Inteligencia artificial al servicio de los Derechos humanos: ¿oportunidad o riesgo?[*]

Artificial intelligence in the service of human rights: opportunity or risk?

Filiberto Eduardo R. Manrique Molina[**]

Fecha de recepción: 21 de Enero de 2025
Fecha de aceptación: 3 de Febrero de 2025

El artículo "Inteligencia Artificial al Servicio de los Derechos Humanos: ¿Oportunidad o Riesgo?" Analiza el impacto dual de la inteligencia artificial (IA) en la defensa y protección de los derechos humanos en la era digital. Por un lado, destaca el potencial de la IA para fortalecer mecanismos de monitoreo, transparencia, acceso a la justicia y prevención de violaciones a los derechos fundamentales. Herramientas como el reconocimiento facial, la minería de datos y el análisis predictivo han permitido a gobiernos y organizaciones detectar abusos, agilizar procesos judiciales y optimizar la formulación de políticas públicas. Por otro lado, el artículo advierte sobre los riesgos inherentes a estas tecnologías, como la discriminación algorítmica,

[1] Producto resultado de investigación entre el grupo de investigación "Ciudadanía, Educación y Derecho UNIR" de la Fundación Universitaria Internacional de la Rioja, código COL0234101, Reconocido y clasificado en C MINCIENCIAS 2024,"Grupo de Investigación Red Internacional de Política Criminal Sistémica Extrema Ratio UNAL", Reconocido y Clasificado en A MINCIENCIAS 2024 y el Cuerpo Académico "Derechos Humanos, Política y Respectividad" de la Facultad de Ciencias de la Ingeniería, Administrativas y Sociales de la Universidad Autónoma de Baja California, en desarrollo del "Fordismo Academicus Cientificus" Año 3.

[2] Doctor en Derecho y Globalización; Maestro en Derecho e investigación, abogado, realizó estudios de posdoctorado en la Facultad de Derecho, Ciencias Políticas y Sociales de la Universidad Nacional de Colombia-sede Bogotá; Investigador Nacional SNI nivel I- SECIHTI, InvestigadorAsociado (I). MINCIENCIAS, es profesor investigador de tiempo completo en FCIAS de la Universidad Autónoma de Baja California. ID ORCID. https://orcid.org/0000-0001-9407-6548E-, ROR ID https://ror.org/05xwcq167, mail: filiberto.manrique@uabc.edu.mx

la vulneración de la privacidad, el uso indebido de sistemas de vigilancia masiva y la concentración de poder en grandes corporaciones tecnológicas. Estos peligros, si no se acompañan de marcos regulatorios sólidos y principios éticos claros, pueden traducirse en nuevas formas de desigualdad y violaciones a los derechos humanos. La investigación se basa en una revisión crítica de literatura y estudios de caso internacionales, resaltando que el equilibrio entre innovación tecnológica y respeto por la dignidad humana dependerá de la capacidad de las sociedades para desarrollar normativas que promuevan la transparencia, la rendición de cuentas y la equidad en el uso de la IA. Como conclusión, se enfatiza la necesidad de garantizar que el avance tecnológico esté al servicio de la justicia social y la protección de los derechos fundamentales, mediante la colaboración entre gobiernos, sociedad civil y desarrolladores tecnológicos.

Palabras clave: Inteligencia Artificial; Derechos Humanos; Sesgo Algorítmico; Privacidad Digital; Regulación Tecnológica.

"Artificial Intelligence in the Service of Human Rights: Opportunity or Risk?"

This article examines the dual impact of artificial intelligence (AI) on the protection and promotion of human rights in the digital age. On the one hand, AI technologies offer significant opportunities for enhancing monitoring systems, improving transparency, facilitating access to justice, and preventing human rights violations. Tools such as facial recognition, data mining, and predictive analytics have enabled governments and organizations to detect abuses, streamline judicial processes, and optimize public policy development. On the other hand, the article warns of the inherent risks associated with AI, including algorithmic bias, privacy breaches, mass surveillance misuse, and the concentration of power within large technology corporations. Without robust regulatory frameworks and clear ethical guidelines, these threats could exacerbate inequality and lead to new forms of human rights violations. The research is based on a critical literature review and international case studies, emphasizing that balancing technological innovation with respect for human dignity requires societies to establish regulations promoting transparency, accountability, and fairness in AI deployment. The article

concludes by highlighting the necessity of ensuring that technological progress serves social justice and human rights protection through coordinated efforts between governments, civil society, and technology developers.

Keywords: Artificial Intelligence; Human Rights; Algorithmic Bias; Digital Privacy; Technology Regulation.

