

CINCO MIRADAS

SOBRE EL PRESENTE Y FUTURO
DE LA INTELIGENCIA ARTIFICIAL
EN COLOMBIA

Editores:

Juan David Gutiérrez
Hartmut Rank



KONRAD
ADENAUER
STIFTUNG



Programa
Estado de Derecho
Latinoamérica

©2025 KONRAD-ADENAUER-STIFTUNG e. V.

FUNDACIÓN KONRAD ADENAUER

Programa Estado de Derecho para Latinoamérica

Calle 93B n.º 18-12, piso 7

Bogotá, Colombia

Tel.: (+571) 743 0947

iusla@kas.de

www.kas.de/iusla

X: [@KASiusLA](#)

Instagram: [@Kasiusla](#)

Facebook: [/kasiusla](#)

Linkedin: [kasiusla](#)

Editores: Juan David Gutiérrez y Hartmut Rank

Coordinación Editorial: Paulo Cañón

Asistencia Editorial: Camila Andrea Sánchez Solórzano

Diseño de cubierta y diagramación: Manuel Alejandro Sáenz Pinilla

Correspondencia, contribuciones, solicitudes de canje o donación email: iusla@kas.de

Cinco miradas sobre el presente y futuro de la inteligencia artificial en Colombia



KONRAD
ADENAUER
STIFTUNG



Programa
Estado de Derecho
Latinoamérica

2025

Contenido

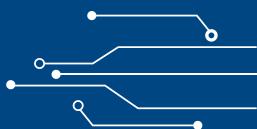
Prólogo: Más preguntas que respuestas9
Desafiando la Creatividad Humana: Propiedad Intelectual en la era de la Inteligencia Artificial. Por Andrea Martínez Devia	15
Introducción.....	16
La Propiedad Intelectual.....	17
La PI como Derecho Humano	18
Titularidad	19
PI e IA.....	20
Desafíos y Oportunidades	21
Casos Actuales.....	23
Conclusión.....	27
Referencias	28
Inteligencia artificial en la justicia: un panorama actual con énfasis en el sistema colombiano. Por Ana María Ramos Serrano y Ana María Giraldo Vargas.....	35
Introducción.....	36
Una aproximación a la inteligencia artificial.....	36
Aplicaciones de la IA en el sistema de justicia.....	41
La regulación de la IA	44
Conclusiones.....	46
Referencias	47
MinerIA de Datos:Inteligencia Artificial y sus Implicaciones para la Privacidad y la Protección de Datos Personales en Colombia. Por María Paula Ángel Arango	53
Introducción.....	54
¿Qué hace a la Inteligencia Artificial diferente?.....	54
Posibles daños	55
¿Cómo ha respondido la ley colombiana?.....	59
Conclusión.....	62
Referencias	63

Los Desafíos Contemporáneos de la Inteligencia Artificial en América Latina: Un Enfoque en el Principio de Transparencia. Por Dra. María Lorena Flórez Rojas	69
Introducción	70
El Principio de Transparencia en la Gobernanza de la IA	72
Operacionalizando el Principio de Transparencia	75
Desafíos y Tensiones en la Implementación de la Transparencia	79
Estrategias Prácticas y Recomendaciones de Políticas	82
Conclusiones.....	86
Referencias	87
Herramientas de toma automatizada de decisiones en el sector público de Colombia. Por Juan David Gutiérrez	97
Introducción.....	98
Contexto regional de la transformación digital en el sector público	99
Sistemas de toma automatizada de decisiones en el sector público de Colombia....	101
Transparencia algorítmica en el sector público colombiano	104
Conclusiones y recomendaciones.....	106
Referencias	108



Hartmut Rank

Director del Programa Estado de Derecho para Latinoamérica de la Fundación Konrad Adenauer desde 2021. De 2017 a 2021 fue jefe del Programa de Estado de Derecho de la KAS en el sureste de Europa. Anteriormente, fue destinado por el Ministerio de Asuntos Exteriores alemán a la Misión de la OSCE en Moldavia en 2014-2017 y trabajó para la GIZ en Ucrania y como abogado en Alemania y Ucrania. Estudió derecho en Leipzig, San Petersburgo, Potsdam y Londres, y economía en Liverpool. El Sr. Rank es abogado, mediador empresarial y traductor público de ucraniano y ruso. Su experiencia profesional se centra en el derecho europeo e internacional, los derechos de las minorías, la política multilateral de desarrollo, la OSCE y las formas extrajudiciales de resolución de conflictos.



Paulo Cañón

Comunicador social y periodista de la Universidad de La Sabana, Colombia. Ha participado en el semillero del Observatorio de Medios de la facultad de Comunicación de la Universidad de La Sabana. Adicionalmente se desempeña como redactor y corrector de estilo en la revista digital Katabasis. Actualmente es Coordinador de proyectos y Network Manager en el Programa Estado de Derecho para Latinoamérica de la Fundación Konrad Adenauer.

Prólogo:

Más preguntas que respuestas

Por: Hartmut Rank¹ y Paulo Cañón²

En muy poco tiempo, la inteligencia artificial ha dejado de ser una promesa tecnológica para convertirse en una realidad cotidiana. Desde que herramientas como ChatGPT irrumpieron en la esfera pública en 2022, el debate sobre la IA se ha intensificado a escala global, y sus efectos ya se sienten tanto en nuestras conversaciones personales como en nuestras instituciones. Sin embargo, las herramientas que incorporan características de inteligencia artificial han existido desde hace varios años. Lo hemos visto en plataformas como Waze o Google Maps, en los algoritmos de recomendación de servicios de *streaming*, en los sistemas de moderación de contenido para las redes sociales o, incluso, en los smartphones, donde asistentes como Siri resuelven preguntas o cuentan chistes desde hace más de una década. La IA lleva tiempo entre nosotros, pero solo recientemente hemos comenzado a reconocerla como tal. Este despertar de la conciencia pública ha transformado nuestra percepción sobre su presencia e impacto.

La rápida acogida de estos desarrollos, así como la facilidad con la que se multiplican, ha impactado no solo a las grandes empresas tecnológicas, sino también a los marcos normativos, las políticas públicas y la forma en que concebimos derechos, deberes y relaciones sociales. La IA ya no es un tema del futuro: está transformando el presente, y con él, los fundamentos sobre los que se construyen nuestras sociedades.

Aunque gran parte del debate ha estado centrado en Europa y Norteamérica, América Latina no ha permanecido al margen. Desde México hasta Argentina, los países han comenzado a posicionarse frente a los retos que plantea la IA, incluso en ámbitos donde su incorporación parecía improbable hace algunos años: la justicia, los derechos de autor, las políticas públicas y el derecho constitucional.

En el Programa Estado de Derecho para Latinoamérica de la Fundación Konrad Adenauer creemos que estos debates deben ser abordados de manera continua y rigurosa. Por eso nace esta publicación: *ColombIA. Cinco miradas nacionales para la inteligencia artificial*. Reunimos aquí voces expertas que, desde distintas disciplinas y enfoques, analizan las implicaciones jurídicas, éticas y políticas de la IA en Colombia y en nuestra región.

1 Director del Programa Estado de Derecho para Latinoamérica de la Fundación Konrad Adenauer

2 Coordinador de proyectos y Network Manager del Programa Estado de Derecho para Latinoamérica de la Fundación Konrad Adenauer

Esta iniciativa busca, más que ofrecer respuestas cerradas, abrir preguntas urgentes. ¿Cómo debe prepararse América Latina para enfrentar estos desafíos? ¿Qué puede aportar a la conversación global sobre ética, transparencia y regulación tecnológica? ¿Cómo garantizar que la IA no profundice las brechas existentes, sino que contribuya a cerrarlas?

Cada artículo de esta publicación aporta una mirada distinta y complementaria, que en conjunto tejen un diálogo rico sobre los desafíos contemporáneos de la IA. Andrea Martínez analiza cómo esta tecnología tensiona los marcos tradicionales de la propiedad intelectual y redefine el concepto de autoría. María Paula Ángel advierte sobre los riesgos que surgen cuando los algoritmos acceden y procesan nuestros datos personales, generando nuevas formas de daño. Ana María Ramos y Ana María Giraldo estudian su impacto en el sistema judicial, señalando tanto sus aportes como los peligros de sesgos automatizados. Por su parte, María Lorena Flórez examina el principio de transparencia como garantía democrática en contextos algorítmicos. Finalmente, Juan David Gutiérrez, quien estuvo a cargo de la coordinación editorial de esta publicación, analiza el uso de sistemas de decisión automatizada en el Estado colombiano y su influencia en los derechos fundamentales.

Además de sus contenidos, esta publicación ofrece una muestra del tipo de análisis que es posible cuando se cruzan el conocimiento jurídico, la perspectiva técnica y la experiencia institucional. Lejos de ofrecer una mirada homogénea, los textos aquí reunidos permiten contrastar desafíos y soluciones desde enfoques diversos, con una mirada crítica y a la vez propositiva. Es precisamente esa pluralidad la que enriquece el conjunto y lo vuelve útil para quienes quieren entender el fenómeno de la IA más allá de las respuestas rápidas o las lecturas tecnocéntricas.

La IA, como cualquier tecnología, no es buena ni mala en sí misma. Pero sí transforma, de manera profunda, los marcos donde opera: desde la manera en que concebimos la justicia, hasta las nociones de privacidad, autoría o transparencia en la gestión pública. Esta transformación es rápida, disruptiva y, muchas veces, difícil de anticipar. Por eso exige reflexión constante, regulación flexible y una ciudadanía informada.

Frente a este escenario, no hay respuestas simples ni universales. La complejidad de los desafíos que plantea la inteligencia artificial requiere marcos flexibles, pensamiento crítico y una constante revisión de nuestros principios jurídicos, éticos y sociales. Por eso proponemos este análisis colectivo, interdisciplinar y contextual, como una hoja de ruta para pensar los desafíos que enfrentamos desde Colombia —el país donde se ubica nuestra oficina—, pero también desde una perspectiva latinoamericana más amplia. Reconocer nuestras particulari-

dades no nos exime de participar activamente en el debate global; al contrario, nos permite enriquecerlo con miradas situadas, experiencias diversas y propuestas propias.

Agradecemos al grupo de especialistas que contribuyó con esta publicación por su generosa participación y compromiso. Sus aportes demuestran que América Latina tiene no solo el interés, sino también la capacidad y el conocimiento para contribuir al desarrollo de marcos éticos y jurídicos que acompañen los avances tecnológicos. Con esta publicación, reafirmamos nuestro compromiso con la democracia y el Estado de Derecho, dos principios que hoy también se disputan en el terreno de los algoritmos.

Esperamos que esta publicación no solo sea leída, sino verdaderamente interrogada. Que cada artículo inspire reflexión y provoque conversaciones dentro y fuera de los ámbitos especializados. Aspiramos a que este texto sirva como punto de encuentro para quienes creen que la inteligencia artificial debe ser pensada críticamente, con responsabilidad, y desde las realidades concretas de nuestra región. Solo así podremos construir un debate más robusto, plural y necesario para el presente latinoamericano. Porque el gran debate sobre la IA no consiste únicamente en observar sus avances o discutir sus efectos, sino también en recuperar la agencia en torno a ella. Nos parece clave recordar que nada está definido por completo, y, por eso mismo, cada persona, organización y gobierno que se enfrente al desafío sobre pensar en qué tipo de tecnologías se usan o desarrollan, también tiene capacidad para actuar frente a ellas. Por eso, el ejercicio crítico y analítico es necesario para que este actuar sea informado y efectivo.

Que este sea, entonces, un punto de partida. Una invitación a seguir pensando colectivamente cómo queremos que se diseñen, regulen y utilicen estas tecnologías en nuestras sociedades. El futuro de la inteligencia artificial debe ser una construcción política, cultural y ética, que cada día se vaya renovando. Pero esta construcción no se hace más sencilla, y paulatinamente abarca más aspectos de nuestras vidas. Por eso, está en nuestras manos —como región, como ciudadanía, como comunidad académica— participar activamente en su escritura.



○

Desafiando la Creatividad Humana: Propiedad Intelectual en la era de la Inteligencia Artificial

ANDREA MATÍNEZ DEVIA



Andrea Martínez Devia

Abogada de la Universidad de los Andes, Master en Derecho Comparado enfocado a la Propiedad Intelectual de la Universidad California Western School of Law de San Diego en los Estados Unidos. Certificada como Auditor Interno de Sistemas de Gestión de la Seguridad de la Información – ISO 27001:2013. Es directora del Área de Propiedad Intelectual y Protección de Datos de la firma de Asesoría Legal Martínez Devia & Asociados.

Cuenta con más de doce años de experiencia en sector privado y público en temas relacionados con Protección de Datos, Privacidad, Seguridad de la Información y Propiedad Intelectual, en Colombia y Estados Unidos.

Hace parte de la Asociación Colombiana de Legaltech ALT+CO, la Asociación Internacional de Profesionales de la Privacidad (IAPP), la Asociación Latinoamérica de privacidad (ALAP) y la Asociación Internacional de Marcas (INTA).

Desafiando la Creatividad Humana: Propiedad Intelectual en la era de la Inteligencia Artificial

Andrea Matínez Devia³

Resumen

Los sistemas de Propiedad Intelectual (PI) han sido diseñados para fomentar la innovación y las creaciones humanas, características que históricamente han definido a nuestra especie. Sin embargo, estamos viviendo un momento crucial en el que la inteligencia artificial (IA), con sus capacidades autónomas, se ha integrado en el ámbito creativo. Esta convergencia presenta desafíos significativos para los sistemas tradicionales de PI, cuyas bases se establecieron en un contexto exclusivamente humano.

Este artículo tiene como objetivo analizar diversos aspectos de la PI, incluyendo su definición, su reconocimiento como un derecho humano, su alcance y quienes son considerados actualmente como titulares de estos derechos. Asimismo, se explorará la relación entre la PI y la IA, así como los retos y beneficios que esta última representa para los sistemas de PI. A través de ejemplos y casos de estudio, se buscará ofrecer una comprensión integral de cómo la IA está transformando el panorama de la PI y sus implicaciones para el futuro de la creatividad y la innovación.

Palabras clave: Propiedad Intelectual, PI, Inteligencia Artificial, IA, creaciones, autoría, titularidad, derecho de autor, derechos patrimoniales, derechos morales, innovación, originalidad, derecho humano.

Abstract

Intellectual property (IP) systems have been designed to foster human innovation and creativity, characteristics that have historically defined our species. However, we are currently experiencing a crucial moment in which artificial intelligence (AI), with its autonomous capabilities, has integrated into the creative realm. This convergence presents significant challenges for IP systems, whose foundations were established in an exclusively human context.

³ Socia y Directora del Área de Propiedad Intelectual y Privacidad de la Firma Martínez Devia y Asociados. Abogada de la Universidad de los Andes, Master en Derecho Comparado enfocado a la Propiedad Intelectual de la Universidad California Western School of Law de San Diego en los Estados Unidos. Certificada como Auditor Interno de Sistemas de Gestión de la Seguridad de la Información – ISO 27001:2013 y Certificada como Profesional Certificado en Privacidad de la Información para Europa (CIPP/E).

This article aims to analyze various aspects of IP, including its definition, its recognition as a human right, its scope, and those who are currently considered the holders of these rights. Additionally, the relationship between IP and AI will be explored, along with the challenges and benefits that the latter represents for IP systems. Through examples and case studies, the article seeks to provide a comprehensive understanding of how AI is transforming the landscape of IP and its implications for the future of creativity and innovation.

Keywords: Intellectual Property, IP, Artificial Intelligence, AI, creations, authorship, ownership, copyright, economic rights, moral rights, innovation, originality, human right.

1. Introducción

Los sistemas de PI surgieron con el propósito de impulsar el ingenio y la creatividad humana. No obstante, nos encontramos en una coyuntura clave, ya que la IA ha comenzado a jugar un papel activo en la generación de bienes intangibles. Esta realidad introduce retos significativos para los sistemas de PI, los cuales fueron concebidos bajo el supuesto de que solo los seres humanos eran los únicos capaces de ser creativos e innovadores.

El surgimiento de la IA como generador de creaciones desafía las nociones tradicionales de autoría, titularidad y derechos de PI, generando preguntas como: ¿Quién debe ser considerado el titular de una creación o innovación generada por un sistema de IA? ¿El programador que diseñó el algoritmo, el usuario que solicitó la creación o la propia IA? ¿Si se reconoce a la IA como el titular de la creación, sobre quién podrían recaer los derechos patrimoniales para efectos prácticos? Estas cuestiones no solo complican la asignación de derechos, sino que también invitan a reconsiderar cómo valoramos el aporte intelectual en un mundo donde las máquinas (IA) también participan en la producción intelectual.

Además, la capacidad de la IA para analizar vastos conjuntos de datos e información y aprender de ellos introduce interrogantes sobre la originalidad de producciones intelectuales creadas. Si una IA genera una creación intelectual basándose en patrones aprendidos de creaciones previas, ¿dónde debe trazarse la línea para definir si se trata de inspiración legítima o, por el contrario, una simple copia? Esta problemática requiere una revisión urgente de las leyes de PI, que en su mayoría no contemplan las implicaciones de una inteligencia que puede crear, innovar y competir con los seres humanos en el ámbito creativo.

La adaptación de los sistemas de PI al contexto actual es esencial para asegurar un entorno que no solo proteja los derechos de los creadores humanos, sino que también regule y reconozca las contribuciones de la IA. Para lograrlo, es indispensable promover un diálogo multidisciplinario que involucre diferentes actores claves como instituciones educativas, legisladores, gobiernos, expertos en tecnología y la sociedad en general. Este esfuerzo conjunto debe culminar en un marco legal flexible, capaz de responder a las rápidas y constantes transformaciones en el ámbito de la IA.

En este contexto, es fundamental no solo proteger los derechos de autor, patentes, diseños industriales y marcas, sino también explorar nuevas formas de incentivos que reconozcan el papel de la IA. De esta manera, se podrá asegurar un equilibrio que fomente tanto la innovación humana como la artificial, garantizando un futuro donde la creatividad siga siendo un motor clave para el desarrollo cultural y tecnológico.

2. La Propiedad Intelectual

Como se ha mencionado, la PI ha sido reconocida históricamente como un medio para proteger las creaciones del intelecto humano. Según lo establecido por la **Organización Mundial de la Propiedad Intelectual (OMPI)** en el artículo *¿Qué es la propiedad intelectual?*

La propiedad intelectual (PI) se relaciona con las creaciones de la mente, como las invenciones, las obras literarias y artísticas, y los símbolos, nombres e imágenes utilizados en el comercio. La PI está protegida por la legislación, por ejemplo, en el ámbito de las patentes, el derecho de autor y las marcas, que permiten obtener reconocimiento o ganancias por las invenciones o creaciones. Al equilibrar el interés de los innovadores y el interés público, el sistema de PI procura fomentar un entorno propicio para que prosperen la creatividad y la innovación. (Organización Mundial de la Propiedad Intelectual, s.f.).

Así mismo, la **Organización Mundial del Comercio (OMC)** en el artículo titulado “*¿Qué se entiende por derechos de propiedad intelectual?*” publicado en la página web de la entidad establece que:

Los derechos de propiedad intelectual son aquellos que se confieren a las personas sobre las creaciones de su mente. Suelen dar al creador derechos exclusivos sobre la utilización de su obra por un plazo determinado. (Organización Mundial de la Propiedad Intelectual, s.f.).

De las anteriores definiciones, se puede concluir que la PI es reconocida como un conjunto de derechos destinados a proteger las invenciones y las obras derivadas del intelecto humano. En el contexto colombiano y a nivel internacional, se destaca que este conjunto de derechos se fundamenta en las capacidades inventivas y la originalidad del ser humano. Al salvaguardar estas creaciones o invenciones, se fomenta un entorno propicio para la innovación y la cultura, garantizando que los creadores y/o inventores disfruten de derechos exclusivos sobre sus desarrollos, al mismo tiempo que se promueve el bienestar social y económico de las comunidades.

Así, la PI no solo valora el esfuerzo individual, sino que también impulsa el progreso colectivo. Este enfoque se manifiesta en su reconocimiento como un derecho humano esencial para el bienestar social. Si bien este reconocimiento ha sido objeto de debate por la aparición de la

IA, es fundamental entender por qué históricamente se ha considerado un derecho humano y su importancia para el desarrollo de la sociedad.

3. La PI como Derecho Humano

La Declaración Universal de los Derechos Humanos y el Pacto Internacional de Derechos Económicos, Sociales y Culturales consagra el derecho de toda persona a la protección de los intereses morales y materiales que le correspondan por razón de sus creaciones científicas, literarias o artísticas (Naciones Unidas, 1948).

Este enfoque subraya que la creatividad es una cualidad inherente al ser humano y que el reconocimiento de los derechos sobre las obras creadas no es únicamente un asunto económico o comercial, sino que está profundamente relacionado con la dignidad humana y el desarrollo integral de las personas. La PI garantiza que los creadores puedan disfrutar del fruto de su trabajo y obtener reconocimiento por sus innovaciones, lo que fomenta un entorno donde la creatividad y la innovación florezcan y generen un fomento a la diversidad cultural.

En la era del conocimiento, la PI se ha convertido en un recurso crucial y un activo intangible comparable a las materias primas durante la revolución industrial, y constituye la base de la economía actual. Reconocer la PI como un derecho humano se fundamenta en la idea de que su protección puede contribuir al desarrollo y bienestar, alineándose con principios que respaldan la dignidad humana y el derecho a la creatividad.

En el contexto colombiano, la propiedad intelectual es reconocida como un derecho fundamental inherente a la persona, de acuerdo con la Constitución Política de 1991. El artículo 61 establece:

El Estado protegerá la propiedad intelectual por el tiempo y mediante las formalidades que establezca la ley.

Este mandato subraya el compromiso del Estado colombiano de fomentar un entorno donde la creatividad y la innovación sean motores del progreso nacional. La protección de la PI en Colombia abarca tanto la propiedad industrial (patentes, signos distintivos, modelos de utilidad y diseños industriales) como los derechos de autor (que protegen obras literarias, artísticas y musicales).

Sin embargo, a pesar de su reconocimiento como derecho humano, la relación entre la PI y los derechos humanos ha sido objeto de debate. Algunos críticos argumentan que, a diferencia de los derechos humanos, que son universales, inalienables y permanentes, los derechos de propiedad intelectual son temporales y territoriales. Esto implica que no son absolutos ya que están sujetos a restricciones y limitaciones según las leyes nacionales e internacionales, y pueden ser ejercidos tanto por personas físicas como jurídicas.

Otro punto de controversia radica en el enfoque comercial que se ha dado a la PI en ciertos contextos. Los derechos de PI, como las patentes, diseños industriales, modelos de utilidad, los signos distintivos y los derechos de autor, pueden ser objeto de explotación comercial. En este sentido, algunos críticos ven la PI más como una herramienta de mercado que como un derecho fundamental.

Este debate ha cobrado aún más relevancia con la aparición de tecnologías, en particular la inteligencia artificial. La capacidad de las IA para generar contenido original ha planteado preguntas sobre quién es el verdadero autor y/o titular de los bienes intangibles creados por estas herramientas. Esta situación cuestiona la concepción tradicional de la PI como un derecho exclusivo de las personas naturales, y plantea la posibilidad de que en el futuro se deban redefinir ciertos conceptos fundamentales en torno a la autoría y la titularidad de las creaciones.

La creciente capacidad de la IA para aprender, generar e innovar también plantea nuevos retos sobre cómo se protege y gestiona la propiedad intelectual en este contexto. A medida que las máquinas comienzan a desempeñar un papel más activo en la creación de bienes intangibles, surge la necesidad de que los marcos legales evolucionen para adaptarse a estos cambios y por tanto la concepción de derecho humano también.

4. Titularidad

Como se expuso previamente, la propiedad intelectual se divide en dos grandes categorías: derecho de autor y propiedad industrial. El derecho de autor protege las obras literarias, artísticas y científicas de los creadores, abarcando una amplia variedad de producciones que incluyen libros, música, pintura, cine, programas informáticos, bases de datos, anuncios publicitarios, mapas y dibujos técnicos (Organización Mundial de la Propiedad Intelectual, s.f.).

Dentro del marco de los derechos de autor, se distinguen dos tipos principales: los derechos patrimoniales y los derechos morales. La Declaración Universal de los Derechos Humanos en su artículo 27.2 señala que “Toda persona tiene derecho a la protección de los intereses morales y materiales que le correspondan por sus producciones científicas, literarias o artísticas”. Según el autor Eduardo de la Parra en su libro “Derechos de los autores, artistas e inventores”, los derechos morales abarcan el reconocimiento de la paternidad de la obra y la protección de su integridad, lo que implica que cualquier modificación debe contar con la autorización expresa del autor. En contraste, los derechos patrimoniales se refieren al valor económico de la obra (Eduardo Parra Trujillo, 2015).

A nivel mundial, el titular inicial del derecho de autor es generalmente el autor, esto es, una persona natural que ha creado la obra. Posteriormente, este puede celebrar contratos con terceros para permitir la explotación de su obra, ya sea con personas naturales o jurídicas.

Por otro lado, la propiedad industrial se define por el autor Juan Pablo Canval en su libro Manual de Propiedad Intelectual como:

[...] el conjunto de derechos exclusivos y temporales que el Estado concede para usar y explotar económicamente aquellas invenciones o innovaciones aplicables a la industria y el comercio que sean producto del ingenio y la capacidad intelectual del hombre. Recae sobre las cosas imperceptibles e inmateriales, como las creaciones que proceden del ingenio humano susceptibles de beneficio comercial o de utilización industrial (Canaval, 2008, p. 78).

En este contexto, una patente es un título de propiedad otorgado en un país que permite al titular impedir que otros fabriquen, vendan o usen comercialmente la invención durante un período determinado. Aunque el derecho de patente pertenece inicialmente al inventor, que siempre se reconoce como una persona natural, este derecho puede ser transferido a otras personas naturales o jurídicas, siempre respetando el derecho moral del inventor.

En lo que respecta a los signos distintivos, se protege la identificación gráfica, denominativa, sonora, olfativa, entre otros que distingue productos y servicios similares ofrecidos dentro del mercado. Al registrar un signo distintivo, el propietario obtiene el derecho exclusivo sobre la misma y puede emprender acciones legales contra cualquier uso no autorizado. Este derecho exclusivo también le permite conceder licencias de uso a terceros, generando así beneficios económicos a través de franquicias o licencias de marca.

Es relevante destacar que, a diferencia del derecho de autor y las patentes, los signos distintivos pueden ser registrados tanto por personas naturales como jurídicas inicialmente. Esta distinción es significativa, ya que los derechos de autor y las patentes suelen atribuirse principalmente a individuos. Sin embargo, como se ha mencionado reiteradamente en este artículo, la realidad actual plantea un debate sobre si, además de las personas físicas, las máquinas podrían considerarse autoras o inventoras de creaciones intelectuales.

5. PI e IA

La IA es un campo de la informática dedicado a resolver problemas cognitivos que comúnmente se asocian con la inteligencia humana o con seres inteligentes, entendidos como aquellos que pueden adaptarse a situaciones cambiantes (Consejo Nacional de Política Económica y Social, 2019). Su integración en diversos sectores, desde la salud y la educación hasta el entretenimiento y la industria, está transformando nuestro mundo. En este contexto, la PI no es la excepción.

Actualmente, existen algoritmos de IA que, de forma “autónoma” como CHATGPT, Midjourney, Stable Diffusion, Flow GPT, DALL-E entre otros, son capaces de generar creaciones que podrían considerarse PI, lo que subraya la necesidad de comprender que la creatividad también puede ser impulsada por máquinas.

El impacto de la IA en la PI no se limita al reconocimiento de los creadores, sino que también afecta a la manera en que se gestionan y distribuyen los derechos. La IA puede facilitar la creación, distribución y consumo de contenido, lo que podría llevar a un aumento en las infracciones de derechos de autor, patentes y marcas, además de desafíos en la protección de estos bienes en un entorno digital. La facilidad con la que la IA puede replicar y modificar bienes intangibles existentes también genera preocupaciones sobre el plagio y la desvalorización de la creatividad humana.

Este panorama subraya la importancia de un diálogo constante entre creadores, expertos tecnológicos, legisladores, gobiernos y la sociedad en su conjunto. Solo a través de una colaboración multidisciplinaria será posible encontrar soluciones que protejan los derechos de los creadores humanos y fomenten la innovación, mientras se navega por los desafíos que la IA presenta en el ámbito de la PI.

6. Desafíos y Oportunidades

La falta de protección sobre las obras generadas por IA puede generar que estas sean utilizadas libremente, desincentivando la inversión en desarrollo tecnológico y amenazando el sustento de los profesionales creativos como inventores, desarrolladores de software, artistas digitales, músicos y compositores, escritores, guionistas, entre muchos otros.

Un ejemplo claro de este conflicto fue la reciente huelga en Hollywood, donde actores y guionistas denunciaron el uso no autorizado de su trabajo para entrenar modelos de IA (Los Angeles Times, 2023). La incertidumbre generada por la falta de directrices claras sobre el uso de la IA en la creación de contenido ha llevado a un llamado urgente por regulaciones que garanticen la originalidad y el reconocimiento adecuado para los artistas, pero también para quienes crean los bienes intangibles a través de la IA.

Muchos profesionales reclaman regulaciones que garanticen la originalidad de las obras y el reconocimiento y la compensación adecuada para los artistas cuyas creaciones puedan ser replicadas por máquinas. Además, el temor a la pérdida de oportunidades laborales ha llevado a un debate urgente sobre la necesidad de establecer marcos legales que protejan a los trabajadores en esta actual realidad.

