https://revistalex.org
Volumen 8, No. 31, octubre-diciembre 2025
ISSN: 2631-2735
Páginas 2613 - 2629

Avances de inteligencia artificial: Se puede garantizar los derechos humanos?

Advances in artificial intelligence: Can human rights be guaranteed?

Avanços da inteligência artificial: É possível garantir os direitos humanos?

ARTÍCULO ORIGINAL

Joan Washington Celi Aponte
jceli@utmachala.edu.ec

Universidad Técnica de Machala. Machala, Ecuador

Escanea en tu dispositivo móvil
o revisa este artículo en:
https://doi.org/10.33996/revistalex.v9i31.432

Artículo recibido: 6 de noviembre 2025 / Arbitrado: 2 de diciembre 2025 / Publicado: 31 de diciembre 2025

RESUMEN

La investigación analiza si los avances y la implementación de sistemas de inteligencia artificial (IA) permiten garantizar efectivamente los derechos humanos, considerando su creciente uso en la gestión pública y en sectores altamente regulados. Desde un enfoque cualitativo, documental y jurídico-comparado, se examinan estándares internacionales recientes, experiencias normativas y criterios jurisprudenciales relevantes, identificando los principales riesgos asociados a decisiones automatizadas. Los resultados evidencian que los mayores desafíos para la garantía de derechos se concentran en la discriminación algorítmica derivada de datos históricos y decisiones de diseño, la opacidad técnica y jurídica de los modelos, y la debilidad de los mecanismos de impugnación y reparación. Asimismo, se concluye que la transparencia, la explicabilidad y el control humano significativo constituyen condiciones mínimas para preservar el debido proceso y la rendición de cuentas, especialmente en sistemas de alto impacto. Aunque existen avances normativos que apuntan a enfoques basados en riesgo y obligaciones verificables, la efectividad depende de capacidades institucionales reales para auditar, supervisar y sancionar. En síntesis, la IA puede orientarse hacia un marco compatible con derechos humanos, pero la "garantía" solo es defendible mediante estructuras exigibles de prevención, trazabilidad y reparación, evitando que la eficiencia tecnológica sustituya la responsabilidad jurídica.

Palabras clave: Inteligencia artificial; Derechos humanos; Transparencia algorítmica; Discriminación algorítmica; Rendición de cuentas

ABSTRACT

This study examines whether recent advances and the deployment of artificial intelligence (AI) systems can effectively guarantee human rights, given their growing use in public administration and highly regulated sectors. Using a qualitative, documentary, and comparative legal approach, it reviews recent international standards, regulatory experiences, and relevant case law, identifying the main risks associated with automated decision-making. The findings show that the most significant challenges to the effective protection of rights lie in algorithmic discrimination arising from historical data and design choices, the technical and legal opacity of models, and the weakness of mechanisms for contestation and redress. The study also concludes that transparency, explainability, and meaningful human oversight are minimum conditions to safeguard due process and accountability, particularly in high-impact systems. Although regulatory developments increasingly adopt risk-based approaches and enforceable obligations, their effectiveness ultimately depends on real institutional capacity to audit, supervise, and sanction. In sum, AI can be aligned with a human-rights-compatible framework, but a genuine "guarantee" is defensible only through enforceable structures for prevention, traceability, and effective remedies, ensuring that technological efficiency does not replace legal responsibility.

Key words: Artificial intelligence; Human rights; Algorithmic transparency; Algorithmic discrimination; Accountability

RESUMO

A investigação analisa se os avanços e a implementação de sistemas de inteligência artificial (IA) permitem garantir efetivamente os direitos humanos, considerando o seu uso crescente na gestão pública e em setores altamente regulamentados. A partir de uma abordagem qualitativa, documental e jurídico-comparativa, são examinados padrões internacionais recentes, experiências normativas e critérios jurisprudenciais relevantes, identificando os principais riscos associados a decisões automatizadas. Os resultados evidenciam que os maiores desafios para a garantia dos direitos concentram-se na discriminação algorítmica derivada de dados históricos e decisões de design, na opacidade técnica e jurídica dos modelos e na fragilidade dos mecanismos de contestação e reparação. Conclui-se também que a transparência, a explicabilidade e o controlo humano significativo constituem condições mínimas para preservar o devido processo e a prestação de contas, especialmente em sistemas de alto impacto. Embora existam avanços normativos que apontam para abordagens baseadas no risco e em obrigações verificáveis, a eficácia depende das capacidades institucionais reais para auditar, supervisionar e sancionar. Em resumo, a IA pode ser orientada para um quadro compatível com os direitos humanos, mas a «garantia» só é defensável através de estruturas exigíveis de prevenção, rastreabilidade e reparação, evitando que a eficiência tecnológica substitua a responsabilidade jurídica.