Zusammenfassung

Der Artikel „Künstliche Intelligenz im Dienste der Menschenrechte: Chance oder Risiko?“ analysiert die doppelte Wirkung künstlicher Intelligenz (KI) auf die Verteidigung und den Schutz der Menschenrechte im digitalen Zeitalter. Einerseits wird das Potenzial von KI hervorgehoben, Überwachungsmechanismen, Transparenz, Zugang zur Justiz und die Prävention von Grundrechtsverletzungen zu stärken. Tools wie Gesichtserkennung, Data Mining und prädiktive Analysen haben es Regierungenund Organisationen ermöglicht, Missbräuche aufzudecken, Gerichtsverfahren zu rationalisieren und die Formulierung öffentlicher Richtlinien zu optimieren. Andererseits warnt der Artikel vor den mit diesen Technologien verbundenen Risiken wie algorithmischer Diskriminierung, Verletzung der Privatsphäre, Missbrauch von Massenüberwachungssystemen und Machtkonzentration in großen Technologiekonzernen. Wenn diese Gefahren nicht durch solide Regulierungsrahmen und klare ethische Grundsätze einhergehen, können sie zu neuen Formen der Ungleichheit und Menschenrechtsverletzungen führen. Die Forschung basiert auf einer kritischen Durchsicht von Literatur und internationalen Fallstudien und betont, dass das Gleichgewicht zwischen technologischer Innovation und Achtung der Menschenwürde von der Fähigkeit der Gesellschaften abhängt, Vorschriften zu entwickeln, die Transparenz, Rechenschaftspflicht und Gerechtigkeit beim Einsatz von KI fördern. Abschließend wird die Notwendigkeit betont, sicherzustellen, dass der technologische Fortschritt im Dienste der sozialen Gerechtigkeit und des Schutzes der Grundrechte steht, und zwardurchdie Zusammenarbeit zwischen Regierungen, der Zivilgesellschaft und Technologieentwicklern.

Schlüsselwörter: Künstliche Intelligenz; Menschenrechte; Algorithmischer Bias; Digitaler Datenschutz; Technologische Regulierung

Resumo

O artigo "Inteligência Artificial a Serviço dos Direitos Humanos: Oportunidade ou Risco?" analisa o duplo impacto da inteligência artificial (IA) na defesa e proteção dos direitos humanos na era digital. Por um lado, destaca o potencial da IA para reforçar os mecanismos de monitorização, a transparência, o acesso à justiça e a prevenção de violações dos direitos fundamentais. Ferramentas como reconhecimento facial, mineração de dados e análise preditiva têm permitido que governos e organizações detectem abusos, agilizem processos judiciais e otimizem a formulação de políticas públicas. Por outro lado, o artigo alerta sobre os riscos inerentes a estas tecnologias, como a discriminação algorítmica, a violação da privacidade, o uso indevido de sistemas de vigilância em massa e a concentração de poder em grandes corporações tecnológicas. Estes perigos, se não forem acompanhados por quadros regulamentares sólidos e princípios éticos claros, podem traduzir-se em novas formas de desigualdade e violações dos direitos humanos. A investigação baseia-se numa revisão crítica da literatura e de estudos de casos internacionais, destacando que o equilíbrio entre a inovação tecnológica e o respeito pela dignidade humana dependerá da capacidade das sociedades de desenvolverem regulamentações que promovam a transparência, a responsabilização e a equidade na utilização da IA. Em conclusão, sublinha-se a necessidade de garantir que o avanço tecnológico esteja ao serviço da justiça social e da protecção dos direitos fundamentais, através da colaboração entre governos, sociedade civil e promotores tecnológicos.

Palavras-chave: Inteligência Artificial; Direitos humanos; Viés Algorítmico; Privacidade Digital; Regulação Tecnológica

Introducción

En las últimas décadas la revolución digital ha desencadenado transformaciones profundas en todas las esferas de la sociedad, esta ha tenido un impacto importante en los procesos de socialización (Herrera-Ortiz, Peña-Avilés, Herrera-Valdivieso, & Moreno-Morán, 2024, p. 278). La inteligencia artificial (IA) se ha convertido en uno de los cambios tecnológicos más disruptivos y prometedores de la era contemporánea, especialmente la desarrollada en los últimos 24 meses, pues este tipo de tecnologías data de la época de los 50s del siglo pasado (Hardy, 2001, p. 3).

En estos tiempos, su avance tecnológico y su aplicación abarca desde la optimización de diversas áreas del conocimiento y profesión hasta la mejora de la actividad de gobierno en la prestación de servicios públicos, tareas administrativas que impactan en derechos civiles, políticos, sociales culturales y hasta ambientales, pasando por la transformación en la forma en que se gestionan y defienden los derechos humanos. Sin embargo, la dualidad inherente a la IA en este vertiginoso avance tecnológico plantea un debate crucial en el cual nos preguntamos si: ¿La IA se erige como una herramienta que fortalece la protección y garantía de los derechos humanos o, por el contrario, representa un riesgo potencial al expandir mecanismos de vigilancia, discriminación, control, la violación de derechos de autor, privacidad, expresión?

Este artículo explora ambas aristas del debate. Por un lado, se evidencian las oportunidades que la IA ofrece para el monitoreo, la denuncia y la promoción de los derechos humanos; por el otro, se analizan los desafíos éticos, legales y sociales que surgen de su implementación sin las debidas garantías. La discusión se fundamenta en estudios recientes y casos de aplicación en diferentes contextos, lo que permite identificar no solo los beneficios potenciales de la integración de la IA en el ámbito de los derechos humanos, sino también los riesgos inherentes a una utilización desmesurada o mal orientada de estas tecnologías con impacto negativo en la vida de las personas.