Además, los problemas de transparencia y sesgos en los algoritmos de IA, evidenciados en casos como el algoritmo COMPAS⁴ (Correctional Offender Management Profiling for Alternative Sanctions) en el ámbito judicial, subrayan la necesidad de abordar la opacidad en las

⁴ **COMPAS- Correctional Offender Management Profiling for Alternative Sanctions.** Algoritmo desarrollado por la empresa privada Northpointe que lo describe como “un instrumento de evaluación de riesgo de cuarta generación, el cual permite el monitoreo de los individuos a lo largo del tiempo a partir de la información de factores de riesgo dinámicos, diseñados para ayudar en la intervención correcional con el fin de disminuir la probabilidad de que los infractores reincidan”.

decisiones algorítmicas (Joselson, 2012), ya que pueden llevar a decisiones erróneas y violaciones de derechos humanos como se evidenció en el caso State of Wisconsin vs. Loomis en el cual COMPAS generó un informe que clasificaba a Loomis como de alto riesgo de reincidencia. Este informe fue utilizado por el juez durante el proceso de sentencia, influyendo en la decisión de imponer una pena más severa. Sin embargo, Loomis no tuvo acceso a los detalles del funcionamiento del algoritmo, lo que planteó preocupaciones sobre la falta de transparencia y la posibilidad de sesgos en la herramienta (Wisconsin Supreme Court, 2016).

Por último, la responsabilidad en caso de infracciones de PI cometidas por sistemas de IA sigue siendo un tema controvertido, especialmente cuando las decisiones son autónomas y no hay intervención humana directa. La rápida evolución de la IA a menudo supera la capacidad de las leyes para adaptarse, dejando vacíos legales que dificultan la protección efectiva. Ejemplos recientes, como la demanda del New York Times contra OpenAI y Microsoft por el uso no autorizado de sus artículos, subrayan esta problemática. El periódico alega que su contenido ha sido utilizado sin permiso para entrenar modelos de inteligencia artificial, sosteniendo que esta práctica no solo infringe sus derechos de autor (copyright), sino que también amenaza la integridad y sostenibilidad del periodismo de calidad. Este caso destaca la necesidad de establecer reglas claras que regulen cómo se puede utilizar el contenido protegido en el desarrollo de tecnologías emergentes (Grynbaum & Mac, 2023).

No obstante, la IA también ofrece nuevas oportunidades en la gestión de la propiedad intelectual. Los creadores pueden establecer licencias específicas que controlen el uso de sus obras en el entrenamiento de IA, lo que no solo protege sus derechos, sino que también permite explorar modelos de negocio innovadores. La automatización de búsquedas de patentes y procesos administrativos, facilitada por la IA, mejora la eficiencia y reduce errores, permitiendo a los creadores concentrarse en el desarrollo de sus ideas.

Asimismo, la IA puede encargarse de tareas rutinarias, reduciendo significativamente el tiempo requerido para completar trámites relacionados con la creación y protección de patentes y marcas. Esto no solo mejora la eficiencia, sino que también minimiza los errores humanos.

En cuanto al análisis de infracciones, la IA ofrece una forma avanzada de detectar posibles violaciones de derechos de propiedad intelectual. A través de la comparación y análisis de grandes conjuntos de datos e información, puede identificar similitudes y diferencias entre productos y/o procesos que pretenden ser o están patentados, marcas y otros activos intelectuales, lo que ayuda a proteger los derechos de los creadores de manera más efectiva.

Además, los algoritmos de IA son útiles para la detección de plagio. Al analizar grandes cantidades de datos, textos e imágenes, pueden identificar posibles casos de copia indebida que podrían indicar violaciones de derechos de autor. Herramientas como DetectGPT o GTPzero están emergiendo como importantes recursos para determinar la registrabilidad de obras generadas por inteligencia artificial, asegurando así que se mantenga la integridad del sistema

de propiedad intelectual en un mundo en constante evolución.

En este contexto, es fundamental que creadores, tecnólogos y legisladores colaboren para encontrar soluciones que protejan los derechos de los creadores humanos y de las máquinas y fomenten la innovación, mientras se navega por los desafíos que la IA presenta en el ámbito de la propiedad intelectual.

7. Casos Actuales

A continuación, se presentan algunos casos que ilustran algunas de las problemáticas desarrolladas previamente:

a. Caso Dabus

DABUS (Device for the Autonomous Bootstrapping of Unified Sentience) es un sistema de inteligencia artificial (IA) desarrollado por Stephen Thaler, que utiliza redes neuronales para generar nociones simples y complejas. Ha creado invenciones como un contenedor de alimentos y una luz intermitente para atraer una mayor atención (WIPO, 1. WO2020079499 - Food container and devices and methods for attracting enhanced attention, 2020). Sin embargo, la cuestión de si DABUS puede ser considerado un “inventor” ha desafiado a varias oficinas de patentes en todo el mundo, dentro de las que se encuentra Reino Unido, Unión Europea, Estados Unidos, Australia, Sudáfrica, entre otros.

Este debate comenzó cuando Thaler, buscó probar el sistema de patentes actual al solicitar patentes generadas por DABUS en varios países del mundo. La discusión ha llevado a decisiones en varias jurisdicciones:

Sudáfrica. La Comisión de Propiedad Intelectual de Sudáfrica hizo historia al reconocer a DABUS como inventor. En julio de 2021, sin llevar a cabo un examen de fondo de la solicitud, Sudáfrica otorgó una patente a DABUS, una invención generada de forma autónoma por inteligencia artificial. Esta decisión destacó porque la legislación sudafricana no define explícitamente el término “inventor” (Companies and Intellectual Property Commission, 2021). No obstante, su validez ha sido cuestionada, ya que el sistema de patentes del país se limita a verificar requisitos formales sin un análisis sustantivo. Algunos analistas sugieren que esta medida podría formar parte de una estrategia gubernamental para atraer innovación y mejorar el desarrollo socioeconómico, lo que añade complejidad al debate sobre el reconocimiento de la IA como inventor.

Australia. La victoria sudafricana fue seguida rápidamente por un hito en Australia. Tras la negativa inicial de la Oficina Australiana de Patentes, que argumentó que no se designaba a un inventor humano, el Dr. Thaler apeló la decisión. En el año 2021, el Tribunal Federal de

Australia falló a favor de DABUS, afirmando que la ley no excluye explícitamente a la IA como inventor. El juez Jonathan Beach de primera instancia, en el caso Thaler v Commissioner of Patents FCA 879, consideró que el sistema de inteligencia artificial, sí podía ser un inventor a los efectos de la ley, dado que el término inventor no es más que un “agent noun” es decir, puede ser una persona o una cosa, al propio tiempo, el juzgador consideró que no tener en cuenta esta posibilidad implicaría que muchas invenciones no serían patentables y enfatizó el papel innovador de la IA en la creación de nuevos productos. Esta decisión sienta un precedente importante a nivel global, dado el reconocimiento del tribunal australiano en asuntos de propiedad intelectual (Federal Court of Australia , 2021).

No obstante, la decisión judicial previa llevó al comisionado de patentes a apelar, y la Corte Federal de Australia determinó que el juez de primera instancia había interpretado incorrectamente el artículo 15 de la Ley de Patentes, así como la regla 3.2C(2)(aa) del Reglamento. Además, consideró que las conclusiones fácticas del juez excedieron las pruebas presentadas en el tribunal. Esto se debe a que el artículo 15.1 de la Ley de Patentes establece que una patente solo puede ser concedida si el inventor es una persona o si los derechos del inventor se derivan de una persona, por tanto, negó la patente DABUS (Corte Federal de Australia, 2022).

Estados Unidos. En 2021, el Tribunal de Distrito de los Estados Unidos para el Distrito Este de Virginia (Lavrichenko, 2021) y, en 2022, el Tribunal de Apelación del Circuito Federal (CAFC) (Tribunal de Apelación del Circuito Federal, 2022) confirmaron el rechazo de la USPTO a DABUS del año 2020, reafirmando que solo las personas físicas pueden ser consideradas inventores. Ambas instancias sostuvieron que la Ley de Patentes exige que el inventor sea un ser humano, basándose en disposiciones del Título 35 del USC que definen expresamente a los inventores como personas físicas (Comité Permanente sobre el Derecho de Patentes, 2023).

En su fallo de 2022, el CAFC interpretó el término “individual” en el contexto de la ley como “ser humano” y se apoyó en decisiones previas que, aunque no relacionadas con la inteligencia artificial, reafirmaron esta postura. Además, el tribunal desestimó los argumentos que proponían permitir que una IA pudiera ser designada como inventor, considerándolos especulativos. La corte aclaró que no se debatía si las invenciones realizadas con ayuda de la IA podían ser patentadas.

Finalmente, el Tribunal Supremo de los Estados Unidos rechazó el 24 de abril de 2023 la solicitud de un auto de avocación relacionada con este asunto.

Reino Unido. El 20 de diciembre de 2023, la Corte Suprema del Reino Unido, en el expediente 2021/0201 Thaler v Comptroller-General of Patents, Designs and Trade Marks, dictaminó que la inteligencia artificial no puede ser reconocida como inventora en una patente relacionada con un nuevo producto o idea. Este fallo reafirma que, según la Patents Act 1977,

un “inventor” debe ser una persona natural, ya que solo un ser humano puede concebir una invención. La decisión respalda los veredictos previos del Tribunal Superior de Justicia del Reino Unido y de la Oficina de Propiedad Intelectual del Reino Unido (UKIPO) en el caso DABUS (Justice, 2020).

Europa. En este caso, La Oficina Europea de Patentes (OEP) rechazó las solicitudes de patente EP 18275163 y EP 18275174, que designaban a DABUS, un sistema de inteligencia artificial, como inventor. La OEP argumentó que, de acuerdo con la Convención sobre Patentes Europeas (CPE), solo una persona humana puede ser designada como inventor, ya que una IA carece de personalidad jurídica y, por lo tanto, no puede cumplir con los requisitos de la CPE. El solicitante presentó recursos que fueron examinados por la Cámara de Recursos en los casos J 8/20 y J 9/20, que también rechazó la designación de una IA como inventor, afirmando que el inventor debe tener capacidad jurídica. La Cámara aclaró que una IA no puede ceder derechos de patente y que la propiedad sobre una IA no implica automáticamente derechos de patente. Aunque la Cámara reconoció que las invenciones de IA no están excluidas de la patentabilidad, insistió en que debe designarse a un inventor humano conforme a la CPE (WIPO, 2019).

El 21 de diciembre de 2021, la Cámara de Apelaciones confirmó las decisiones de la Sección de Recepción de la OEP de rechazar las solicitudes EP 18275163 y EP 18275174, y también rechazó una solicitud auxiliar que proponía no identificar a ninguna persona como inventor, indicando en su lugar que una persona natural tenía el derecho a la patente por ser propietaria y creadora del sistema de IA DABUS.

Después de la audiencia oral, el solicitante fue informado en una comunicación con fecha del 31 de enero de 2022 de que la solicitud se consideraba retirada debido al impago de tasas. Sin embargo, el 20 de marzo de 2022, el apelante presentó una solicitud de procesamiento adicional, que fue concedida por la División de Examen el 6 de abril de 2022. Esto condujo a la emisión de la decisión escrita por parte de la Cámara de Recursos.

b. Caso Théâtre D'Opéra Spatial – Jason M. Allen

La Oficina de Derechos de Autor de EE. UU. ha rechazado la solicitud del artista Jason M. Allen para obtener derechos sobre su imagen premiada “Theatre D'opera Spatial”, creada con el sistema de inteligencia artificial Midjourney. La oficina argumentó que la obra no era elegible para derechos de autor porque no resultaba de la autoría humana.

Esta imagen, que evoca una corte real futurista, ganó el concurso de arte de la Feria Estatal de Colorado en 2022, convirtiéndose en la primera obra generada por IA en recibir tal reconocimiento. Un examinador de la Oficina solicitó más detalles sobre el proceso creativo, y Allen explicó que realizó al menos 624 revisiones y modificaciones utilizando Adobe Photoshop.

La Oficina pidió a Allen que renunciara a las partes generadas por Midjourney para poder recibir protección de derechos de autor, pero él se negó, lo que llevó al rechazo de su solicitud. Posteriormente, el Comité de Revisión de Derechos de Autor confirmó esta decisión, concluyendo que la imagen no podía ser protegida debido a la significativa cantidad de material generado por IA.

Actualmente, el caso está en un Tribunal Federal de Colorado, ya que Allen ha solicitado que se revoque la decisión de la Oficina. El artista ha manifestado que la decisión lo ha dejado en una “situación terrible”, sin ninguna defensa contra quienes roban su trabajo de manera flagrante y repetida. En su demanda, Allen argumenta que debería recibir derechos de autor para “Théâtre d’Opéra Spatial” como una expresión de su creatividad (United States District Court - District of Colorado, 2024).

c. Caso Zarya of the Dawn' - Kristina Kashtanova:

La Oficina de Derechos de Autor de EE. UU. (Copyright Office) canceló el registro del cómic “Zarya of the Dawn”, creado por la artista Kristina Kashtanova con herramientas de inteligencia artificial (IA), argumentando que no es una creación humana. Esta Oficina determinó que los usuarios no pueden ser considerados autores de las imágenes generadas por IA, incluso si han proporcionado numerosos prompts.⁵ Aunque Kashtanova alegó haber dedicado mucho esfuerzo en su trabajo con Midjourney, la oficina rechazó el argumento de que esto la convierta en autora.

Además, la Oficina de Derechos de Autor descartó la aplicación del régimen de “obra por encargo”. En la Unión Europea, se enfatiza la necesidad de distinguir entre las creaciones humanas y las generadas por IA, ya que estas últimas no deben tener derechos de autor para preservar el principio de originalidad ligado a autores humanos. Aunque los prompts pueden ser creativos, el resultado final de la IA es impredecible y, por lo tanto, no puede ser registrado como obra humana.

d. Caso Conversaciones Poéticas con CHATGPT: Poesía de la Mente Artificial

El 26 de febrero de 2023, se solicitó a la Dirección Nacional de Derecho de Autor Colombiana (DNDA) el registro de la obra “Conversaciones Poéticas con ChatGPT: Poesía de la mente artificial”. La solicitante argumentó que se trataba de una creación original e individual. En su Resolución Núm. 137 del 2 de mayo de 2023, la DNDA enfatizó que, según la Ley 23 de

⁵ **Prompts:** “Un prompt es una instrucción o texto inicial que se le proporciona a una herramienta de IA generativa para guiar su generación de respuestas o resultados, según los formatos en los que se especialice la herramienta. El prompt funciona como una “entrada de información” con la cual el usuario le especifica el contexto y la tarea que se espera que la herramienta complete.” Comunidad de Empresas (CE). (n.d.). Qué es un prompt en IA y para qué sirve (+ Ejemplos). tomado de: <https://ce.entel.cl/articulos/que-es-un-prompt/>

1982 y la Decisión 351 de 1992, los derechos de autor solo protegen obras que sean creación intelectual y original del ingenio humano.

La DNDA analizó que la colección de poemas resultaba de una interacción entre un ser humano y la inteligencia artificial ChatGPT. Sin embargo, concluyó que la respuesta generada por la IA no se puede considerar como una obra creada por un humano. Al revisar el texto presentado, la DNDA determinó que la usuaria solo había proporcionado poemas como inspiración, sin participar en la generación de la respuesta, lo que llevó al rechazo de su solicitud de registro.

Los casos mencionados reflejan cómo la creciente capacidad de la IA para crear bienes intangibles plantea importantes desafíos para las leyes de propiedad intelectual a nivel global. Situaciones como las de DABUS, Théâtre D'Opéra Spatial, Zarya of the Dawn y Conversaciones Poéticas con CHATGPT muestran la falta de claridad y los vacíos de las normativas existentes y de las autoridades a considerar a la IA como autor o inventor, lo que subraya la necesidad de reexaminar conceptos esenciales de titularidad, autoría, originalidad y propiedad, con el fin de poder armonizar posiciones y que los desarrollos jurisprudenciales puedan tener mayores bases para tomar sus decisiones.

8. Conclusión

La intersección entre la PI y la IA plantea desafíos y oportunidades que demandan un análisis profundo y la adaptación de marcos legales y éticos existentes. Es fundamental considerar la creación de una nueva categoría de titularidad para los bienes intangibles generados por máquinas. A medida que la IA se integra en la creación de obras, productos, procedimientos, signos distintivos las líneas entre la autoría humana y la producción automatizada se desdibujan, lo que genera interrogantes sobre la titularidad de los derechos de PI y la revisión de diferentes conceptos alrededor de esta.

La educación en PI y tecnología es esencial para preparar a creativos, legisladores, gobiernos, autoridades, jueces y usuarios en un entorno cada vez más influenciado por la IA. En este contexto, el futuro de la PI dependerá de la colaboración entre múltiples actores para establecer soluciones que equilibren la protección de los derechos sobre bienes inmateriales y la innovación que la IA puede ofrecer. La evolución de la PI en la era de la IA no solo es necesaria, sino inevitable, y requiere un enfoque proactivo para abordar los retos que surgen en este panorama actual.



9. Referencias

Abbott, R. (2019). El Proyecto de inventor artificial. *OMPI Revista*, 6, 8-14. Recuperado el 24 de octubre de 2024, de https://www.wipo.int/export/sites/www/wipo_magazine/es/pdf/2019/wipo_pub_121_2019_06.pdf.

Azuaje, M. (2020). Protección jurídica de los productos de la inteligencia artificial en el sistema de propiedad intelectual. *Revista Jurídica Austral*, 1(1), 319-342.

Canaval, J. (2008). *Manual de propiedad intelectual*. Bogotá D.C.: Editorial Universidad del Rosario.

Colombia. Asamblea Nacional Constituyente. (1991). *Constitución Política de Colombia*. Bogotá D.C.: 4 de julio de 1991.

Comité Permanente sobre el Derecho de Patentes. (16 de octubre de 2023). *La inteligencia artificial (IA) y la calidad de inventor*. Recuperado el 24 de octubre de 2024, de OMPI: https://www.wipo.int/edocs/mdocs/scp/es/scp_35/scp_35_7.pdf

Companies and Intellectual Property Commission. (2021). Patent journal - Including trade marks, designs and copyright in cinematograph films. *Patent journal* -, 54(7), 1-350. Recuperado el 23 de octubre de 2024, de https://iponline.cipc.co.za/Publications/PublishedJournals/E_Journal_July%202021%20Part%201.pdf

Comunidad de Empresas (CE). (s.f.). *Qué es un prompt en IA y para qué sirve (+ Ejemplos)*. Recuperado el 24 de octubre de 2024, de CE: <https://ce.entel.cl/articulos/que-es-un-prompt/>

Consejo Nacional de Política Económica y Social. (2019). *COMPES 3975: Política nacional para la transformación digital e inteligencia artificial*. Recuperado el 24 de octubre de 2024, de Departamento Nacional de Planeación: <https://colaboracion.dnp.gov.co/CDT/Compes/Econ%C3%B3micos/3975.pdf>

Corte Federal de Australia. (2022). *Commissioner of Patents v Thaler*. Australia: Jueces Allsop C. J., Nicholas, Yates, Moshinsky and Burley J. J.

Eduardo Parra Trujillo. (2015). *Derechos de los autores, artistas e inventores*. México D.F.: Talleres Cromo Editores S.A.

Escudero, P. G. (20 de abril de 2012). *La propiedad industrial.- Conceptos Básicos*. Recuperado el 24 de octubre de 2024, de Escuela de Organización Industrial: https://www.eoi.es/wiki/index.php/La_propiedad_industrial.Conceptos_b%C3%A1sicos

Federal Court of Australia . (2021). *Thaler v Commissioner of Patents [2021] FCA 879* . Australia: Stephen L. Thaler.

Grynbaum, M., & Mac, R. (27 de diciembre de 2023). *The New York Times demanda a OpenAI y Microsoft por el uso de obras con derechos de autor en la IA*. Recuperado el 23 de octubre de 2024, de The New York Times: <https://www.nytimes.com/es/2023/12/27/espanol/new-york-times-demanda-openai-microsoft.html>

Guío Español, A., Tamayo Uribe, E., & Gómez Ayerbe, P. (2021). *Marco ético para la inteligencia artificial en Colombia*. Bogotá D.C.: Presidencia de Colombia, Consejo Internacional de Inteligencia Artificial para Colombia.

Joselson, N. (17 de agosto de 2012). *Practitioner's Guide to COMPAS Core - tooltrack*. Recuperado el 23 de octubre de 2024, de https://njoselson.github.io/pdfs/FieldGuide2_081412.pdf

Justice, T. R. (15 de julio de 2020). *En apelación contra la decisión de Huw Jones (Director Adjunto, en funciones del Contralor de la Oficina de Propiedad Intelectual del Reino Unido)*. Recuperado el 23 de octubre de 2024, de The Royal Courts of Justice: <https://www.bailii.org/ew/cases/EWHC/Patents/2020/2412.html>

Lavrichenko, M. (2021). Thaler v. Vidal: Artificial Intelligence—Can the Invented Become the Inventor? *Cardozo Law Review*, 44(2). Recuperado el 24 de octubre de 2024

Moncayo Santacruz, D. M., & Vásquez Osorio, D. (2023). El derecho de la propiedad intelectual puesto a prueba: inteligencia artificial con capacidad inventiva. *Revista La Propiedad Inmaterial*, 35, 147–175. doi:<https://doi.org/10.18601/16571959.n35.06>

Naciones Unidas. (1948). *La Declaración Universal de los Derechos Humanos*. Recuperado el 23 de octubre de 2024, de Naciones Unidas: <https://www.un.org/es/about-us/universal-declaration-of-human-rights>

Naciones Unidas. (16 de diciembre de 1966). *Pacto Internacional de Derechos Económicos, Sociales y Culturales*. Recuperado el 24 de octubre de 2024, de Naciones Unidas: <https://www.ohchr.org/es/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights>

National Archives. (16 de marzo de 2023). *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*. Recuperado el 24 de octubre de 2024, de Federal Register: <https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence>

Ordelin, J. L. (2024). *Los sistemas de inteligencia artificial como inventores: consideraciones para un debate a partir de fallos judiciales*. Recuperado el 24 de octubre de 2024, de Universidad Nacional Autónoma de México-Instituto de Investigaciones Jurídicas: <https://archivos.juridicas.unam.mx/www/bjv/libros/15/7355/12.pdf>

Organización Mundial de la Propiedad Intelectual. (s.f.). *¿Qué es el derecho de autor?* Recuperado el 23 de octubre de 2024, de OMPI: <https://www.wipo.int/copyright/es/>

Organización Mundial de la Propiedad Intelectual. (s.f.). *¿Qué es la propiedad intelectual?* Recuperado el 24 de octubre de 2024, de OMPI: <https://www.wipo.int/about-ip/es/>

Organización Mundial del Comercio. (OMC). (s.f.). *Propiedad intelectual: Un resumen. Recuperado de artículo “¿Qué se entiende por ‘derechos de propiedad intelectual’?”* Recuperado el 24 de octubre de 2024, de Organización Mundial del Comercio. (OMC): https://www.wto.org/spanish/tratop_s/trips_s/intel1_s.htm

Roa Avella, M. d., Sanabria-Moyano, J. E., & DinaS-HurtaDo, K. (2022). Uso del algoritmo compas en el proceso penal y los riesgos a los derechos humanos. *Revista Brasileira de Direito Processual Penal*, 8(1), 275-310. doi:<https://doi.org/10.22197/rbdpp.v8i1.615>

Rodríguez, O. (2 de febrero de 2018). *Inteligencia Artificial: ¿conveniencia o amenaza?* Recuperado el 24 de octubre de 2024, de El nuevo Herald: <https://www.elnuevoherald.com/Noticias/Tecnologia/Article197719359.Html>

Tribunal de Apelación del Circuito Federal. (5 de agosto de 2022). *Thaler v. Vidal, 43 F.4th 1207, 1210 (Fed. Cir. 2022)*, Estados Unidos de América. Recuperado el 24 de octubre de 2024, de OMPI: <https://www.wipo.int/wipolex/es/judgments/details/2098>

United States District Court - District of Colorado. (26 de septiembre de 2024). *Shira perlmuter, in her official capacity as register of copyrights and director of the United States copyright office.* Recuperado el 24 de octubre de 2024, de Civil Action No. 1:24-cv-2665: <https://fingfx.thomsonreuters.com/gfx/legaldocs/gdvzkrrmapw/AI%20COPYRIGHT%20REGISTRATION%20appeal.pdf>

WIPO. (6 de noviembre de 2019). *1. EP3563896 - Devices and methods for attracting enhanced attention.* Recuperado el 23 de octubre de 2024, de WIPO: <https://patentscope.wipo.int/search/en/detail.jsf?docId=EP276033389>

WIPO. (23 de abril de 2020). *1. WO2020079499 - Food container and devices and methods for attracting enhanced attention.* Recuperado el 23 de octubre de 2024, de WIPO: <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2020079499>

Wisconsin Supreme Court. (2016). *Warning Before Use of Algorithmic Risk Assessments in Sentencing. — “State v. Loomis”, 881 N.W.2d 749 (Wis. 2016).* Recuperado el 23 de octubre de 2024, de Harvard Law Review: <https://www.jstor.org/stable/44865547>



Inteligencia Artificial en la Justicia

Un panorama actual con énfasis
en el sistema colombiano

ANA MARÍA RAMOS SERRANO
ANA MARÍA GIRALDO VARGAS



Ana María Ramos Serrano

Abogada e ingeniera industrial de la Universidad de los Andes, con maestría en Derecho Público para la Gestión Administrativa y Especialización en Evaluación Social de Proyectos. Se ha desempeñado como subdirectora de la Corporación Excelencia en la Justicia, viceministra de promoción de la justicia, consultora de la Alcaldía de Bogotá, del Fondo Colombia en Paz y de la Universidad de los Andes, entre otros. Actualmente es magistrada auxiliar en la Rama Judicial de Colombia y es profesora de cátedra de la Universidad de Los Andes.



Ana María Giraldo Vargas

Estadística de la Universidad Nacional de Colombia. Cuenta con una maestría en Políticas Públicas de la Universidad de los Andes y con una maestría en Estadística y Ciencia de Datos de la KU Leuven en Bélgica. Ha sido consultora de la Organización Internacional para las Migraciones (OIM), de la Universidad de los Andes y de la Pontificia Universidad Javeriana. En el sector público ha trabajado para el Ministerio de Educación Nacional, el INVIMA, y el ICFES. Actualmente es profesional especializada en la Rama Judicial de Colombia.

Inteligencia artificial en la justicia: un panorama actual con énfasis en el sistema colombiano

Por Ana María Ramos Serrano⁶ y Ana María Giraldo Vargas⁷

Resumen

Este artículo examina la creciente integración de la inteligencia artificial (IA) en el sistema de justicia, centrándose en el contexto colombiano, y presenta los conceptos básicos de la IA, su evolución a través del procesamiento del lenguaje natural (NLP) y el aprendizaje automático, hasta los modelos grandes de lenguaje (LLM). Destaca cómo la IA facilita la clasificación de información, la consulta de jurisprudencia y la edición de texto, y la transcripción de audiencias, pero también advierte sobre los riesgos de sesgos algorítmicos que pueden afectar derechos fundamentales. El documento subraya la necesidad de abordar los desafíos regulatorios derivados de la expansión de la IA y de mitigar estos sesgos mediante benchmarks, la mejora del curado de datos y la intervención humana, además de exemplificar el uso de herramientas de IA en tribunales de Brasil, España, Colombia, Perú, Estados Unidos y Países Bajos.

Palabras clave: inteligencia artificial, justicia, modelos grandes de lenguaje (LLM), regulación de la IA, legaltech.

Abstract

This article examines the growing integration of artificial intelligence (AI) in the justice system, focusing on the Colombian context, and presents the basic concepts of AI, its evolution through natural language processing (NLP) and machine learning, to large language models (LLM). It highlights how AI facilitates information classification, jurisprudence consultation and text editing, and hearing transcription, but also warns about the risks of algorithmic bias.

6 Abogada e ingeniera industrial de la Universidad de los Andes, con maestría en Derecho Público para la Gestión Administrativa y Especialización en Evaluación Social de Proyectos. Se ha desempeñado como subdirectora de la Corporación Excelencia en la Justicia, viceministra de promoción de la justicia, consultora de la Alcaldía de Bogotá, del Fondo Colombia en Paz y de la Universidad de los Andes, entre otros. Actualmente es magistrada auxiliar en la Rama Judicial de Colombia y es profesora de cátedra de la Universidad de Los Andes.

7 Estadística de la Universidad Nacional de Colombia. Cuenta con una maestría en Políticas Públicas de la Universidad de los Andes y con una maestría en Estadística y Ciencia de Datos de la KU Leuven en Bélgica. Ha sido consultora de la Organización Internacional para las Migraciones (OIM), de la Universidad de los Andes y de la Pontificia Universidad Javeriana. En el sector público ha trabajado para el Ministerio de Educación Nacional, el INVIMA, y el ICFES. Actualmente es profesional especializada en la Rama Judicial de Colombia.

ses that may affect fundamental rights. The paper highlights the need to address regulatory challenges arising from the expansion of AI and to mitigate these biases through benchmarks, improved data curation and human intervention, as well as exemplifying the use of AI tools in courts in Brazil, Spain, Colombia, Peru, the United States and the Netherlands.

Keywords: artificial intelligence, justice, large language models (LLM), AI regulation, legaltech.

1. Introducción

La inteligencia artificial (IA) dejó de ser un concepto técnico reservado para especialistas para convertirse en un asunto de interés en múltiples disciplinas, incluido el sector justicia. Esta discusión enfrenta varios retos, entre ellos, la brecha de conocimiento sobre las herramientas tecnológicas, la cual genera desconfianza y escepticismo, y bloquea su implementación por considerar que son una amenaza para los principios esenciales de la justicia. En contraste, también incentiva posturas excesivamente optimistas que pueden fomentar un uso inadecuado de la IA, viéndola como una solución mágica para problemas fuera de su alcance o que pueden resolverse a través de otros mecanismos. Cabe aquí el refrán “*para un martillo, todo es un clavo*”.