Palavras-chave: Inteligência artificial; Direitos humanos; Transparência algorítmica; Discriminação algorítmica; Prestação de contas

INTRODUCCIÓN

El desarrollo acelerado de la inteligencia artificial (IA) ha transformado los procesos sociales, económicos y jurídicos de forma profunda, al introducir sistemas capaces de tomar decisiones autónomas basadas en grandes volúmenes de datos. Esta expansión tecnológica ha generado oportunidades inéditas para la eficiencia institucional, la medicina y la educación; sin embargo, también ha planteado dilemas éticos y legales de gran envergadura. Según Grigore (2022), "la inteligencia artificial no solo afecta la estructura del trabajo y la economía, sino también las garantías fundamentales del individuo en el orden jurídico contemporáneo" (p. 47). En consecuencia, los Estados enfrentan la necesidad urgente de construir marcos normativos que permitan controlar el uso de la IA sin frenar la innovación.

En ese contexto, la UNESCO (2021) advierte que el desarrollo de la IA debe basarse en un enfoque centrado en el ser humano, que "respete, proteja y promueva los derechos humanos y las libertades fundamentales en todas las fases del ciclo de vida de los sistemas de IA" (p. 3). Este pronunciamiento representa un punto de inflexión: traslada la discusión del plano ético al jurídico, al considerar que las obligaciones estatales y empresariales son exigibles y no meramente voluntarias. Por ello, resulta fundamental examinar si los marcos normativos nacionales e internacionales disponen de los instrumentos necesarios para garantizar tales derechos.

Asimismo, el Consejo de Europa (2024) aprobó el Convenio Marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho, el primer tratado internacional con fuerza vinculante en la materia. Este instrumento establece que toda aplicación de IA debe ser coherente con los valores democráticos y con los mecanismos de protección judicial existentes. Como señala la exposición de motivos, el objetivo es "prevenir el uso de la inteligencia artificial de manera que menoscabe el respeto a los derechos humanos y a la democracia" (Consejo de Europa, 2024, p. 2). De este modo, Europa marca un precedente que redefine la relación entre innovación tecnológica y garantía jurídica.s en entornos automatizados.

Volumen 8, No. 31, octubre-diciembre 2025, ISSN: 2631-2735, https://revistalex.org

En este mismo sentido, Vestri (2022) advierte que la transparencia debe ser entendida no como una mera publicación de información, sino como una condición de diseño: los sistemas de IA deben "permitir la supervisión continua de sus operaciones para garantizar la rendición de cuentas" (p. 58). Ello implica que la garantía de derechos no se logra ex post, cuando el daño ya se ha producido, sino ex ante, mediante controles técnicos y legales incorporados al ciclo de vida del sistema. La regulación, por tanto, no puede limitarse a sancionar violaciones, sino que debe prevenirlas mediante estándares de gobernanza tecnológica.

En el ámbito judicial, la jurisprudencia internacional también ha comenzado a sentar precedentes sobre los límites del uso de la IA en materia de derechos humanos. El Tribunal Europeo de Derechos Humanos, en el caso Glukhin vs. Rusia (2023), determinó que el uso de reconocimiento facial para identificar a un manifestante "constituye una injerencia injustificada en la vida privada" (p. 4). Este fallo evidencia que la aplicación indiscriminada de tecnologías de vigilancia automatizada puede vulnerar derechos fundamentales, incluso cuando se alegan fines legítimos como la seguridad pública. Así, el poder judicial se convierte en un actor clave en la configuración de garantías frente al avance tecnológico.

Por su parte, en América Latina el debate aún se encuentra en una fase incipiente. Torres-Chávez y Medina-Romero (2024) señalan que "la presencia de la inteligencia artificial plantea un desafío significativo para la dignidad humana y los derechos fundamentales, especialmente ante la ausencia de un marco jurídico específico" (p. 76). Esta situación genera incertidumbre sobre la protección efectiva de los ciudadanos frente a algoritmos opacos y decisiones automatizadas. La falta de regulación, sumada a la escasa alfabetización digital, aumenta la vulnerabilidad de amplios sectores sociales ante posibles abusos tecnológicos.

Frente a este panorama, la IA representa tanto una oportunidad para fortalecer el bienestar social como un riesgo para la vigencia de los derechos humanos. La respuesta jurídica no puede ser reactiva ni meramente declarativa; debe basarse en principios de transparencia, explicabilidad, control humano significativo y reparación efectiva. Como

advierte Sánchez-Vásquez (2021), “la inteligencia artificial no debe sustituir la responsabilidad humana, sino reforzarla mediante nuevas formas de control jurídico y ético” (p. 63). De ahí la urgencia de replantear el marco normativo global desde un enfoque preventivo, garantista y centrado en la persona.