La metodología adoptada para este análisis se basa en una revisión crítica de la literatura científica y de informes de organismos inter nacionales, complementada con estudios de caso que evidencian tanto éxitos como fracasos en la implementación de sistemas basados en IA. A lo largo del desarrollo, se presentarán diversos estudios empíricos y teóricos, los cuales ofrecen perspectivas contrastadas sobre el impacto de la inteligencia artificial en la defensa de los derechos humanos.

El objetivo principal es proporcionar un marco de referencia que permita a los legisladores, académicos y sociedad en general reflexionar sobre las implicaciones de integrar la IA en políticas públicas y mecanismos de protección de derechos. Así, se busca identificar estrategias que maximicen las oportunidades y mitiguen los riesgos, promoviendo una convergencia entre innovación tecnológica y

respeto irrestricto de los derechos fundamentales de los seres humanos.

I. CONTEXTUALIZACIÓN DE LA INTELIGENCIA ARTIFICIAL Y LOS DERECHOS HUMANOS EN LA ERA DIGITAL

A. Definición de la IA

La Inteligencia Artificial (IA) se define como la capacidad de las máquinas para emular funciones cognitivas humanas, tales como el aprendizaje, la resolución de problemas, la percepción y el reconocimiento de patrones (Russell & Norvig, 2020). Desde sus inicios, impulsados por la propuesta de Alan Turing sobre máquinas inteligentes (Turing, 1950, 433), la IA ha recorrido un largo camino. En las décadas posteriores, los primeros sistemas emplearon algoritmos simbólicos y reglas lógicas, lo que permitió desarrollar soluciones puntuales para problemas específicos, aunque con limitadas capacidades de adaptación.

El advenimiento de la era del aprendizaje profundo (deep learning) y la disponibilidad masiva de datos e información por parte del internet y la inversión multimillonaria de grandes empresas tecnológicas han supuesto un cambio de paradigma en el campo por dominar la carrera de la IA en el mundo y dicho sea de paso de la computación cuántica, este último es un tema para otro estudio. El avance en el desarrollo de las redes neuronales artificiales, inspiradas en la estructura del cerebro humano, han demostrado un rendimiento excepcional en tareas complejas, desde el desarrollo de contenido digital, al reconocimiento de imágenes y voz hasta la traducción automática (LeCun, Bengio, & Hinton, 2015, p. 437).

Esta revolución tecnológica ha permitido que la IA se aplique en diversas áreas del conocimiento, desde la computación, ingeniería, ciencias sociales, medicina, arquitectura, economía, arte, y, por supuesto su uso en la educación, hasta la seguridad y la administración pública. En este sentido, la IA se posiciona como una herramienta de doble filo: por una parte, su potencial para transformar procesos y mejorar la eficiencia es innegable; por otra, la ausencia de una regulación jurídica robusta puede dar lugar a prácticas que vulneren derechos humanos fundamentales (UNESCO, 2020, p. 8), como lo son

algunos de ellos que derivan del intelecto y la razón humana.

No obstante, el rápido avance de estas tecnologías ha generado importantes debates éticos, jurídicos y sociales. La implementación de la IA sin marcos regulatorios sólidos o bajo normativas débiles y deficientes, puede derivar en prácticas ilegales, discriminatorias, que pueden llevar a la vulneración de derechos humanos y r iesgos relacionados con la privacidad y la identidad (UNESCO, 2020; Cath et al., 2018). Organismos internacionales, como la UNESCO y la Comisión Europea, han enfatizado la necesidad de establecer normativas que aseguren un desarrollo y uso responsable y ético de la IA (European Commission, 2019), promoviendo el respeto a la dignidad, los derechos y libertades de las personas en su implementación.

En síntesis, la evolución de la IA se caracteriza por un crecimiento exponencial en sus capacidades y aplicaciones. Las investigaciones actuales sugieren que esta tendencia continuará, impulsando innovaciones que podrían transformar radicalmente múltiples sectores de la sociedad. Sin embargo, para maximizar sus beneficios y minimizar riesgos, es imprescindible acompañar el avance tecnológico con un sólido debate ético y la elaboración de marcos regulatorios que garanticen un uso responsable y respetuoso de los derechos fundamentales (Floridi et al., 2018, p. 681).

II. DERECHOS HUMANOS EN LA ERA DIGITAL

Los derechos humanos, consagrados en diversos instrumentos internacionales, tales como la Declaración Universal de los Derechos Humanos (1948) y el Pacto Internacional de Derechos Civiles y Políticos (1966), se han visto ampliados y reinterpretados en el contexto de la era digital. La interconexión global de la comunicación, la recopilación masiva de datos y el desarrollo de algoritmos inteligentes han planteado nuevos desafíos en este siglo XXI, algunos relacionados con la falta de privacidad, la libertad de expresión, uso malicioso de datos e información, desinformación, robo de identidad de las personas, discriminación algorítmica, violación a los derechos de las obras intelectuales e industriales, y uno que es centro de los debates científicos en torno al uso de la IA, la seguridad para la especie humana.