Con el fin de aportar en esta conversación, este escrito presenta el panorama actual de la incorporación de la IA en la justicia. Para esto, primero presentará los conceptos básicos para entender la inteligencia artificial y algunos hitos de su evolución. Luego, relacionará algunos de los usos en la administración de justicia y, finalmente, esbozará algunos de los avances en su regulación, que ha tenido un ritmo disonante de la expansión acelerada de la IA en los últimos años. Los ejemplos tendrán énfasis en Colombia, dado que es el contexto que las autoras conocen de manera más directa.

2. Una aproximación a la inteligencia artificial

A pesar de que el término inteligencia artificial parece reciente, su origen se remonta al siglo pasado (ver Ilustración 1). En esencia, la IA se define como la capacidad que tiene una máquina para replicar tareas cognitivas que son realizadas por los humanos (analizar, resumir, realizar operaciones matemáticas, por ejemplo). Estas capacidades de las máquinas se logran mediante algoritmos, esto es, la secuencia de pasos que una máquina sigue para llevar a cabo una tarea específica.

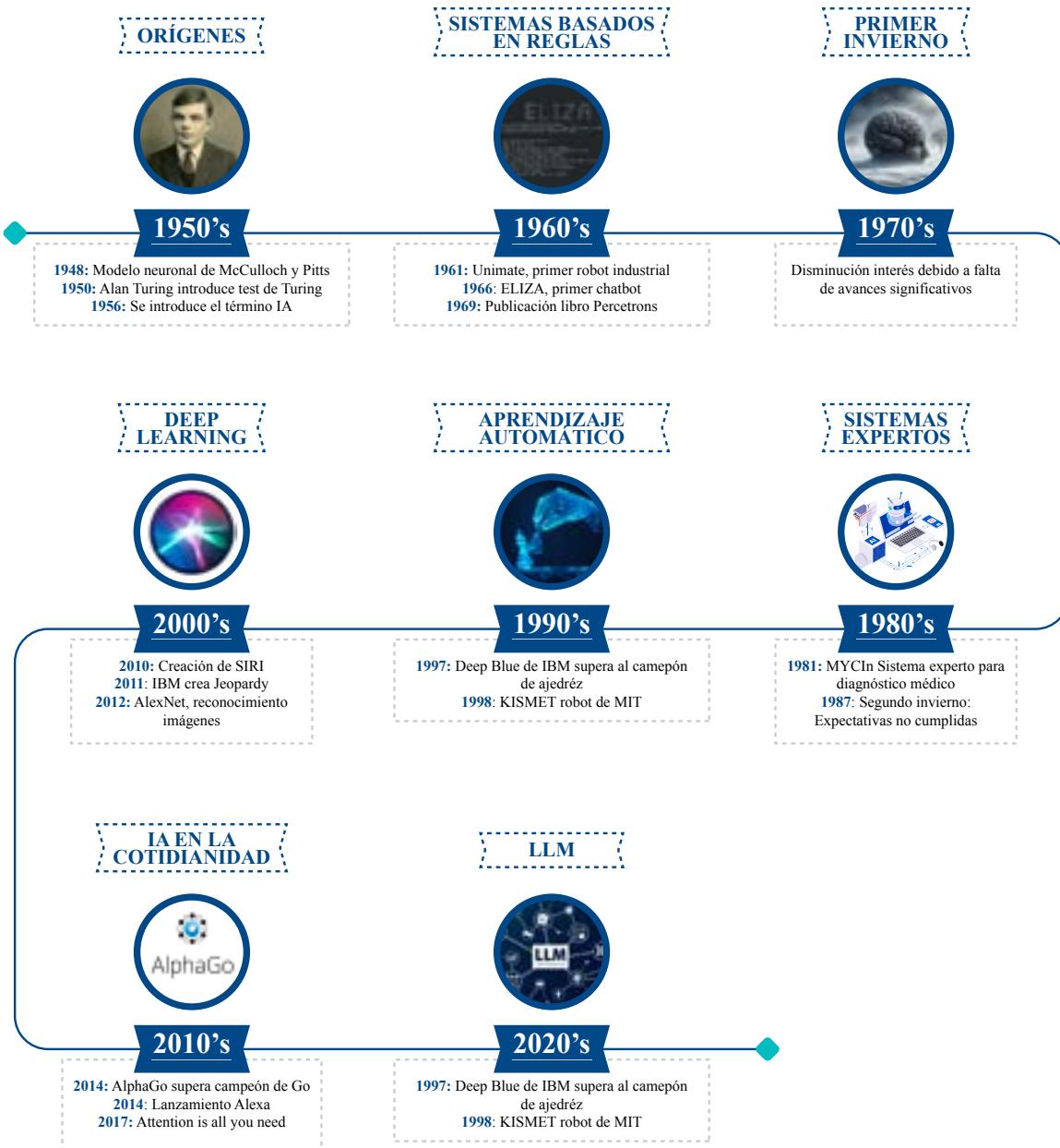


Ilustración 1 – Línea de tiempo de la IA

La IA tiene aplicación en campos diversos, como el procesamiento de lenguaje natural (*Natural Language Processing*, NLP) y la visión por computadora. El NLP, permite a las máquinas “aprender” a entender el lenguaje hablado y escrito por los humanos. Esta capacidad ha permitido, por ejemplo, traducir automáticamente textos o audios de un idioma, la generación de textos coherentes, la categorización de información o la elaboración de resúmenes. Por otro lado, la visión por computadora se centra en analizar y extraer información relevante

de videos o imágenes. Estos avances son aplicados en el reconocimiento facial, diagnósticos médicos basados en imágenes, la conducción de vehículos autónoma, entre otros.

a. Tipos de algoritmos de IA

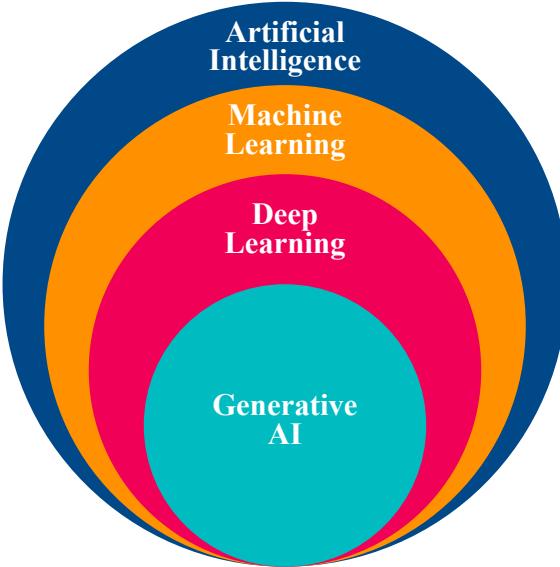
Desde una perspectiva técnica, la IA se basa en diferentes tipos de algoritmos, con características y niveles de complejidad diversos. Para comenzar, en sentido amplio, los sistemas expertos basados en reglas dominaron los primeros años de la IA. Estos sistemas les daban a las máquinas unas reglas lógicas predefinidas para que tomaran decisiones, por lo tanto, dependían completamente de la programación y los pasos indicados por su programador.

Con el tiempo, la aparición de algoritmos matemáticamente más complejos, como los algoritmos de *machine learning*, y una mayor capacidad computacional se amplió significativamente el espectro de posibilidades de la IA. Las máquinas pasaron de simplemente ejecutar instrucciones a “aprender” de los datos, permitiéndoles identificar patrones, realizar predicciones y tomar decisiones. Los algoritmos de *machine learning* se organizan en una taxonomía según el tipo de resultado esperado, siendo las clasificaciones más comunes las siguientes:

Aprendizaje supervisado: el algoritmo se entrena utilizando un conjunto de datos etiquetados, es decir, datos de entrada (input) que están asociados a una salida esperada o variable objetivo (output). Un ejemplo común es la detección de correos *spam*, donde el modelo aprende de ejemplos previos etiquetados.

Aprendizaje no supervisado: no requiere de datos etiquetados. Su objetivo es identificar patrones o estructuras en los datos de manera autónoma. Un ejemplo es la segmentación de documentos en grupos con características similares, sin que se les haya asignado previamente una categoría.

Aprendizaje por refuerzo: en este tipo de aprendizaje la máquina toma decisiones basadas en la interacción con su entorno, recibiendo retroalimentación en forma de recompensas o penalizaciones. Esta tipología es ampliamente usada en soluciones de robótica o videojuegos (Ayodele, 2010).



Fuente: Amol Wagh. (2023)⁸

b. El aprendizaje profundo y los *transformers*

Una subespecialidad del *machine learning* es el aprendizaje profundo, en el que las redes neuronales artificiales, un tipo especial de algoritmo, cobran especial relevancia. Las redes neuronales tienen sus orígenes en la década de 1950⁹, pero con el paso del tiempo, pasaron a ser más complejas, permitiendo procesar enormes cantidades de señales de entrada (inputs), a través de múltiples capas ocultas. Estas capas se conectan por diversas posibilidades de funciones de activación, ampliando la posibilidad de aprendizaje de los algoritmos a representaciones complejas.

Aunque el mundo de las redes neuronales es enorme, uno de los avances más revolucionarios en el área aparece en el año 2017, con el artículo: *Attention is All You Need*. Este artículo introduce una arquitectura conocida como *transformer*, el cual amplió las posibilidades de procesamiento de datos, y sentó las bases requeridas para los modelos de inteligencia artificial generativa actuales, como Copilot, ChatGPT, Gemini entre otros. Los *transformers* han permitido la mejora de tareas como la generación de texto, la traducción automática, o la realización de tareas semánticas complejas (Vaswani, A. et al., 2023).

⁸ <https://medium.com/@amol-wagh/whats-generative-ai-explore-underlying-layers-of-machine-learning-and-deep-learning-8f99272e0b0d>

⁹ Esto ocurrió con la aparición del Perceptrón, basado en el trabajo de McCulloch y Pitts.

c. Modelos grandes de lenguaje (LLM)

Los modelos grandes de lenguaje o LLM por sus siglas en inglés (*Large Language Models*) son una reciente evolución de los modelos basados en la arquitectura de *transformers*. Estos modelos son entrenados con volúmenes enormes de datos textuales, conocidos como *corpus*, que pueden incluir millones de documentos provenientes de diversas fuentes, como contenido de internet, libros, artículos, etc. Los datos deben pasar por un proceso riguroso de “curado” con el propósito de que los datos sean relevantes, diversos, no sesgados, entre otras características. Este proceso busca identificar y eliminar contenido inapropiado, para tratar de mitigar los riesgos asociados a la reproducción de sesgos de los datos originales.

La infraestructura computacional necesaria para entrenar uno de estos modelos es enorme, debido a que implica estimar miles de millones de parámetros. Esto exige el uso de hardware especializado, como unidades de procesamiento gráfico (GPUs) o unidades tensoriales (TPUs), junto con semanas o incluso meses de procesamiento en servidores de altísima capacidad. Por esta razón, el desarrollo y entrenamiento de modelos de lenguaje a gran escala (LLM) suele estar limitado a grandes empresas tecnológicas o instituciones de investigación que cuentan con recursos significativos para su implementación. Sin embargo, estas tecnologías se han puesto al alcance de los usuarios finales a través de herramientas de IA generativa, como ChatGPT, disponibles en versiones gratuitas o de pago, lo que ha facilitado su masificación.

Los modelos probabilísticos en los que se basan estas tecnologías pretenden crear una secuencia de palabras que siempre parecerá tener un sentido y lógica, pero que puede ser errónea o puede ser una alucinación, es decir, información inventada o ficticia. La calidad del *prompt*, es decir, la instrucción que se le da a la herramienta para que arroje una respuesta, puede ayudar a tener mejores resultados; pero no garantiza que sean correctos.

d. Sesgos algorítmicos

Un sesgo algorítmico es la tendencia de un sistema de IA a generar sistemáticamente resultados desbalanceados o que desfavorecen a ciertos grupos de personas. Esto puede ser el resultado de las posturas, estereotipos y prejuicios que hay en los datos de entrenamiento o en el diseño mismo del algoritmo. Un ejemplo es un sistema que desfavorece a la población afrodescendiente en el otorgamiento de créditos o a las mujeres en los procesos de selección laboral.

Existen medidas para mitigar los sesgos algorítmicos, entre ellas: (i) la creación benchmarks que evalúen la neutralidad, ausencia de lenguaje potencialmente dañino, entre otros criterios éticos de los diferentes modelos, (ii) mejorar el curado de los datos, eliminando o equilibrando fuentes que generen sesgos y (iii) intervención humana, con el fin de incluir auditores que revisen las salidas de los modelos.



3. Aplicaciones de la IA en el sistema de justicia

Este breve recuento de la IA desde el punto de vista técnico se puede resumir en cuatro puntos: i) La IA no es un concepto nuevo, pero su aplicación se ha venido incrementando debido a una mayor disponibilidad de datos, mejoras en la capacidad computacional y la evolución de algoritmos. ii) Existen varias técnicas de IA, pero ninguna de ellas es infalible, tienen el riesgo de incurrir en errores. iii) Los modelos de IA pueden incorporar sesgos que afecten derechos fundamentales. iv) Con el lanzamiento de los LLM, la IA dejó de ser un espacio restringido a las herramientas que dispongan los poderes judiciales, existe un abanico de alternativas a disposición de todos los servidores judiciales y abogados.

A medida que ha ocurrido esta evolución de la IA, las expectativas en el uso de la inteligencia artificial en la justicia también se han incrementado. Estas se han orientado a facilitar el acceso, mejorar la seguridad jurídica y, principalmente, a lograr una mayor eficiencia. Este abanico de posibilidades no está solo en el plano teórico. En los últimos años varios sistemas judiciales han comenzado a implementar sistemas de IA en su gestión, bien sea con herramientas desarrolladas por la misma organización (*in-house*) o adquiridas en el mercado. Estos son algunos ejemplos:

a. Clasificación y agrupación de información

Para facilitar y mejorar la precisión en la revisión de los miles de documentos que cada día se tramitan en los despachos judiciales algunos tribunales han creado herramientas basadas en IA que ayudan en la clasificación de los contenidos, antes de que sean leídos por una persona. Así, por ejemplo, el Tribunal Superior de Brasil desarrolló la herramienta Víctor, que ayuda a identificar dentro de los recursos que envían los ciudadanos aquellos que cumplen con el requisito de repercusión general y, posteriormente, VictorIA, una herramienta que ayuda a identificar patrones entre las demandas para tramitarlas de manera conjunta (Supremo Tribunal Federal, 2023).

En España, se integró a LexNET, la plataforma para el intercambio de información entre órganos judiciales y abogados, una funcionalidad para la catalogación de documentos (Ministerio de la Presidencia, Justicia y Relaciones con las Cortes de España, 2024). Por su parte, la Corte Constitucional de Colombia desarrolló PretorIA, una herramienta que ayuda a identificar ciertas categorías (derechos tutelados, sujetos de especial protección, región de origen, por ejemplo) dentro de las cerca de 4.000 tutelas que deben analizarse diariamente para decidir cuáles seleccionará para su revisión (Catellanos Padilla, 2022).

b. Consulta de jurisprudencia

La IA también se viene aplicando para facilitar la investigación jurídica, a través de búsquedas el lenguaje natural, relacionamiento de sentencias o sugerencias de jurisprudencia

para resolver casos específicos. En el sector privado existen varias, como Vincent¹⁰, Ariel¹¹ o Westlaw¹², pero también el Poder Judicial ha dado pasos en este sentido. Por ejemplo, el buscador de jurisprudencia de la Corte Constitucional de Colombia¹³ cuenta con funcionalidades de IA para identificar relaciones de citación entre providencias, recomendar sentencias similares y graficar esas conexiones.

c. Edición de texto

La inteligencia artificial puede apoyar la edición de textos y, por esa vía, contribuir al lenguaje claro. El sector público ya cuenta con aplicaciones específicas para esta funcionalidad. Así, por ejemplo, la Ciudad de Buenos Aires desarrolló “Clappi” para adaptar textos legales y administrativos a lenguaje claro (Gobierno de la Ciudad Autónoma de Buenos Aires, 2023). En el sector justicia no se identificaron modelos acotados específicamente para este fin, pero es una funcionalidad que se puede realizar a través de cualquiera de los LLMs disponibles en el mercado (ChatGPT, Claude, Gemini, etc.). Muchos servidores públicos ya están utilizando esta tecnología, aunque no siempre se haga explícito su uso. En Colombia, se volvió famoso un caso de la Corte Suprema de Justicia en el que, debido a un error, quedó en evidencia el uso de IA en una sentencia, ya que no se eliminó una parte de la interacción con la herramienta antes de publicarla (Forero, 2025).

d. Transcripción y borradores de actas de audiencias

Las reformas judiciales de las últimas décadas incentivaron la oralidad como estrategia para fortalecer principios como la inmediación y la contradicción. Esta transformación trajo consigo la necesidad de elaborar actas y, en algunos casos, de hacer transcripciones para surtir las segundas instancias, acarreando una carga de trabajo importante en los despachos.

Con la pandemia en 2020 y la consecuente masificación de las audiencias virtuales estas tareas se han facilitado, pues los mismos aplicativos en los que se realizan las audiencias permiten realizar transcripciones y resúmenes que facilitan la elaboración de las actas. En Colombia, por ejemplo, este ha sido uno de los casos de uso de la herramienta Copilot de Microsoft, en el marco de ejercicios de experimentación liderados por el Consejo Superior de la Judicatura (Quiñones, 2023).

e. Cálculos matemáticos

Otro de los usos de la IA sido el cálculo de plazos, pensiones o cesantías, usos de los cuales

10 <https://vlex.es/vincent-ai>

11 <https://www.arielapp.co>

12 <https://legal.thomsonreuters.com/en/products/westlaw-edge>

13 <https://www.corteconstitucional.gov.co/relatoria/>

han reportado los jueces en sus decisiones. Tal es el caso de un juzgado civil de Perú, que consultó a ChatGPT la forma de calcular la cuota alimentaria de una menor de edad teniendo en cuenta los ingresos de los padres (Gutierrez, 2024). En Colombia un despacho del Tribunal Superior de Medellín empleó la herramienta Copilot para contabilizar los días hábiles transcurrieron en un periodo de tiempo (Tribunal Superior de Distrito Judicial de Medellín, Sala Civil de Decisión, Auto 2024-117, 2024).

f. Búsqueda de términos y conceptos

Las herramientas generativas han sido consideradas como un recurso para buscar conceptos o interpretar términos relevantes en la solución de casos concretos. En un voto concurrente, un juez Tribunal de Apelaciones del 11º Circuito de Estados Unidos analizó el uso de herramientas basadas en modelos de lenguaje a gran escala (LLMs) para interpretar el significado del uso común de las palabras, como el caso del término “paisajismo”, que era clave para resolver un conflicto con una aseguradora. El juez destacó la utilidad de estas herramientas, pero también señaló los riesgos asociados a su uso, como las alucinaciones y errores (Weiss, 2024).

En Países Bajos, un juez utilizó ChatGPT para calcular la vida útil de paneles solares y el precio de la electricidad, contribuyendo así a la resolución de una disputa económica (Flórez Rojas, 2024). Y, en Colombia, el Tribunal Superior de Distrito de Pereira, que estudiaba el caso de una muerte en accidente de tránsito, empleó ChatGPT para consultar a cuántas copas de vino, aguardiente, ron o cerveza equivalen 20 miligramos sobre 100 mililitros de etanol en la sangre. La pregunta estaba orientada a determinar si era una presencia de alcohol significativa, elemento que era relevante para determinar la autopuesta en peligro de la víctima (Boada Acosta, J. C., & Márquez, M. L., 2024).

g. Predicción de riesgo de reincidencia

En el ámbito penal, la inteligencia artificial se ha utilizado para evaluar la probabilidad de reincidencia en delitos o el incumplimiento de términos de libertad condicional. Un ejemplo destacado es COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) en Estados Unidos, cuyo algoritmo considera factores como el historial criminal y el entorno social. Con un propósito similar, la Fiscalía de Colombia desarrolló en 2017 un sistema llamado PRISMA¹⁴, que no llegó a escalar de su fase piloto.

Como se puede observar, la inteligencia artificial ya es una realidad y se ha ido incorporando progresivamente en diferentes etapas del proceso judicial. Sin embargo, su uso también

14 Fiscalía General de la Nación. Herramienta Prisma: Perfil de riesgo de reincidencia para la solicitud de medidas de aseguramiento. <https://www.fiscalia.gov.co/colombia/wp-content/uploads/Perfil-de-riesgo-de-reincidencia-para-solicitudes-de-medida-de-aseguramiento.pdf>

conlleva diversos riesgos. Uno de los más significativos está relacionado con las aplicaciones que influyen en decisiones que pueden impactar derechos fundamentales. Un ejemplo destacado es el uso de herramientas para la predicción del riesgo de reincidencia, donde los sesgos algorítmicos pueden perjudicar a poblaciones históricamente discriminadas. Esto quedó evidenciado en el caso de COMPAS, un sistema que, aunque asegura no incluir información sobre raza en su algoritmo, investigaciones han demostrado que tiende a clasificar erróneamente a una mayor proporción de personas negras como de alto riesgo de reincidencia.

Pero incluso otros usos de la inteligencia artificial que parecen inofensivos pueden presentar complicaciones. Un ejemplo es la aplicación en la búsqueda de información jurídica. Por un lado, la IA no siempre logra captar las sutilezas del derecho, que resultan esenciales al momento de construir una línea jurisprudencial. Por otro, pueden existir decisiones contradictorias sobre un mismo tema que la IA no identifica, lo que conlleva el riesgo de priorizar una posición mientras se invisibilizan otras. Además, en la IA generativa —al menos hasta el momento— persiste la imposibilidad de eliminar por completo las alucinaciones, incluso cuando la base de datos ha sido rigurosamente depurada y controlada. Un estudio reciente de profesores de la Universidad de Standford dejó en evidencia esta situación, al analizar algunas herramientas que, aun cuando contaban con datos acotados, alucinaron entre un 17% y un 33% (Varun Magesh et al., 2024).

También está en duda el potencial de la IA para ahorrar tiempos. Un ejemplo de esto es su uso para la elaboración de cálculos matemáticos, en el que se pueden generar resultados erróneos que obliguen a probar distintos *prompts* antes de lograr un resultado correcto, por lo que pue-
de ser más confiable y eficiente el uso de herramientas tradicionales como Microsoft Excel¹⁵.

En conclusión, la inteligencia artificial tiene el potencial de fortalecer el sistema de justicia mediante diversas herramientas que optimizan tareas administrativas, facilitan el análisis de información y permiten abordar estratégicamente la carga judicial. Sin embargo, cada aplicación debe ser evaluada minuciosamente, ya que un uso inadecuado podría comprometer la calidad y la imparcialidad de la justicia o no generar el impacto esperado en la eficiencia.

4. La regulación de la IA

El florecimiento de la IA en el mundo ha surgido en medio de una regulación escasa. Si bien existen desde hace varios años instrumentos de *soft law* con principios generales, aún son pocos los avances legislativos. Además, el tiempo que toma un trámite de esta naturaleza la posibilidad de que se produzcan legislaciones obsoletas, que no respondan a las necesidades

15 Al respecto se recomienda ver el Foro: Chat GPT en el Derecho, organizado por el Instituto Colombiano de Derecho Procesal, en el que la Mónica Hidalgo, Magistrada de Tribunal de Distrito Judicial de Cali (Colombia) mostró varios ejemplos de errores en los que puede incurrir la herramienta en realización de cálculos: https://www.youtube.com/watch?v=OidyVkJFshY&list=PL6sXY0gYuQtAotFvkUqE_sfzzn151jS8K&index=2

y retos del contexto, pues la tecnología evoluciona a un ritmo que es difícil de seguir.

Estos debates no son nuevos ni se circunscriben a la administración de justicia. Sin embargo, con la llegada de los LLM y la posibilidad de que cualquier funcionario judicial utilice herramientas de IA, sin depender de desarrollos propios y controlados de los poderes judiciales, han puesto sobre la mesa la urgencia de una regulación específica para la administración de justicia.

Varios ejemplos corroboran esta necesidad. Así, al poco tiempo de que ChatGPT se puso a disposición del público y mostró su “magia”, comenzaron a aparecer casos de uso de varios tribunales que pusieron en evidencia sus riesgos. Algunos de los más sonados se relacionaron con la búsqueda de jurisprudencia para sustentar posiciones en documentos jurídicos. Los resultados no siempre han sido afortunados, pues salieron a la luz pública casos en los que se incluyeron sentencias inventadas o con contenidos distintos a los citados. Uno de estos fue el caso Mata Vs. Avianca en Estados Unidos, en el que uno de los abogados presentó un escrito que se sustentaba en decisiones judiciales inexistentes (Forbes Colombia, 2023).

Para enfrentar estas malas prácticas, tribunales y órganos de gobierno del Poder Judicial han comenzado a elaborar directrices sobre el uso de la IA, especialmente la de tipo generativo. Dentro de estos se encuentra el Reino Unido¹⁶ que fue pionero en expedir una guía para el uso responsable de la IA en cortes y juzgados en 2023. En otros casos han sido los jueces los que han empezado a definir algunas pautas para el uso de la IA en el sistema judicial. Este fue la situación de Colombia, que se concretó en la novedosa Sentencia T-323 de 2024¹⁷ de la Corte Constitucional.

El problema jurídico que resolvió la Corte fue la presunta vulneración del debido proceso por el uso de ChatGPT por parte de un juez para consultar jurisprudencia sobre la exoneración de cuotas moderadoras en las terapias de un menor con espectro autista. La Corte concluyó que, en el caso concreto, no se configuró esta vulneración, ya que el uso de la IA ocurrió después de que la decisión había sido tomada. Sin embargo, para llegar a esta conclusión, realizó un análisis general sobre el tema, estableció criterios orientadores para el uso de la IA y ordenó al Consejo Superior de la Judicatura elaborar lineamientos específicos.

La columna vertebral de la argumentación de la Corte fue la garantía del derecho al debido proceso, que incluye la figura del juez natural, que debe ser humano y no una máquina. Por ello, esta garantía se vulnera cuando un “funcionario judicial utiliza la IA para sustituir el razonamiento lógico y humano que le compete realizar a efectos de interpretar los hechos, las pruebas, motivar la decisión e incluso adaptarla”. Por el contrario, se consideraron admisi-

16 Courts and Tribunals Judiciary of United Kingdom. Artificial Intelligence (AI). Guidance for Judicial Office Holders 12 December 2023. <https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf>

17 <https://www.corteconstitucional.gov.co/relatoria/2024/T-323-24.htm>

bles otros usos de la IA, como el apoyo en la gestión administrativa y documental, la gestión judicial (siempre que no implique la generación de contenido, interpretación o resolución de casos) y la corrección y síntesis de textos. En estos eventos, se estableció que deben cumplirse requisitos como la supervisión humana de los resultados, la protección del *habeas data* y la garantía de transparencia e intimidad. En cumplimiento de la orden dada en la sentencia, el Consejo Superior de la Judicatura expidió, en diciembre de 2024, un Acuerdo que regula el uso responsable de la IA en la Rama Judicial (Consejo Superior de la Judicatura, Acuerdo PCSJA24-12243, 2024).

5. Conclusiones

Aunque la IA no es un concepto nuevo, experimenta un crecimiento vertiginoso. Este auge ha sido impulsado por el avance del *machine learning*, facilitado inicialmente por la acumulación masiva de datos en Internet y, más recientemente, por la democratización del uso a través de modelos de lenguaje a gran escala (LLM).

Esta evolución ha ocurrido a un ritmo desacompasado de la formación y la regulación, generando riesgos potenciales para el debido proceso, la intimidad y la calidad de la administración de justicia. Los reguladores, tanto legislativos como judiciales, han comenzado a asumir este desafío mediante proyectos de ley y lineamientos de uso, que aún se encuentran en fases tempranas de desarrollo.

El panorama actual implica un reto local y global: avanzar en el aprovechamiento de la tecnología con responsabilidad. Este desafío requiere un diálogo entre tecnooptimistas y tecnófobos y entre las distintas disciplinas que intervienen en este proceso, para lograr un equilibrio que permita aprovechar el potencial transformador de la inteligencia artificial, sin afectar los fundamentos esenciales del sistema judicial.



6. Referencias

Al-Mekhlal, M., & Ali Khwaja, A. (2019). A synthesis of big data definition and characteristics. *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 314–322. <https://doi.org/10.1109/cse/euc.2019.00067>

Ayodele, T. O. (2010). Types of machine learning algorithms. *New advances in machine learning*, 3 (19-48), 5-1.

Boada Acosta, J. C. ., & Márquez, M. L. . (2024). Un buen ejemplo de cómo no usar ChatGPT en la administración de justicia. Comentario a la Sentencia de la Sala Penal del Tribunal Superior de Pereira del 3 de abril de 2024. Radicado 76 001 60 00193 2013 80734 01. Nuevo Foro Penal, 20(103). Recuperado a partir de <https://publicaciones.eafit.edu.co/index.php/nuevo-foro-penal/article/view/8213>

Castellanos Padilla, R. (2022). Inteligencia artificial en el sector de la justicia. Universidad de los Andes. Disponible en: <http://hdl.handle.net/1992/64175>

Cepeda, M. J. y Otálora, G. (2020). *Modernización de la administración de justicia través de la inteligencia artificial*. Bogotá: Fedesarrollo, julio, 67 p. Disponible en: <https://repository.fedesarrollo.org.co/handle/11445/3980#.XynSvUzObZQ.twitter>

Consejo Superior de la Judicatura. Acuerdo PCSJA24-12243 de 2024. https://actosadministrativos.ramajudicial.gov.co/GetFile.ashx?url=%7e%2fApp_Data%2fUpload%2fPCS-JA24-12243.pdf

Corte Constitucional de Colombia. (2024). Sentencia T-323-24 <https://www.corteconstitucional.gov.co/relatoria/2024/T-323-24.htm>

Forero, P. [@PamelaForeroB]. (2025, 14 de abril). [Contenido del tuit] [Publicación en X]. <https://x.com/pamelaforerob/status/1846345000243662883>

Flórez Rojas, M. L. (2024). La IA en la toma de decisiones judiciales: Un comentario sobre dos decisiones recientes (Colombia/Países Bajos). Foro Administración, Gestión y Política Pública. <https://forogpp.com/2024/08/15/la-ia-en-la-toma-de-decisiones-judiciales-un-comentario-sobre-dos-decisiones-recientes-colombia-paises-bajos/>

<https://forogpp.com/2024/08/15/la-ia-en-la-toma-de-decisiones-judiciales-un-comentario-sobre-dos-decisiones-recientes-colombia-paises-bajos/>

Fiscalía General de la Nación. Herramienta Prisma: Perfil de riesgo de reincidencia para la solicitud de medidas de aseguramiento. <https://www.fiscalia.gov.co/colombia/wp-content/uploads/Perfil-de-riesgo-de-reincidencia-para-solicitudes-de-medida-de-aseguramiento.pdf>

Forbes Colombia. Abogado usó ChatGPT para demandar a Avianca, pero todo salió mal. <https://forbes.co/2023/06/09/actualidad/abogado-uso-chatgpt-para-demandar-a-avianca-pero-todo-salio-mal>

Gutiérrez, J. D. (2024). “[Critical appraisal of Large Language Models in judicial decision-making](#)”. In R. Paul, E. Carmel, and J. Cobbe (eds), *[Handbook on Public Policy and Artificial Intelligence](#)*, Edward Elgar Publishing, UK.