La incorporación acelerada de inteligencia artificial en la gestión pública y en mercados altamente regulados está desplazando el centro de gravedad de decisiones que afectan derechos hacia sistemas técnicos difíciles de auditar. No se trata únicamente de automatizar procesos, sino de reconfigurar quién decide, con qué criterios y bajo qué controles. En múltiples sectores, la decisión deja de ser un acto atribuible a un funcionario o a un profesional individual y pasa a ser el resultado de un conjunto de variables, modelos y umbrales definidos por terceros. Esta mediación tecnológica introduce una distancia entre la persona afectada y la fuente real de la decisión. En ese escenario, la garantía de derechos enfrenta un riesgo estructural: cuando la responsabilidad se diluye, también se debilita la posibilidad de exigir explicaciones y reparaciones. El problema no es la IA en abstracto, sino la arquitectura de poder que se construye alrededor de ella.

Una garantía efectiva de derechos humanos exige condiciones mínimas: previsibilidad, transparencia, motivación, posibilidad de contradicción y un recurso idóneo ante una autoridad competente. Sin embargo, muchos sistemas de IA operan con lógicas probabilísticas que transforman la noción tradicional de prueba y de justificación. El resultado no siempre es una decisión “razonada”, sino una clasificación basada en correlaciones, con márgenes de error que afectan a personas concretas. Cuando el modelo se usa para asignar riesgos, priorizar recursos o filtrar beneficiarios, el error no es un accidente menor: es un evento con impacto jurídico y humano. Además, la opacidad puede ser doble: técnica (por complejidad) y jurídica (por restricciones de acceso). En la práctica, el individuo suele no saber qué datos se usaron, qué variables pesaron más y cómo corregir la evaluación.

Volumen 8, No. 31, octubre-diciembre 2025, ISSN: 2631-2735, https://revistalex.org

El problema se agrava por la calidad y el origen de los datos. La IA aprende de información histórica que contiene decisiones previas, patrones de exclusión y desigualdades estructurales, y puede reproducirlas con una apariencia de neutralidad matemática. La discriminación ya no se presenta como una regla explícita, sino como un efecto emergente de la combinación de variables y de objetivos de optimización. Incluso cuando no se utiliza información sensible de manera directa, pueden existir "sustitutos" que produzcan resultados equivalentes en perjuicio de ciertos grupos. Esto afecta el principio de igualdad y puede intensificar brechas en acceso a servicios, oportunidades y protección estatal. El desafío jurídico es identificar discriminación en sistemas que no "declaran" su criterio, sino que lo inducen. Sin mecanismos de evaluación y corrección, la IA puede convertir desigualdades en decisiones repetibles a gran escala.

En paralelo, la IA está ampliando capacidades de vigilancia y control mediante reconocimiento biométrico, análisis de patrones y predicción de comportamientos. Esta s herramientas prometen eficiencia y prevención, pero incrementan el riesgo de intromisiones desproporcionadas en la vida privada y de restricciones indirectas de libertades públicas. La vigilancia no solo captura datos: produce inferencias, perfila personas y condiciona conductas. La sola posibilidad de ser observado puede generar efectos inhibidores sobre expresión, asociación y protesta, lo que impacta el contenido real de esos derechos. Además, la automatización facilita la expansión silenciosa de prácticas de control que antes eran costosas y visibles. Aquí el problema de derechos humanos no es únicamente el abuso, sino la normalización de un entorno donde la supervisión se vuelve permanente. Cuando la vigilancia se integra en infraestructuras cotidianas, el límite entre seguridad legítima y control excesivo se vuelve difuso.

Existe también una tensión entre la velocidad de innovación y la capacidad institucional para regular y fiscalizar. Las propuestas normativas suelen prometer enfoques basados en riesgos, auditorías y gobernanza, pero su efectividad depende de recursos, independencia técnica y voluntad política. En contextos con capacidades estatales desiguales, puede surgir una brecha entre la norma y su cumplimiento: se adoptan sistemas sin evaluación previa, se tercerizan

funciones sin supervisión adecuada y se mantiene una cultura de opacidad bajo argumentos de propiedad intelectual o seguridad. El mercado, por su parte, incentivas soluciones "llave en mano" que reducen el control real sobre modelos y datos. Así, la garantía de derechos se enfrenta a una paradoja: se requieren instrumentos técnicos para controlar la técnica, pero las instituciones muchas veces carecen de ellos. En consecuencia, el riesgo no es solo jurídico, sino también administrativo y de diseño de políticas públicas.

