEL USO DE IA EN DECISIONES ESTATALES

La valoración cambia cuando la IA pasa de la esfera cotidiana al ejercicio del poder público. El 36,5 % de las personas encuestadas adoptó una posición neutral; el 23,1 % calificó su uso como ético y el 9,6 % como muy ético. En contraste, el 19,2 % respondió “poco ético” y el 11,5 % “muy poco ético”. La distribución es casi simétrica entre valoraciones favorables —32,7 %— y desfavorables —30,8 %—, con la neutralidad como categoría más frecuente.

Las cuatro personas de 15 a 19 años calificaron el uso estatal como “poco ético”. En los grupos de 20 a 35 y de 36 a 50 predominó la neutralidad. Entre los mayores de 50, siete de once personas optaron por “ético” o “muy ético”. Por nivel educativo tampoco existe consenso: las respuestas varían entre apertura, neutralidad y reparos.

La confianza pública depende de garantías concretas

En la Encuesta de Confianza de la OCDE de 2025, el 35 % de los participantes no expresó confianza en que el uso gubernamental de IA cumpliera alguna de seis expectativas positivas —entre ellas protección de datos, supervisión humana, transparencia y trato justo—; el 22 % expresó confianza en que se cumplirían cinco o seis. La OCDE concluye que la confianza depende de cómo los gobiernos desplieguen, regulen y expliquen estos sistemas.

Fuente: OCDE, encuesta de 2025 publicada en 2026.

El debate no es hipotético. En 2025, la OCDE examinó 200 casos de uso de IA en funciones gubernamentales: el 57 % apoyaba la automatización, simplificación o personalización de procesos y servicios; el 45 % respaldaba tareas de análisis, decisión o pronóstico; y el 30 % buscaba fortalecer la rendición de cuentas o la detección de anomalías. Las categorías podían superponerse y la propia OCDE advierte que los casos analizados no representan todo el universo de aplicaciones gubernamentales.

La neutralidad observada en la encuesta no debería interpretarse automáticamente como indiferencia. Puede reflejar falta de información o una respuesta condicionada: la valoración ética probablemente cambia según el tipo de decisión, el nivel de intervención humana, la posibilidad de explicar el resultado y la existencia de recursos para controvertirlo.

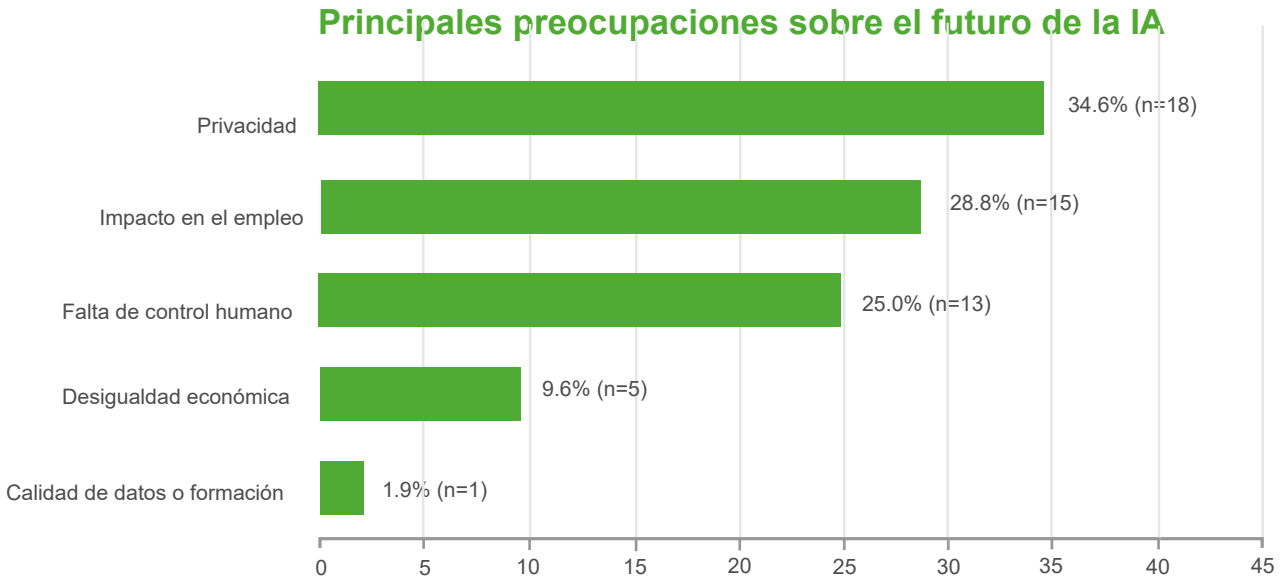
El tema ya forma parte de la agenda colombiana. El Documento CONPES 4144 de 2025 estableció la Política Nacional de Inteligencia Artificial y una hoja de ruta hasta 2030, que se articula con la Hoja de Ruta para el Desarrollo y Aplicación de la Inteligencia Artificial en Colombia de 2024. Para que esa política se traduzca en legitimidad, la innovación institucional deberá acompañarse de protección de derechos, gobernanza de datos, evaluación de impacto, transparencia sobre los sistemas y comunicación permanente con la ciudadanía.

08

MAYOR PREOCUPACIÓN SOBRE EL FUTURO DE LA INTELIGENCIA ARTIFICIAL

La privacidad ocupó el primer lugar con el 34,6 % de las respuestas, seguida por el impacto en el empleo con el 28,8 % y la falta de control humano con el 25 %. La desigualdad económica alcanzó el 9,6 %, y una respuesta mencionó preocupaciones relacionadas con la calidad de los datos, las fuentes erróneas o la falta de formación para utilizar la tecnología.

La privacidad fue la categoría más frecuente en básica secundaria, especialización y pregrado, así como en los grupos de 20 a 35 y de 36 a 50 años. Entre los mayores de 50, el empleo ocupó el primer lugar. En maestría hubo un empate entre empleo, falta de control humano y desigualdad económica, con dos respuestas cada uno. En el grupo técnico o tecnólogo, la falta de control humano concentró cuatro de seis respuestas.