Gobierno de la Ciudad Autónoma de Buenos Aires. (2023). Clappi: Corrector de Lenguaje Administrativo para Políticas Públicas Inclusivas. <https://buenosaires.gob.ar/que-es-lenguaje-claro/clappi>

Julia Dressel, Hany Farid , The accuracy, fairness, and limits of predicting recidivism.Sci. Adv.4,eaao5580(2018).DOI:10.1126/sciadv.aa05580

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>

Ministerio de la Presidencia, Justicia y Relaciones con las Cortes de España. (2024, 7 de mayo). Implementación de la Inteligencia Artificial en LexNET. Servicio Público de Justicia. <https://www.administraciondejusticia.gob.es/w/implementacion-de-la-inteligencia-artificial-en-lexnet>

Quiñones, M. V. [@MariaVictoriaQuiñones]. (2023, 18 de diciembre). Copilot en audiencias judiciales celebradas por Microsoft Teams [Video]. YouTube. <https://www.youtube.com/watch?v=jiCG7yf-EBs>

Supremo Tribunal Federal. (2023, 14 de abril). STF finaliza testes de nova ferramenta de Inteligência Artificial. <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=507120&ori=1>

Tribunal Superior de Distrito Judicial de Medellín, Sala Civil de Decisión. Auto 2024-117. Proceso: 05001220300020240015100.

Urueña, R. y Ángel N. (editores). *Derecho, poder y datos : aproximaciones críticas al derecho y las nuevas tecnologías*. Primera edición., Universidad de los Andes, Facultad de Derecho, Ediciones Uniandes, 2024.

Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning, and Daniel E. Ho. 2024. Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools. https://dho.stanford.edu/wp-content/uploads/Legal_RAG_Hallucinations.pdf

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention is all you need*. arXiv. <https://arxiv.org/abs/1706.03762>

Weiss, D. C. (2024, 6 de junio). In concurrence confession, appeals judge says ChatGPT research ‘less nutty’ than feared. ABA Journal. <https://www.abajournal.com/news/article/appeals-judge-makes-a-confession-he-consulted-chatgpt-and-found-the-results-less-nutty-than-i-feared>

*Las opiniones aquí expresadas solo comprometen a las autoras.

**En la elaboración de este artículo se utilizaron ChatGPT 4.0 y Claude para la edición de texto.



MinerIA de Datos

Inteligencia Artificial y sus
Implicaciones para la
Privacidad y la Protección de
Datos Personales en Colombia

MARÍA PAULA ÁNGEL ARANGO



Maria Paula Ángel

María Paula Ángel Arango es Doctora en Derecho por la Universidad de Washington, Seattle. Es abogada y politóloga de la Universidad de los Andes, y cuenta con una Maestría en Derecho Administrativo de la Universidad del Rosario. Actualmente se desempeña como Resident Fellow del Information Society Project (ISP) en la Universidad de Yale.

Ha trabajado como investigadora en la línea de Transparencia e Intimidación de Dejusticia y como asistente de investigación en el Tech Policy Lab de la Universidad de Washington. También ha realizado pasantías de verano en el Open Technology Institute (OTI) de la organización New America y en el Social Media Collective de Microsoft Research.

En el ámbito académico, ha sido docente en programas de posgrado como la Especialización en Derecho Administrativo de la Universidad del Rosario, la Especialización en Derecho de las Fintech de la Universidad Externado de Colombia y la Maestría en Innovación en Derecho Digital y Legal Tech de la Universidad Sergio Arboleda.

María Paula fue becaria Fulbright-Colciencias entre 2019 y 2023. En 2020 recibió el Westin Scholar Award otorgado por la Asociación Internacional de Profesionales de la Privacidad (IAPP) y fue seleccionada como Public Voices Fellow on Technology in the Public Interest durante 2023-2024. En 2023 fue reconocida por Women in AI Ethics como una de las ocho “Estrellas en ascenso en ética de la IA” para 2024.

MinerIA de Datos: Inteligencia Artificial y sus Implicaciones para la Privacidad y la Protección de Datos Personales en Colombia

Por María Paula Ángel Arango¹⁸

Resumen

El artículo explora las implicaciones de la Inteligencia Artificial (IA) en la privacidad y la protección de datos personales en Colombia, en el contexto de los retos de gobernanza que surgieron de la Cumbre Ministerial Latinoamericana y del Caribe por la IA “ColombIA” de 2024. Se argumenta que la IA facilita e incentiva la recopilación masiva de datos personales y la generación de datos inferidos, lo que amenaza los derechos a la privacidad y la protección de datos personales, así como otros derechos fundamentales. Se destaca el papel de la Superintendencia de Industria y Comercio (SIC) en la búsqueda de mecanismos para responsabilizar a los actores públicos y privados que tratan datos personales, y se exploran los posibles daños a la privacidad, la protección de datos, la igualdad, la autonomía y la democracia, derivados del uso de la IA.

Palabras clave: Inteligencia Artificial, Privacidad, Protección de Datos, Daños, SIC.

Abstract

The article explores the implications of Artificial Intelligence (AI) on privacy and personal data protection in Colombia, in the context of the governance challenges that emerged from the Latin American and Caribbean Ministerial Summit for AI “ColombIA” of 2024. It is argued that AI facilitates and encourages the massive collection of personal data and the ge-

18 Doctora en Derecho de la Universidad de Washington (UW). Resident Fellow en el *Information Society Project (ISP)* de la Universidad de Yale. Entre 2015 y 2019 fue investigadora de la línea de Transparencia e Intimidad de Dejusticia. Entre 2019 y 2024 fue investigadora asistente del *UW Tech Policy Lab*. Durante su doctorado, María Paula también trabajó en el *Open Technology Institute (OTI)* de la organización *New America* y en el *Social Media Collective of Microsoft Research*. Ha sido profesora de cátedra de la Especialización en Derecho Administrativo de la Universidad del Rosario, de la Especialización en Derecho de las Fintech de la Universidad Externado de Colombia y de la Maestría en Innovación en Derecho Digital y Legal Tech de la Universidad Sergio Arboleda. Fue becaria Fulbright-MinCiencias 2019, ganadora del premio *IAPP Westin Scholar Award 2021* y *Public Voices Fellow on Technology in the Public Interest 2023-2024*. *Women in AI Ethics* la seleccionó como una de las ocho *Rising Stars in AI Ethics-2024*.

neration of inferred data, which threatens the rights to privacy and personal data protection, as well as other fundamental rights. It highlights the role of the Superintendency of Industry and Commerce (SIC) in seeking mechanisms to hold public and private actors that process personal data accountable, and explores the potential harms to privacy, data protection, equality, autonomy and democracy arising from the use of AI.

Keywords: Artificial Intelligence, Privacy, Data Protection, Tort, SIC.

1. Introducción

En agosto de 2024, Colombia fue el país anfitrión de la Cumbre Ministerial Latinoamericana y del Caribe por la Inteligencia Artificial (IA) “ColombIA”, durante la que se buscó consolidar un marco de colaboración regional en tecnologías de IA (MinTIC, 2024). Para el efecto, los ministros, legisladores, expertos en tecnología y líderes de la industria en la región que se reunieron en Cartagena discutieron tanto las oportunidades de productividad y eficiencia como los retos de gobernanza que la IA plantea para la región.

Este artículo pretende abordar uno de esos retos de gobernanza, a saber: los posibles impactos de la IA en la privacidad y la protección de datos personales. Como se verá, la IA no solo facilita, sino que incentiva de manera significativa la recopilación masiva de datos personales y la generación de datos inferidos. En esa medida, su implementación amenaza tanto los derechos a la privacidad y la protección de datos personales como otros derechos que se pueden ver afectados debido a la disponibilidad masiva de datos personales en manos de terceros y al poder informático reforzado (Angel, 2020) que estos le otorgan a quienes los tratan.

Ante este panorama, un régimen de protección de datos personales exclusivamente dirigido a empoderar a los ciudadanos para que tomen el control de sus datos parece quedarse corto. Por ello, en los últimos años la Superintendencia de Industria y Comercio (SIC)—en su calidad de autoridad de protección de datos de Colombia—ha buscado establecer obligaciones de responsabilidad demostrada que hagan rendir cuentas a los actores públicos y privados que tratan datos personales. Así mismo, la SIC podría estar abriendole la puerta a un tercer nivel de medidas—aún incipientes pero prometedor—, encaminado a restarle poder a quienes tratan los datos personales.

2. ¿Qué hace a la Inteligencia Artificial diferente?

Evidentemente, la IA no es la primera tecnología digital a la que nos enfrentamos. En general, las nuevas tecnologías—piénsese por ejemplo en los computadores, los teléfonos inteligentes, o el Internet—presentan retos importantes para la privacidad y la protección de datos personales, en la medida en la que *hacen más fácil y eficiente la recolección de información personal*. Gracias a estas tecnologías, y más específicamente a los sensores, sistemas de GPS, cookies, etc. que ellas incluyen, hoy en día no requerimos del esfuerzo humano para

recolectar datos personales. Mientras escribo estas páginas en Microsoft Word, por ejemplo, el software está recopilando información sobre mi ubicación, el idioma en el que escribo, las herramientas de Word que utilizo, entre otros (Microsoft, 2024). En alguna medida, se podría llegar a decir que, en la era digital, la producción masiva de datos personales se da como una “externalidad” o efecto secundario de nuestras actividades diarias.

Sin embargo, la IA representa un cambio adicional en la relación entre tecnologías digitales, privacidad y protección de datos personales. Como sabemos, el rápido desarrollo de la IA “moderna” se basa principalmente en el uso de algoritmos de “machine learning” o aprendizaje automático (Kissinger & Schmidt, 2022). A diferencia de los algoritmos clásicos, que codifican la experiencia humana en reglas, los algoritmos de “machine learning” requieren de grandes bases de datos—conocidas comúnmente como datos de entrenamiento—para poder encontrar patrones e inferir así sus propias reglas. En consecuencia, la IA no sólo depende de la existencia de grandes cantidades de datos digitales, sino que *incentiva su recolección*. Si no se recogen datos, los algoritmos de machine learning en los que se basa la IA no tienen cómo “aprender” y de esa manera, mejorar su rendimiento. Es por esa razón por la que, hoy aún más que antes, las grandes compañías de tecnología están permanentemente buscando nuevas formas de recoger más datos (Shrivastava, 2024).

Adicionalmente, las capacidades de procesamiento de datos con las que cuentan los sistemas de IA hacen que sea cada vez *más fácil combinar varios datos personales, crear perfiles digitales de sus titulares, y con base en ellos, inferir nuevos datos personales*. El escándalo de Cambridge Analytica es un ejemplo ilustrativo al respecto. Por medio de un “*inocente*” test de personalidad en la red social Facebook, Aleksandr Kogan, profesor la Universidad de Cambridge, recolectó datos personales de 265.000 usuarios de Facebook, así como de sus respectivas redes de amigos. Luego de venderle esta información a la empresa Cambridge Analytica, esta última cruzó “los datos del test de Kogan con la información de Facebook para inferir perfiles psicológicos de cada usuario. Así, Cambridge Analytica logró saber cuál debía ser el contenido, tema y tono de un mensaje para cambiar la forma de pensar de los votantes de forma casi individualizada” (BBC Mundo, 2018). Entonces, a partir de datos aparentemente inocuos Cambridge Alanytica pudo inferir datos más sensibles “que fueron luego utilizados para manipular psicológicamente a los votantes en las elecciones de EE.UU. de 2016, donde Donald Trump resultó electo presidente” (BBC Mundo, 2018). Y, según lo reconoció Brittany Kaiser, exdirectora de desarrollo comercial de la compañía y quien luego que se convertiría en “whistleblower”, la AI jugó un papel fundamental en esta tarea (Delcker, 2020).

3. Posibles daños

a. Daños a la privacidad y la protección de datos personales

Como vimos, el desarrollo de la IA moderna ha facilitado e incentivado significativamente la

recopilación masiva de datos personales y la generación de datos inferidos. Estos dos fenómenos plantean amenazas graves para los derechos a la privacidad y a la protección de datos personales.

De acuerdo con la Corte Constitucional de Colombia (1993), el derecho a la privacidad “implica amparo positivo a la vida privada”. Con base en la jurisprudencia de esta alta Corte (Corte Constitucional, 1995), en una publicación pasada mis coautoras y yo describimos la vida privada como “una esfera, ámbito o espacio ontológico –no material– del individuo, que está *sustraído a la injerencia o al conocimiento de terceros, y en el que se encuentran fenómenos, comportamientos, datos y situaciones que solo le interesan al titular del derecho*” (Newman et al., 2018, p. 17). Partiendo de esta definición, parece claro que, al darle acceso a terceros (e.g., quienes recolectan y procesan la información) a fenómenos, comportamientos, datos y situaciones que el titular de los datos hubiera preferido mantener como parte de su vida privada, tanto la recopilación masiva de datos personales como la generación de datos inferidos afectan el derecho a la privacidad.

Asimismo, la disponibilidad de estas grandes cantidades de información en manos de terceros facilita la vigilancia masiva, que puede ser tanto estatal como comercial. Por ejemplo, de acuerdo con un reporte publicado por Dejusticia en 2022, en Colombia las autoridades de inteligencia, tales como la Policía Nacional y la Dirección Nacional de Inteligencia, obtienen información de inteligencia en Internet y redes sociales bajo la narrativa de que, por encontrarse disponibles en Internet, los datos personales son datos públicos (Dejusticia, 2022). Por su parte, en su libro *La Era del Capitalismo de la Vigilancia*, Shoshana Zuboff describió de manera extensa el surgimiento de un nuevo mercado en el que se capitalizan los datos personales de los consumidores para comercializar predicciones sobre su comportamiento (Zuboff, 2019). Sin importar quien lo ejerza, este monitoreo indiscriminado a los titulares de los datos resulta violatorio del derecho a la privacidad, en la medida en que inhibe a las personas de la posibilidad de contar con espacios de su vida libres de injerencias externas.

De igual forma, la recopilación masiva de datos personales y la generación de datos inferidos también amenazan el derecho a la protección de datos personales. De acuerdo con la Corte Constitucional, este derecho es

“aquel que otorga la facultad al titular de datos personales, de exigir a las administradoras de datos personales el acceso, inclusión, exclusión, corrección, adición, actualización, y certificación de los datos, así como la limitación en las posibilidades de divulgación, publicación o cesión de los mismos, conforme a los principios que informan el proceso de administración de bases de datos personales”. (Corte Constitucional, 2002).

Desafortunadamente, y tal como veremos en la sección 4 de este artículo, dada la misma naturaleza y lógica de los sistemas de analítica de datos propios de la era digital, muchos de los derechos que componen el derecho al habeas data (e.g., el derecho a que mis datos sean

recogidos con una finalidad específica; el derecho a la caducidad de mis datos personales después de un tiempo determinado) no son respetados por los sistemas de IA actualmente en uso. En otros casos, la falta de respeto a estos derechos (e.g., el derecho a conocer los datos que se han inferido sobre mí; las limitaciones en la posibilidad de transferir mis datos a terceros) no se desprenden de la naturaleza propia de la analítica de datos, sino de la falta de capacidades operativas o del simple desprecio de la ley por parte de quienes los recolectan y explotan.

b. Daños a otros derechos

El panorama se vuelve aún más complejo cuando nos percatamos de que, cuando se violan los derechos a la privacidad y a la protección de datos personales, se abre la puerta a un conjunto adicional de violaciones de derechos. En otras palabras, aunque los daños que se expondrán a continuación no son daños a la privacidad o a la protección de datos personales, los mismo se dan gracias a la disponibilidad masiva de datos personales en manos de terceros. En ese sentido, se pueden describir como daños *basados en datos* (“data-driven harms”).

¿Qué hace a los datos tan “peligrosos”? Los datos son una fuente importante de información y conocimiento sobre el titular de los mismos. Tal y como lo resaltó la Corte Constitucional desde los años noventa,

“quien tiene la posibilidad de acopiar, ordenar, utilizar y difundir datos personales adquiere un poder de facto, denominado “*poder informático*”, en ejercicio del cual puede influir decisivamente, por ejemplo, en la definición de perfiles poblacionales que servirán de base para decisiones de política económica, o en la clasificación de una persona, según criterios predeterminados, a fin de definir si debe ser sujeto de una determinada acción pública o privada. Como puede advertirse, *el abuso o la negligencia* en el ejercicio de este enorme poder, apareja un serio riesgo, entre otros, para los derechos fundamentales a la personalidad, a la identidad, a la igualdad, a la intimidad, a la honra, al buen nombre o al debido proceso del sujeto concernido” (Corte Constitucional, 1999)

En el mundo del Big Data y la IA, el “poder informático” que otorgan los datos se ha potencializado, convirtiéndose en un poder informático potenciado o reforzado (Angel, 2020). Hoy en día, la IA se utiliza para tomar todo tipo de decisiones sobre los titulares de datos personales, las cuales pueden ir desde qué contenidos se incluyen en nuestro *feed* de Instagram hasta si se nos debe o no autorizar un préstamo bancario. Y, como se verá a continuación, en estos procesos de toma de decisiones se pueden presentar riesgos para varios derechos.

Antes de abordar estos riesgos, es importante aclarar que los riesgos que plantea la IA son múltiples e incluyen, entre otras cosas, posibles violaciones a la libertad de expresión y de pensamiento (e.g., censura en Internet por medio de tecnologías más avanzadas), el derecho a una vida libre de violencia (e.g., acoso en línea), los derechos de autor (e.g., por medio de la IA generativa), y riesgos en materia de ciberseguridad (e.g., ransomwares que atentan contra

infraestructura crítica). Así mismo, la implementación de la IA supone costos humanos (e.g., costos psicológicos para quienes etiquetan datos) y ambientales (e.g., uso masivo de recursos naturales para el funcionamiento de los *data centers*). Sin embargo, los riesgos a derechos que abordaré a continuación se refieren exclusivamente a los que se derivan concretamente de la recolección y el procesamiento masivo de datos personales que la IA incentiva/permite.

Por un lado, los sistemas de decisión automatizada—que basan sus decisiones en el procesamiento masivo de datos—pueden afectar los *derechos a la igualdad y no discriminación*, al tomar decisiones discriminatorias. Esto puede suceder por varias razones, incluyendo la inclusión de sesgos en el diseño de un modelo algorítmico, o la utilización de datos de entrenamiento que no son representativos de la población en la que se va a utilizar el sistema o que traen incorporados los sesgos racistas, clasistas, etc. propios de nuestra sociedad (Barocas & Selbst, 2016).

Los efectos desiguales de un procesamiento de datos aparentemente neutral son graves, pudiendo conducir a la pérdida de oportunidades en materia de educación, empleo, vivienda, seguros asequibles, etc. Así mismo, pueden aumentar de manera injustificada la exposición de ciertos grupos poblacionales a ciertos tipos de acciones, tales como vigilancia policial o publicidad dirigida. Por ejemplo, en 2021 la organización Global Witness (2021) denunció que el algoritmo de Facebook les mostraba a los usuarios ofertas de trabajo según estereotipos de género. En esa medida, mientras el 96% de las personas a las que se les mostró el anuncio de empleos de mecánico eran hombres, el 95% de las personas a las que se les mostró el anuncio de empleos de enfermeras preescolares eran mujeres.

Por otro lado, los datos personales también pueden ser utilizados para tomar decisiones que atentan contra nuestro *derecho a la autonomía*. En conjunto con herramientas de minería de datos y tácticas de la ciencia cognitiva y del comportamiento, actores públicos y privados pueden explotar nuestros datos para influir de manera inaceptable en nuestras decisiones o comportamientos (Calo, 2014). Estas tácticas, comúnmente conocidas como manipulación algorítmica, se implementan por medio de la identificación y el aprovechamiento de las limitaciones cognitivas o vulnerabilidades contextuales individuales de las personas (Spencer, 2020). Por ejemplo, en 2018, Amazon patentó una tecnología que le permitiría a Alexa leer el estado emocional de sus usuarios y con base en ello, presentarles anuncios publicitarios (Fussell, 2018).

Adicionalmente, hay decisiones basadas en datos personales que también pueden poner en riesgo *la democracia*. En primer lugar, varias investigaciones han demostrado cómo, la personalización de contenidos en redes sociales—posible gracias a los datos personales y al perfilamiento de las personas que se puede hacer con ellos—disminuye la exposición a diferentes perspectivas. Como resultado, los ciudadanos terminamos en “cámaras de eco,” en las que sólo interactuamos con personas que tienen visiones muy parecidas a las nuestras. Como consecuencia, las opiniones se radicalizan y la sociedad se polariza (Levin et al., 2021). En

segundo lugar, los datos personales son utilizados para distribuir, de manera dirigida y por ende más efectiva, *deepfakes*¹⁹ y otros contenidos de desinformación. Así, al tener un mayor conocimiento de su público objetivo, los emisores de estos mensajes pueden saber con mayor facilidad a quién deben dirigirlos.

Finalmente, los datos personales también pueden ser utilizados para entrenar sistemas de IA que amenazan los *derechos de asociación y de protesta*. En particular, las tecnologías de reconocimiento facial están siendo utilizadas alrededor del mundo para, entre otras cosas, identificar manifestantes durante protestas. Así sucedió en Colombia durante las protestas del Paro Nacional de 2019, cuando la policía anunció que, con el fin de monitorear las manifestaciones, utilizaría un helicóptero equipado con cámaras de reconocimiento facial capaces de identificar a personas a una distancia de 15 km y reconocer incluso rostros cubiertos (Murrillo, 2019). En el informe *In Focus*, publicado por la Red Internacional de Organizaciones de Libertades Civiles (INCLO) en 2021, se argumenta que este anuncio fue un intento de las autoridades colombianas de disuadir a los ciudadanos para que no salieran a protestar (INCLO, 2021).

4. ¿Cómo ha respondido la ley colombiana?

Colombia cuenta con un régimen constitucional y legal de protección de datos personales. Establecido tanto en el artículo 15 de la Constitución Política como en las leyes estatutarias 1266 de 2008 (Ley de Habeas Data Financiero) y 1581 de 2012 (Ley de Protección de Datos Personales),²⁰ este régimen está estructurado sobre un conjunto de principios orientadores que buscan empoderar a los titulares de datos para que puedan tener control sobre sus datos. Tal y como lo explica la Corte Constitucional,

“producto de una tarea de sistematización de lo dispuesto por la jurisprudencia constitucional en la materia, la Ley Estatutaria 1581 de 2012, en su artículo 4º, ‘establece los principios a los que está sujeto cualquier tipo de tratamiento de datos en Colombia’. En concreto, consagra los principios de (i) *legalidad*, (ii) *finalidad*, (iii) *libertad*, (iv) *veracidad o calidad*, (v) *transparencia*, (vi) *acceso y circulación restringida*, (vii) *seguridad* y (viii) *confidencialidad*, los cuales son ‘oponibles a todos los sujetos involucrados en el tratamiento del dato, entiéndase en la recolección, circulación, uso, almacenamiento, supresión, etc., sin importar la denominación que los sujetos adquieran (...)’.” (Corte Constitucional, 2022)

A su vez, estos principios responden a una tendencia internacional que tuvo su origen en un informe de 1973 presentado en Estados Unidos por el Comité Asesor del Departamento de

19 Imágenes, videos, o audios editados con IA y utilizados para imitar la realidad y engañar.

20 Así como en el Decreto Reglamentario 1377 de 2013, por medio del cual se reglamenta parcialmente la Ley 1581 de 2012 (incorporado en el Decreto 1074 de 2015).

Salud, Educación y Bienestar titulado *Registros, Computadoras y los Derechos de los Ciudadanos* (1973). En él, los autores del informe incluyeron un conjunto de Principios de Prácticas Justas de Información (“Fair Information Practices Principles” (FIPPs, por sus siglas en inglés), los cuales tenían como objetivo regular el uso, recopilación y privacidad de los datos personales. Luego de ser incorporados en las *Directrices de la OCDE que Regulan la Protección de la Privacidad y los Flujos Transfronterizos de Datos Personales* (1980), adoptadas el 23 de septiembre de 1980, los FIPPs fueron rápidamente adoptados en varias jurisdicciones nacionales, incluida Colombia.

A pesar de la popularidad de estos principios, las nuevas tecnologías digitales han hecho que varios de ellos parezcan hoy ineficaces y en algunos casos obsoletos. De acuerdo con el *principio de libertad*, por ejemplo, el tratamiento de datos personales sólo puede ejercerse con el consentimiento, previo, expreso e informado del titular de los datos. En un mundo de Small Data, el cumplimiento de este requisito parece viable. Sin embargo, en el marco del Big Data y la IA, los datos personales no solo son tratados por quienes los recolectaron inicialmente, sino por un conjunto masivo de actores de segundo nivel, comúnmente conocidos como *data brokers*, que están constantemente comprando datos personales, infiriendo nuevos datos, construyendo perfiles y vendiendo esos perfiles a otros actores (FTC, 2014). En la práctica, lo que termina sucediendo es que las plataformas y aplicaciones con las que interactuamos comúnmente solicitan nuestro consentimiento no sólo para tratar ellas nuestros datos personales, sino para “compartir información con terceros” (Newman & Angel, 2019, p. 39). Sin embargo, al dar nuestro consentimiento previo y expreso, muchos de nosotros no estamos realmente “informados” sobre ni somos conscientes del complejo ecosistema digital, los actores, y los usos que existen detrás de esos actores de primer nivel (Tene & Polonetsky, 2012).

Igualmente, la era digital ha hecho evidente lo que se conoce como el “valor de opción” de los datos, a saber, “su potencial aparentemente ilimitado de reutilización” (Mayer-Schönberger & Cukier, 2013, p. 122). Dado que en el marco del Big Data y la IA la mayor parte del valor de los datos radica en sus posibles, pero aún indeterminados usos, el cumplimiento de principios como el *principio de finalidad* también se torna ilusorio (Hirsch, 2014). De acuerdo con este principio, la finalidad con la que se pretende tratar los datos personales debe ser informada al titular. ¿Qué sucede entonces, con las finalidades que surgen tiempo después, cuando los datos ya están, por ejemplo, en manos de *data brokers*? En nuestro día a día, no es usual recibir mensajes de *data brokers* como Acxiom, Corelogic, Datalogix informándonos sobre los distintos usos que le van a dar a los datos que han comprado sobre nosotros.

Dadas estas nuevas circunstancias, y el hecho de que las leyes estatutarias 1266 de 2008 y 1581 de 2012 fueron redactadas neutralmente, para ser aplicables a cualquier herramienta manual o tecnológica (Corte Constitucional, 2024), la Delegatura para la Protección de Datos Personales de la Superintendencia de Industria y Comercio (SIC), en su calidad de Autoridad de Protección de Datos de Colombia, se ha puesto en la tarea de instruir a los ciudadanos en la aplicación del régimen de protección de datos aquí descrito en el marco de la era digital. Para

el efecto, desde al menos 2015 la SIC ha emitido varias guías que desarrollan los contenidos de las leyes vigentes a la luz de varias de las tecnologías emergentes. Entre ellas, pueden resaltarse la *Guía de Protección de Datos en los Servicios de Computación en la Nube* (SIC, 2015), la *Guía para la Implementación del Principio de Responsabilidad Demostrada* (SIC, 2016) (así como la publicada en 2019 para la implementación del principio en las transferencias internacionales de datos personales), la *Guía sobre el Tratamiento de Datos Personales para Fines de Comercio Electrónico* (SIC, 2019a), y la *Guía sobre el Tratamiento de Datos Personales para Fines de Marketing y Publicidad* (SIC, 2019b). Adicionalmente, en 2024 la SIC publicó la *Circular Externa 002 de 2024*, por medio de la cual se establecen Lineamientos sobre el Tratamiento de Datos personales en Sistemas de Inteligencia Artificial.

Tomando inspiración de instrumentos desarrollados en otras jurisdicciones, tales como el *Reglamento General de Protección de Datos Personales* (GDPR, por sus siglas en inglés) de la Unión Europea y la versión revisada de las *Directrices de la OCDE* más arriba mencionadas, las anteriores guías y la Circular 002 de la SIC incluyen o desarrollan al menos cuatro herramientas innovadoras que buscan hacer frente a los desafíos de la era digital. En primer lugar, y como su mismo título lo indica, la *Guía para la Implementación del Principio de Responsabilidad Demostrada* hace énfasis en y explica en detalle el contenido de este principio. De acuerdo con el artículo 26 del Decreto reglamentario 1377 de 2013 (incorporado en el Decreto 1074 de 2015), “los responsables del tratamiento de datos personales deben ser capaces de demostrar, a petición de la Superintendencia de Industria y Comercio, que han implementado medidas apropiadas y efectivas para cumplir con las obligaciones establecidas en la Ley 1581 de 2012 (...”). Partiendo de esta obligación, la guía les recomienda a los responsables del tratamiento de datos personales el desarrollo de un programa integral de gestión de datos personales que les permita demostrar adecuadamente el cumplimiento de la ley.