Aunque también es una oportunidad para la garantía de los derechos, como lo es el acceso a la justicia, avance de la medicina, potenciar la educación a un nivel nunca antes visto, como algunos de los grandes cambios sociales en la era de la Internet (Castells, 2019, p. 10). En este escenario, la IA puede actuar como un amplificador tanto de derechos como de violaciones, dependiendo de su implementación y regulación.

Uno de los aspectos más prometedores del uso de la IA en el ámbito de los derechos humanos es su capacidad para monitorizar, medir, evaluar, dar seguimiento y detectar las violaciones. Por ejemplo, el uso de algoritmos de análisis han sido utilizados para la toma de decisiones judiciales, mediante el empleo de los sistemas expertos jurídicos SEJs (Martínez Bahena, 2013, p. 833). La automatización en el análisis de grandes volúmenes de datos permite que tribunales, ONG's y organismos internacionales detecten irregularidades de manera oportuna, facilitando intervenciones tempranas y la movilización de la instancias nacionales e internacionales para darle el tratamiento adecuado a esas violaciones.

Un caso ilustrativo también es la prevención de delitos, mediante el uso de sistemas de reconocimiento facial y análisis de vídeo en plataformas sociales, que, al ser configurados para detectar discursos de odio y mensajes incitadores a la violencia (Buolamwini & Gebru, 2018, p. 2), han permitido a organizaciones como Human Rights Watch recopilar evidencia de violaciones a los derechos humanos en regiones afectadas por conflictos (Smith, 2021. p. 2). Esta aplicación, sin embargo, debe manejarse con cautela, ya que la precisión y los sesgos inherentes a los algoritmos pueden afectar la integridad de la información recopilada o invadir la privacidad de las personas.

La IA también tiene el potencial de democratizar el acceso a la justicia, pues las herramientas basadas en algoritmos pueden ayudar a analizar grandes bases de datos legales y jurisprudenciales, facilitando la identificación de precedentes y apoyando a abogados y defensores en la elaboración de estrategias legales. En países donde el sistema judicial es lento o ineficiente, estas tecnologías pueden contribuir a reducir los tiempos de respuesta y ofrecer soluciones

prontas a las demandas de las personas (Pérez & López, 2023).

Además, plataformas y aplicaciones que integran la IA han sido desarrolladas para asesorar a los ciudadanos sobre sus derechos y las vías disponibles para denunciar abusos. Estas iniciativas, impulsadas tanto por el sector privado como por organizaciones sin fines de lucro, representan un paso importante hacia la creación de sociedades más informadas y empoderadas en materia de derechos humanos. Un ejemplo de ello es el trabajo de las universidades, que organizan concursos de hackathon para que los estudiantes de derecho propongan y exploren el uso de este tipo de tecnologías en la asesoría de casos o en la educación sobre los derechos de las personas.

Otra área beneficiada del uso de la IA es la que se refiere a la transparencia y la rendición de cuentas, la cuales se ven reforzadas cuando los procesos de toma de decisiones se apoyan en datos objetivos y verificables. Iniciativas como el " Big Data for Human Rights" han demostrado que el uso de información masiva y algoritmos de IA puede fomentar la colaboración entre distintos actores sociales, incluyendo instituciones públicas, organizaciones no gubernamentales y la sociedad civil, en la búsqueda de soluciones integrales a problemas complejos (International Crisis Group, 2022).

III. RIESGOS E IMPLICACIONES NEGATIVAS DE LA IA EN DERECHOS HUMANOS

A. Sesgos y discriminación algorítmica

Uno de los principales r iesgos identificados en la investigación de la información que generan las IA más populares o de mayor uso en la actualidad, son los que se encuentran asociados a la posibilidad de que los datos reproduzcan y amplifiquen sesgos existentes en la sociedad, que nos puede llevar a obtener información incorrecta. Esto se debe no a una programación, sino al entrenamiento del modelo con datos erróneos para que aprenda patrones y tome decisiones distorsionadas (Sánchez Cabrera, Fajardo Sandoval, & Trujillo González, 2024, p. 95). Estudios recientes han demostrado que los sistemas de reconocimiento facial y clasificación automatizada pueden presentar tasas

de error significativamente mayores al identificar a personas de minorías étnicas, color de piel, raza o géneros no hegemónicos (Buolamwini & Gebru, 2018, p. 3). Esta problemática no solo compromete la calidad de las decisiones automatizadas, sino que también puede derivar en prácticas discriminatorias que vulneran el derecho a la igualdad.

La falta de información objetiva, diversidad de datos en los equipos de desarrollo y la carencia de datos representativos en los conjuntos de entrenamiento son factores críticos que contribuyen a la perpetuación de estos sesgos. Por ello, es fundamental que tanto el sector público como el privado implementen mecanismos de seguimiento, auditoría, revisión y corrección de esos datos, que garanticen en el uso inadecuado de la información de la IA con sesgos (García et al., 2022).