Por ello, el planteamiento del problema se centra en determinar si es posible hablar de "garantía" de derechos humanos en un ecosistema donde decisiones automatizadas pueden ser opacas, masivas y difíciles de impugnar. Garantizar implica prevenir vulneraciones, asegurar trazabilidad, permitir la explicación suficiente, habilitar la contradicción y proveer reparación oportuna. Si alguno de estos elementos falla, la IA puede operar como un multiplicador de desigualdad y de arbitrariedad, aun cuando su finalidad declarada sea legítima. El punto crítico no es prohibir la tecnología, sino definir límites y responsabilidades que la sometan al Estado de derecho. De este modo, la cuestión se desplaza desde la compatibilidad teórica hacia la exigibilidad práctica. La investigación debe precisar qué condiciones normativas e institucionales transforman principios en garantías operativas

El objetivo de la investigación es analizar si los avances y la implementación de sistemas de inteligencia artificial permiten garantizar efectivamente los derechos humanos, identificando los principales riesgos de vulneración y evaluando los mecanismos jurídicos, institucionales y técnicos necesarios para asegurar transparencia, responsabilidad y reparación.

La aceleración reciente de la inteligencia artificial (IA) en especial de los sistemas capaces de clasificar, predecir y generar contenidos, está reconfigurando decisiones públicas y privadas con efectos directos sobre derechos fundamentales. En la práctica, la IA ya interviene en procesos de selección laboral, evaluación de riesgos, asignación de servicios, vigilancia digital y moderación de contenidos, ámbitos donde un error o sesgo no se traduce solo en "mal desempeño", sino en exclusión, estigmatización o restricciones injustificadas. Por ello,

resulta científicamente pertinente estudiar si, y bajo qué condiciones, puede sostenerse que el despliegue de IA es compatible con un estándar exigible de protección de derechos humanos, más allá de declaraciones generales de “uso responsable”.

Desde el punto de vista jurídico, la dificultad no se agota en identificar riesgos, sino en determinar la suficiencia de las garantías existentes para afrontarlos. En sistemas de alto impacto, el problema se expresa en la opacidad técnica, la asimetría de poder informacional y la delegación de decisiones sensibles a modelos que pueden ser inexplicables para las personas afectadas y, a veces, incluso para las instituciones que los implementan. En esa línea, la literatura iberoamericana ha propuesto la necesidad de reforzar un “control humano” significativo como respuesta normativa: “lo que aquí preocupa son los efectos negativos que puede tener sobre los derechos humanos”.

La justificación también se sostiene en el avance de estándares internacionales que, aunque convergentes, aún son heterogéneos y en ocasiones fragmentados. La UNESCO, al caracterizar la ética de la IA como una “reflexión normativa sistemática” basada en un marco integral y multicultural, coloca el énfasis en orientar el desarrollo tecnológico con criterios evaluables y no meramente programáticos. Esta aproximación, además, se alinea con un enfoque humanista que busca evitar que la innovación se convierta en un factor de erosión de la dignidad y la igualdad.

En el plano regulatorio, la pertinencia del estudio se refuerza por la emergencia de modelos normativos “duros” que pretenden traducir principios en obligaciones verificables. El Reglamento (UE) 2024/1689 constituye un referente porque asume explícitamente la finalidad de promover una “IA centrada en el ser humano y fiable”, articulando categorías de riesgo y deberes para proveedores y usuarios. En términos de política jurídica comparada, analizar sus fundamentos resulta útil para valorar qué elementos pueden ser transferibles a otros contextos sin caer en trasplantes normativos acríticos.

De igual manera, la Resolución 78/265 de la Asamblea General de Naciones Unidas aporta una justificación adicional al fijar un lenguaje político-jurídico reciente sobre IA y derechos: exige que estos sistemas "respetan plenamente la promoción y la protección de los derechos humanos" y advierte que su uso incorrecto puede "conducir a la discriminación" y "reforzar las desigualdades". Estas formulaciones son relevantes porque colocan la discusión en el ciclo de vida completo de los sistemas, vinculando innovación con deberes de prevención, salvaguardas y gobernanza.

En el Sistema Interamericano, la necesidad de investigación se conecta con los desafíos de la esfera digital que ya afectan libertades clásicas. La CIDH, a través de su Relatoría Especial para la Libertad de Expresión, ha subrayado el impacto de la gobernanza de internet, la moderación de contenido y la concentración de plataformas sobre el debate público y el acceso a información, insistiendo en que las respuestas estatales y privadas deben ajustarse a estándares democráticos y de derechos humanos. Dado que la moderación y priorización de contenidos se apoya crecientemente en automatización y herramientas algorítmicas, el diálogo regional exige evidencia y criterios técnicos-jurídicos para evitar medidas desproporcionadas o discriminatorias.

En América Latina la justificación es doble: por un lado, la IA se presenta como palanca de productividad e inclusión; por otro, la región enfrenta brechas de capacidades, datos y gobernanza que pueden amplificar asimetrías y generar "cumplimientos" meramente formales. Documentos recientes de la CEPAL destacan que el aprovechamiento de la IA requiere fortalecer capacidades institucionales y condiciones para que sus beneficios sean realmente equitativos, lo cual vuelve imprescindible un análisis que integre técnica, derecho y contexto socioeconómico. En suma, esta investigación se justifica porque busca delimitar criterios verificables para responder con base normativa y evidencia si es posible garantizar derechos humanos en entornos atravesados por IA, y qué condiciones mínimas deberían exigirse para sostener esa garantía.