En segundo lugar, tanto la *Guía para la Implementación del Principio de Responsabilidad Demostrada en las Transferencias Internacionales de Datos Personales* (SIC, 2019c) como las subsiguientes guías de la SIC (incluida la Circular 002 de 2024) sugieren la implementación de “estudios de impacto de privacidad” así como el principio de “privacidad, ética y seguridad desde el diseño y por defecto” como medidas proactivas para cumplir con el Principio de Responsabilidad Demostrada. Con la primera herramienta, se busca poner en funcionamiento un sistema efectivo de previsión y manejo de riesgos que, por medio de controles internos, permita garantizar que los riesgos se mitigarán y los datos se tratarán debidamente y conforme con la regulación existente. Por su parte, la implementación del principio de privacidad, ética y seguridad desde el diseño y por defecto pretende que, en cualquier proyecto, lo referente a la protección de datos, la ética y la seguridad de la información se involucre desde el inicio. La *Circular Externa 002 de 2024*, por su parte, propone la utilización de técnicas de privacidad diferencial para cumplir con dicho principio.

Finalmente, la *Circular Externa 002 de 2024* también exige que en el tratamiento de datos personales por medio de sistemas de IA se realice una especie de test de proporcionalidad—al estilo de los utilizados por la Corte Constitucional—en el que se ponderen los siguientes

cuatro criterios: (i) idoneidad; (ii) necesidad; (iii) razonabilidad; y (iv) proporcionalidad en sentido estricto. Así mismo, en la circular la SIC parece sugerir un principio de precaución según el cual, “[e]n caso de presentarse *falta de certeza* frente a los potenciales desafíos que puede causar el Tratamiento de Datos personales, y con miras a evitar que se cause un daño grave e irreversible, los Administradores de Datos personales deberán abstenerse de realizar dicho Tratamiento o adoptar medidas precautorias o preventivas para proteger los derechos del Titular del dato, su dignidad y otros derechos humanos.”

Antes de terminar, es necesario mencionar que, en mayo de 2021, el Gobierno Nacional de Colombia publicó un documento titulado *Marco Ético para la Inteligencia Artificial en Colombia* (Gobierno Nacional de Colombia, 2021). Este documento incluye nueve principios éticos, que en su mayoría hacen eco de principios ya propuestos alrededor del mundo por otros gobiernos, organizaciones multilaterales, el sector privado y la sociedad civil (Fjeld et al., 2020) y entre los cuales se encuentra el principio de privacidad. Desafortunadamente, el mismo documento deja claro que se trata sólo de una guía de derecho blando y únicamente aplicable a las entidades públicas. En esa medida “no es de obligatorio cumplimiento por parte de las entidades, sino que se presenta como recomendaciones” (Gobierno Nacional de Colombia, 2021, p. 20). Adicionalmente, varias de las herramientas incluidas en el documento para la implementación de los principios propuestos (e.g., dashboard que servía como registro ético de algoritmos) dejaron de funcionar luego del cambio de gobierno nacional.

5. Conclusión

Como se ve, las nuevas tecnologías digitales—y en particular la IA—han facilitado e incentivado significativamente la recopilación masiva de datos personales y la generación de datos inferidos. En consecuencia, tanto los derechos a la privacidad y la protección de datos personales como otros derechos y valores tales como la igualdad y no discriminación, la autonomía, los derechos de asociación y de protesta, y la democracia se están viendo amenazados con estos nuevos desarrollos. Dado que el régimen actual de protección de datos personales de Colombia—encaminado a empoderar a los ciudadanos para que tomen el control de sus datos—se ha tornado en muchos casos ineficaz y obsoleto, en los últimos diez años la Autoridad de Protección de Datos de Colombia ha buscado proponer mecanismos que, en línea con prácticas internacionales, buscan, además de empoderar a los ciudadanos, establecer obligaciones de responsabilidad demostrada que hagan rendir cuentas a los actores públicos y privados que tratan datos personales. Lo anterior, por medio de mecanismos como los “estudios de impacto de privacidad” y el test de proporcionalidad propuesto por la SIC en la *Circular Externa 002 de 2024*.

Sin embargo, la SIC también le ha abierto la puerta a mecanismos como el principio de “privacidad, ética y seguridad desde el diseño y por defecto” y un posible principio de precaución en materia de explotación de datos. A diferencia de las demás herramientas, estas propuestas dan pie para pensar que, en un futuro, el país se puede encaminar hacia un tercer objetivo,

adicional a los de empoderar a los ciudadanos y hacer rendir cuentas a quienes tratan los datos personales, a saber: restarle poder a estos últimos actores. Esto, por medio de medidas que le apunten a intervenir tanto sus modelos de negocio basados en datos como el diseño de las tecnologías digitales emergentes.

6. Referencias

Angel, M.P. (2020). *La administración pública datificada Gobernando a través del Big Data*. Editorial Legis.

Barocas, S. & Selbst, A. (2016). Big Data's Disparate Impact. *California Law Review*, 104, 671-732.

BBC Mundo (2018). 5 claves para entender el escándalo de Cambridge Analytica que hizo que Facebook perdiera US\$37.000 millones en un día. *BBC News Mundo*. Disponible a través de: <https://www.bbc.com/mundo/noticias-43472797>

Calo, R. (2014). Digital Market Manipulation, *George Washington Law Review*, 82, 995-1051.

Camacho, L., Ospina, D., & Upegui, J.C. (2022). *Inteligencia estatal en internet y redes sociales: el caso colombiano*. Editorial Dejusticia.

Comité Asesor del Departamento de Salud, Educación y Bienestar. (1973). *Registros, Computadoras y los Derechos de los Ciudadanos*.

Corte Constitucional de Colombia. Sentencia T-349 de 1993. Expediente: T-13211. M. P. José Gregorio Hernández Galindo.

Corte Constitucional de Colombia. Sentencia SU-056 de 1995. Expedientes: T-40754 y T-44219. M. P. Antonio Barrera Carbonell.

Corte Constitucional de Colombia. Sentencia T-307 de 1999. M.P. Eduardo Cifuentes Muñoz.

Corte Constitucional de Colombia. Sentencia T-729 de 2002. Expediente: T-467467. M. P. Eduardo Montealegre Lynett.

Corte Constitucional de Colombia. Sentencia T-143 de 2022. Expediente T-8.197.643. M.P. Alejandro Linares Cantillo.

Corte Constitucional de Colombia. Sentencia T-323 de 2024. Expediente T-9.301.656. M.P. Juan Carlos Cortés González.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M. (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center Research Publication No. 2020-1.

Fussell, S. (2018). Alexa Wants to Know How You're Feeling Today. *The Atlantic*. Disponible a través de: <https://www.theatlantic.com/technology/archive/2018/10/alexa-emotion-detection-ai-surveillance/572884/>

FTC. (2014). *DATA BROKERS. A Call for Transparency and Accountability*.

Global Witness. (2021). *How Facebook's ad targeting may be in breach of UK equality and data protection laws*. Disponible a través de: <https://www.globalwitness.org/en/campaigns/digital-threats/how-facesbooks-ad-targeting-may-be-in-breach-of-uk-equality-and-data-protection-laws/>

Gobierno Nacional de Colombia. (2021). *Marco Ético para la Inteligencia Artificial en Colombia*.

Hirsch, D. D. (2014a). The Glass House Effect: Big Data, The New Oil, and the Power of Analogy. *Maine Law Review*, 66(2), 373–395.

INCLO. (2021). *In Focus: Facial Recognition Tech Stories And Rights Harms From Around The World*. Disponible a través de: <https://inclo.net/publications/in-focus-facial-recognition-tech-stories-and-rights-harms-from-around-the-world/>

Kissinger, H.A. & Schmidt, E. (2022). *The Age of AI And Our Human Future*. Little, Brown and Company.

Levin, S.A., Milner, H.V., Perrings, C. (2021). The dynamics of political polarization. *Proceedings of the National Academy of Sciences*, 118(50).

Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data. A revolution that will transform how we live, work and think*. Mariner Books.

Microsoft (2024). *Declaración de privacidad de Microsoft*. Disponible a través de: <https://www.microsoft.com/es-mx/privacy/privacystatement#mainpersonaldatawecollectmodule>

MinTIC. (2024). *Contexto y Objetivos de la Cumbre Ministerial Latinoamericana y del Caribe de IA*. Disponible a través de: <https://mintic.gov.co/cumbre-ia/secciones/Cumbre-Ministerial-de-IA/>

Murillo, O. (2019). Helicóptero halcón de la Policía estrena identificación facial. *El Tiempo*. Disponible a través de: <https://www.eltiempo.com/bogota/estrenan-helicoptero-halcon-con-reconocimiento-facial-en-paro-del-21-de-noviembre-435766>

- Newman, V., Angel, M.P., & Dávila, M.X. (2018). *Víctimas y Prensa Despues de la Guerra. Tensiones entre intimidad, verdad histórica y libertad de expresión*. Editorial Dejusticia.
- OCDE. (1980). *Directrices de la OCDE que Regulan la Protección de la Privacidad y los Flujos Transfronterizos de Datos Personales*.
- Pasquale, F. (2016). *The Black Box Society. The Secret Algorithms That Control Money and Information*. Harvard University Press.
- Delcker, J. (2020). POLITICO AI: Decoded: How Cambridge Analytica used AI — No, Google didn't call for a ban on face recognition — Restricting AI exports. *Politico* Disponible a través de: <https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-how-cambridge-analytica-used-ai-no-google-didnt-call-for-a-ban-on-face-recognition-restricting-ai-exports/>
- Shrivastava, R. (2024). The Prompt: What Happens When We Hit The ‘Data Wall’? *Forbes*. Disponible a través de: <https://www.forbes.com/sites/rashishrivastava/2024/07/30/the-prompt-what-happens-when-we-hit-the-data-wall/>
- SIC. (2015). *Guía de Protección de Datos en los Servicios de Computación en la Nube*.
- SIC. (2016). *Guía para la Implementación del Principio de Responsabilidad Demostrada*
- SIC. (2019a). *Guía sobre el Tratamiento de Datos Personales para Fines de Comercio Electrónico*
- SIC. (2019b). *Guía sobre el Tratamiento de Datos Personales para Fines de Marketing y Publicidad*.
- SIC. (2019c). *Guía para la Implementación del Principio de Responsabilidad Demostrada en las Transferencias Internacionales de Datos Personales*.
- SIC. (2024). *Circular Externa 002 de 2024*.
- Spencer, S.B. (2020). The Problem of Online Manipulation. *University of Illinois Law Review*, 2020, 959-1006.
- Tene, O., & Polonetsky, J. (2012a). To Track Or “Do Not Track”: Advancing Transparency and Individual Control in Online Behavioral Advertising. *Minnesota Journal of Law, Science & Technology*, 13, 281-357.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.



Desafiando Contemporáneos de la Inteligencia Artificial:

**Un Enfoque en el Principio de
Transparencia**

MARÍA LORENA FLÓRES ROJAS



María Lorena Florez

Abogada colombiana y profesora asistente de Derecho en la Universidad de Groninga, especializada en la intersección entre derecho, tecnología y ética. Con un doctorado del programa de Persona Individual y Protección Legal en la Scuola Superiore Sant'Anna (Italia), su trabajo se centra en cómo las tecnologías emergentes, especialmente la inteligencia artificial (IA), impactan a los consumidores y al mercado. Actualmente lidera un proyecto de investigación sobre la responsabilidad en IA, que busca desarrollar una taxonomía comprensiva de prácticas y directrices éticas para su implementación.

Anteriormente, Lorena fue profesora asistente en la Universidad de Los Andes, donde impartió cursos desde derecho comercial hasta tecnología y derecho. Allí, participó en la creación de herramientas ético-legales para la adopción de IA en organizaciones y colaboró en proyectos educativos innovadores, como el MOOC fAIR LAC sobre el uso responsable de IA en el sector público y el diseño de módulos de ética para la Maestría en Inteligencia Artificial (MaIA). Además, es docente del MOOC “UNESCO AI and the State of Law for Judges and Judicial Operators”, aportando su experiencia en la formación de operadores judiciales sobre el impacto ético y regulatorio de la IA en el derecho.

Los Desafíos Contemporáneos de la Inteligencia Artificial en América Latina: Un Enfoque en el Principio de Transparencia

Dra. María Lorena Flórez Rojas²¹

Resumen

Este white paper analiza los desafíos contemporáneos que enfrenta América Latina en la implementación del principio de transparencia en la gobernanza de la inteligencia artificial (IA). A pesar de que las estrategias nacionales frecuentemente se inspiran en paradigmas internacionales, la región enfrenta barreras normativas, técnicas y sociales que complican su aplicación efectiva. El documento aborda cómo las tensiones entre estándares internacionales y marcos legales locales generan opacidad en sectores críticos, como la administración pública y los programas sociales, afectando la confianza y la rendición de cuentas. A través de casos concretos en Colombia, se evidencia la desconexión entre la adopción tecnológica y los principios de transparencia, equidad y privacidad. Además de diagnosticar problemas, el white paper propone estrategias para fortalecer la transparencia en la IA, incluyendo el desarrollo de estándares claros, la adopción de IA explicable (IAE) y la implementación de auditorías independientes. Estas propuestas buscan garantizar un uso ético, inclusivo y responsable de la IA, alineando la tecnología con valores democráticos y los derechos humanos.

Palabras clave: Transparencia, Inteligencia Artificial Explicable (IAE), Gobernanza de IA, Rendición de cuentas, América Latina.

21 Abogada colombiana y profesora asistente de Derecho en la Universidad de Groninga, especializada en la intersección entre derecho, tecnología y ética. Con un doctorado del programa de Persona Individual y Protección Legal en la Scuola Superiore Sant'Anna (Italia), su trabajo se centra en cómo las tecnologías emergentes, especialmente la inteligencia artificial (IA), impactan a los consumidores y al mercado. Actualmente lidera un proyecto de investigación sobre la responsabilidad en IA, que busca desarrollar una taxonomía comprensiva de prácticas y directrices éticas para su implementación.

Anteriormente, Lorena fue profesora asistente en la Universidad de Los Andes, donde impartió cursos desde derecho comercial hasta tecnología y derecho. Allí, participó en la creación de herramientas ético-legales para la adopción de IA en organizaciones y colaboró en proyectos educativos innovadores, como el MOOC fAIr LAC sobre el uso responsable de IA en el sector público y el diseño de módulos de ética para la Maestría en Inteligencia Artificial (MaIA). Además, es docente del MOOC “UNESCO AI and the State of Law for Judges and Judicial Operators”, aportando su experiencia en la formación de operadores judiciales sobre el impacto ético y regulatorio de la IA en el derecho.

Abstract

This white paper analyzes the contemporary challenges faced by Latin America in the implementation of the principle of transparency in the governance of artificial intelligence (AI). Although national strategies are often inspired by international paradigms, the region faces regulatory, technical and social barriers that complicate their effective implementation. The paper addresses how tensions between international standards and local legal frameworks generate opacity in critical sectors, such as public administration and social programs, affecting trust and accountability. Through concrete cases in Colombia, it shows the disconnection between technological adoption and the principles of transparency, equity and privacy. In addition to diagnosing problems, the white paper proposes strategies to strengthen AI transparency, including the development of clear standards, the adoption of explainable AI (EAI) and the implementation of independent audits. These proposals seek to ensure ethical, inclusive and responsible use of AI, aligning the technology with democratic values and human rights.

Keywords: Transparency, Artificial Intelligence Explainable (AI), AI Governance, Accountability, Latin America.

1. Introducción

¿Qué nivel de confianza, del 1 al 5, tendrías en un juez que utiliza ChatGPT para analizar y fundamentar su decisión en tu caso?

Este cuestionamiento refleja una de las muchas inquietudes que emergen en el debate sobre el uso de inteligencia artificial (IA) en la toma de decisiones. Un ejemplo paradigmático es el caso revisado por la Corte Constitucional de Colombia, donde un juez de circuito utilizó ChatGPT para asistir en la redacción de un fallo relacionado con una acción de tutela presentada en nombre de un menor con trastorno del espectro autista (Case T-323 of 2024, 2024). Este caso subraya las tensiones éticas, legales y técnicas en la integración de IA en la administración de justicia, y plantea interrogantes fundamentales sobre la transparencia, la imparcialidad y la confianza en el sistema judicial.

Aunque el juez afirmó que la herramienta no reemplazó su capacidad de análisis, sino que buscaba complementar su labor, su decisión de incorporar respuestas generadas por ChatGPT en el texto de la sentencia provocó un intenso debate (Flórez Rojas, 2024). La Corte Constitucional, al revisar el caso, enfatizó la necesidad de marcos normativos sólidos que guíen el uso de estas tecnologías, subrayando la importancia de preservar la independencia judicial y garantizar que la IA complemente, pero no sustituya, el juicio humano. Este episodio marca un hito en la discusión sobre la regulación de IA en América Latina, resaltando la necesidad de establecer salvaguardas éticas, técnicas y legales.

El caso colombiano no es un hecho aislado, sino parte de una tendencia más amplia en la

región. Diversos países de América Latina han comenzado a integrar herramientas de IA en diversos sectores, como parte de estrategias nacionales de transformación digital (Hernández et al., 2022). Sin embargo, estas iniciativas enfrentan desafíos significativos, especialmente en la implementación del principio de transparencia, que es fundamental para garantizar la confianza pública y la rendición de cuentas (Gutiérrez Rodríguez, 2023). El principio de transparencia emerge como un pilar esencial, ya que permite que los ciudadanos comprendan y cuestionen cómo las decisiones automatizadas afectan sus vidas (Addink, 2019). Estudios recientes han identificado una proliferación de sistemas de toma de decisiones automatizadas en la región, pero pocos cuentan con marcos regulatorios adecuados para abordar los riesgos inherentes relacionados con la privacidad, la seguridad y los sesgos algorítmicos (CEPAL & Naciones Unidas, 2012; Cumbre Ministerial y de Altas Autoridades de América Latina y el Caribe, 2023; Hernández et al., 2022).

Este white paper aborda los desafíos contemporáneos que enfrenta América Latina en la implementación del principio de transparencia en la gobernanza de la IA. Si bien las estrategias nacionales suelen inspirarse en paradigmas anglosajones y europeos, impulsados por organismos internacionales como el Banco de Desarrollo de América Latina y el Caribe (CAF) y el Banco Interamericano de Desarrollo (BID), su aplicación en la región revela barreras significativas (Guío Español, 2021; Magrani, 2021; Ortiz & Iglesias, 2018). Estas barreras incluyen vacíos normativos, desigualdades sociales y limitaciones técnicas que dificultan la operatividad de los estándares internacionales en contextos locales.

A través de ejemplos concretos, el documento analiza cómo los principios globales de transparencia frecuentemente chocan con los marcos legales, culturales y administrativos existentes en la región, evidenciando la desconexión entre las aspiraciones internacionales y las capacidades locales. También se examinan casos en los que estas tensiones han agravado problemas de opacidad, exacerbando desafíos preexistentes en sectores críticos como la justicia, la administración pública y los programas sociales.

Este análisis no se limita a identificar problemas, sino que busca trazar una hoja de ruta hacia soluciones prácticas. A partir de estrategias efectivas y recomendaciones de políticas públicas, el documento propone enfoques concretos para superar los desafíos y avanzar hacia una implementación responsable de la IA en América Latina. En particular, se destacan iniciativas como el desarrollo de estándares claros, el fortalecimiento de capacidades locales, la adopción de sistemas de IA explicables (IAE) y la creación de marcos regulatorios adaptados a las realidades de la región.

Estructurado en cuatro secciones principales, el documento comienza explorando el concepto de transparencia en IA, incluyendo las características esenciales requeridas por estándares internacionales. Luego, examina en detalle los objetivos de la IAE, describiendo sus procesos y técnicas más prominentes. Posteriormente, analiza casos específicos de implementación de IA en la región, particularmente en Colombia, identificando puntos de falla relacionados

con la transparencia y su impacto en los derechos ciudadanos. Finalmente, ofrece recomendaciones y observaciones concluyentes que buscan facilitar el camino hacia una gobernanza tecnológica más explicable.

2. El Principio de Transparencia en la Gobernanza de la IA

La transparencia en la IA se refiere a la provisión de información significativa sobre los procesos y decisiones generados por sistemas automatizados(Contreras Caballol & Pefaur Dendal, 2020). Este principio no solo permite a los usuarios comprender cómo funcionan los sistemas de IA, sino que también es esencial para garantizar la rendición de cuentas y la confianza pública (Felzmann et al., 2020). Según el **Reglamento de IA de la Unión Europea** (EU Reg. AI), por transparencia, se entiende que “*los sistemas de IA se desarrollan y utilizan de un modo que permita una trazabilidad y explicabilidad adecuadas, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos*”(Regulation EU 2024/1689 (AI Act), 2024, Part 27). De esta forma, los elementos clave de la transparencia en la IA incluyen varios aspectos fundamentales que garantizan su implementación efectiva.

En primer lugar, el **acceso a la información** es esencial, ya que permite a los ciudadanos conocer cómo se recopilan y utilizan sus datos, lo que facilita una supervisión adecuada (Varošanec, 2022, p. 97). Asimismo, la **comprendibilidad de la información** es crucial; esta debe presentarse de manera clara y accesible, evitando tecnicismos que puedan dificultar su interpretación por parte de personas no expertas (Sharma et al., 2019). La **puntualidad** también juega un papel importante, asegurando que los datos relevantes estén disponibles en tiempo real para respaldar decisiones oportunas (Felzmann et al., 2020). Además, la **exactitud y completitud** de los datos y procesos divulgados son indispensables, ya que deben ser precisos, confiables y abarcar todos los aspectos necesarios para una evaluación adecuada. Por último, la **accesibilidad** garantiza que esta información esté fácilmente disponible para todos los usuarios, eliminando barreras tecnológicas o sociales que puedan limitar su alcance (Balasubramaniam et al., 2023). Estos principios en conjunto forman la base para fomentar la confianza pública y la rendición de cuentas en el uso de sistemas de IA.

Asegurando la Transparencia en la IA

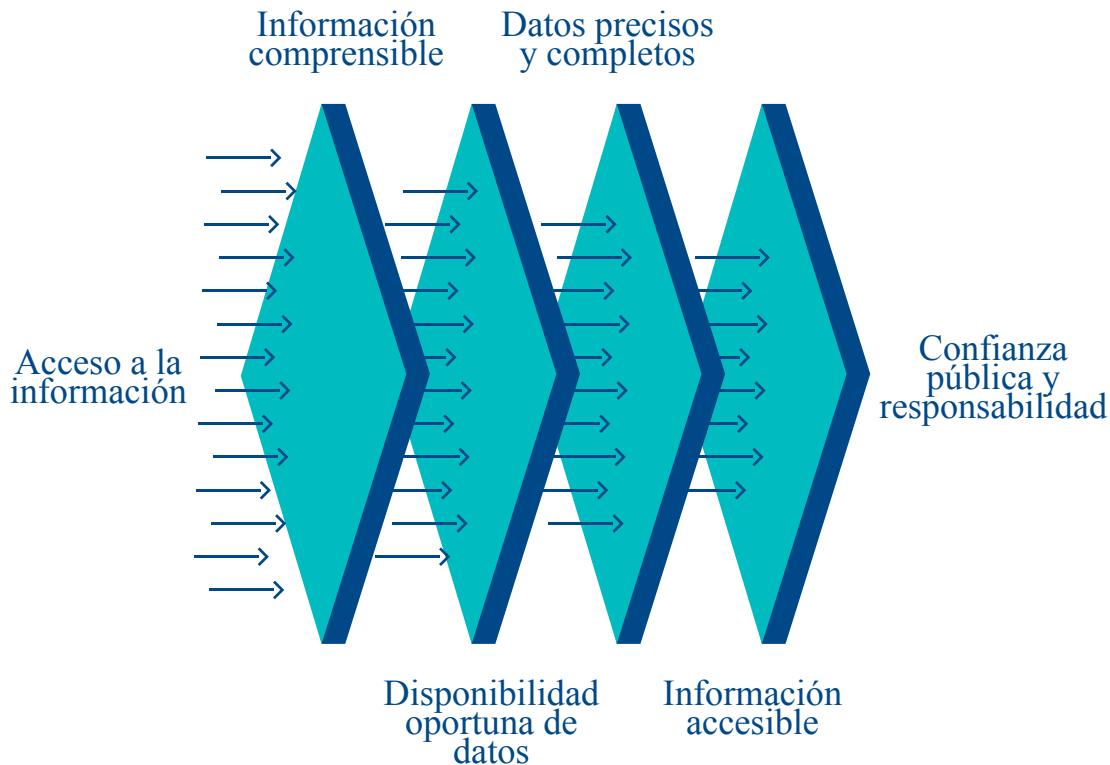


Figura 1. Asegurando la Transparencia en la IA. Elaborado por la autora²².

La transparencia juega un papel crucial en diversas disciplinas, adaptándose a las necesidades específicas de cada área de conocimiento. En economía, es considerada un requisito fundamental para el funcionamiento eficiente de los mercados, ya que fomenta la confianza entre inversores y consumidores (Kaufmann & Weber, 2010, p. 780). La disponibilidad de información clara y precisa permite a los participantes del mercado tomar decisiones informadas, lo que contribuye a la estabilidad y la eficiencia económica. Por otro lado, en ciencias políticas, la transparencia es esencial para la participación democrática, ya que facilita el acceso ciudadano a las decisiones y actividades gubernamentales (Kim & Lee, 2019, p. 1030). Esto no solo empodera a los ciudadanos para influir en las políticas públicas, sino que también fortalece la rendición de cuentas de los representantes políticos, promoviendo gobiernos más abiertos y responsables (Cheong, 2024). En el ámbito jurídico, la transparencia asegura

²² La gráfica fue elaborada con el apoyo de Napkin AI utilizando las características y el prompt proporcionados por la autora.

la equidad y la seguridad jurídica al permitir que las decisiones sean revisables (Romaní et al., 2024). Los individuos pueden impugnar decisiones cuando se presentan razones claras y fundamentadas, lo que refuerza el acceso a la justicia y la protección de los derechos fundamentales(Contini et al., 2024). Finalmente, en ciencias de la computación, la transparencia se relaciona estrechamente con la explicabilidad de los sistemas de IA (Bücker et al., 2022; Sharma et al., 2019). Este enfoque busca que los usuarios puedan comprender cómo funcionan los algoritmos y evaluar las decisiones automatizadas. Esto es especialmente importante en aplicaciones críticas como la salud, la justicia o la seguridad, donde la confianza en los sistemas depende de su capacidad para justificar sus resultados.

En este sentido, han surgido marcos de gobernanza de la IA para orientar el desarrollo y la implementación responsables de sistemas de IA, siendo la transparencia un aspecto clave en muchas de estas iniciativas. La transparencia en la IA generalmente se refiere a la provisión de información significativa sobre la lógica, el funcionamiento y los procesos de toma de decisiones de los sistemas automatizados. Este concepto es central para varios marcos importantes, cada uno de los cuales ofrece su propia perspectiva sobre cómo se debe implementar y mantener la transparencia en los sistemas de IA. La siguiente tabla describe las características clave de la transparencia en la IA en las principales organizaciones internacionales y Estados que han desarrollado marcos legales:

Organizaciones Internacionales/ Instrumentos Estatales	Definición de Transparencia en los Sistemas de IA
Principios OECD	Destacan la transparencia como un principio central para una IA confiable. Los actores involucrados en la IA deben proporcionar información significativa para fomentar la comprensión, hacer que las partes interesadas sean conscientes de las interacciones con la IA y permitir que las personas afectadas puedan questionar los resultados generados por estos sistemas (OECD, 2021; OECD/CAF, 2022).
Recomendaciones de la UNESCO sobre la Ética de la IA	Hacen un llamado a la transparencia multinivel a lo largo del ciclo de vida de los sistemas de IA, desde los procesos de toma de decisiones hasta las evaluaciones de impacto ético. La UNESCO ha desarrollado directrices preliminares para el uso de la IA en tribunales y juzgados, destacando la importancia de la transparencia para garantizar la rendición de cuentas y la confianza en las decisiones judiciales (UNESCO, 2021).

Reglamento de IA de la Unión Europea	Exige que los sistemas de IA de alto riesgo proporcionen información clara y suficiente a los usuarios, incluyendo documentación detallada sobre las capacidades y limitaciones del sistema. Este reglamento enfatiza la importancia de la supervisión humana y la explicabilidad de las decisiones de la IA. Así, los usuarios deben ser informados siempre que estén interactuando con un sistema de IA (Regulation EU 2024/1689 (AI Act), 2024).
Carta de Derechos de la IA de EE. UU.	Promueve la transparencia mediante una comunicación clara sobre cuándo y cómo se están utilizando los sistemas de IA. Aboga por explicaciones de las decisiones de los sistemas automatizados en un lenguaje comprensible (Proyecto de Ley- Carta de Derechos: Hacer Que Los Sistemas Automatizados Trabajen al Servicio Del Pueblo Americano, 2022).
Directiva de Canadá sobre la Toma de Decisiones Automatizada	Se centra en la transparencia para garantizar que los sistemas de IA operen de manera ética y en alineación con los valores sociales. Esto incluye divulgar cuándo se utiliza la IA y proporcionar explicaciones claras sobre sus decisiones . La directiva exige que los departamentos gubernamentales brinden explicaciones comprensibles de los sistemas de decisión automatizada (Directive on Automated Decision-Making, 2021).

Figura 2. Principales Marcos Éticos en relación con el principio de transparencia. Elaborado por la autora.