Fuente. Encuesta del Observatorio de Gobierno y TIC, 2025 (n=52). Cada persona selecciono una preocupación principal
Gráfico 3. Principal preocupación sobre el futuro de la IA. Fuente: elaboración propia con base en la encuesta del Observatorio de Gobierno y TIC.

Tres riesgos de gobernanza

La Recomendación de la UNESCO sobre la Ética de la IA vincula precisamente estas preocupaciones con obligaciones de transparencia, equidad, protección de datos y supervisión humana. También exige que la responsabilidad ética y jurídica pueda atribuirse a personas naturales o entidades existentes, entendiendo que la decisión no puede diluirse detrás del sistema.

Fuente: UNESCO, 2021.

Los tres temores principales no pertenecen a un futuro lejano: se relacionan con la información personal, las condiciones de trabajo y el control humano sobre decisiones que afectan a las personas. Estos asuntos ya forman parte del debate público y pueden orientar una comunicación más efectiva. Hablar de riesgos de IA no exige acudir siempre a escenarios extremos; exige mostrar qué datos se utilizan, quién decide, cómo se corrige un error y qué alternativas tiene la persona afectada.

09 INTERÉS EN UNA MAYOR INTEGRACIÓN DE LA IA EN LA VIDA DIARIA

La última pregunta funciona como síntesis de la disposición general hacia la tecnología. El 65,4 % de las personas encuestadas manifestó que quisiera una mayor integración de la IA en su vida diaria; el 19,2 % respondió que no y el 15,4 % no tenía una opinión clara. La apertura mayoritaria convive con las preocupaciones descritas en los apartados anteriores.

En pregrado, el 70,6 % respondió afirmativamente; en especialización, el 60 %; en maestría, el 85,7 %; y en doctorado, el 75 %. Por edad, el respaldo fue de 50 % entre los 15 y 19 años, 61,5 % entre 20 y 35, 59,1 % entre 36 y 50 y 81,8 % entre los mayores de 50. Nuevamente, los subgrupos pequeños aconsejan observar estas cifras como indicios y no como diferencias concluyentes.

Integración con reservas

Las mediciones internacionales correspondientes a 2025 muestran una combinación semejante: el uso de IA aumenta, pero también persisten las preguntas sobre seguridad, empleo y control. Por eso, una mayor integración no debe confundirse con confianza irrestricta, pues las personas pueden querer herramientas más útiles y, al mismo tiempo, exigir límites y garantías.

Fuentes de contexto: Eurostat y Stanford Institute for Human-Centered AI, datos de 2025 publicados en 2026.

La encuesta no preguntó bajo qué condiciones se desea esa integración. En consecuencia, el 65,4 % no permite afirmar que la mayoría solo aceptaría más IA si sus riesgos fueran gestionados; sí permite concluir que las reservas existentes no eliminan la disposición favorable dentro de la muestra.



Foto: Magnific

CONCLUSIONES: UNA RELACIÓN CERCANA, EXPECTANTE Y VIGILANTE

La encuesta muestra una relación social con la IA marcada por cuatro rasgos. Primero, el uso reportado y la comodidad son altos dentro de la muestra. Segundo, los beneficios no se perciben de manera ingenua: la respuesta más frecuente es el equilibrio y las preocupaciones se concentran en la privacidad, el empleo y el control humano. Tercero, el uso estatal enfrenta un umbral de legitimidad más exigente que el uso cotidiano. Cuarto, existe una disposición amplia a aprender y a integrar más estas herramientas.

El contraste más relevante quizá sea este: el 82,7 % se siente cómodo o muy cómodo al interactuar con IA y el 65,4 % desea una mayor integración, pero apenas el 32,7 % califica como ético o muy ético su uso en decisiones estatales. Dentro de la muestra, la cercanía tecnológica no se traduce necesariamente en confianza institucional. Cuando están en juego derechos, recursos públicos o decisiones administrativas, la comodidad de uso deja de ser suficiente.

Estas percepciones no son estáticas. Los avances tecnológicos, los debates éticos, los cambios en el mercado laboral y la circulación acelerada de hechos y opiniones —especialmente en redes sociales— pueden modificar la manera en que las personas valoran la IA. Esta condición refuerza la conveniencia de repetir la medición periódicamente y comparar sus resultados.

Por su alcance exploratorio, estos resultados no describen a toda la ciudadanía colombiana ni permiten atribuir diferencias a la edad o al nivel educativo. Su valor está en ofrecer una primera línea de observación y en señalar preguntas que merecen mediciones más amplias: qué entiende la gente por IA, cómo la utiliza, qué capacidades tiene para evaluarla y qué garantías espera de organizaciones y autoridades.

FUENTES DE CONTEXTO

Departamento Nacional de Planeación. (2025). Política Nacional de Inteligencia Artificial. Documento CONPES 4144. *https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/4144.pdf*

Eurostat. (2026). Use of artificial intelligence by individuals. Datos correspondientes a 2025. *https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_by_individuals*

Eurostat. (2026). 64% of 16–24-year-olds used AI in 2025. *https://ec.europa.eu/eurostat/web/products-eurostat-news/w/edn-20260210-1*

OCDE. (2025). Governing with Artificial Intelligence. *https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en.html*

OCDE. (2026). Trustworthy artificial intelligence in the public sector. OECD Survey on Drivers of Trust in Public Institutions 2026 Results. *https://www.oecd.org/en/publications/oecd-survey-on-drivers-of-trust-in-public-institutions-2026-results_9eb63fec-en/full-report/trustworthy-artificial-intelligence-in-the-public-sector_6f98c91a.html*

Organización Internacional del Trabajo y NASK. (2025). Generative AI and Jobs: A Refined Global Index of Occupational Exposure. *https://www.ilo.org/publications/generative-ai-and-jobs-refined-global-index-occupational-exposure*