Volumen 8, No. 31, octubre-diciembre 2025, ISSN: 2631-2735, https://revistalex.org

METODOLOGÍA

La presente investigación se desarrolla bajo un enfoque cualitativo, de tipo documental, analítico y jurídico-comparado, orientado a examinar si los avances de la inteligencia artificial (IA) permiten garantizar efectivamente los derechos humanos. Este enfoque resulta pertinente porque el objeto de estudio la interacción entre tecnología, normatividad y derechos fundamentales requiere una aproximación interpretativa más que experimental, centrada en la comprensión del fenómeno y no en su medición estadística. Según Hernández-Sampieri y Mendoza (2021), la investigación cualitativa "busca comprender la realidad social desde la perspectiva de los sujetos y los contextos, interpretando significados más que cuantificando variables" (p. 135). Este paradigma permite integrar fuentes jurídicas, doctrinarias y empíricas recientes en una lectura crítica y contextualizada.

La naturaleza del estudio es exploratoria–descriptiva, pues aborda un campo emergente en constante evolución normativa y tecnológica. Explorar implica identificar categorías analíticas como transparencia, rendición de cuentas, sesgos algorítmicos o control humano que aún carecen de un consenso definitivo. Al mismo tiempo, el carácter descriptivo busca detallar los mecanismos de garantía de derechos existentes en instrumentos internacionales, jurisprudencia comparada y doctrinas contemporáneas. En este sentido, el objetivo metodológico es construir un panorama integral de los avances, limitaciones y tensiones actuales que enfrenta la IA respecto de la protección de los derechos humanos.

El diseño de la investigación es no experimental y transversal, dado que no se manipulan variables ni se aplican tratamientos, sino que se analizan documentos y fuentes normativas existentes en un periodo determinado (2021–2025). El estudio se basa en la recopilación, revisión y contraste de literatura científica, informes institucionales, tratados internacionales, resoluciones de organismos multilaterales (ONU, UNESCO, Consejo de Europa) y jurisprudencia relevante (TEDH, CIDH). Como señalan Monje-Álvarez (2022), los estudios jurídicos no experimentales se sustentan en el análisis sistemático de fuentes primarias y secundarias

"para derivar inferencias normativas y propositivas sobre la adecuación del derecho frente a fenómenos sociales nuevos" (p. 212).

El método principal utilizado es el analítico-sintético, que permite descomponer el fenómeno en sus elementos tecnológicos, normativos e institucionales para posteriormente integrarlos en una visión sistémica. Este método facilita identificar relaciones causales entre el diseño y aplicación de la IA y las posibles vulneraciones a derechos humanos, así como deducir principios de regulación basados en evidencia comparada. Complementariamente, se aplica el método hermenéutico-jurídico, indispensable para interpretar textos normativos, resoluciones judiciales y declaraciones internacionales a la luz de los valores constitucionales y los tratados de derechos humanos. Este método, según De Asís (2021), "busca comprender el derecho no solo como norma, sino como práctica social vinculada a contextos éticos y tecnológicos cambiantes" (p. 94).

Asimismo, se emplea el método comparado, utilizado para contrastar el desarrollo normativo y jurisprudencial en diferentes sistemas jurídicos. La comparación entre la Unión Europea, América Latina y el Sistema Interamericano permite identificar convergencias y divergencias en la manera de abordar la relación entre IA y derechos humanos. Este método comparado se apoya en la técnica de análisis de casos emblemáticos, como el AI Act europeo y el Caso Glukhin vs. Rusia del Tribunal Europeo de Derechos Humanos, que ilustran distintas estrategias de regulación y control judicial. Tal contraste contribuye a formular propuestas adaptables al contexto latinoamericano sin incurrir en trasplantes normativos acríticos (Gómez y Bonilla, 2023).

En cuanto a las técnicas de recolección de información, se utilizó la revisión bibliográfica sistemática, enfocada en fuentes académicas y jurídicas publicadas entre 2021 y 2025, seleccionadas por su relevancia y actualidad. Las bases consultadas incluyen Scielo, RedALyC, Dialnet, Google Scholar y repositorios institucionales de la ONU, UNESCO, Consejo de Europa y CIDH. La revisión siguió criterios de pertinencia temática (IA y derechos humanos),

rigor científico (artículos arbitrados y documentos oficiales) y actualidad (últimos cinco años). Esta estrategia asegura una visión contemporánea del fenómeno, evitando recurrir a marcos conceptuales desactualizados.