Si bien los requisitos y las implementaciones específicas pueden diferir, los marcos mencionados anteriormente enfatizan la necesidad de una comunicación y comprensión claras de los sistemas de IA. Este enfoque en la transparencia tiene como objetivo generar confianza, mejorar la rendición de cuentas y facilitar una mejor toma de decisiones en el campo de la IA. En este sentido, la realización ‘ideal’ de la transparencia en los sistemas de IA dentro de la administración pública, aunque aparentemente simple, es un concepto complejo y multifaético (Ananny & Crawford, 2018). Su definición y aplicación varían según las disciplinas y los contextos, pero sus elementos centrales giran en torno a la disponibilidad y la comprensibilidad de la información. Mientras las sociedades lidian con cuestiones de gobernanza, tecnología y derechos individuales, la transparencia sigue desempeñando un papel crucial en el fomento de **la rendición de cuentas, la confianza y la toma de decisiones informada**. Sin embargo, es esencial equilibrar la transparencia con otros valores importantes y reconocer sus limitaciones.

3. Operacionalizando el Principio de Transparencia

Tal y como se evidenció en la sección anterior, la transparencia en los sistemas de IA actúa como un concepto macro que engloba principios fundamentales para garantizar la compren-

sión, confianza y rendición de cuentas en el uso de estas tecnologías. Dentro de este marco general, la **explicabilidad** y la **interpretabilidad** emergen como subconceptos complementarios que permiten de cierta forma operacionalizar la transparencia (Schraagen et al., 2021; Sharma et al., 2019). Mientras que la interpretabilidad se refiere al nivel de entendimiento que una persona puede tener sobre un modelo y sus procesos internos —es decir, qué tan claro y accesible es el funcionamiento del sistema (Tan et al., 2016)—, la explicabilidad va un paso más allá, enfocándose en las acciones y mecanismos específicos utilizados para detallar y aclarar los resultados del sistema (Barredo Arrieta et al., 2020). De este modo, la explicabilidad facilita la justificación de decisiones, identifica los factores clave que influyen en los resultados y expone posibles sesgos, asegurando que la transparencia no sea solo un principio abstracto, sino una práctica tangible y medible. Juntos, estos conceptos forman la base para construir sistemas de IA que no solo sean técnicamente avanzados, sino también **accesibles, auditables y comprensibles** para los diferentes actores involucrados (Ebermann et al., 2023; Zhu et al., 2018).

La IA explicable (IAE; XAI, por sus siglas en inglés) se refiere al conjunto de métodos y técnicas que permiten comprender y justificar el funcionamiento de los sistemas de IA, facilitando la interpretación de los procesos y resultados que generan(ICO & The Alan Turing Institute, 2022). Según Montavon et al. (2018, p. 5), XAI busca proporcionar explicaciones claras y comprensibles para los humanos, permitiendo identificar cómo y por qué un modelo de IA llegó a una decisión específica Esta capacidad es crucial para abordar la “opacidad algorítmica” presente en los sistemas de aprendizaje automático, particularmente en los modelos de tipo “caja negra” donde ni siquiera los desarrolladores pueden rastrear las razones detrás de ciertos resultados (Palmiotto, 2021; Paudyai & Wong, 2018). Por su parte, Adadi y Berrada (2018) destacan que IAE no solo fomenta la transparencia y la confianza pública en las decisiones automatizadas, sino que también facilita la detección de sesgos, errores y vulnerabilidades en los sistemas. De esta forma, expertos en el campo de la IAE han identificado diversos objetivos(Balasubramaniam et al., 2023; Bücker et al., 2022; Coeckelbergh, 2019; ICO & The Alan Turing Institute, 2022; Schraagen et al., 2021)—los cuales se describen en la figura 3— que estos sistemas deberían cumplir para garantizar que es un sistema explicable:

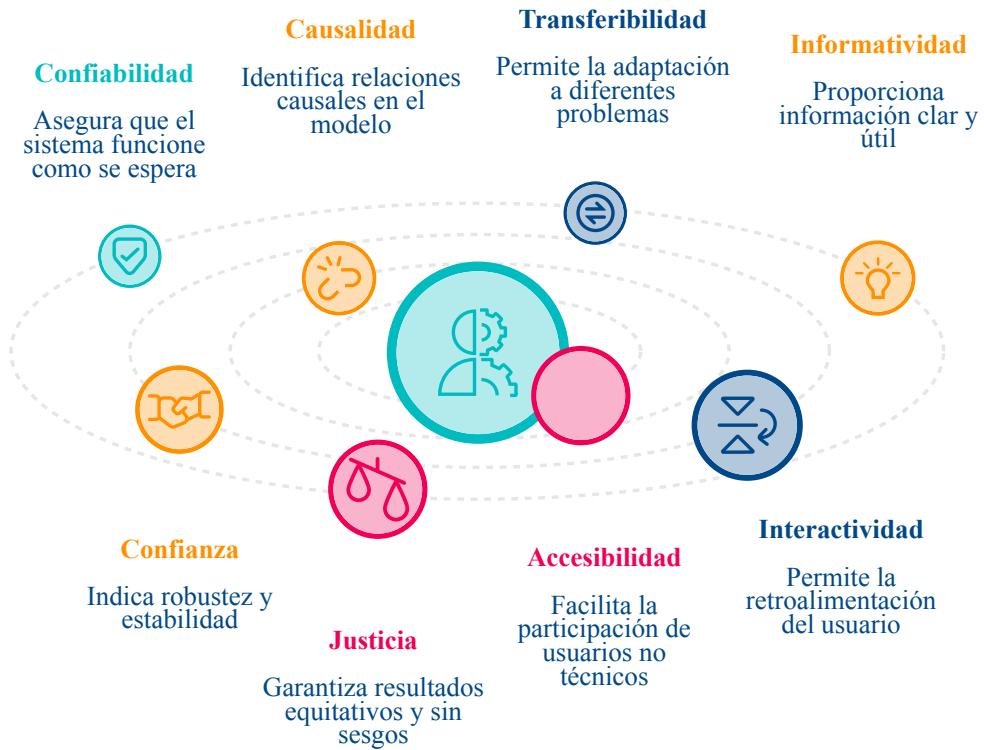


Figura 3. Objetivos de los sistemas de IA explicables. Elaborado por la autora²³.

Estos objetivos reflejan la necesidad de contar con sistemas de IA que, además de eficientes, sean transparentes, comprensibles y justos para generar confianza social. De esta forma, la explicabilidad puede centrarse en dos aspectos: la explicación de procesos y la explicación de resultados(ICO & The Alan Turing Institute, 2022). La explicación de procesos se enfoca en las etapas de diseño, desarrollo e implementación de un sistema de IA, demostrando las mejores prácticas empleadas(Vilone & Longo, 2021, p. 93). Se puede abordar en tres niveles:

- I. Transparencia algorítmica:** Permite seguir y entender el proceso mediante el cual los datos de entrada generan resultados.
- II. Descomponibilidad:** Explicar cada componente del modelo, incluyendo entradas, parámetros y cálculos.
- III. Simulabilidad:** La capacidad de una persona para reproducir y razonar sobre el funcionamiento del modelo.

²³ La gráfica fue elaborada con el apoyo de Napkin AI utilizando las características y el prompt proporcionados por la autora.

La explicación de resultados tiene como propósito aclarar cómo un sistema de IA llega a un resultado específico, utilizando diversas técnicas que facilitan su comprensión. Una de las estrategias más comunes son las *explicaciones textuales y visuales*, que permiten representar el funcionamiento del sistema mediante mapas de texto, gráficos o visualizaciones (Cascone et al., 2023; Vilone & Longo, 2021, p. 96). Por ejemplo, en un modelo que predice la probabilidad de otorgar un préstamo, se pueden usar gráficos para mostrar qué variables, como los ingresos o el historial crediticio, influyeron más en la decisión.

Otra técnica clave son las *explicaciones locales*, que se centran en analizar segmentos específicos del modelo. Esto permite entender cómo el sistema funciona en escenarios particulares, como explicar por qué a un solicitante de crédito específico se le rechazó un préstamo mientras a otro con características similares no (Linardatos et al., 2020, p. 18). Asimismo, el uso de *ejemplos representativos* ayuda a simplificar la complejidad del modelo al extraer subconjuntos de datos que muestran las correlaciones internas más relevantes (Hassija et al., 2024, p. 62). Por ejemplo, en un sistema de recomendación de películas, estos ejemplos podrían ilustrar cómo la edad y las preferencias anteriores de los usuarios influyen en las sugerencias.

Las *explicaciones simplificadas* son otra técnica útil, ya que reconstruyen versiones más simples del modelo original sin perder su desempeño (Barredo Arrieta et al., 2020, p. 85). Esto sería como usar un modelo más pequeño para explicar cómo se toman las decisiones de un sistema de diagnóstico médico complejo. Finalmente, la *identificación de características relevantes* cuantifica qué variables tienen mayor impacto en los resultados del modelo (Barredo Arrieta et al., 2020, p. 90). Por ejemplo, en un sistema que determina la contratación de un candidato, esta técnica podría mostrar que la experiencia profesional tuvo más peso que la educación académica.

Estas técnicas, ya sea aplicadas individualmente o de manera combinada, son fundamentales para auditar, mejorar y comprender los sistemas de IA, especialmente para usuarios finales que no tienen conocimientos técnicos. Facilitan la confianza en las decisiones automatizadas y garantizan que las tecnologías sean más accesibles y transparentes.

Con todo lo anterior, se puede concluir que la falta de explicabilidad representa un obstáculo crítico para generar confianza en los sistemas de IA, especialmente en contextos sensibles como la medicina, la justicia y las finanzas, donde las decisiones pueden tener un impacto significativo en la vida de las personas —o generar daños económicos o morales—(De Brito Duarte et al., 2023). La IAE procura abordar este desafío al proporcionar razones claras y comprensibles detrás de los resultados, permitiendo a los usuarios comprender cómo y por qué se tomaron ciertas decisiones. Esto no solo refuerza la confianza en que los sistemas operan de manera justa y robusta, sino que también asegura que los usuarios puedan participar activamente en la supervisión y el perfeccionamiento de estas tecnologías. Además, la IAE ayuda a identificar las limitaciones del sistema, su propósito y los datos que lo sustentan,

estableciendo una base sólida para su implementación ética y responsable. Al ofrecer explicaciones claras y accesibles, la IAE beneficia tanto a los desarrolladores como a los usuarios finales, fortaleciendo la relación entre la tecnología y la sociedad.

4. Desafíos y Tensiones en la Implementación de la Transparencia

La implementación del principio de transparencia en los sistemas de IA enfrenta desafíos significativos en América Latina, donde las realidades normativas, culturales y políticas no siempre están alineadas con los estándares internacionales de gobernanza transparente. Estos retos no solo limitan la rendición de cuentas y la equidad, sino que también generan tensiones que dificultan la adopción ética y efectiva de tecnologías de IA en el sector público(Contreras Caballol & Pefaur Dendal, 2020). A continuación, se describen los principales desafíos y tensiones, complementados con ejemplos concretos, para ilustrar cómo estas dinámicas se manifiestan en la región.

a. Choques Normativos y la Falta de Marco Regulatorio Específico

Un desafío recurrente en América Latina es la “disonancia normativa”, en la que los principios globales de transparencia no se traducen adecuadamente en las prácticas locales debido a vacíos regulatorios y falta de alineación con las realidades nacionales. Por ejemplo, en Colombia, aunque la Ley 1712 de 2014 promueve el acceso a la información pública, no contempla las especificidades técnicas y éticas asociadas a los sistemas de IA, limitando su aplicabilidad efectiva en contextos donde la tecnología juega un rol determinante en la toma de decisiones automatizadas. Según un informe realizado por Dejusticia como parte del proyecto de investigación “Inteligencia Artificial e Inclusión en América Latina,” esta regulación enfrenta múltiples desafíos en su implementación práctica (Saavedra & Upegui, 2021).

La Ley 1712 de 2014 establece como regla general la máxima publicidad de la información en posesión de las entidades del Estado, incluyendo juzgados, tribunales y altas cortes. Sin embargo, esta publicidad tiene limitaciones significativas. Por ejemplo, aunque el artículo 123 del Código General del Proceso permite cierto acceso a expedientes judiciales, las deliberaciones y actas de las salas de selección, según el artículo 55 del Acuerdo 02 de 2015, son de carácter reservado(Saavedra & Upegui, 2021, p. 15). Estas excepciones introducen opacidad en procesos clave, lo que dificulta garantizar la transparencia y la rendición de cuentas, especialmente en contextos donde la tecnología, como la IA, está siendo utilizada para apoyar decisiones críticas. Además, según Gutiérrez y Muñoz-Cadena (2022), el país cuenta con 113 sistemas automatizados de toma de decisiones en el sector público, pero muchos de ellos operan sin una supervisión adecuada y con niveles preocupantes de opacidad, como en el caso de **Fiscal Watson**, utilizado por la Fiscalía General para analizar patrones delictivos. La falta de información sobre los procesos de adquisición, criterios de diseño y auditoría evidencia una desconexión entre el marco normativo y su aplicación práctica (Flórez Rojas & Vargas Leal, 2020).

Esta brecha normativa es común en la región, donde los estándares internacionales se adoptan de manera superficial, sin mecanismos claros de adaptación y cumplimiento. Como señalan diversas organizaciones de la sociedad civil, es urgente que los países latinoamericanos desarrollen marcos regulatorios específicos para la IA que incluyan principios de transparencia, rendición de cuentas y equidad (AlSur, 2024).

b. Desafíos Culturales y Barreras de Comprensión

La transparencia efectiva no solo implica acceso a la información, sino también la capacidad de comprenderla. En América Latina, las brechas culturales y educativas limitan la capacidad de los ciudadanos para participar activamente en la supervisión de sistemas de IA. La falta de alfabetización digital, combinada con la complejidad técnica de los algoritmos, genera exclusión en sectores clave de la población, especialmente en comunidades rurales y marginadas (Katz, 2018).

El caso del Sistema de Identificación y Clasificación de Potenciales Beneficiarios de Programas Sociales (SISBEN) en Colombia ilustra este problema. Según López y Castañeda de la Fundación Karisma, aunque SISBEN utiliza un algoritmo para determinar la elegibilidad de las personas para programas sociales, opera como una “caja negra,” dejando a los beneficiarios sin capacidad de comprender cómo se calculan sus puntajes ni de cuestionar decisiones potencialmente inexactas o injustas (2020, p. 24). Esto genera una desconexión entre los ciudadanos y los sistemas que impactan directamente sus derechos sociales, erosionando la confianza pública. Para enfrentar este desafío, se requiere no solo diseñar sistemas más transparentes, sino también invertir en programas educativos que capaciten a los ciudadanos para entender y participar en la gobernanza de la IA (Felzmann et al., 2020; Zhu et al., 2018).

c. Tensiones Políticas y la Falta de Voluntad Gubernamental

La transparencia en la IA también enfrenta barreras políticas, donde la falta de voluntad gubernamental para compartir información clave obstaculiza la rendición de cuentas (Botero, 2023). En algunos casos, los gobiernos utilizan la opacidad para mantener el control sobre la tecnología, evitando que los ciudadanos y las organizaciones de la sociedad civil supervisen su implementación (Flórez Rojas, 2023a). Un ejemplo relevante es el sistema PretorIA, utilizado por la Corte Constitucional de Colombia para priorizar casos relacionados con derechos fundamentales. Aunque su diseño responde a un objetivo ético importante, la falta de claridad sobre los criterios utilizados para entrenar el modelo y la ausencia de información sobre el proceso de adquisición y las propuestas rechazadas limitan la capacidad de supervisión (Flórez Rojas & Vargas Leal, 2020; Saavedra & Upégui, 2021). Esto refleja una tendencia generalizada en la región, donde las tensiones políticas dificultan la implementación efectiva de principios de transparencia, perpetuando la falta de rendición de cuentas. Además, en algunos casos, los gobiernos pueden utilizar la opacidad como estrategia para mantener el control sobre la tecnología, limitando el acceso a información clave sobre cómo se imple-

mentan y auditán los sistemas de IA. Esto no solo dificulta la rendición de cuentas, sino que también perpetúa estructuras de poder desiguales, como lo han documentado (Hernández et al., 2022). Además, la falta de transparencia en la contratación de tecnologías de IA —*incluyendo la ausencia de mecanismos para evaluar su impacto ético y social*— pone en riesgo la legitimidad de estas tecnologías en la esfera pública (Flórez Rojas & Vargas Leal, 2020).

d. La Tensión Entre Transparencia y Privacidad

Otro desafío clave es el equilibrio entre transparencia y privacidad, especialmente en sistemas que procesan grandes volúmenes de datos personales sensibles. Por ejemplo, la Ley de Habeas Data en Colombia (Ley 1581 de 2012) busca proteger los derechos de privacidad de los ciudadanos, pero a menudo entra en conflicto con los principios de máxima divulgación de la Ley 1712 de 2014. En el contexto de la IA, esta tensión se amplifica. Por ejemplo, los algoritmos del SISBEN recolectan y procesan datos sensibles, pero la falta de claridad sobre cómo se manejan estos datos genera preocupación entre los ciudadanos, quienes temen violaciones de privacidad (López & Castañeda, 2020, p. 13,26).

Newman ya destacaba la importancia de aplicar la prueba de daño y el concepto de divulgación parcial para equilibrar el derecho de acceso a la información pública y la protección de los datos personales (2015, p. 48). Este enfoque permite analizar la naturaleza de los datos, clasificándolos como privados, semiprivados o sensibles, y considerar si están sujetos a legislaciones específicas. Además, propone evaluar las bases de datos según su contexto, como las de seguridad, justicia, inteligencia o aquellas en manos de periodistas y registros públicos. La implementación de estos criterios podría ayudar a mitigar las tensiones entre ambos derechos.

Para abordar este problema, es esencial establecer mecanismos como auditorías algorítmicas y herramientas de anonimización de datos que permitan gestionar esta dualidad de manera equilibrada (Bücker et al., 2022; Flórez Rojas, 2023b; Metcalf et al., 2021). Sin embargo, este desafío se ve agravado por interpretaciones fragmentadas y, en ocasiones, una preferencia por proteger datos sensibles a expensas de la rendición de cuentas pública (Alianza Más Información Más Derechos, 2024).

e. Desafíos Comerciales y Propiedad Intelectual

Otro obstáculo significativo proviene de los intereses comerciales que rodean a los sistemas de IA. De esta forma, la viabilidad económica de muchas empresas tecnológicas depende de algoritmos patentados, lo que genera tensiones con el objetivo de transparencia (Wischmeyer, 2020). Por ejemplo, en el caso del sistema Fiscal Watson en Colombia, la falta de información pública sobre los criterios utilizados para su implementación y los mecanismos de auditoría revela cómo los acuerdos comerciales pueden restringir la supervisión ciudadana y gubernamental (Morales Higuita et al., 2021). Sin un marco que permita equilibrar la pro-

piedad intelectual con la transparencia, estos sistemas corren el riesgo de operar en un limbo ético, donde los derechos de los ciudadanos quedan subordinados a los intereses privados.

Para abordar estos desafíos, se requiere un enfoque integral que combine legislación, educación, auditorías independientes y participación ciudadana. Autores como Barocas, et al., (2023) proponen implementar evaluaciones de impacto algorítmico para garantizar que los sistemas de IA sean diseñados y operados de manera ética y transparente. Además, diversas organizaciones enfatizan la importancia de empoderar a la sociedad civil para supervisar activamente el uso de tecnologías de IA, garantizando que los ciudadanos tengan acceso a herramientas significativas para cuestionar y, si es necesario, corregir decisiones automatizadas (Alianza Más Información Más Derechos, 2024; AlSur, 2024).

En conclusión, los desafíos asociados con la implementación del principio de transparencia en la IA en América Latina son complejos y multidimensionales, pero no insuperables. Adoptar enfoques que equilibren intereses comerciales, derechos ciudadanos y valores éticos no solo fortalecerá la confianza pública en estas tecnologías, sino que también asegurará que se implementen de manera justa, inclusiva y responsable. Los casos de Colombia ofrecen lecciones valiosas que pueden guiar el desarrollo de políticas más robustas y adaptadas al contexto regional.

5. Estrategias Prácticas y Recomendaciones de Políticas

La implementación efectiva del principio de transparencia en la IA en América Latina enfrenta desafíos complejos debido a las brechas normativas, técnicas y sociales presentes en la región. No obstante, existen estrategias prácticas y recomendaciones de políticas públicas que pueden guiar la integración de la transparencia en los sistemas de IA, asegurando no solo su implementación legal, sino su operación efectiva en contextos locales. Las recomendaciones puntuales que se hacen son:

Estrategias de Transparencia en IA

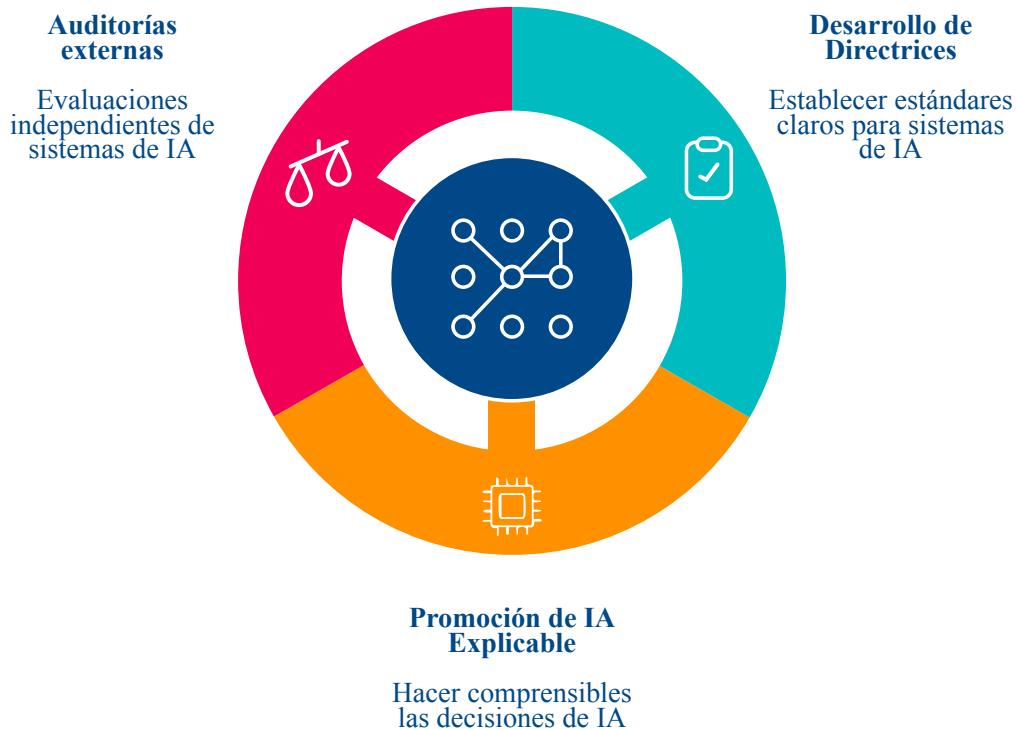


Figura 3. Estrategias para operacionalizar el Principio de Transparencia. Elaborado por la autora²⁴.

a. Desarrollo de Directrices y Estándares Claros

Una de las estrategias clave es desarrollar directrices y estándares claros para el diseño e implementación de sistemas de IA transparentes (Gutiérrez Rodríguez & Muñoz-Cadena, 2022). En este sentido, los marcos legales y técnicos deben ser elaborados con el objetivo de garantizar la publicación de documentación, la claridad sobre las fuentes de datos y la definición de métricas de evaluación. Estos estándares deben ser accesibles para que los actores del sector público y privado, así como los ciudadanos, puedan entender cómo funcionan los sistemas de IA y qué datos influyen en sus decisiones (Soler Garrido et al., 2023).

En Colombia, por ejemplo, la Ley 1712 de 2014 sobre acceso a la información pública re-

²⁴ La gráfica fue elaborada con el apoyo de Napkin AI utilizando las características y el prompt proporcionados por la autora.

presenta un avance significativo hacia la promoción de la transparencia, pero su aplicación efectiva en el contexto de la IA sigue enfrentando importantes desafíos. Aunque la ley establece el principio de máxima publicidad, también contempla una serie de **excepciones** que pueden limitar la divulgación de información (Newman, 2015, p. 35). Estas excepciones, si bien necesarias en algunos casos, pueden obstaculizar la transparencia en el uso de sistemas de IA cuando no se aplican de manera estricta y justificada.

De acuerdo con los artículos 18 y 19 de la ley, las entidades públicas pueden rechazar solicitudes de acceso a información si consideran que su divulgación puede causar daño a derechos de personas naturales o jurídicas, como la intimidad, la vida, la salud, la seguridad o los secretos comerciales e industriales. Asimismo, el acceso puede ser negado si afecta intereses públicos relacionados con la defensa y seguridad nacional, la seguridad pública, las relaciones internacionales, la investigación y persecución de delitos, o la estabilidad macroeconómica y financiera del país (Fundamedios & Thomson Reuters Foundation, 2020). Aunque estas excepciones tienen una base legítima, su aplicación puede volverse problemática si no se acompaña de un razonamiento claro y una motivación escrita, tal como lo exige la normativa.

Por otro lado, el artículo 21 introduce un mecanismo que busca equilibrar la necesidad de proteger información sensible con la obligación de garantizar la transparencia. Según esta disposición, cuando solo una parte de la información está protegida por las excepciones, las entidades están obligadas a generar una versión pública del documento, ocultando únicamente los datos indispensables para mantener la reserva. Sin embargo, en la práctica, la implementación de este mecanismo es limitada debido a la falta de recursos y capacidades técnicas en muchas entidades públicas, así como a la falta de supervisión efectiva sobre su cumplimiento.

Un aspecto relevante es la obligación de las entidades públicas de mantener un índice actualizado de los documentos clasificados o reservados, según lo estipula el artículo 20. Este índice debería facilitar el monitoreo ciudadano sobre qué información está siendo protegida y bajo qué justificaciones. No obstante, en la realidad, muchas entidades no cumplen con esta obligación, lo que dificulta evaluar si las excepciones están siendo utilizadas de manera adecuada o arbitraria.

En el contexto de la IA, estas excepciones pueden no solo convertirse en un obstáculo importante para la transparencia, sino que pueden potencializar la opacidad. Por ejemplo, en la adquisición de cualquier sistema de IA se plantean preguntas sobre cómo se manejan los datos procesados por estos sistemas y qué criterios se utilizan para proteger información bajo la ley. La falta de divulgación sobre los procesos de recolección de datos y la metodología de decisión no solo dificulta la rendición de cuentas, sino que también refuerza la percepción de opacidad en la gobernanza de la IA.

Finalmente, aunque las excepciones temporales (artículo 22) limitan la reserva de informa-

ción a un máximo de 15 años en ciertas circunstancias, estas disposiciones no son suficientes para garantizar un equilibrio adecuado entre la transparencia y la privacidad o la seguridad nacional. Es fundamental que estas excepciones se apliquen únicamente cuando exista un daño evidente que supere el interés público de acceder a la información, tal como lo estipula el principio de proporcionalidad de la ley.

Así, si bien la Ley 1712 de 2014 establece un marco para la transparencia, su efectividad depende de la correcta implementación y supervisión de las excepciones. En el contexto de la IA, es urgente que las entidades públicas adopten mecanismos más robustos para justificar y limitar el uso de estas excepciones, asegurando que no se conviertan en un pretexto para perpetuar la opacidad en sistemas que impactan directamente los derechos de los ciudadanos. Además, será clave establecer directrices claras para que la información asociada a los sistemas de IA sea accesible y útil, promoviendo la rendición de cuentas y fortaleciendo la confianza pública.

b. Promoción de la IA Explicable (IAE)

Otra estrategia central es promover el uso de técnicas de IAE. En donde se promueva el uso de sistemas que permiten a los humanos comprender el razonamiento detrás de las decisiones y predicciones de la IA, lo cual facilita las auditorías y depuración más efectivas(Shin, 2021; Vilone & Longo, 2021). Por ejemplo, en el sector de la salud, donde los sistemas de IA se utilizan para hacer diagnósticos o proponer tratamientos, es esencial que los profesionales puedan entender cómo se llegaron a las recomendaciones del sistema para poder verificarlas (Ada Lovelace Institute, 2022; Ng et al., 2022). En América Latina, muchos sistemas de IA aún operan como “cajas negras”, lo que dificulta la supervisión y genera desconfianza. Los gobiernos deben fomentar la integración de XAI en los sistemas implementados, garantizando que no solo los expertos, sino también los ciudadanos, puedan acceder a explicaciones claras sobre las decisiones automatizadas que afectan su vida diaria(Ada Lovelace Institute, 2022; Kaminski & Malgieri, 2020; Selbst, 2021).

c. Implementación de Mecanismos de Auditoría Externa y Evaluación Independiente

Los gobiernos deben crear marcos de auditoría algorítmica que evalúen el impacto y la equidad de los sistemas de IA, asegurando que las decisiones tomadas por estos sistemas no estén sesgadas ni violen los derechos de los ciudadanos (Adler et al., 2018; Raji et al., 2020). En países como Brasil y Argentina, los sistemas de IA utilizados en la justicia o en la asignación de recursos sociales deben someterse a auditorías periódicas realizadas por entidades independientes que puedan verificar su cumplimiento con las normativas de transparencia y justicia (Aguerre & Daniel, 2021; Gaspar & Curzi de Mendonça, 2021; Ministry of Science-Technology and Innovation, 2021; Ortiz & Iglesias, 2018).

De esta forma, la implementación de la transparencia en los sistemas de IA en América Latina requiere un enfoque integral que combine legislación, educación, cooperación internacional y auditoría constante. A medida que los sistemas de IA se integran más en las políticas públicas, es esencial que la región desarrolle marcos regulatorios robustos y efectivos que aseguren que estas tecnologías se implementen de manera ética y responsable. Adoptar marcos no solo como ideales, sino que se pueden efectivamente materializar en el contexto regional como y fomentar la IAE son pasos cruciales para avanzar hacia una gobernanza tecnológica más transparente y accesible para todos los ciudadanos.