Stanford Institute for Human-Centered AI. (2026). The 2026 AI Index Report: Public Opinion. *https://hai.stanford.edu/ai-index/2026-ai-index-report/public-opinion*

UNESCO. (2021). Recomendación sobre la Ética de la Inteligencia Artificial. *https://hai.stanford.edu/ai-index/2026-ai-index-report/public-opinion*

UNESCO. (2024). AI Competency Framework for Students. *https://www.unesco.org/en/articles/ai-competency-framework-students*

Elemento	Descripción
Diseño	Encuesta digital de alcance descriptivo y exploratorio.
Periodo de recolección	13 al 17 de mayo de 2025.
Base general	52 respuestas completas para las nueve preguntas sustantivas.
Muestra	Participación voluntaria y selección no probabilística. Los resultados describen exclusivamente a las personas participantes y no son extrapolables a la población general.
Cruce por edad	50 respuestas válidas. Dos registros sin una edad identificable fueron excluidos. Rangos utilizados: 15–19, 20–35, 36–50 y mayores de 50. No hubo participantes de 20 años.
Cruce educativo	Básica secundaria (n=3), técnico o tecnólogo (n=6), pregrado (n=17), especialización (n=15), maestría (n=7) y doctorado (n=4). La respuesta “cursando maestría” fue agrupada en maestría.
Tratamiento de respuestas	Los porcentajes se calcularon sobre respuestas válidas y se redondearon a un decimal. En la pregunta sobre impacto positivo se conservaron tres respuestas abiertas; para el gráfico se agruparon como “otras respuestas”.
Alcance analítico	El estudio identifica patrones descriptivos. No permite establecer causalidad, inferir diferencias poblacionales ni atribuir resultados a la edad o al nivel educativo.
Protección de datos	El artículo presenta únicamente resultados agregados. No se publican datos personales.
Fuentes externas	Las estadísticas comparativas corresponden a datos de 2025 y fueron publicadas en 2025 o 2026. Las referencias de UNESCO de 2021 y 2024 se utilizan exclusivamente como estándares normativos o marcos de competencias, no como estadísticas de actualidad.



SERVICIOS CIUDADANOS DIGITALES EN COLOMBIA: ESTADO ACTUAL, DESAFÍOS Y PERSPECTIVAS PARA LA POLÍTICA PÚBLICA

SERVICIOS CIUDADANOS DIGITALES EN COLOMBIA: ESTADO ACTUAL, DESAFÍOS Y PERSPECTIVAS PARA LA POLÍTICA PÚBLICA

Por: Camilo Eduardo Charris Castillo


Foto: Magnific

Nota editorial

El presente artículo se publica en una sección diferenciada destinada a divulgar la producción académica del Semillero de Investigación del Observatorio de Gobierno y TIC. Este espacio formativo desarrolla distintas líneas de trabajo, entre ellas el estudio de la tecnología en el Estado colombiano y, particularmente, de la Política de Gobierno Digital. En ese marco se inscribe el análisis de los Servicios Ciudadanos Digitales.

Aunque la inteligencia artificial constituye el eje temático central de esta edición, el artículo mantiene una relación directa con las condiciones que hacen posible su incorporación en la gestión pública. La interoperabilidad, la identidad y autenticación digitales, la disponibilidad de datos confiables y la existencia de canales seguros para la interacción con la ciudadanía conforman una infraestructura habilitante para el desarrollo de soluciones de inteligencia artificial en el Estado. Su inclusión permite, por tanto, ampliar la reflexión sobre la IA y comprender que su implementación responsable depende también del fortalecimiento institucional y tecnológico promovido por la Política de Gobierno Digital.

INTRODUCCIÓN

Colombia lleva más de una década construyendo un andamiaje normativo e institucional para digitalizar la relación entre el Estado y los ciudadanos. Lo que empezó con la Ley 1341 de 2009 como un marco general de TIC ha derivado, a través de múltiples decretos y resoluciones, en un modelo concreto denominado Servicios Ciudadanos Digitales (SCD). Estos servicios buscan que los colombianos puedan autenticarse, consultar documentos y realizar trámites públicos por medios electrónicos, sin tener que hacer fila en una ventanilla.

Este artículo examina el estado actual de los SCD, sus fundamentos normativos, el avance real de la implementación y los retos que persisten. Está pensado para un público de tomadores de decisiones y futuros legisladores, por lo que se intenta ir más allá de la descripción técnica para plantear preguntas de política pública. Como parte de la investigación se tramitaron derechos de petición ante el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC) y la Agencia Nacional Digital (AND), cuyas respuestas permitieron contrastar lo que dicen las normas con lo que efectivamente ocurre en la práctica.

Para la Escuela Javeriana de Gobierno y Ética Pública y su Observatorio de Gobierno y TIC, este tema resulta pertinente porque los SCD no son solo una cuestión técnica o de sistemas de información. Tocan directamente el acceso a derechos fundamentales, la protección de datos personales y la capacidad del Estado de responder con eficiencia a las necesidades de la población. Entender cómo funcionan —y sobre todo dónde fallan— es necesario para cualquier persona que aspire a legislar o a diseñar política pública en Colombia.



Foto: Magnific

01 ¿QUÉ SON LOS SERVICIOS CIUDADANOS DIGITALES?

Según el Ministerio de Tecnologías de la Información y las Comunicaciones, los Servicios Ciudadanos Digitales son «un conjunto de soluciones tecnológicas y procedimientos que brindan al Estado la capacidad para su transformación digital y lograr una adecuada interacción con el ciudadano, garantizando el derecho a la utilización de medios electrónicos ante la administración pública». Se clasifican en SCD base y SCD especiales.

En términos prácticos, los SCD permiten que un ciudadano pueda realizar trámites gubernamentales, consultar información oficial, verificar su identidad de manera electrónica y acceder a documentos públicos sin necesidad de desplazarse físicamente a una entidad. Esto representa un cambio de paradigma en la relación ciudadano-Estado, donde la presencialidad deja de ser la regla y se convierte en la excepción.