B. Robo de identidad, violaciones a la privacidad y la libertad individual

Los nuevos modelos de IA pueden facilitar formas de suplantación de identidad debido a su capacidad para procesar, analizar y replicar datos personales, tienen la capacidad de generar videos, fotos, o audios falsos de una persona. Es posible el robo de identidad en el contexto de la inteligencia artificial (IA), lo cual es un problema creciente que implica el uso indebido de datos personales mediante tecnologías avanzadas. Esto puede derivar en fraudes, suplantación de identidad y violaciones a la privacidad para la obtención de un lucro ilícito.

Por otro lado, el uso extensivo de tecnologías de IA en la vigilancia masiva y la recopilación de datos personales plantea serias preocupaciones en torno a la privacidad y la protección de datos. En regímenes autoritarios, por ejemplo, el empleo de sistemas de reconocimiento facial y análisis de redes sociales se ha utilizado para reprimir la disidencia, protestas y controlar a la población (Zuboff, 2019, p. 19). La integración de estas tecnologías en sistemas de seguridad puede derivar en un estado de vigilancia permanente, afectando la libertad de expresión, de manifestación y el derecho a la intimidad.

Algunos informes de organismos internacionales han advertido sobre la “sociedad de control” en la que la

tecnología se utiliza para monitorear cada acción de los ciudadanos, eliminando cualquier espacio para la autonomía personal (UN Special Rapporteur on the Right to Privacy, 2020). La implementación de estas prácticas, por medio de sistemas avanzados de grabación con cámaras inteligentes de reconocimiento, muchas veces sin el consentimiento informado de la ciudadanía, representa un grave riesgo para la democracia y la protección de los derechos fundamentales.

C. Concentración de poder y desigualdad digital

Otro riesgo inherente al uso de la IA es la concentración de poder en manos de grandes corporaciones tecnológicas y gobiernos que controlan los datos y las infraestructuras digitales. Esta concentración puede llevar a una asimetría en el acceso a la información y a la toma de decisiones, generando desigualdades que se traducen en barreras para la protección de los derechos humanos en sectores vulnerables (Morozov, 2013, p. 28). La "brecha digital" se amplifica cuando el desarrollo y la implementación de tecnologías de IA se restringen a contextos de alto poder adquisitivo, dejando de lado a comunidades y países en vías de desarrollo. Por eso, hoy en día la carrera de la IA en el mundo se pelea en las grandes industrias tecnológicas con fuertes inversiones tanto en USA como en China, en donde se busca asegurar el liderazgo en esta materia, lo cual es considerado un tema de seguridad nacional. Esto no solo limita las oportunidades de estas regiones para aprovechar los beneficios de la innovación tecnológica (OECD, 2021), sino que también las expone a riesgos mayores en términos de dependencia tecnológica, seguridad, privacidad y derechos básicos.

IV. REGULACIÓN JURÍDICA Y MARCOS ÉTICOS EN EL USO DE LA IA

A. Normativas internacionales y propuestas legislativas

Ante la rápida expansión de la IA y sus implicaciones en derechos humanos, diversas organizaciones internacionales han comenzado a elaborar marcos regulatorios y éticos para orientar su

desarrollo y aplicación. La UNESCO, por ejemplo, ha propuesto una serie de recomendaciones que buscan armonizar la innovación tecnológica con el respeto a la dignidad humana y la justicia social (UNESCO, 2020). Asimismo, la Unión Europea ha impulsado iniciativas como la "Ley de Inteligencia Artificial" (European Commission, 2021), conocida coloquialmente como la EU AI Act, es una regulación orientada a establecer estándares que aseguren la transparencia, la seguridad y la no discriminación en el uso de estas tecnologías. Esta normativa busca equilibrar la innovación con la protección de los derechos y la seguridad de los ciudadanos, mediante un escalonamiento del riesgo que representan los distintos sistemas de IA para las personas.

Estas propuestas legislativas pretenden no solo regular el desarrollo de la IA, sino también fomentar la colaboración entre distintos actores sociales para crear un entorno en el que la innovación tecnológica se convierta en un motor de progreso y bienestar social. Sin embargo, la implementación efectiva de estas normativas enfrenta desafíos importantes, tales como la rápida evolución de las tecnologías y la dificultad para prever todas las posibles aplicaciones y consecuencias, dejando rápidamente enormes vacíos jurídicos que van a requerir de la reforma o de la interpretación judicial que permita resolver todas las controversias que se susciten de las nuevas tecnologías.

B. Principios éticos y responsabilidad social

La formulación de principios éticos en el desarrollo y uso de la IA es esencial para garantizar que estas tecnologías se orienten hacia la protección y promoción de los derechos humanos. Entre los principios fundamentales se encuentran la transparencia, la rendición de cuentas, la justicia, la privacidad y la no discriminación (Floridi et al., 2018, p. 701). Estos principios deben ser incorporados desde las etapas iniciales del diseño de sistemas de IA y mantenerse a lo largo de todo su ciclo de vida.