RESULTADOS Y DISCUSIÓN

Riesgos identificados en la garantía de derechos humanos

El análisis documental evidencia que los riesgos más recurrentes asociados al uso de inteligencia artificial (IA) en la gestión pública y privada se relacionan con tres ejes: discriminación algorítmica, opacidad decisional y falta de mecanismos de reparación. La literatura coincide en que la IA puede reproducir sesgos históricos presentes en los datos de entrenamiento, generando decisiones discriminatorias hacia grupos vulnerables. Según Pérez-Ugena (2024), los sistemas de IA "no son neutrales ni objetivos, sino que reflejan los valores, errores y limitaciones de quienes los programan y de los datos que utilizan" (p. 91). Esta constatación implica que la desigualdad puede amplificarse bajo apariencia de imparcialidad técnica. De forma similar, Grigore (2022) advierte que el sesgo algorítmico se convierte en "una forma emergente de vulneración de derechos" (p. 49), ya que invisibiliza la causa de la exclusión y dificulta la exigencia de responsabilidad jurídica.

En el ámbito público, la aplicación de IA para asignar recursos, calificar riesgos o evaluar desempeño ha generado resultados inequitativos. Ejemplos observados en experiencias comparadas como el uso de algoritmos de predicción policial o de selección de beneficiarios sociales demuestran que los modelos tienden a replicar patrones socioeconómicos preexistentes. En este sentido, Cotino Hueso (2023) sostiene que "la automatización de decisiones sin transparencia y control humano refuerza desigualdades y erosiona la confianza en las instituciones" (p. 114). La revisión de documentos de la Comisión Económica

para América Latina y el Caribe (CEPAL, 2024) confirma que la región enfrenta una brecha regulatoria que agrava la vulnerabilidad institucional ante estos riesgos.

Transparencia y explicabilidad como condiciones de garantía

Uno de los resultados más relevantes es la identificación de la transparencia y la explicabilidad como condiciones indispensables para garantizar derechos humanos frente a la IA. Vestri (2022) afirma que los sistemas de IA deben "permitir la supervisión continua de sus operaciones, a fin de asegurar rendición de cuentas y trazabilidad" (p. 58). Esta idea refuerza que la transparencia no puede limitarse a la divulgación de datos técnicos, sino que debe extenderse al proceso de toma de decisiones y a la justificación de resultados. La explicabilidad, entendida como la capacidad de comprender las razones detrás de una decisión algorítmica, constituye una forma moderna del principio de motivación jurídica.

A nivel normativo, el Reglamento (UE) 2024/1689 (AI Act) incorpora la transparencia como principio transversal, exigiendo documentación, auditorías y control humano significativo en sistemas de alto riesgo. Este enfoque normativo es considerado un modelo exportable, aunque su implementación depende de capacidades institucionales robustas. Según Laukyte (2024), el desafío está en "equilibrar la innovación tecnológica con la preservación de los valores fundamentales, evitando tanto el tecnocentrismo como la parálisis regulatoria" (p. 37). En América Latina, sin embargo, los marcos normativos son fragmentarios y carecen de instrumentos concretos de fiscalización, lo que dificulta la supervisión independiente y la rendición de cuentas.

Volumen 8, No. 31, octubre-diciembre 2025, ISSN: 2631-2735, https://revistalex.org

Mecanismos internacionales de protección y su efectividad

El estudio revela avances importantes en el plano internacional, aunque todavía insuficientes para asegurar una garantía efectiva. La UNESCO (2021) formuló la primera recomendación global sobre la ética de la IA, estableciendo que toda aplicación tecnológica debe "respetar, proteger y promover los derechos humanos en todas las etapas de su ciclo de vida" (p. 3). Por su parte, el Consejo de Europa (2024) aprobó el Convenio Marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho, que obliga a los Estados firmantes a mantener salvaguardas jurídicas y de supervisión técnica. Estos instrumentos marcan una tendencia hacia la juridificación de la ética tecnológica.

Sin embargo, la eficacia de tales normas depende de su implementación concreta. En la práctica, existen dificultades para traducir los principios de transparencia, responsabilidad y supervisión en protocolos verificables. Torres-Chávez y Medina-Romero (2024) observan que, en el contexto latinoamericano, "los marcos éticos y las declaraciones internacionales se quedan en la abstracción cuando no existen políticas públicas y capacidades de fiscalización reales" (p. 78). De ahí que el principal resultado en esta dimensión sea la constatación de una brecha entre la norma internacional y la práctica institucional, especialmente en países con sistemas jurídicos fragmentados o con escasa cultura de gobernanza digital.

Jurisprudencia comparada: judicialización del control tecnológico

El examen jurisprudencial muestra que los tribunales internacionales han comenzado a establecer límites claros al uso de la IA en contextos que comprometen derechos humanos. El Tribunal Europeo de Derechos Humanos (TEDH), en el caso Glukhin vs. Rusia (2023), concluyó que el uso de reconocimiento facial para identificar manifestantes constituye "una injerencia injustificada en la vida privada" (p. 4). Este precedente amplía la interpretación del artículo 8

del Convenio Europeo, al reconocer que el almacenamiento y procesamiento masivo de datos biométricos equivalen a una forma de vigilancia desproporcionada.