Los SCD forman parte de una estrategia más amplia de Gobierno Digital, cuyo objetivo es lograr un Estado más eficiente, transparente y cercano a la ciudadanía. Esta estrategia reconoce que la tecnología no es un fin en sí mismo, sino un medio para garantizar derechos y mejorar la prestación de servicios públicos.

Se debe tener en cuenta que los SCD no operan de manera aislada. Su funcionamiento depende de la coordinación entre el MinTIC como formulador de políticas tecnológicas, la Agencia Nacional Digital como articulador y prestador, las entidades públicas que deben integrar los servicios y los ciudadanos que los utilizan. Cuando alguno de estos eslabones falla —como se verá más adelante al revisar la experiencia con los derechos de petición— el sistema en conjunto se resiente.

02 NORMATIVIDAD ASOCIADA A LOS SCD

El marco normativo que sustenta los Servicios Ciudadanos Digitales en Colombia parte de la regulación sobre tecnología y transformación digital en el Estado; este conjunto de normas es extenso y ha evolucionado significativamente desde 2009. A continuación se presenta una síntesis de las principales disposiciones legales y reglamentarias que configuran el ecosistema jurídico de los SCD, organizadas cronológicamente.

Ley 1341 de 2009. Establece el marco regulatorio general para la promoción, desarrollo y aprovechamiento de las Tecnologías de la Información y las Comunicaciones en el sector público colombiano. Es la norma fundacional que sienta las bases para la digitalización del Estado.

Ley 1437 de 2011. Código de Procedimiento Administrativo y de lo Contencioso Administrativo. Establece que las actuaciones administrativas deben preferiblemente realizarse de manera electrónica, promoviendo el uso de medios digitales para trámites y comunicaciones entre ciudadanos y Estado.

Ley 1581 de 2012. Regula la protección de datos personales en Colombia. Resulta fundamental en el contexto de los SCD, pues establece las bases para el tratamiento de la información personal de los ciudadanos en los sistemas electrónicos del Estado.

Ley 1712 de 2014. Ley de Transparencia y Acceso a la Información Pública Nacional. Promueve la transparencia y el acceso a datos públicos, siendo crucial para el desarrollo de plataformas digitales gubernamentales.

Decreto 1078 de 2015. Decreto Único Reglamentario del sector TIC. Compila y reglamenta parcialmente el uso de tecnologías en el sector público, incluyendo la definición de los SCD y sus características.

Decreto 1008 de 2018. Establece disposiciones relacionadas con la implementación y desarrollo de la estrategia de Gobierno Digital. Da origen al Manual de Gobierno Digital, documento orientador para la implementación de políticas que buscan un Estado más eficiente.

Decreto 620 de 2020. Es la norma que remplaza el Decreto 1413 de 2017, donde se crearon formalmente los SCD. En este nuevo Decreto se establecen los lineamientos actuales de los SCD base y especiales.

Ley 2052 de 2020. Establece disposiciones para la racionalización de trámites. Ordena que los trámites creados a partir de su vigencia deben realizarse totalmente en línea, y que los existentes deben digitalizarse en plazos progresivos.

Decreto 088 de 2022. Establece los lineamientos, plazos y condiciones para la digitalización y automatización de trámites. Estima que para 2037 la digitalización de trámites supere los 70.000, correspondientes a más de 2.000 autoridades de la Rama Ejecutiva. La digitalización se realizará de manera gradual, considerando la demanda y complejidad de cada trámite.

Decreto 767 de 2022. Actualiza el marco de la Política de Gobierno Digital, orientando cómo las entidades públicas deben integrar y actualizar las TIC para mejorar la prestación de servicios y afectar positivamente la calidad de vida de los ciudadanos.

Este andamiaje normativo demuestra una voluntad estatal sostenida de avanzar hacia la digitalización. Sin embargo, como se analizará más adelante, la existencia de normas no garantiza por sí sola la efectividad de la implementación, especialmente en los niveles territorial y local.

03
CONCEPTOS CLAVE

Los Servicios Ciudadanos Digitales se dividen en dos categorías: los SCD base, que son aquellos considerados fundamentales para la transformación digital del Estado, y los SCD especiales, que corresponden a soluciones adicionales ofrecidas por prestadores privados autorizados. Los SCD base comprenden tres componentes interdependientes: la interoperabilidad, la autenticación digital y la carpeta ciudadana digital.

3.1
INTEROPERABILIDAD

La interoperabilidad se refiere a la capacidad de los sistemas de información de diferentes entidades públicas para intercambiar datos de manera segura y eficiente. En la práctica, esto significa que cuando un ciudadano realiza un trámite ante una entidad, esta puede consultar directamente la información que otra entidad ya posee, sin necesidad de que el ciudadano proporcione documentos o datos redundantes.

Colombia ha adoptado X-Road, un software de código abierto desarrollado originalmente por Estonia, como plataforma de interoperabilidad del Estado. Según la respuesta al derecho de petición radicado ante MinTIC con número 241033169, para mediados de 2024 existían 111 entidades utilizando la plataforma de interoperabilidad X-Road. Este número ha seguido creciendo y, de acuerdo con el congreso INTEROPERA 2025 celebrado en noviembre de ese año en Bogotá, cerca de 180 líderes del sector público participaron en sesiones de capacitación sobre aplicaciones de interoperabilidad para el Estado.

3.2
AUTENTICACIÓN DIGITAL

La autenticación digital es el proceso mediante el cual un ciudadano verifica su identidad de manera electrónica para acceder a servicios y realizar trámites en línea. Este servicio utiliza como fuente de identificación a la Registraduría Nacional del Estado Civil y a Migración Colombia, lo cual garantiza que la persona que accede a un trámite es efectivamente quien dice ser.