Además, la responsabilidad social y la participación ciudadana son elementos cruciales para asegurar que la IA beneficie a la mayoría y no se convierta en una herramienta

de control y exclusión. En el estudio de Floridi et al. (2018, p. 700), donde se resalta la necesidad de que la gobernanza de la inteligencia artificial se base en marcos éticos sólidos de los desarrolladores y en la colaboración activa entre gobiernos, empresas y la sociedad civil. Según estos autores, la creación de mecanismos de auditoría independientes y la participación ciudadana son fundamentales para supervisar en tiempo real el impacto de las tecnologías de IA y garantizar que se empleen de manera que beneficien a la mayoría de la población, evitando que se conviertan en instrumentos de control o exclusión. La colaboración entre gobiernos, empresas desarrolladoras y sociedad civil puede facilitar la creación de mecanismos que evalúen el impacto de las tecnologías y propongan ajustes necesarios para mitigar riesgos presentes y futuros.

C. Desafíos en la implementación y perspectivas futuras

A pesar de los esfuerzos normativos y éticos, la implementación de marcos regulatorios en el contexto global enfrenta obstáculos significativos. La disparidad en los niveles de desarrollo tecnológico y la diversidad de contextos culturales y políticos dificultan la adopción de políticas universales. Por ejemplo, mientras algunos países avanzados pueden permitirse inversiones en sistemas de seguimiento de IA, naciones en vías de desarrollo carecen de los recursos y la infraestructura necesaria para garantizar la aplicación de estos estándares (Castells, 2019, p. 65). Asimismo, la naturaleza dinámica de la tecnología implica que las regulaciones deben ser flexibles y adaptables, lo que exige una constante actualización y colaboración internacional entre países o estados, como lo es California, por ser una de las regiones del mundo epicentro de los avances tecnológicos y punta en la regulación de éstas. En este sentido, iniciativas de cooperación internacional y el intercambio de mejores prácticas regulatorias, se convierten en estrategias indispensables para enfrentar los desafíos futuros y asegurar que la IA actúe como un aliado en la defensa de los derechos humanos y no como una amenaza que pone en riesgo a la humanidad.

IV. ESTUDIOS DE CASO: APLICACIONES Y CONTROVERSIAS EN LA PRÁCTICA

A. Prevención de conflictos y protección de civiles

En diversas regiones en conflicto, organizaciones internacionales han implementado sistemas basados en IA para identificar patrones de violencia, monitorio de discursos de odio, amenazas, terrorismo, prever crisis humanitarias, o daños provocados por la naturaleza o el ser humano (Huertas Díaz, 2023, p. 28). Un ejemplo destacado es el empleo análisis de datos en la región del Sahel, donde la recopilación de información a través de satélites y redes sociales ha permitido anticipar movimientos de grupos armados y planificar intervenciones humanitarias (International Crisis Group, 2022). Esta experiencia demuestra que, con un diseño adecuado y una implementación responsable, la IA puede contribuir significativamente a la prevención de violaciones masivas de derechos humanos en situaciones de conflicto, así como a atender a la población civil, tanto a quienes se encuentran en condición de desplazamiento como a quienes están en el centro del conflicto.

No obstante, la aplicación de estas tecnologías también ha generado controversia, especialmente en relación con su uso en la guerra y en la defensa militar, lo cual representa un riesgo para la humanidad, pues en esos conflictos bélicos no sólo se afecta a los ejércitos, sino que también pueden verse afectadas comunidades inocentes. Los desafíos técnicos y éticos que emergen de estos casos subrayan la necesidad de establecer protocolos rigurosos y transparentes que aseguren que su empleo se dirija únicamente a fines humanitarios y a la protección de la vida y la dignidad humana, y no al desarrollo de armamento de destrucción con autonomía completa.

B. Experiencias en la transparencia gubernamental y la lucha contra la corrupción

Otro campo en el que la IA ha mostrado un gran potencial es en la promoción de la transparencia y la lucha contra la corrupción. Países como Estonia han implementado sistemas digitales

que integran algoritmos de IA para supervisar la gestión pública y detectar irregularidades en tiempo real (Estonian e- Governance Academy, 2020). Estas herramientas han permitido reducir significativamente los casos de malversación de fondos y han fortalecido la confianza ciudadana en las instituciones. Además, la IA puede ser de gran utilidad en la implementación de gobiernos 100% digitales para algunos de los servicios públicos que se ofrecen, lo que contribuiría a reducir el presupuesto destinado al aparato gubernamental físico.

Sin embargo, la centralización de datos y el uso de tecnologías avanzadas también plantean riesgos en cuanto a la privacidad y la seguridad de la información. La experiencia de algunos países muestra que, sin mecanismos adecuados de protección y supervisión, la concentración de información puede derivar en abusos y en la creación de espacios de control que, paradójicamente, socavar los mismos principios de transparencia que se pretenden fortalecer.