6. Conclusiones

La implementación de la IA en América Latina, y particularmente el principio de transparencia, se posiciona como un tema crucial en la búsqueda de una gobernanza tecnológica responsable, ética y accesible. A lo largo de este white paper, se han identificado los desafíos contemporáneos que enfrenta la región, destacando las tensiones normativas, políticas, culturales y comerciales que dificultan la materialización efectiva de estándares internacionales de transparencia en contextos locales.

Uno de los puntos clave de este análisis es la necesidad de cerrar la brecha entre los ideales globales y las realidades regionales. En América Latina, la opacidad inherente a los sistemas de IA y la falta de marcos regulatorios específicos exacerbaban problemas preexistentes de inequidad y desconfianza hacia las instituciones públicas. Casos como Fiscal Watson, Pretoria y SISBEN en Colombia ilustran cómo la falta de claridad en los procesos de adquisición, la ausencia de auditorías independientes y la desconexión entre los ciudadanos y los sistemas automatizados pueden perpetuar estas tensiones.

El principio de transparencia no debe limitarse a la divulgación técnica de algoritmos y datos. Por el contrario, debe integrarse como un valor fundamental en todas las etapas del desarrollo e implementación de la IA. Para lograrlo, es indispensable adoptar un enfoque integral que combine legislación clara, educación inclusiva, auditorías independientes y la promoción de la IAE. Estos esfuerzos no solo facilitarán la rendición de cuentas, sino que también permitirán a los ciudadanos participar activamente en la supervisión y el perfeccionamiento de estas tecnologías.

Asimismo, el equilibrio entre transparencia y privacidad sigue siendo un desafío central que requiere soluciones adaptadas al contexto regional. Las excepciones previstas en marcos legales como la Ley 1712 de 2014 en Colombia deben aplicarse con criterios estrictos y justificados, evitando que se conviertan en herramientas para perpetuar la opacidad. A través de estrategias como el uso de herramientas de anonimización de datos, auditorías algorítmicas y divulgaciones parciales, es posible garantizar que la transparencia coexista con la protección de derechos fundamentales.

Finalmente, este documento resalta la importancia de fomentar la colaboración internacional, tomando como referencia marcos regulatorios como el Reglamento de IA de la Unión Europea y las recomendaciones de la UNESCO, mientras se adaptan a las necesidades y capacidades locales. La transparencia en IA no es solo una obligación técnica o legal; es una herramienta esencial para construir confianza pública, promover la equidad y garantizar que la tecnología sirva a los intereses de toda la sociedad.

En conclusión, América Latina tiene una oportunidad única de avanzar hacia una gobernanza tecnológica inclusiva y ética. Esto requerirá voluntad política, compromiso intersectorial y un enfoque constante en la protección de derechos humanos. La adopción de sistemas de IA transparentes y explicables permitirá no solo abordar los desafíos actuales, sino también sentar las bases para un futuro donde la tecnología se integre de manera equitativa y responsable en nuestras sociedades.

7. Referencias

Ada Lovelace Institute. (2022). *Algorithmic impact assessment: a case study in healthcare*. <https://www.adalovelaceinstitute.org/resource/aia-user-guide/>

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>

Addink, H. (2019). The Principle of Transparency. *Good Governance*, 111–128. <https://doi.org/10.1093/OSO/9780198841159.003.0008>

Adler, P., Falk, C., Friedler, S. A., Nix, T., Rybeck, G., Scheidegger, C., Smith, B., & Venkatasubramanian, S. (2018). Auditing black-box models for indirect influence. *Knowledge and Information Systems*, 54(1), 95–122. <https://doi.org/10.1007/S10115-017-1116-3>

Aguerre, C., & Daniel, M. L. (2021). *Public policy reports: Argentina*. <https://www.empatia.la/wp-content/uploads/2021/12/Policy-report-Argentina-version-final.pdf>

Proyecto de Ley- Carta de Derechos: Hacer Que Los Sistemas Automatizados Trabajen al Servicio Del Pueblo Americano (2022). <https://www.whitehouse.gov/ostp/ai-bill-of-rights>

Alianza Más Información Más Derechos. (2024). *Acceso a la información pública, en riesgo: nuestra alerta al Relator Especial para la Libertad de Expresión*. https://www.dejusticia.org/alerta_por_acceso_informacion_publica_en_riesgo/

AlSur. (2024). *Los caminos regulatorios para la IA en América Latina: Recopilación de es-*

tudios de caso en Brasil, México, Perú y Colombia (A. Ovelar, Trans.). [https://www.alsur.lat/sites/default/files/2024-09/ALSUR%20-%20IA%20en%20Latam%20\[ESP\].pdf](https://www.alsur.lat/sites/default/files/2024-09/ALSUR%20-%20IA%20en%20Latam%20[ESP].pdf)

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. <https://doi.org/10.1177/1461444816676645>

Balasubramiam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K., & Kujala, S. (2023). Transparency and explainability of AI systems: From ethical guidelines to requirements. *Information and Software Technology*, 159, 107197. <https://doi.org/10.1016/J.INFSOF.2023.107197>

Barocas, Solon., Hardt, Moritz., & Narayanan, Arvind. (2023). *Fairness and machine learning : limitations and opportunities*.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/J.INFFUS.2019.12.012>

Botero, C. (2023). The shortcomings of Artificial Intelligence policies. *El Espectador*. <https://www.elespectador.com/opinion/columnistas/carolina-botero-cabrera/las-deficiencias-de-las-politicas-de-inteligencia-artificial/>

Bücker, M., Szepannek, G., Gosiewska, A., & Biecek, P. (2022). Transparency, auditability, and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society*, 73(1), 70–90. <https://doi.org/10.1080/01605682.2021.1922098>

Cascone, L., Pero, C., & Proença, H. (2023). Visual and textual explainability for a biometric verification system based on piecewise facial attribute analysis. *Image and Vision Computing*, 132, 104645. <https://doi.org/10.1016/J.IMAVIS.2023.104645>

Case T-323 of 2024 (2024).

CEPAL, & Naciones Unidas. (2012). *Las tecnologías digitales frente a los desafíos de una educación inclusiva en América Latina Algunos casos de buenas prácticas* (G. Sunkel & D. Trucco, Eds.). Naciones Unidas, CEPAL. https://repositorio.cepal.org/bitstream/handle/11362/35386/S2012809_es.pdf

Cheong, B. C. (2024). Transparency and accountability in AI systems: safeguarding well-being in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6, 1421273. <https://doi.org/10.3389/FHUMD.2024.1421273/BIBTEX>

Coeckelbergh, M. (2019). Artificial Intelligence, Responsibility Attribution, and a Relational

Justification of Explainability. *Science and Engineering Ethics*, 1–18. <https://doi.org/10.1007/s11948-019-00146-8>

Contini, F., Ontanu, E. A., & Velicogna, M. (2024). AI Accountability in Judicial Proceedings: An Actor–Network Approach. *Laws 2024, Vol. 13, Page 71, 13(6)*, 71. <https://doi.org/10.3390/LAWS13060071>

Contreras Caballol, D., & Pefaur Dendal, D. (2020). *Transparencia algorítmica buenas prácticas y estándares de transparencia en el proceso de toma de decisiones automatizadas*. <https://doi.org/ISSN 0719-4609>

Cumbre Ministerial y de Altas Autoridades de América Latina y el Caribe. (2023). *Declaration of Santiago to promote ethical Artificial Intelligence in Latin America and the Caribbean*. https://minciencia.gob.cl/uploads/filer_public/40/2a/402a35a0-1222-4dab-b090-5c81bbf34237/declaracion_de_santiago.pdf

De Brito Duarte, R., Correia, F., Arriaga, P., & Paiva, A. (2023). AI Trust: Can Explainable AI Enhance Warranted Trust? *Human Behavior and Emerging Technologies*, 2023(1), 1–12. <https://doi.org/10.1155/2023/4637678>

Directive on Automated Decision-Making (2021).

Ebermann, C., Selisky, M., & Weibelzahl, S. (2023). Explainable AI: The Effect of Contradictory Decisions and Explanations on Users' Acceptance of AI Systems. *International Journal of Human–Computer Interaction*, 39(9), 1807–1826. <https://doi.org/10.1080/10447318.2022.2126812>

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards Transparency by Design for Artificial Intelligence. *Science and Engineering Ethics*, 26(6), 3333–3361. <https://doi.org/10.1007/S11948-020-00276-4/FIGURES/1>

Flórez Rojas, M. L. (2023a). Pensamiento de diseño y marcos éticos para la Inteligencia Artificial: una mirada a la participación de las múltiples partes interesadas. *Desafíos*, 35(1), 1–31. <https://doi.org/10.12804/revistas.urosario.edu.co/desafios/a.12183>

Flórez Rojas, M. L. (2023b). Vista de Las evaluaciones de impacto algorítmico: Una visión más allá de la ética. *GIGAPP Estudios Working Papers*, 10(267), 335–350. <https://www.gigapp.org/ewp/index.php/GIGAPP-EWP/article/view/327/328>

Flórez Rojas, M. L., & Vargas Leal, J. (2020). *El impacto de las herramientas de Inteligencia Artificial: Un análisis en el sector público en Colombia*. <https://guia.ai/wp-content/uploads/2020/05/GECTI-El-impacto-de-herramientas-de-inteligencia-artificial.pdf>

Fundamedios, & Thomson Reuters Foundation. (2020). *Analisis comparativo de acceso a la informacion publica en America Latina*. <https://www.fundamedios.org.ec/wp-content/>

[uploads/2020/10/Informe-Ana%CC%81lisis-Acceso-Informaci%C3%B3n-P%C3%BAblica.pdf](https://cyberbrics.info/wp-content/uploads/2020/10/Informe-Ana%CC%81lisis-Acceso-Informaci%C3%B3n-P%C3%BAblica.pdf)

Gaspar, W., & Curzi de Mendonça, Y. (2021, May). Artificial Intelligence In Brazil Still Needs A Strategy. *Center for Technology and Society at FGV Law School, Cyberbrics*, 1–8. <https://cyberbrics.info/wp-content/uploads/2021/05/EBIA-en-2.pdf>

Guío Español, A. (2021). *Task Force for the development and implementation of Artificial Intelligence in Colombia*. <https://dapre.presidencia.gov.co/TD/CONSEJO-INTERNACIONAL-INTELIGENCIA-ARTIFICIAL-COLOMBIA.pdf>

Gutiérrez Rodríguez, J. D. (2023). AI Technologies in the Judiciary: Critical Appraisal of Large Language Models in Judicial Decision-making. In P. Regine, E. Carmel, & J. Cobbe (Eds.), *Handbook on public policy and artificial intelligence* (pp. 323–339). Edward Elgar Publishing. <https://papers.ssrn.com/abstract=4667572>

Gutiérrez Rodríguez, J. D., & Muñoz-Cadena, S. M. (2022). Sistemas de decisión automatizada implementados por el Estado Colombiano (Automated Decision-Making Systems Implemented By The Colombian State). *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4294032>

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, 16(1), 45–74. <https://doi.org/10.1007/S12559-023-10179-8/FIGURES/14>

Hernández, L., Canales, M. P., & De Souza, M. (2022). *Artificial Intelligence and Participation in Latin America: National AI Strategies*. <https://www.derechosdigitales.org/publicaciones/inteligencia-artificial-y-participacion-en-america-latina-las-estrategias-nacionales-de-ia/>

ICO, & The Alan Turing Institute. (2022). *Explaining decisions made with AI*. <https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf>

Kaminski, M. E., & Malgieri, G. (2020). Multi-layered Explanations from Algorithmic Impact Assessments in the GDPR. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. <https://doi.org/10.1145/3351095>

Katz, R. L. (2018). *Capital humano para la transformación digital en América Latina | Publicación | Comisión Económica para América Latina y el Caribe*. <https://doi.org/1680-8754>

Kaufmann, C., & Weber, R. H. (2010). The Role of Transparency in Financial Regulation. *Journal of International Economic Law*, 13(3), 779–797. <https://doi.org/10.1093/JIEL/JGQ037>

Kim, S., & Lee, J. (2019). Citizen Participation, Process, and Transparency in Local Government: An Exploratory Study. *Policy Studies Journal*, 47(4), 1026–1047. <https://doi.org/10.1111/PSJ.12236>

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy*, 23(1), 1–45. <https://doi.org/10.3390/E23010018>

López, J., & Castañeda, J. D. (2020). *Experimentando con la pobreza: El SISBÉN y los proyectos de analítica de datos en Colombia*. <https://web.karisma.org.co/wp-content/uploads/download-manager-files/Experimentando%20con%20la%20pobreza.pdf>

Magrani, E. (2021). *Estrategia Brasileña de Inteligencia Artificial: comentarios sobre la Resolución 4.617/2021 del Ministerio de Ciencia, Tecnología e Innovación (MCTI)*. <https://www.empatia.la/wp-content/uploads/2021/12/ART.-DE-OPINION-Brasil-ES-final.pdf>

Metcalf, J., Anne Watkins, E., Singh, R., Clare Elish, M., & Moss, E. (2021). Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts CCS CONCEPTS. *ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 735–746. <https://doi.org/10.1145/3442188>

Ministry of ScienceTechnology and Innovation. (2021). *Brazilian Strategy of Artificial Intelligence (EBIA) and Public Policies*. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-summary_brazilian_4-979_2021.pdf

Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing: A Review Journal*, 73, 1–15. <https://doi.org/10.1016/J.DSP.2017.10.011>

Morales Higuita, L., Agudelo Londoño, S., Montoya Raigosa, M., & Montoya Vidales, A. M. (2021). Inteligencia artificial en el proceso penal: Análisis a la luz del fiscal Watson. *Pensamiento jurídico*, 54, 147–164. <https://dialnet.unirioja.es/servlet/articulo?codigo=8252942&info=resumen&idioma=SPA>

Newman, V. (2015). *Datos personales en información pública: Oscuridad en lo privado y luz en lo público*. Dejusticia. <https://www.dejusticia.org/publication/datos-personales-en-informacion-publica-oscuridad-en-lo-privado-y-luz-en-lo-publico/>

Ng, M. Y., Kapur, S., Blizinsky, K. D., & Hernandez-Boussard, T. (2022). The AI life cycle: a holistic approach to creating ethical AI for health decisions. *Nature Medicine*, 28(11), 2247. <https://doi.org/10.1038/S41591-022-01993-Y>

OECD. (2021). *Artificial Intelligence, Machine Learning and Big Data in Finance Opportunities, Challenges and Implications for Policy Makers*. <https://www.oecd.org/finance/financial-markets/Artificial-intelligence-machine-learning-big-data-in-finance.pdf>

OECD/CAF. (2022). *The Strategic and Responsible Use of Artificial Intelligence in the Public Sector of Latin America and the Caribbean*. <https://doi.org/10.1787/1f334543-en>

Ortiz, F. J., & Iglesias, C. (2018). *Algorithms and Artificial Intelligence in Latin America: A Study of Implementation by Governments in Argentina and Uruguay*. https://webfoundation.org/docs/2018/09/WF_AI-in-LA_Report_Screen_AW.pdf

Palmiotto, F. (2021). The Black Box on Trial: The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings. In M. Ebers & M. Cantero Gamito (Eds.), *Algorithmic Governance and Governance of Algorithms* (pp. 49–70). Springer, Cham. https://doi.org/10.1007/978-3-030-50559-2_3

Paudyai, P., & Wong, B. L. W. (2018). Algorithmic Opacity: Making Algorithmic Processes Transparent through Abstraction Hierarchy. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1(62), 192–196. <https://doi.org/10.1177/1541931218621046>

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44. <https://doi.org/10.1145/3351095.3372873>

Regulation EU 2024/1689 (AI Act), Pub. L. No. 2024/1689, OJ L, 2024/1689 (2024).

Flórez Rojas, M. L. (2024). Artificial Intelligence in Judicial Decision-Making: A Comparative Analysis of Recent Rulings in Colombia and The Netherlands. *Journal of AI Law and Regulation*, 1(3), 361–366. <https://doi.org/10.21552/AIRE/2024/3/11>

Romaní, M. S., Korving, J., & Eliantonio, M. (2024). Transparency and foreseeable relevance in exchange of information procedures. In *Exchange of Information in the EU: Taxpayers' Rights, Transparency and Effectiveness*. Edward Elgar Publishing. <https://www.elgaronline.com/edcollchap/book/9781035314560/book-part-9781035314560-9.xml>

Saavedra, V., & Upegui, J. C. (2021). *PretorIA y la automatización del procesamiento de causas de derechos humanos*. <https://ia.derechosdigitales.org/>.

Schraagen, J. M., Lopez, S. K., Schneider, C., Schneider, V., Tönjes, S., & Wiechmann, E. (2021). The Role of Transparency and Explainability in Automated Systems. *Pro-*

ceedings of the Human Factors and Ergonomics Society, 65(1), 27–31. <https://doi.org/10.1177/1071181321651063>

Selbst, A. D. (2021). An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology (Harvard JOLT)*, 35.

Sharma, S., Henderson, J., & Ghosh, J. (2019). CERTIFAI: Counterfactual Explanations for Robustness, Transparency, Interpretability, and Fairness of Artificial Intelligence models. In A. Markham, J. Powles, T. Walsh, & A. L. Washington (Eds.), *AIES '20: AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery. <https://doi.org/10.1145/3375627.3375812>

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551. <https://doi.org/10.1016/J.IJHCS.2020.102551>

Soler Garrido, J., Tolan, S., Hupont Torres, I., Fernández Llorca, D., Charisi, V., Gómez, E., Junklewitz, H., Hamon, R., Fano, D., & Panigutti, C. (2023). *AI Watch: Artificial Intelligence Standardisation Landscape Update*. <https://doi.org/10.2760/131984>

Tan, S., Sim, K. C., & Gales, M. (2016). Improving the interpretability of deep neural networks with stimulated learning. *2015 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2015 - Proceedings*, 617–623. <https://doi.org/10.1109/ASRU.2015.7404853>

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

Varošanec, I. (2022). On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI. *International Review of Law, Computers & Technology*, 36(2), 95–117. <https://doi.org/10.1080/13600869.2022.2060471>

Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76, 89–106. <https://doi.org/10.1016/J.INFFUS.2021.05.009>

Wischmeyer, T. (2020). Artificial Intelligence and Transparency: Opening the Black Box. *Regulating Artificial Intelligence*, 75–101. https://doi.org/10.1007/978-3-030-32361-5_4

Zhu, J., Liapis, A., Risi, S., Bidarra, R., & Youngblood, G. M. (2018). Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. *IEEE Conference on Computational Intelligence and Games, CIG, 2018-August*. <https://doi.org/10.1109/CIG.2018.8490433>



○

Herramientas de Toma Automatizada

de decisiones en el sector
público de Colombia

JUAN DAVID GUTIÉRREZ



Juan David Gutiérrez

Profesor asociado de la Escuela de Gobierno de la Universidad de los Andes, en donde enseña e investiga sobre políticas públicas, gestión pública, e inteligencia artificial. Doctor (PhD) y máster (MSc) en política pública de la Universidad de Oxford. Máster en derecho y economía (LLM) otorgado conjuntamente por la Universidad de Bolonia y la Universidad Erasmus de Rotterdam. Abogado de la Universidad Javeriana. Colidera la Mesa de Trabajo Multiactor sobre Regulación de Inteligencia Artificial coorganizada por la Universidad de los Andes y la Universidad Externado. Es integrante del grupo de expertos del Global Partnership on Artificial Intelligence – GPAI en donde colideró el proyecto sobre Transparencia Algorítmica en el Sector Público. Juan David tiene casi 20 años de experiencia como consultor en asuntos legales y de políticas públicas, asesorando a gobiernos de América Latina y el Caribe, organizaciones multilaterales, organizaciones de la sociedad civil, asociaciones gremiales y empresas.

Herramientas de toma automatizada de decisiones en el sector público de Colombia

Juan David Gutiérrez²⁵

Resumen

El texto analiza el uso de sistemas de toma automatizada de decisiones (SDA), incluidas herramientas de inteligencia artificial (IA), en el sector público colombiano, destacando su papel en la transformación digital del Estado para mejorar la eficiencia y garantizar derechos fundamentales. Se documentan 383 SDA en diversas entidades y niveles del gobierno, con aplicaciones en áreas como educación, protección social y seguridad pública. Además, se aborda la importancia de la transparencia algorítmica, definida como el acceso a información sobre estos sistemas en todas sus fases, y se examinan iniciativas del Gobierno Nacional, la Rama Judicial y el Congreso para regular su uso mediante políticas públicas, sentencias judiciales y proyectos legislativos. Finalmente, se proponen recomendaciones para fortalecer la gobernanza de estos sistemas y asegurar su alineación con los mandatos constitucionales.

Palabras clave: sistemas de toma automatizadas de decisiones, inteligencia artificial, sector público, transparencia algorítmica, Colombia

Abstract

The text analyzes the use of automated decision-making systems (ADS), including artificial intelligence (AI) tools, in the Colombian public sector, highlighting their role in the digital transformation of the state to improve efficiency and guarantee fundamental rights. It do-

25 Profesor asociado de la Escuela de Gobierno de la Universidad de los Andes, en donde enseña e investiga sobre políticas públicas, gestión pública, e inteligencia artificial. Doctor (PhD) y máster (MSc) en política pública de la Universidad de Oxford. Máster en derecho y economía (LLM) otorgado conjuntamente por la Universidad de Bolonia y la Universidad Erasmus de Rotterdam. Abogado de la Universidad Javeriana. Dirige el proyecto Sistemas de Algoritmos Públicos en la Universidad de los Andes. Colabora en la Mesa de Trabajo Multiactor sobre Regulación de Inteligencia Artificial coorganizada por la Universidad de los Andes y la Universidad Externado. Es integrante del grupo de expertos del Global Partnership on Artificial Intelligence – GPAI en donde lidera el proyecto sobre Transparencia Algorítmica en el Sector Público. Juan David tiene casi 20 años de experiencia como consultor en asuntos legales y de políticas públicas, asesorando a gobiernos de América Latina y el Caribe, organizaciones multilaterales (UNESCO, CE-PAL, BID, APEC), organizaciones de la sociedad civil (Ford Foundation, NRGI, ILDA, Fundación Corona, SOMO, CEJ), asociaciones gremiales y empresas.

cuments 383 SDAs in various entities and levels of government, with applications in areas such as education, social protection and public safety. In addition, the importance of algorithmic transparency, defined as access to information on these systems in all their phases, is addressed, and initiatives by the National Government, the Judicial Branch and Congress to regulate their use through public policies, court rulings and legislative projects are examined. Finally, recommendations are proposed to strengthen the governance of these systems and ensure their alignment with constitutional mandates.

Key words: automated decision-making systems, artificial intelligence, public sector, algorithmic transparency, Colombia.

1. Introducción²⁶

El poder —entendido como la capacidad de actuar— está en el corazón del debate sobre los sistemas computacionales que el Estado adopta y cómo los utiliza para cumplir sus funciones. Estos sistemas tecnológicos otorgan al Estado capacidades que de otro modo serían imposibles o significativamente más lentas de ejecutar. Esto resulta particularmente evidente cuando hablamos de sistemas algorítmicos que automatizan tareas y poseen cierto grado de autonomía operativa, es decir, cuando incorporan herramientas de inteligencia artificial (IA).

En un Estado Social de Derecho y una sociedad democrática como Colombia, el ejercicio del poder no solo debe tener límites claros, sino que debe orientarse específicamente hacia el cumplimiento de las funciones que la ley asigna a cada entidad pública. Cuando el Estado colombiano incorpora sistemas computacionales en sus procesos de toma de decisiones —especialmente aquellos que automatizan funciones— las preguntas sobre los límites de ese poder estatal continúan siendo fundamentales y adquieren nuevas dimensiones.

A la luz de la Constitución colombiana, cualquier respuesta sobre el papel de los sistemas computacionales en nuestro sector público debe contemplar al menos dos elementos esenciales. Primero, su implementación debe estar orientada a promover, respetar y proteger los derechos fundamentales y colectivos. Segundo, en coherencia con nuestra aspiración de consolidar un Estado Social de Derecho, no basta con que estos sistemas cumplan formalmente el marco legal durante su ciclo de vida; es imperativo que su uso sirva como instrumento efectivo para materializar los derechos y garantías consagrados en la Constitución.

Este texto busca contribuir al entendimiento de cómo el Estado colombiano utiliza sistemas de toma automatizada de decisiones para apoyar la realización de sus funciones y de las implicaciones en relación con los derechos fundamentales. Para tal efecto, primero explicaré

26 Agradezco a Sarah Muñoz-Cadena y Michelle Castellanos-Sánchez por su asistencia de investigación y por su apoyo en la elaboración de las gráficas incluidas.

cuál es el contexto en el que se ha ocurrido este proceso: las iniciativas de transformación digital del sector público.

Luego, presento los principales resultados de una nueva base de datos construida en el marco del proyecto Sistemas de Algoritmos Públicos de la Universidad de los Andes que identifica y documenta más de 300 sistemas de toma automatizada de decisiones del Estado Colombiano, lo cual incluye herramientas de inteligencia artificial (IA) (Gutiérrez et al., 2025).²⁷

Posteriormente, expondré los aportes desde las tres ramas del poder público respecto de la transparencia en el uso de sistemas computacionales por parte del Estado colombiano. Finalmente, cerraré con conclusiones y recomendaciones relacionadas con la adopción de los sistemas computacionales que apoyan la toma automatizada de decisiones en el sector público en Colombia.

2. Contexto regional de la transformación digital en el sector público

El proceso de adopción de herramientas computacionales que automatizan procesos en el sector público no es nuevo, hace parte de los planes de transformación digital que los Estados latinoamericanos, incluido el colombiano, emprendieron hace al menos dos décadas. Los Estados de la región han mejorado en ese lapso de tiempo en materia de los servicios ofrecidos en línea y en la construcción de infraestructura tecnológica para digitalización gubernamental (CEPAL, 2024).

Sin embargo, según el reporte de la encuesta de gobierno electrónico de 2024 de las Naciones Unidas, dicho progreso ha sido heterogéneo. El índice de desarrollo de gobierno electrónico de 2024 presenta países con alto nivel como Uruguay, Chile y Argentina (con resultados por encima de 0,85), mientras que en otro extremo se ubican países como Honduras y Haití (0,49 y 0,21, respectivamente). Colombia aparece en la novena posición del índice de gobierno electrónico (0,78) pero está por encima del promedio latinoamericano (0,66) (UN, 2024).

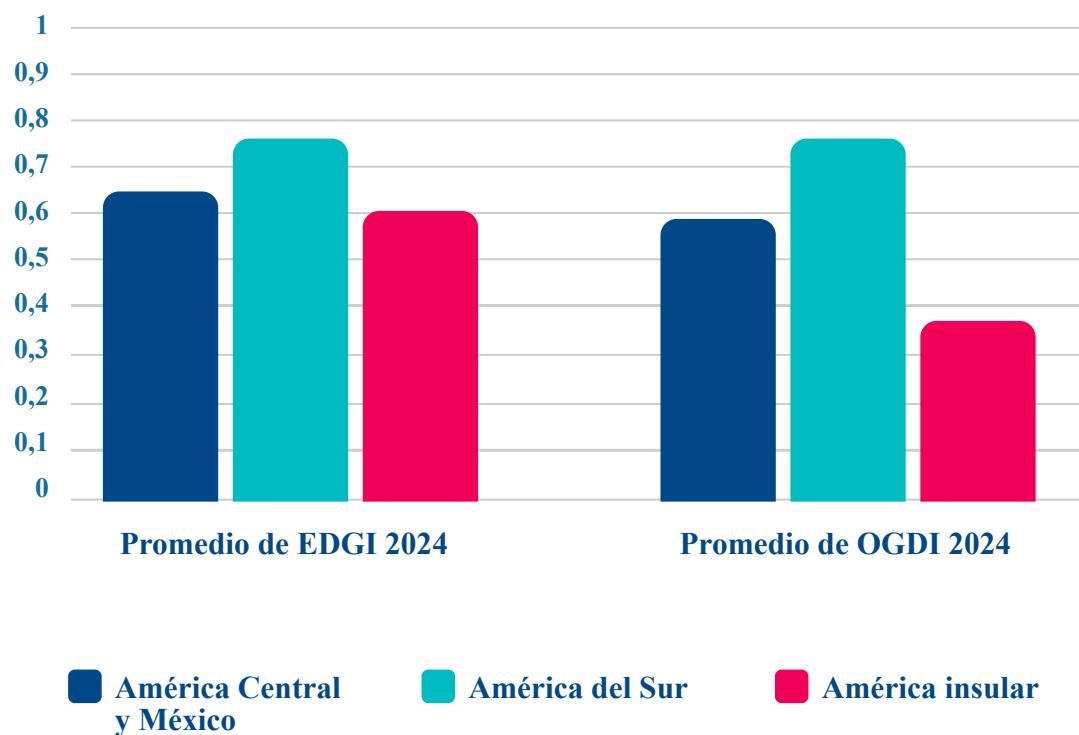
Por otra parte, la pandemia del COVID-19 incentivó la aceleración en la adopción de las Tecnologías de la Información y las Comunicaciones (TIC) por parte las entidades del sector público de Colombia y nuestra región. Las TIC fueron adoptadas para automatizar procesos, brindar canales de información a la ciudadanía a través de agentes virtuales, predecir posibles casos de corrupción, facilitar el análisis de grandes volúmenes de información, entre muchas otras funciones.

27 La base de datos caracteriza cada sistema a partir de 59 variables. Se construyó con información públicamente disponible en los sitios web de las entidades, repositorios de algoritmos gubernamentales, con las respuestas obtenidas del Departamento Nacional de Planeación a través de oficios que envió a entidades públicas, derechos de petición enviados a las entidades públicas, observación directa, en reportes de empresas de tecnología, reportes de prensa, artículos académicos, y reportes de organizaciones de la sociedad civil.