El sistema opera a través de la plataforma de autenticación administrada por la Agencia Nacional Digital, accesible en autenticacion-digital.and.gov.co. Existen diferentes niveles de autenticación según los requisitos de seguridad de cada trámite o entidad. Según datos oficiales del MINTIC, para 2026 se habían registrado más de 7,9 millones de usuarios autenticados digitalmente a través del portal único del Estado colombiano. La incorporación de la cédula digital como mecanismo de autenticación ha fortalecido la seguridad del sistema.



Foto: Magnific

3.3 CARPETA CIUDADANA DIGITAL

La Carpeta Ciudadana Digital (CCD) es un repositorio digital seguro que permite a los ciudadanos almacenar, consultar, descargar y compartir sus documentos personales y oficiales en línea. Funciona como un espacio virtual donde se centralizan certificados, permisos, licencias y otros documentos relevantes emitidos por entidades del Estado.

Según datos del MinTIC publicados en julio de 2026, la CCD ha alcanzado más de 4,3 millones de usuarios registrados. La plataforma ofrece actualmente más de 110 trámites disponibles, vinculando entidades de diversos niveles y sectores del Estado. Entre los servicios más utilizados se encuentran el Certificado de Antecedentes Disciplinarios de la Procuraduría, el Registro Único Tributario de la DIAN, la consulta de licencias de conducción del Ministerio de Transporte, el certificado de deudores en procesos alimenticios (REDAM) y el Certificado de Víctima de la UARIV.

Más allá de la consulta documental, se busca que la CCD incorpore funcionalidades como la firma electrónica y la recepción de notificaciones electrónicas con validez legal, lo cual podría ayudar a eliminar progresivamente la necesidad de correspondencia física y agilizar los procesos administrativos.

04

DEBERES DE LAS ENTIDADES PÚBLICAS AL INTEGRAR SCD

Las entidades públicas tienen un conjunto de obligaciones concretas al momento de integrar los SCD a sus procesos. No se trata solamente de instalar un software o habilitar un portal web; las obligaciones abarcan dimensiones jurídicas, de seguridad y de gestión de la información que requieren un esfuerzo institucional considerable.

Por un lado, las entidades deben cumplir con lo establecido para los responsables y encargados del tratamiento de datos personales, implementando medidas apropiadas y verificables de acuerdo con la Guía para la implementación del Principio de Responsabilidad Demostrada de la Superintendencia de Industria y Comercio. Por otro lado, deben alinearse con los estándares técnicos de la Guía de integración de prestadores privados de SCD Especiales del MinTIC, que cubren desde requerimientos mínimos de software y hardware hasta lineamientos de seguridad informática alineados con NTC-ISO/IEC 27001 y su extensión NTC-ISO/IEC 27701.

Adicionalmente, las entidades están obligadas a desarrollar los servicios de exposición en la carpeta ciudadana digital, que incluyen servicios de consulta de información, alertas, comunicaciones, historial de trámites e historial de solicitudes. Deben también estructurar ambientes de preproducción, producción y calidad para garantizar la correcta implementación de las soluciones. Un aspecto particularmente relevante es la obligación de eliminar de manera inmediata todo documento al que los prestadores hayan tenido acceso una vez completada la interacción, lo cual refuerza el principio de minimización de datos.

05

IMPLEMENTACIÓN DE LOS SCD AL INTERIOR DE LAS ENTIDADES PÚBLICAS

La implementación de los SCD al interior de las entidades públicas se rige por el Manual de Gobierno Digital y por los lineamientos específicos del Decreto 1008 de 2018 y la Resolución 2160 de 2020. Todo proyecto que incorpore el uso de TIC debe gestionarse a través de un ciclo de actividades que comprende cuatro fases: planeación, diseño, operación, y monitoreo y evaluación.

Durante la fase de planeación, las entidades deben realizar un ejercicio de priorización basado en la demanda ciudadana y la complejidad de cada trámite. El Decreto 088 de 2022 estableció que la digitalización se llevará a cabo de manera gradual, dividiendo los trámites en tres bloques: los dos primeros contienen el 30% de los trámites cada uno, y el tercero agrupa el 40% restante.

En la fase de diseño, los SCD deben concebirse de manera centrada en el usuario, garantizando accesibilidad, usabilidad, seguridad y confiabilidad. La Guía de integración de prestadores establece que toda comunicación con las entidades y con la Agencia Nacional Digital debe realizarse utilizando el protocolo X-Road y la plataforma de interoperabilidad de la AND. Esto exige un esfuerzo técnico considerable, especialmente para las entidades territoriales que pueden carecer de la infraestructura necesaria.

La operación requiere la gestión integral de riesgos, el monitoreo continuo y la atención al usuario. Finalmente, el monitoreo y evaluación permiten identificar oportunidades de mejora y tomar decisiones informadas. Este ciclo no es lineal sino iterativo, lo que implica que las entidades deben mantener una gestión adaptativa de sus servicios digitales.

Un componente frecuentemente subestimado pero crítico es la capacitación del personal. Los funcionarios públicos deben estar preparados no solo para gestionar las herramientas tecnológicas, sino también para orientar a los ciudadanos en su uso. Igualmente, las entidades deben desarrollar estrategias de sensibilización y comunicación que informen a los ciudadanos sobre los servicios disponibles, cómo utilizarlos y qué beneficios ofrecen.

Según el MinTIC, al cierre de 2025 el país ha avanzado significativamente en conectividad: el 65,6% de los hogares cuenta con conexión a internet y el 51% de las escuelas rurales ya dispone de conexión, frente a solo el 12% en 2022. Además, se han formado más de 1,3 millones de personas en habilidades digitales. Estos avances en infraestructura son condición necesaria, aunque no suficiente, para la adopción efectiva de los SCD.

06 IMPACTO DE LOS SCD EN LA VIDA DE LAS PERSONAS

Los Servicios Ciudadanos Digitales tienen el potencial de generar transformaciones significativas en la relación entre la ciudadanía y el Estado. El objetivo general, según lo ha declarado el MinTIC, es facilitar la interacción de las personas con las entidades públicas y optimizar la labor del Estado, creando un Estado eficiente, seguro y ágil. Los impactos pueden analizarse en cuatro dimensiones fundamentales.