C. Impacto en la participación ciudadana y en la rendición de cuentas

Con relación a este tema, diversos estudios han analizado cómo la inteligencia artificial puede facilitar la participación ciudadana al ofrecer plataformas digitales que permiten la denuncia anónima de corrupción, abusos de servidores y la colaboración en la vigilancia de instituciones públicas. Por ejemplo, Wirtz, Weyerer y Geyer (2019, p. 596) destacan que la integración de herramientas digitales basadas en IA en el sector público no solo potencia la eficiencia en la gestión administrativa, sino que también puede fortalecer la rendición de cuentas al facilitar el acceso a canales de denuncia y la difusión de información legal entre los ciudadanos.

También se han desarrollado proyectos con apoyo de la IA, es el ejemplo de "AI for Justice"[2], que han implementado esta tecnología para orientar a organizaciones de asistencia legal y la población sobre sus derechos y mecanismos de denuncia, ilustran

[2] Véase: https://lawschool.thomsonreuters.com/aiforjustice/

cómo estas tecnologías democratizan el acceso a la información y promueven una cultura de transparencia. Sin embargo, se ha señalado que la eficacia de estas plataformas depende en gran medida de la infraestructura digital y de la alfabetización tecnológica de la población. En regiones con bajos índices de conectividad o en contextos de alta vulnerabilidad social, el riesgo de exclusión digital podría contrarrestar los beneficios de estas iniciativas, subrayando la necesidad de políticas integrales que incluyan la capacitación y el acceso equitativo a las tecnologías de la información, internet y computadoras.

VI. RETOS Y RECOMENDACIONES PARA UN USO RESPONSABLE DE LA IA EN DERECHOS HUMANOS

Para mitigar los riesgos inherentes a la IA, es crucial fomentar una mayor diversidad en los equipos de desarrollo y en los conjuntos de datos utilizados para entrenar los algoritmos. La inclusión de perspectivas diversas puede contribuir a identificar y corregir sesgos antes de que se traduzcan en prácticas discriminatorias (Buolamwini & Gebru, 2018, p. 4). Las políticas corporativas y gubernamentales deben incentivar la contratación de expertos de distintas disciplinas y orígenes, garantizando que el diseño de las tecnologías tenga en cuenta la pluralidad cultural y social.

También se requiere de la creación de marcos legales robustos y flexibles capaces de adaptarse a los avances de tecnología, lo cual es fundamental para asegurar que la IA se desarrolle en un entorno que respete los derechos humanos, no se emplee este tipo de tecnologías en la destrucción y en la guerra. Se recomienda la instauración de organismos de auditoría independientes que supervisen la aplicación de la IA en diversos sectores, asegurando la transparencia, la rendición de cuentas y su uso responsable. Además, la cooperación internacional resulta esencial para establecer estándares jurídicos globales y mecanismos de intercambio de buenas prácticas (European Commission, 2021).

No podemos dejar pasar la importancia del fortalecimiento del capital humano en competencias digitales y en derechos humanos es otro pilar clave para aprovechar las oportunidades de la IA y mitigar sus riesgos. La

educación y la capacitación con responsabilidad del buen uso de la IA deben orientarse tanto a los desarrolladores como a la ciudadanía en general, promoviendo una cultura de uso ético, respetuoso y responsable de las tecnologías de inteligencia avanzada. Las iniciativas de educación digital pueden contribuir a reducir la brecha digital y a empoderar a los ciudadanos en el ejercicio de sus derechos (OECD, 2021).

Por último, la convergencia entre tecnología y derechos humanos requiere un enfoque interdisciplinario que combine conocimientos de diversas áreas para asegurar su protección, tales como la informática, la ética, el derecho, la sociología y las ciencias políticas, entre otras. Se recomienda la creación de centros de pensamiento e investigación, además de la generación de redes colaborativas que permitan abordar de manera integral los desafíos y oportunidades que presenta la IA. La inversión en investigación y desarrollo no solo fomentará la innovación, sino que también permitirá anticipar problemas y diseñar soluciones que protejan los derechos fundamentales en un mundo cada vez más digitalizado (Floridi et al., 2018, p. 702), el cual se ve en riesgo por el avance de esta tecnología y su combinación con otras, como la computación cuántica.

Conclusión

La inteligencia artificial representa, sin duda, una de las herramientas más poderosas y prometedoras de la era contemporánea, con el potencial de transformar significativamente la forma en que se protegen y promueven los derechos humanos. Como se ha evidenciado a lo largo de este análisis, la IA puede ser una aliada en la detección temprana de violaciones, en la mejora del acceso a la justicia, el fortalecimiento de la transparencia gubernamental y la participación ciudadana. Sin embargo, este potencial viene acompañado de riesgos considerables: desinformación, la posibilidad de reproducir y amplificar sesgos, la vulneración de la privacidad, la censura en la libertad de expresión, uso malicioso de datos e información, robo de identidad de las personas, violación a los derechos de las obras intelectuales e industriales la concentración de poder, las guerras, y la exclusión digital y uno que es centro de los debates científicos

en torno al uso de la IA, la seguridad para la especie humana.