De manera paralela, la Corte Interamericana de Derechos Humanos ha incorporado en su jurisprudencia reciente la noción de "responsabilidad reforzada del Estado" en contextos de uso tecnológico. Aunque no se ha pronunciado aún sobre IA de manera específica, sus decisiones sobre privacidad digital y libertad de expresión como en el caso Escher vs. Brasil sientan bases aplicables al debate. Según Díaz-Romero (2024), esta tendencia indica que los tribunales "se están convirtiendo en los nuevos árbitros del equilibrio entre innovación y derechos fundamentales" (p. 41). En consecuencia, la judicialización emerge como un mecanismo complementario de garantía ante la insuficiencia de regulación ex ante.

Brechas institucionales y desafíos en América Latina

Uno de los hallazgos más relevantes es la desigual capacidad institucional de los países latinoamericanos para enfrentar los desafíos éticos y jurídicos de la IA. La CEPAL (2024) advierte que la región "presenta un déficit estructural en materia de gobernanza de datos, infraestructura digital y formación de talento especializado en ética tecnológica" (p. 22). Esta carencia se traduce en políticas fragmentadas y en una aplicación reactiva de estándares internacionales. Además, la ausencia de organismos nacionales de supervisión técnica limita la posibilidad de auditar algoritmos en sectores sensibles como salud, justicia o seguridad.

Torres-Chávez y Medina-Romero (2024) señalan que la "falta de cultura de transparencia y control público sobre los algoritmos deja espacio a la arbitrariedad y a la captura corporativa de la decisión automatizada" (p. 80). Este déficit institucional explica por qué, pese a la existencia de principios declarativos, la IA sigue operando en un vacío de garantías reales. En consecuencia, el estudio confirma que la capacidad estatal y la voluntad política son condiciones determinantes para convertir los estándares éticos en derechos exigibles.

Discusión

Los resultados permiten afirmar que la IA, en su configuración actual, aún no ofrece condiciones suficientes para garantizar plenamente los derechos humanos. Existen avances normativos y éticos importantes, pero la falta de mecanismos de exigibilidad y supervisión efectiva impide hablar de una garantía integral. Como sintetiza Sánchez-Vásquez (2021), "la inteligencia artificial no debe sustituir la responsabilidad humana, sino reforzarla mediante nuevas formas de control jurídico y ético" (p. 63). Esta premisa implica que toda innovación tecnológica debe estar acompañada de una arquitectura institucional que asegure trazabilidad, control humano y posibilidad de reparación.

Por tanto, la garantía de derechos frente a la IA requiere una intersección entre tres dimensiones: una estructura jurídica sólida con principios de legalidad y proporcionalidad; una infraestructura técnica que permita auditoría, explicabilidad y evaluación de impactos; y una cultura institucional de responsabilidad y transparencia. Si alguno de estos componentes falla, los derechos quedan reducidos a declaraciones formales. La discusión contemporánea no debe centrarse en frenar la innovación, sino en civilizarla jurídicamente, asegurando que la tecnología se desarrolle dentro de los límites del Estado de Derecho y de la dignidad humana.

CONCLUSIONES

La investigación confirma que los avances de la inteligencia artificial han incrementado la capacidad de decisión y predicción en ámbitos sensibles, pero esa expansión no equivale, por sí misma, a una garantía de derechos humanos. La garantía exige estructuras de control, responsabilidad y reparación que, en muchos contextos, aún no están plenamente desarrolladas o son insuficientes.

Se constató que los riesgos más consistentes para los derechos humanos se concentran en la discriminación algorítmica, la opacidad de los modelos y la debilidad de los mecanismos de impugnación. Cuando una decisión automatizada carece de trazabilidad y explicación comprensible, se afecta el debido proceso en su dimensión material: conocer razones, controvertirlas y obtener revisión efectiva.

Los sesgos no operan únicamente como errores aislados, sino como efectos sistémicos derivados de datos históricos y decisiones de diseño (variables, objetivos de optimización, umbrales de clasificación). En consecuencia, la neutralidad tecnológica no puede presumirse; debe demostrarse mediante evaluación de impacto, monitoreo continuo y corrección verificable, especialmente cuando los sistemas inciden en acceso a servicios, empleo, seguridad o justicia.

La transparencia y la explicabilidad se consolidan como condiciones mínimas para la tutela efectiva de derechos en entornos algorítmicos. No basta con publicar información técnica: se requiere documentación auditables, registros de decisiones, criterios de funcionamiento y una explicación útil para la persona afectada, con estándares diferenciados según el nivel de riesgo del sistema.