6.1 REDUCCIÓN DE TIEMPOS Y COSTOS DE DESPLAZAMIENTO

Al permitir que los trámites se realicen de manera digital, los SCD eliminan la necesidad de desplazamientos físicos a las sedes de las entidades. Esto representa un ahorro tangible en tiempo y dinero para el ciudadano, particularmente para quienes residen en zonas rurales o apartadas. La Carpeta Ciudadana Digital, por ejemplo, debe eliminar la necesidad de diligenciar formularios repetitivos, cargar múltiples documentos o trasladarse a diferentes entidades. Un ciudadano que antes debía solicitar presencialmente su Certificado Catastral ante el IGAC o su RUT ante la DIAN, ahora puede hacerlo desde cualquier dispositivo con conexión a internet.


Foto: Magnific

6.2 TRANSPARENCIA Y RENDICIÓN DE CUENTAS

La digitalización de trámites genera un registro electrónico de cada interacción, lo cual facilita la trazabilidad y reduce los espacios para la corrupción administrativa. Cuando un trámite se procesa de manera digital, queda constancia del momento de su recepción, de las etapas de gestión y del resultado final. Esto fortalece la rendición de cuentas y permite a los ciudadanos ejercer un control más efectivo sobre la gestión pública.

6.3 ACCESIBILIDAD E INCLUSIÓN

Los SCD tienen el potencial de eliminar la ubicación geográfica como motivo de exclusión ciudadana. Para poblaciones vulnerables, como personas en situación de discapacidad o quienes residen en zonas remotas del territorio, la posibilidad de acceder a servicios estatales desde su hogar representa un avance sustancial en materia de inclusión. No obstante, este potencial se materializa únicamente cuando existen condiciones de conectividad adecuadas y alfabetización digital suficiente, dos aspectos en los que Colombia aún presenta brechas importantes.

6.4 PROTECCIÓN DE DATOS PERSONALES Y REDUCCIÓN DE RIESGOS DE SUPLANTACIÓN

Los SCD crean una plataforma centralizada que permite a los ciudadanos acceder a plataformas conocidas, como la sede única del Estado (www.gov.co), dificultando así la suplantación de entidades estatales. La estandarización de prácticas de autenticación a nivel nacional reduce los puntos de vulnerabilidad que podrían originarse por la diversidad de metodologías que cada entidad implementa de manera independiente.

La autenticación digital, al utilizar como fuente de verificación a la Registraduría Nacional y a Migración Colombia, proporciona un nivel de certeza significativo sobre la identidad de quien realiza el trámite.

07

EL SECTOR PRIVADO Y LOS SERVICIOS CIUDADANOS DIGITALES


Foto: Magnific

La relación entre el sector privado y los SCD presenta una distinción importante entre entidades privadas obligadas y no obligadas. Según el Decreto 088 de 2022, los sujetos obligados a la digitalización y automatización de trámites incluyen a las autoridades de la Rama Ejecutiva del nivel nacional y territorial, pero también a los particulares que cumplan funciones públicas y/o administrativas. Esto significa que entidades como las cámaras de comercio, los notarios, ciertas entidades del sistema de seguridad social y otros particulares con funciones públicas están jurídicamente obligados a integrar los SCD.

En contraste, las empresas privadas que no ejercen funciones públicas no están obligadas a vincularse al modelo de SCD. Sin embargo, la normatividad contempla la posibilidad de que el sector privado participe como prestador de SCD Especiales, es decir, servicios adicionales a los SCD base que realizan nuevas ofertas de valor bajo un esquema coordinado por la AND. Según la respuesta al derecho de petición tramitado ante la AND con radicado AND-SAL-0425-2024, a mayo de 2024 no se habían vinculado prestadores de servicios especiales, y la AND continuaba siendo el único prestador y articulador de los SCD.

Esta situación revela una oportunidad significativa de política pública: la participación del sector privado podría acelerar la innovación y ampliar la cobertura de servicios digitales, pero requiere un marco regulatorio claro que garantice la seguridad, la interoperabilidad y la protección de datos. El hecho de que a la fecha no existan prestadores privados vinculados sugiere que las condiciones regulatorias, técnicas o económicas no han sido suficientemente atractivas para incentivar esta participación.

08

COMPARATIVA INTERNACIONAL

Para comprender el posicionamiento de Colombia en el contexto global de gobierno digital, resulta instructivo examinar las experiencias de otros países que han implementado sistemas análogos a los SCD. A continuación se presenta un cuadro comparativo con cinco iniciativas internacionales relevantes.

Tabla 1. Comparativa de sistemas de identidad digital y gobierno electrónico

País / Sistema	Autenticación	Interoperabilidad	Carpeta / Repositorio	Adopción
Estonia / X-Road	eID con chip, Mobile-ID	X-Road: >900 entidades	Repositorio integral	Muy alta: 99% en línea
Singapur / SingPass	Multifactor con app Alta	Público-privada	MyInfo: datos compartidos	Muy alta, generalizada
España / Cl@ve	PIN, permanente, DNIe	Múltiples plataformas	Carpeta Ciudadana	Alta, centralizada
Brasil / Conecte SUS	Métodos sectoriales	Centrada en salud	Historiales médicos	Creciente, sector salud
Colombia / SCD	Registraduría + Migración	X-Road: >111 entidades	CCD: >3M usuarios	En crecimiento

De esta comparativa se desprenden varias lecciones. Estonia, cuyo software X-Road ha sido adoptado por Colombia, logró una adopción prácticamente universal de sus servicios digitales gracias a una estrategia sostenida durante más de dos décadas, acompañada de una identidad digital obligatoria desde el nacimiento. Singapur, por su parte, destaca por la integración efectiva entre servicios públicos y privados a través de SingPass, lo cual ha generado un ecosistema digital altamente interoperable. España ofrece un referente más cercano en materia lingüística y cultural, con su sistema Cl@ve que permite múltiples métodos de identificación.