Adicionalmente, la apertura de datos del sector público es un elemento clave tanto en el desarrollo del gobierno electrónico como en el desarrollo de herramientas de IA. Al respecto, el índice de datos abiertos del reporte de gobierno electrónico de 2024 de las Naciones Unidas también muestra tendencias heterogéneas en los países latinoamericanos. Colombia se ubica en la quinta posición entre los países de la región del *índice de datos abiertos* de 2024, incluso por encima del promedio Europeo (UN, 2024).

La Figura 1 resume las disparidades regionales que aparecen tanto en el índice de desarrollo de gobierno electrónico como en la apertura de datos en el sector.

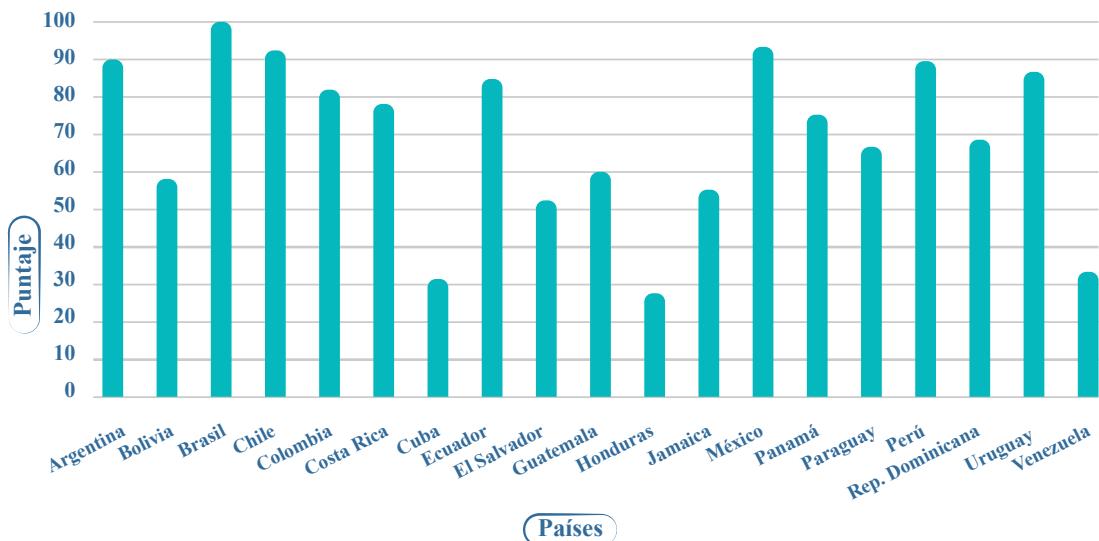
Figura 1 –Puntaje promedio en 2024 de los países latinoamericanos y del Caribe agrupados por subregiones en el índice de datos abierto y en el índice de desarrollo de gobierno electrónico



Fuente: EDGI (UN, 2024)

Finalmente, el Índice Latinoamericano de Inteligencia Artificial (ILIA) 2024 también da cuenta de la heterogeneidad en los 19 países la región medidos en términos del subíndicador “Gobierno Digital” (figura 2).

Figura 2 – Puntaje del subíndicador “Gobierno Digital” en el ILIA 2024



Fuente: CENIA (2024)

Como se puede observar en las estadísticas presentadas en esta subsección, en los países de la región persisten contrastes marcados en cuanto a su nivel de desarrollo en gobierno electrónico. Sin embargo, Colombia hace parte de los países que - como Uruguay, Chile, Brasil - Colombia se destacan porque suelen encabezar los puntajes más altos en la región en temas asociados al gobierno electrónico.

3. Sistemas de toma automatizada de decisiones en el sector público de Colombia

El Estado colombiano también utiliza tanto sistemas de IA como sistemas de algoritmos basados en reglas (sin la autonomía que tiene la IA) para automatizar procesos. En adelante nos referirnos colectivamente a estas herramientas como sistemas de toma automatizada de decisiones.²⁸

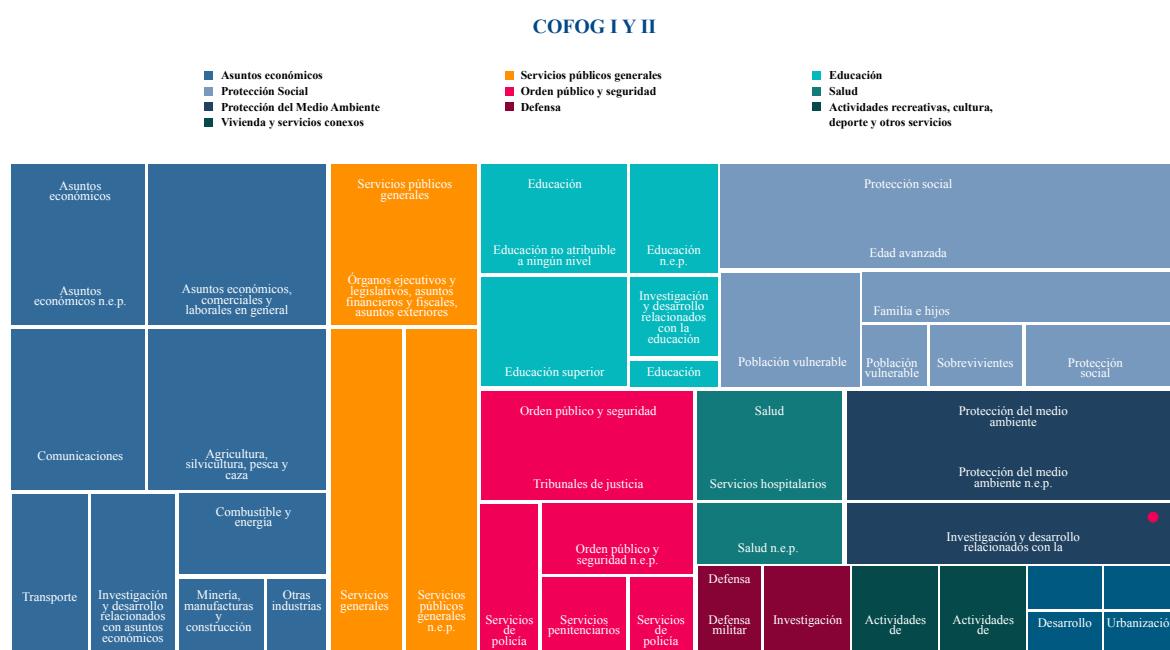
La nueva base de datos creada en el marco del proyecto Sistemas de Algoritmos Públicos identifica y documenta 383 sistemas de toma automatizada de decisiones en el sector público colombiano, de los cuales 289 en ejecución y 49 en fase de pilotaje. Del total de las herramientas, el 65% (220) son sistemas de IA²⁹ y 35% (118) son sistemas de automatización robótica de procesos (Gutiérrez et al., 2025).

28 Para una explicación más detallada sobre el concepto de “sistemas de toma automatizada de decisiones”, véase Gutiérrez & Castellanos-Sánchez (2023) y Gutiérrez & Muñoz-Cadena (2023).

29 Colombia está a la vanguardia regional en la adopción de herramientas de IA en el sector público, con 176 sistemas en operación, superando a Brasil (81) y México (72). Además, hay 44 herramientas de IA en etapa de pilotaje en el sector público del país, entonces en el futuro próximo su rol en los procesos públicos aumentará (Muñoz-Cadena et al., 2025).

Estas herramientas apoyan diversos sectores del Estado Colombiano, como lo ilustra la Figura 3. A partir de la Clasificación de las Funciones del Gobierno (COFOG), esta figura representa el tipo y proporción de las entidades públicas que han adoptado o piloteado estas herramientas. Los cinco tipos de entidades públicas que más aportan sistemas son las que realizan funciones de asuntos económicos, servicios públicos generales, educación, protección social, y orden público y seguridad.

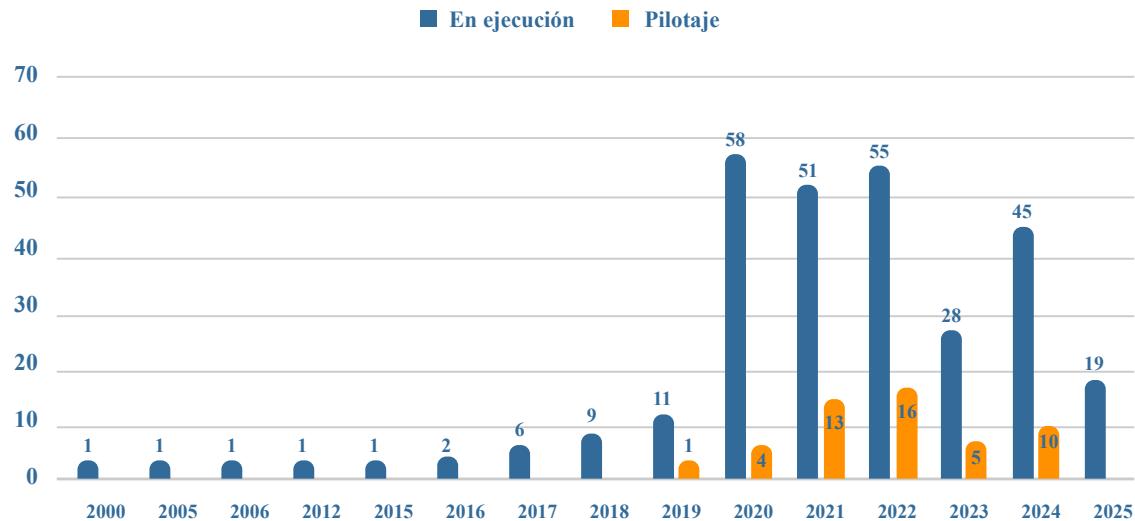
Figura 3 – Tipos de funciones de las entidades públicas que pilotean o ejecutan sistemas (COFOG I y II)



Fuente: Gutiérrez et al. (2025)

Por otra parte, como se observa en la Figura 4, el inicio de los pilotos y de la puesta en marcha de los sistemas presentó un crecimiento significativo a partir de 2020 y posteriormente tuvo picos en los dos años subsiguientes. Futuras investigaciones podrían explorar las razones por las cuales entre 2023 y 2024 el número de sistemas que entraron en ejecución o que iniciaron la fase de pilotaje se redujeron respecto de los tres años precedentes.

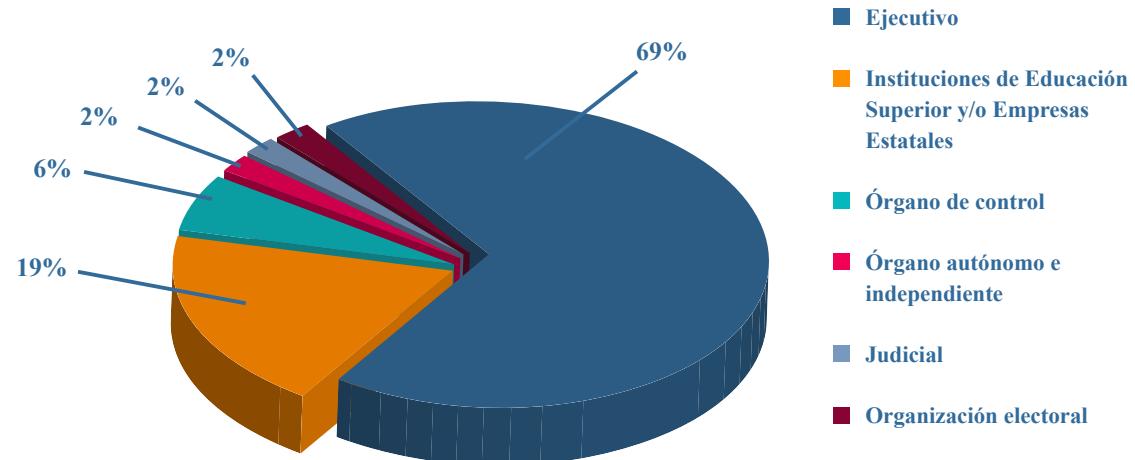
Figura 4 – Evolución temporal del inicio de los proyectos de sistemas de toma automatizada de decisiones en el sector público colombiano (pilotaje y ejecución)



Fuente: Gutiérrez et al. (2025)

En cuanto a la distribución de los sistemas de toma automatizada de decisiones en la estructura del Estado colombiano, las Figuras 4 y 5 muestran su ubicación según en las ramas del poder público y su nivel geográfico. La Figura 4 muestra que el 69% de las herramientas pertenecen al poder ejecutivo, seguidas por el 19% en instituciones de educación superior o empresas estatales, 6% en órganos de control, 2% en órganos autónomos, 2% en el poder judicial y 2% en organizaciones electorales.

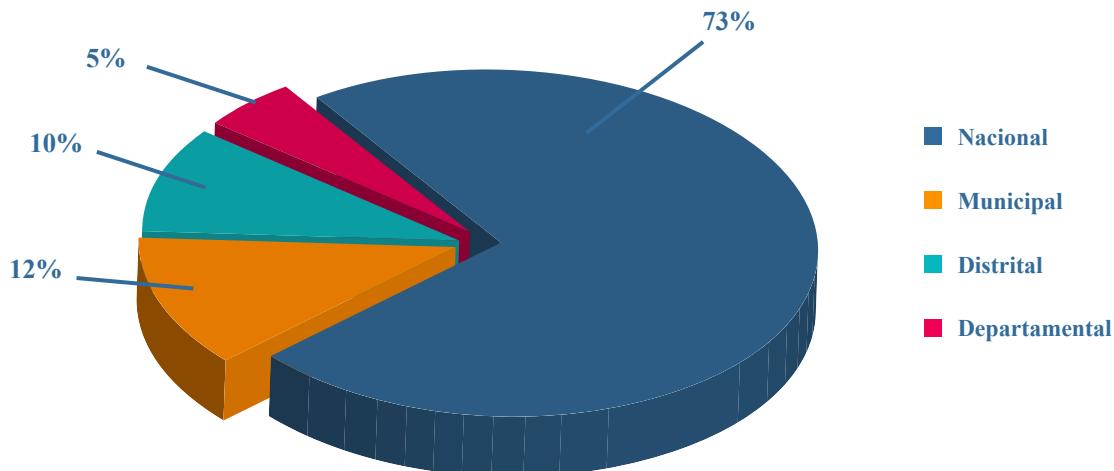
Figura 4 – Distribución de los sistemas en la estructura del Estado



Fuente: Gutiérrez et al. (2025)

Por su parte, la Figura 5 muestra que el 73% de las herramientas son de entidades públicas del nivel nacional, 12% del nivel municipal, 10% del nivel distrital y 5% a nivel departamental.

Figura 5 – Distribución de los sistemas en el nivel del Estado



Fuente: Gutiérrez et al. (2025)

Finalmente, en cuanto al tipo de tareas realizadas por los sistemas, a partir de la Clasificación OCDE, las cinco más comunes son: reconocimiento de objetos y personas (178), detección de eventos (117), soporte de interacción humano-máquina (114), predicción (77) y optimización orientada a objetivos (74 sistemas) (Gutiérrez et al., 2025).

4. Transparencia algorítmica en el sector público colombiano

Las tres ramas del poder público en Colombia no solo han aportado sistemas de toma automatizada de decisiones sino que también han desarrollado los marcos legales, de política pública y éticos en relación con desarrollo y uso de estas herramientas. En este caso nos centraremos en un punto fundamental que permitió la creación de las bases de datos del proyecto Sistemas de Algoritmos Públicos: la transparencia algorítmica por parte del Estado colombiano.

Si bien pueden haber muchas definiciones del término “transparencia algorítmica”, en este texto la entenderemos como la disponibilidad de información o la posibilidad de acceder a información sobre los sistemas algorítmicos en cualquiera de las fases del ciclo de vida de dichos sistemas (GPAI, 2024; Gutiérrez & Castellanos-Sánchez, 2023; Gutiérrez & Muñoz-Cadena, 2025).

a. La respuesta del Gobierno Nacional: principios, principios y principios

El Gobierno Nacional fue el primer poder público que levantó la mano en relación con la transparencia algorítmica. Lo hizo liderando el proceso de adopción formal de la “Recomendación sobre la IA” del Consejo de la OCDE en mayo de 2019 y mediante su adhesión a la “Recomendación sobre la ética de la inteligencia artificial” de la UNESCO en noviembre de 2021 (COMPES, 2025; MINTIC, 2019).

En los dos instrumentos internacionales acogidos por el Estado colombiano, la transparencia es un principio y deber del Estado en relación con los sistemas de IA. En los últimos seis años, estas recomendaciones influyeron en las políticas públicas e instrumentos de política pública adoptados el Gobierno Nacional: el documento COMPES 3975 de noviembre 2019; el Marco Ético para la IA que publicó el Departamento Administrativo de la Presidencia de marzo de 2021; la Hoja de Ruta para el desarrollo y aplicación de la IA en Colombia del Ministerio de Ciencia, Tecnología e Innovación; y el documento COMPES 4144 de febrero de 2025, por medio del cual se adoptó la nueva Política Nacional de Inteligencia Artificial. En todos estos documentos se resalta que la transparencia es un principio clave para la gobernanza responsable y ética de la IA.

Además, el Gobierno Nacional implementó al menos tres repositorios de algoritmos públicos, dos que están en funcionamiento (Datos Abiertos y el repositorio MINTIC) y uno que dejó de funcionar en el 2023. Los dos primeros, que siguen en operación, presentan un número muy limitado de sistemas identificados (menos de veinte). El tercero, se trataba de un tablero de control asociado al monitoreo del cumplimiento del Marco Ético para la IA que estaba alojado en la página web www.inteligenciaartificial.gov.co y que presentaba información detallada sobre seis sistemas de IA, pero dicha página web ya no está disponible (Gutiérrez & Castellanos-Sánchez, 2023; Gutiérrez & Muñoz-Cadena, 2025).

b. La Rama Judicial toma la palabra

La Rama Judicial también ha contribuido en este asunto. El año pasado, la Corte Constitucional abordó el tema a través de la sentencia T-323 de 2024, que decidió sobre el caso del juez de Cartagena que utilizó ChatGPT para redactar una tutela de segunda instancia (Gutiérrez, 2023, 2024a).

Se trata de una sentencia pionera, no solo en el contexto latinoamericano sino también a nivel global, al definir los principios y límites que deben observar los jueces cuando utilizan herramientas de IA, particularmente las generativas, en el ejercicio de sus funciones relacionadas con la administración de justicia (Gutiérrez, 2024c).

En la sentencia la Corte Constitucional establece que los jueces deben observar el criterio de transparencia cuando decidan utilizar herramientas de IA. Este criterio es definido por la Corte como “la obligación de evidenciar con claridad y precisión el uso, alcances y ubicación en las actuaciones o decisiones de los resultados obtenidos por la utilización de tales

herramientas, que permita a los usuarios e interesados su pleno conocimiento y la posibilidad efectiva de contradicción” (*Sentencia T- 323, 2024*).

Posteriormente, el Consejo Superior de la Judicatura, en cumplimiento de la sentencia, desarrolló este punto a finales del año pasado, a través de sus lineamientos de uso de IA en la Rama Judicial adoptados en el Acuerdo PCSJA24-12243 de 2024 que aplican a todos los magistrados, jueces y empleados judiciales (Consejo Superior de la Judicatura, 2024).

Más recientemente, Corte Constitucional le haló las orejas a varias entidades de la rama ejecutiva del nivel nacional por incumplir con el derecho fundamental al acceso a la información pública respecto de la aplicación *CoronApp*. Lo hizo a través de la sentencia T-067 de 2025, un precedente jurisprudencial sin paralelo en América Latina. El fallo aborda por primera vez el concepto de transparencia algorítmica, establece su relación con los derechos fundamentales, y define sus alcances cuando el Estado implementa sistemas computacionales en sus funciones (*Sentencia T-067, 2025*).

c. El turno del Congreso de la República

Vivimos un auge regulatorio en materia en relación con los sistemas de IA. En el proyecto Sistemas de Algoritmos Públicos de la Universidad de los Andes, mapeamos y documentamos más de 300 instrumentos regulatorios sobre IA en América Latina y el Caribe (Gutiérrez & Hurtado, 2025).

En Colombia, identificamos 37 instrumentos regulatorios, de los cuales 11 corresponden a proyectos de ley actualmente tramitados en el Congreso de la República. Cuatro de dichos proyectos de ley incluyen la transparencia algorítmica como uno de los principios orientadores y uno de ellos, el proyecto de ley 091/2023S, expresamente trata sobre “el deber de información para el uso responsable de la Inteligencia Artificial en Colombia y se dictan otras disposiciones” (Gutiérrez & Hurtado, 2025).



5. Conclusiones y recomendaciones

Este texto da cuenta de los avances del sector público colombiano en el desarrollo y adopción de sistemas de toma automatizada de decisiones como en la construcción de una gobernanza básica respecto de la transparencia algorítmica. Para que el Estado colombiano pueda continuar avanzando en la adopción de herramientas fiables y que cumplan con los mandatos constitucionales mencionados al comienzo del texto considero imprescindible avanzar en al menos tres frentes.

Primero, necesitamos más información sobre las barreras y dificultades del Estado colombiano para abordar los desafíos asociados con la “investigación, desarrollo, adopción, y apro-

vechamiento ético y sostenible de sistemas de IA”, como lo plantea el documento CONPES 4144. Un vacío evidente que dicho documento CONPES refleja es que el Gobierno Nacional carece de información básica para entender dichos retos. En múltiples ocasiones el propio documento CONPES reconoce que no cuentan con ciertos datos y evidencia para conocer la dimensión y estado de algunos de los desafíos de política pública. Ahí la tarea consiste en que el Estado se asegure de contar los insumos necesarios para tomar decisiones informadas.

Segundo, como lo solicita la sentencia T-067 de 2025, el Estado colombiano debe emitir lineamientos vinculantes sobre cómo cumplir con sus deberes de transparencia algorítmica, tanto en fase activa como en fase pasiva. En el plano regional, vale la pena destacar el caso del Consejo para la Transparencia de Chile (CPLT), una corporación autónoma de derecho público que vela por el cumplimiento de la Ley de Transparencia de la Función Pública y de Acceso a la Información de la Administración del Estado, que recientemente publicó sus recomendaciones para la transparencia algorítmica para todos los sujetos obligados a cumplir la ley de transparencia (CPLT, 2024).³⁰ El ejemplo de Chile podría ser considerado por la Defensoría del Pueblo, la Procuraduría General de la Nación y la Agencia Nacional Digital para construir sus propios lineamientos como lo dispuso la Corte Constitucional.

Finalmente, para que la adopción de sistemas de toma automatizada de decisiones en el sector público colombiano esté a la altura de lo que se espera de un Estado Social de Derecho, es necesario construir un marco regulatorio que garantice la promoción, respeto y protección de los derechos fundamentales y colectivos a lo largo de todo el ciclo de vida de estas herramientas. Son muchos los derechos fundamentales que pueden estar en juego cuando el Estado despliega estos sistemas: derecho a la no discriminación, debido proceso, presunción de inocencia, protección y de datos y privacidad, acceso a información pública, y libertad de expresión, entre otros.³¹

Actualmente hay miles de proyectos de ley estudiados por órganos legislativos alrededor de los continentes y hay diversos enfoques emergentes de regulación sobre IA a nivel global (Gutiérrez, 2024b). Es esencial que el debate regulatorio avance a una nueva fase de maduración, en la cual los reguladores colombianos: identifiquen qué problemas públicos deben y pueden ser abordados por instrumentos regulatorios; establezcan qué derechos fundamentales podrían estar en riesgo de vulneración, y qué futuros deseables podrían ser facilitados o catalizados; exploren diferentes enfoques regulatorios a partir de nuestro contexto y necesidades locales; e, identifiquen qué instrumentos regulatorios son eficientes, efectivos y equitativos para alcanzar los objetivos que debe procurar un Estado Social de Derecho.

30 El objeto de las recomendaciones del CPLT es “orientar y promover la adopción de buenas prácticas en torno a la transparencia y publicidad de los sistemas de decisiones automatizadas y semiautomatizadas que se utilizan en el sector público, tanto desde el respeto al ejercicio del derecho de acceso a la información pública, como de la publicación proactiva de información relevante sobre dichos sistemas.”

31 Por ejemplo, el 66.27% de los sistemas de toma de decisiones automatizadas del sector público colombiano requiere tratar datos personales para su funcionamiento, lo cual evidencia el deber acentuado de cumplir con los preceptos constitucionales y legales sobre protección de datos personales (Gutiérrez et al., 2025).



6. Referencias

Cenia. (2024). *Índice Latinoamericano de Inteligencia Artificial (ILIA) 2024*. 1 Centro Nacional de Inteligencia Artificial de Chile (Cenia). https://indicelatam.cl/wp-content/uploads/2024/09/ILIA_2024.pdf

CEPAL. (2024). *Superar las trampas del desarrollo de América Latina y el Caribe en la era digital: El potencial transformador de las tecnologías digitales y la inteligencia artificial* (LC/CMSI.9/3). Comisión Económica para América Latina y el Caribe (CEPAL).

CONPES. (2025). *Política Nacional de Inteligencia Artificial* (Documento CONPES No. 4144). <https://colaboracion.dnp.gov.co/CDT/Conpes/Económicos/4144.pdf>

Consejo Superior de la Judicatura. (2024). *Lineamientos para el uso y aprovechamiento respetuoso, responsable, seguro y ético de la inteligencia artificial en la Rama Judicial* (No. Acuerdo PCSJA24-12243). Consejo Superior de la Judicatura. https://actosadministrativos.ramajudicial.gov.co/GetFile.ashx?url=~%2fApp_Data%2fUpload%2fPCSJA24-12243.pdf

CPLT. (2024). *Resolución 372 de 2024 | Aprueba texto de las recomendaciones del Consejo para la Transparencia sobre Transparencia Algorítmica*. Consejo para la Transparencia de Chile (CPLT),. <https://www.consejotransparencia.cl/instrucion/aprueba-texto-de-las-recomendaciones-del-consejo-para-la-transparencia-sobre-transparencia-algoritmica/>

GPAI. (2024, mayo). *Algorithmic transparency in the public sector: A state-of-the-art report of algorithmic transparency instruments*. Global Partnership on Artificial Intelligence. <https://gpai.ai/projects/responsible-ai/algorithmic-transparency-in-the-public-sector/algorithmic-transparency-in-the-public-sector.pdf>

Gutiérrez, J. D. (2023, febrero 26). Hablemos sobre el uso de ChatGPT para redactar decisiones judiciales. *La Silla Vacía*. <https://www.lasillavacia.com/historias/historias-silla-llena/hablemos-sobre-el-uso-de-chatgpt-para-redactar-decisiones-judiciales/>

Gutiérrez, J. D. (2024a). Chapter 24: Critical appraisal of large language models in judicial decision-making. En *Handbook on Public Policy and Artificial Intelligence* (pp. 323–338). Edward Elgar Publishing. <https://doi.org/10.4337/9781803922171.00033>

Gutiérrez, J. D. (2024b). *Consultation paper on AI regulation: Emerging approaches across the world*. The United Nations Educational, Scientific and Cultural Organization (UNESCO). <https://unesdoc.unesco.org/ark:/48223/pf0000390979>

Gutiérrez, J. D. (2024c). *Document for consultation: Draft UNESCO Guidelines for the use*

of AI systems in Courts and Tribunals (No. CI/DIT/2024/GL/01). United Nations Educational, Scientific and Cultural Organization (UNESCO). <https://unesdoc.unesco.org/ark:/48223/pf0000390781>

Gutiérrez, J. D., & Castellanos-Sánchez, M. (2023). Transparencia algorítmica y Estado Abierto en Colombia. *Revista Reflexión Política*, 25(52). <https://revistas.unab.edu.co/index.php/reflexion/issue/archive>

Gutiérrez, J. D., Castellanos-Sánchez, M., & Muñoz-Cadena, S. (2025). *Sistemas automatizados de toma de decisiones en el sector público colombiano* (Versión 2) [Dataset]. <https://sistemaspublicos.tech/sistemas-automatizados-de-toma-de-decisiones-en-el-sector-publico-de-colombia/>

Gutiérrez, J. D., & Hurtado, S. (2025). *Regulación sobre IA en América Latina y el Caribe* (Versión 1) [Dataset]. Sistemas de Algoritmos Públicos, Universidad de los Andes. <https://sistemaspublicos.tech/regulacion-sobre-ia-en-america-latina/>

Gutiérrez, J. D., & Muñoz-Cadena, S. (2023). Adopción de sistemas de decisión automatizada en el sector público: Cartografía de 113 sistemas en Colombia. *GIGAPP Estudios Working Papers*, 10(270), 365–395.

Gutiérrez, J. D., & Muñoz-Cadena, S. (2025). Proactive algorithmic transparency in government: The case of the Colombian repositories of public algorithms. En *Handbook on Governance and Data Science* (1a ed., pp. 222–239). Edward Elgar Publishing. <https://www.elgaronline.com/edcollchap/book/9781035301348/chapter13.xml>

MINTIC. (2019, mayo 22). *Colombia se adhiere a acuerdo de Inteligencia Artificial ante la OCDE*. Ministerio de Tecnologías de la Información y las Comunicaciones (MINTIC). <https://www.mintic.gov.co/portal/715/w3-article-100826.html>

Muñoz-Cadena, S., Gutiérrez, J. D., Castellanos-Sánchez, M., & Peralta, D. S. (2025). *Sistemas de IA en el sector público de América Latina y el Caribe* (Versión 2) [Dataset]. <https://sistemaspublicos.tech/sistemas-de-ia-en-america-latina/>

Sentencia T- 323, No. Expediente T-9.301.656 (Corte Constitucional el 2 de agosto de 2024). <https://www.corteconstitucional.gov.co/relatoria/2024/T-323-24.htm>

Sentencia T-067, No. Expediente T-8.202.533 (Corte Constitucional el 26 de febrero de 2025). <https://www.corteconstitucional.gov.co/Relatoria/2025/T-067-25.htm>

UN. (2024). *UN E-Government Survey 2024*. <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2024>



KONRAD
ADENAUER
STIFTUNG



Programa
Estado de Derecho
Latinoamérica