El equilibrio entre estas dos caras de la moneda dependerá en gran medida de las decisiones que tomen los actores políticos, tecnológicos y sociales en los próximos años. La implementación de marcos éticos y regulatorios robustos, la promoción de la diversidad y la inclusión en el desarrollo de tecnologías, así como el fortalecimiento de la educación y la participación ciudadana, son medidas indispensables para garantizar que la IA se convierta en un instrumento de progreso y no en una amenaza para los derechos humanos. El camino hacia un uso responsable y ético de la inteligencia artificial pasa por la colaboración intersectorial y el compromiso con valores fundamentales que trascienden fronteras y disciplinas. La innovación tecnológica debe ser siempre evaluada desde la perspectiva del respeto a la dignidad humana y los derechos de las personas y otras especies. Solo a través de un enfoque integrado y multidimensional será posible aprovechar al máximo las oportunidades que ofrece la IA, minimizando sus riesgos y garantizando que su desarrollo beneficie a toda la humanidad.

En definitiva, la inteligencia artificial se presenta tanto como una oportunidad transformadora como un riesgo que no puede ser subestimado, incluso pensado como uno de los riesgos más importantes para la especie humana. La clave reside en establecer mecanismos de control que permitan detectar y corregir desviaciones, y en fomentar una cultura ética de innovación responsable en la que el respeto por los derechos humanos sea el pilar fundamental. Con una visión comprometida y colaborativa, es posible que la IA se erija en uno de los mayores avances para la consolidación de sociedades respetuosas de la dignidad humana, de los derechos y libertades de la población mundial.

Referencias

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research, 81:1–15.

Castells, M. (2019). Redes de indignación y esperanza: Los movimientos sociales en la era de internet. Alianza Editorial.

European Commission. (2021). Proposal for a Regulation on Artificial Intelligence. Disponible en: https://ec.europa.eu/info/index_en.htm

European Commission. (2019). Ethics Guidelines for Trustworthy AI. Recuperado de https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Estonian e-Governance Academy. (2020). Digital Governance in Estonia: A Case Study. Disponible en: https://e-estonia.com/

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689–707.

García, M., Rodríguez, L., & Sánchez, P. (2022). Inteligencia Artificial y Derechos Humanos: Un Análisis Crítico. Revista Iberoamericana de Tecnología y Sociedad, 14(2), 45–68.

Huertas Díaz, O. (2023). El principio de legalidad en crímenes de lesa humanidad en el tribunal europeo de derechos humanos. En O. Huertas Díaz, & Á. C. Sánchez Cabrera, Derecho y política criminal sistémica. Transdisciplinariedad y sistema social (pp. 27-43). Bogotá: Ibáñez.

International Crisis Group. (2022). Leveraging Big Data for Human Rights Monitoring. Disponible en: https://www.crisisgroup.org/

Morozov, E. (2013). To Save Everything, Click Here: The Folly of Technological Solutionism. PublicAffairs.

OECD. (2021). AI in Societ y: Opportunities and Risks. Disponible en: https://www.oecd.org/

Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.

Sánchez Cabrera, Á. C., Fajardo Sandoval, F., & Trujillo González, J. S. (2024). Neurociencia y derecho penal, aportes para la discusión. En O. Huertas Díaz, A. L. Ruíz Herrera, & N. P. Jiménez Rodríguez, Voces polifónicas del derecho en la construcción de una sociedad en paz a través de la educación (págs. 86-131). Bogotá: Ibáñez.

Smith, A. (2021). Artificial Intelligence in Human Rights Advocacy. Journal of Human Rights and Technology, 7(1), 34–56.

Turing, A. M. (1950). Computing Machinery and Intelligence. Mind, 59(236), 433–460.

UNESCO. (2020). Recommendation on the Ethics of Artificial Intelligence. Disponible en: https://unesdoc.unesco.org/

UN Special Rapporteur on the Right to Privacy. (2020). Report on Privacy in the Digital Age. Disponible en: https://www.ohchr.org/

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Strategies. International Journal of Public Administration, 42(7), 596–615.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.

Nota: forma de citación del artículo.

APA (Ciencias Sociales, 7ª edición)

R. Manrique Molina, F. E. (2025). Inteligencia artificial al servicio de los derechos humanos: ¿Oportunidad o riesgo? *Revisitus Academicus Cientificus, 1*(1), 47-66.

MLA (Humanidades, 9ª edición)

Manrique Molina, Filiberto Eduardo R. "Inteligencia artificial al servicio de los derechos humanos: ¿Oportunidad o riesgo?" *Revisitus Academicus Cientificus*, vol. 1, no. 1, 2025, pp. 47-66.

Chicago (Notas y Bibliografía o Author-Date, según la disciplina) *Opción Author-Date:*

R. Manrique Molina, Filiberto Eduardo, 2025. "Inteligencia artificial al servicio de los derechos humanos: ¿Oportunidad o riesgo?" *Revisitus Academicus Cientificus* 1, no. 1 (2025): 47-66.

Opción de Notas y Bibliografía:

R. Manrique Molina, Filiberto Eduardo "Inteligencia artificial al servicio de los derechos humanos: ¿Oportunidad o riesgo?" *Revisitus Academicus Cientificus*, vol. 1, no. 1 (2025): 47-66.