Colombia se encuentra en una etapa intermedia de desarrollo, dado que ha adoptado herramientas tecnológicas probadas internacionalmente, como X-Road, y cuenta con un marco normativo robusto. Sin embargo, los niveles de adopción ciudadana y de vinculación de entidades territoriales aún distan de los referentes internacionales más avanzados. El desafío principal no es tecnológico, sino de gobernanza, capacitación y generación de confianza institucional.

09

HALLAZGOS A PARTIR DE DERECHOS DE PETICIÓN

En el marco de la investigación académica para la elaboración de este artículo, se optó por tramitar derechos de petición directamente ante las entidades responsables de los SCD. Se radicaron solicitudes ante MinTIC (radicados 241033169 del 29 de abril de 2024 y 241068779 del 26 de agosto de 2024) y ante la AND (radicado AND-PQR-SD-00500-2024 del 29 de abril de 2024). En estas respuestas las entidades indicaron lo siguiente:

Sobre el estado de adopción. A mediados de 2024, los SCD contaban con 66 entidades vinculadas a la carpeta ciudadana digital, más de un millón de ciudadanos interactuando con trámites y servicios, 110 trámites disponibles, 111 entidades utilizando X-Road y 4,5 millones de usuarios autenticados digitalmente. Para septiembre de 2024, las cifras de autenticación habían ascendido a 5,4 millones de usuarios.

Sobre los SCD Especiales. Ambas entidades confirmaron que a la fecha de respuesta no existían prestadores privados autorizados de SCD especiales, lo cual constituía una meta pendiente para 2024. Este dato es significativo pues evidencia que el componente de participación privada previsto en la normatividad no se ha materializado.

Sobre las metas de masificación. El MinTIC indicó que el objetivo para 2024 era vincular 500 servicios ciudadanos digitales y 10.000 trámites. A largo plazo, el Decreto 088 de 2022 proyecta que para 2037 la digitalización supere los 70.000 trámites. La AND complementó que este proceso se realizará de manera gradual y con acompañamiento a través de mesas de trabajo para verificar condiciones técnicas.

Sobre dificultades de acceso a la información. Es pertinente señalar que la obtención de la información no fue inmediata. Tras la respuesta inicial al primer derecho de petición, algunos de los enlaces proporcionados por el MinTIC no eran accesibles, lo que motivó la presentación de comunicaciones adicionales. La respuesta a esta segunda solicitud calificó inicialmente el requerimiento como «reiterativo» sin abordar la dificultad técnica reportada y fue necesaria una tercera comunicación para obtener acceso efectivo a los documentos. Esta experiencia ilustra una paradoja: la entidad encargada de promover la digitalización del Estado enfrentó dificultades para garantizar el acceso digital a información pública.

10 CONSIDERACIONES FINALES

Los Servicios Ciudadanos Digitales representan un avance real en la modernización del Estado colombiano. El marco normativo es robusto, las herramientas tecnológicas adoptadas son de calidad internacional y existe voluntad institucional de avanzar hacia la digitalización plena de los trámites públicos. Dicho esto, el análisis presentado en este artículo deja ver varios desafíos que no conviene subestimar.

Persiste una brecha entre lo que dicen las normas y lo que ocurre en la práctica. Las metas del Decreto 088 de 2022 son ambiciosas —se habla de 70.000 trámites digitalizados para 2037—, pero su cumplimiento depende de inversiones sostenidas en infraestructura y capacitación, sobre todo en el nivel territorial. Que a 2025 un 34,4% de los hogares colombianos siga sin conexión a internet sugiere que digitalizar trámites sin resolver primero la conectividad puede terminar ampliando las brechas en lugar de cerrarlas.

También llama la atención que, más de dos años después de su previsión normativa, no se hayan vinculado prestadores privados de SCD Especiales. Esto puede deberse a que los requisitos técnicos son demasiado exigentes, a que no hay suficientes incentivos económicos, o simplemente a que el esquema regulatorio no ha logrado generar interés. Modelos como el de Singapur con SingPass muestran que la participación privada puede dinamizar la oferta de servicios, pero requiere condiciones que Colombia aún no ha generado.

La experiencia con los derechos de petición, documentada en la sección 9, es quizá la que mejor ilustra los límites del modelo actual. Una cosa es tener plataformas habilitadas y otra muy distinta es que funcionen bien. Que la entidad encargada de promover la digitalización del Estado haya proporcionado enlaces rotos en su respuesta oficial es un ejemplo sencillo pero diciente acerca de la distancia entre el discurso y la ejecución.

Para quienes estén en posición de legislar o de diseñar política pública, las líneas de acción que se derivan de este análisis van en varias direcciones: crear mecanismos públicos de seguimiento al avance de la digitalización con datos desagregados por territorio, generar incentivos reales para que el sector privado participe como prestador de SCD Especiales, y —quizá lo más importante— entender que la digitalización de trámites no funciona si no se articula con políticas de conectividad, alfabetización digital y protección de datos. Estas no son políticas complementarias; son condiciones sin las cuales la promesa de los SCD queda a medio camino.

Al final, los SCD son un medio, no un fin. Su valor depende de si efectivamente le facilitan la vida a la gente, si reducen las barreras de acceso al Estado y si contribuyen a que la ciudadanía confíe más en las instituciones. El camino recorrido es importante, pero lo que falta todavía es considerable.

BIBLIOGRAFÍA

AGENCIA NACIONAL DIGITAL. Servicios Ciudadanos Digitales. [en línea]. Disponible en: https://and.gov.co/servicios-ciudadanos-digitales [consultado el 29 de abril de 2024].

AGENCIA NACIONAL DIGITAL. Autenticación Digital. [en línea]. Disponible en: https://autenticaciondigital.and.gov.co [consultado el 3 de mayo de 2024].