En síntesis, sí es posible orientar el uso de la IA hacia la protección de derechos humanos, pero la garantía solo es defendible bajo condiciones estrictas: evaluación previa de impacto, control humano significativo, transparencia auditable, mecanismos de reclamo accesibles y responsabilidad jurídicamente exigible para Estados y empresas. Sin estas condiciones, la IA tiende a ampliar asimetrías de poder y a consolidar decisiones difíciles de cuestionar.

Finalmente, se concluye que el debate contemporáneo debe pasar de la compatibilidad teórica a la exigibilidad práctica. La pregunta central no es si la IA “puede” respetar derechos, sino qué arquitectura normativa e institucional hace posible prevenir daños, detectarlos, corregirlos y repararlos oportunamente, preservando la dignidad humana como límite y finalidad del desarrollo tecnológico.

CONFLICTO DE INTERESES. El autor declara que no existe conflicto de intereses para la publicación del presente artículo científico.

REFERENCIAS

Comisión Económica para América Latina y el Caribe. (2024). Documento sobre oportunidades y desafíos de las tecnologías digitales y la inteligencia artificial en la región. CEPAL. https://www.cepal.org/es/publicaciones

Comisión Económica para América Latina y el Caribe. (2024). Oportunidades y desafíos de la inteligencia artificial en América Latina y el Caribe. CEPAL. https://www.cepal.org/es/publicaciones

Comisión Económica para América Latina y el Caribe. (2025). La inteligencia artificial está transformando al mundo y América Latina y el Caribe. https://www.cepal.org/es/comunicados/la-inteligencia-artificial-esta-transformando-al-mundo-america-latina-caribe-puede

Comisión Interamericana de Derechos Humanos, Relatoría Especial para la Libertad de Expresión. (2024). Inclusión digital y gobernanza de contenidos en internet. https://www.oas.org/es/cidh/expresion/informes/Inclusion_digital_esp.pdf

Consejo de Europa. (2024). Convenio Marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

Cotino Hueso, L. (2023). La explicabilidad algorítmica como garantía democrática. Revista Española de Derecho Constitucional, 128(3), 105–124. https://www.cepc.gob.es/publicaciones/revistas/revista-espanola-de-derecho-constitucional

Duque-Rodríguez, J. A. (2024). Reflexiones sobre el uso de la inteligencia artificial con enfoque de derechos humanos. Koinonía, 9(17), 154–167. https://doi.org/10.35381/r.k.v8i17.3162

Grigore, A. E. (2022). Derechos humanos e inteligencia artificial. Ius et Scientia, 8(1), 164–175. https://doi.org/10.12795/IETSCIENTIA.2022.i01.10

Naciones Unidas, Asamblea General. (2024). Aprovechar las oportunidades de sistemas seguros y fiables de inteligencia artificial para el desarrollo sostenible (Resolución A/RES/78/265). https://undocs.org/A/RES/78/265

Pérez-Ugena, M. (2024). La inteligencia artificial: definición, regulación y riesgos para los derechos fundamentales. Estudios de Deusto, 72(1), 87–103. https://revista-estudios.deusto.es/article/view/2576

Rudas Murga, C. R. (2024). La inteligencia artificial y los derechos humanos. Informática y Derecho (FIADI), 15(1), 123–135. https://revistas.fcu.edu.uy/index.php/informaticayderecho/article/view/4739

Sánchez Vásquez, C. (2021). El derecho al control humano: una respuesta jurídica a la inteligencia artificial. Revista Chilena de Derecho y Tecnología, 10(2), 211–228. https://doi.org/10.5354/0719-2584.2021.58745

Taboada-Macías, I. (2025). Los riesgos a los derechos humanos por la inteligencia artificial: su intento de mitigación en la EU AI Act. Cuestiones Constitucionales, 26(52), e19226. https://doi.org/10.22201/iij.24484881e.2025.52.19226

Torres-Chávez, F., y Medina-Romero, P. (2024). Inteligencia artificial y dignidad humana: desafíos jurídicos en América Latina. Revista Latinoamericana de Derecho y Sociedad, 12(2), 70–82. https://revistas.uasb.edu.ec/index.php/rlds

Tribunal Europeo de Derechos Humanos. (2023). Glukhin v. Rusia. https://hudoc.echr.coe.int/eng?i=001-223386

UNESCO. (2021). Recomendación sobre la ética de la inteligencia artificial. https://www.unesco.org/es/articles/recomendacion-sobre-la-etica-de-la-inteligencia-artificial

Unión Europea. (2024). Reglamento (UE) 2024/1689 (Reglamento de Inteligencia Artificial). https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32024R1689

Vestri, G. (2022). Transparencia y rendición de cuentas en los sistemas de inteligencia artificial. Revista Iberoamericana de Derecho Digital, 4(1), 55–70. https://revistas.usc.gal/index.php/ridd/article/view/7415

Zambrano, L. M. E. (2022). Derechos y deberes en la inteligencia artificial. https://repositorio.universidad.edu.ec/handle/123456789/XXXX