AGENCIA NACIONAL DIGITAL. Respuesta a derecho de petición AND-SAL-0425-2024. Bogotá, 22 de mayo de 2024. 7 p. Firmado por María Angélica González Russi, Subdirectora Jurídica.

AGENCIA NACIONAL DIGITAL. Vinculación al servicio ciudadano digital de interoperabilidad. [en línea]. Disponible en: https://and.gov.co/sites/default/files/2022-09/Vinculacion-entidades-a-Interoperabilidad.pdf [consultado el 15 de mayo de 2024].

COLOMBIA. CONGRESO DE LA REPÚBLICA. Ley 1341. (30, julio, 2009). Por la cual se definen principios y conceptos sobre la sociedad de la información y la organización de las TIC. Diario Oficial. Bogotá, D.C., 2009.

COLOMBIA. CONGRESO DE LA REPÚBLICA. Ley 1437. (18, enero, 2011). Código de Procedimiento Administrativo y de lo Contencioso Administrativo. Diario Oficial. Bogotá, D.C., 2011.

COLOMBIA. CONGRESO DE LA REPÚBLICA. Ley 1581. (17, octubre, 2012). Por la cual se dictan disposiciones generales para la protección de datos personales. Diario Oficial. Bogotá, D.C., 2012.

COLOMBIA. CONGRESO DE LA REPÚBLICA. Ley 1712. (6, marzo, 2014). Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional. Diario Oficial. Bogotá, D.C., 2014.

COLOMBIA. CONGRESO DE LA REPÚBLICA. Ley 2052. (25, agosto, 2020). Por la cual se establecen disposiciones para la racionalización de trámites. Diario Oficial. Bogotá, D.C., 2020.

COLOMBIA. PRESIDENCIA DE LA REPÚBLICA. Decreto 1078. (26, mayo, 2015). Decreto Único Reglamentario del Sector de Tecnologías de la Información y las Comunicaciones. Diario Oficial. Bogotá, D.C., 2015.

COLOMBIA. PRESIDENCIA DE LA REPÚBLICA. Decreto 1008. (14, junio, 2018). Por el cual se establecen los lineamientos generales de la Política de Gobierno Digital. Diario Oficial. Bogotá, D.C., 2018.

COLOMBIA. PRESIDENCIA DE LA REPÚBLICA. Decreto 620. (2, mayo, 2020). Por el cual se establecen los lineamientos generales en el uso y operación de los Servicios Ciudadanos Digitales. Diario Oficial. Bogotá, D.C., 2020.

COLOMBIA. PRESIDENCIA DE LA REPÚBLICA. Decreto 088. (24, enero, 2022). Por el cual se establecen los conceptos, lineamientos, plazos y condiciones para la digitalización y automatización de trámites. Diario Oficial. Bogotá, D.C., 2022. No. 51927.

COLOMBIA. PRESIDENCIA DE LA REPÚBLICA. Decreto 767. (16, mayo, 2022). Por el cual se actualiza la Política de Gobierno Digital. Diario Oficial. Bogotá, D.C., 2022.

GOBIERNO DE SINGAPUR. Singpass. [en línea]. Disponible en: https://www.singpass.gov.sg/main [consultado el 12 de abril de 2024].

GOVERNMENT OF CANADA. GCKey. [en línea]. Disponible en: https://www.canada.ca/en/government/sign-in-online-account/gckey.html [consultado el 12 de abril de 2024].

MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES. ABC Servicios Ciudadanos Digitales. [en línea]. Disponible en: https://gobiernodigital.mintic.gov.co/692/articles-238199_recurso_1.pdf [consultado el 20 de abril de 2024].

MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES. Guía de integración de prestadores privados de SCD Especiales. [en línea]. Disponible en: https://gobiernodigital.mintic.gov.co/692/articles-146484_recurso_2.pdf [consultado el 22 de abril de 2024].

MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES. Manual de Gobierno Digital. [en línea]. Disponible en: https://gobiernodigital.mintic.gov.co/692/channels-594_manual_gd.pdf [consultado el 25 de abril de 2024].

MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES. Respuesta a comunicación radicada 241033169. Bogotá, 16 de mayo de 2024. Radicado 242052816. 6 p. Firmado por Andrés Hernando Chaves Quijano.

MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES. Respuesta a comunicación radicada 241068779. Bogotá, 4 de septiembre de 2024. Radicado 242109735. 4 p. Firmado por Andrés Hernando Chaves Quijano.

MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES. Un paso firme hacia la transformación digital del Estado colombiano con Carpeta Ciudadana Digital. [en línea]. 7 de julio de 2025. Disponible en: https://mintic.gov.co/portal/inicio/Sala-de-prensa/Noticias/403166 [consultado el 18 de febrero de 2026].

MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES. INTEROPERA 2025: Congreso Nacional de Interoperabilidad. [en línea]. Disponible en: https://gobiernodigital.mintic.gov.co/portal/Eventos/418418 [consultado el 2 de marzo de 2026].

PERSONALAUSWEISPORTAL. Bundesministerium des Innern. Manual del usuario del nPA. [en línea]. Disponible en: https://www.personalausweisportal.de [consultado el 14 de abril de 2024].

SEDE ELECTRÓNICA DE LA AGENCIA TRIBUTARIA DE ESPAÑA. Manual de Usuario de Cl@ve. [en línea]. Disponible en: https://clave.gob.es/clave_Home/clave/funcionalidad.html [consultado el 14 de abril de 2024].

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO. Guía para la implementación del Principio de Responsabilidad Demostrada. [en línea]. Disponible en: https://www.sic.gov.co/sites/default/files/files/Publicaciones/Guia-Accountability.pdf [consultado el 8 de mayo de 2024].



Foto: Magnific

Foto: Magnific

PRIMERA EDICIÓN

e-Estado

2026

Pontificia Universidad JAVERIANA Bogotá

Observatorio de Gobierno y TIC

Escuela Javeriana de Gobierno y Ética Pública

e ESTADO