

La equidad en la inteligencia artificial: un desafío ineludible

Matias Benavides Sandoval

Instituto tecnológico de Costa Rica

La inteligencia artificial (IA) se ha convertido en una transformación que ha impactado desde el ámbito médico y la aplicación de justicia hasta la concesión de créditos y la gestión financiera de una manera muy rápida “Modern AI is transforming multiple industries, redefining processes, and creating new possibilities” (L. Inglada Galiana, 2024, 179) Su capacidad para procesar cantidades de datos gigantescos y aprender de patrones complejos promete una mayor eficiencia y personalización de muchos servicios. Sin embargo, esto también puede conllevar riesgos cuando esos datos contienen prejuicios históricos o carecen de representatividad. “...apenas se discute su utilización, el enfoque ético que debe tener y sus posibles repercusiones económicas, políticas y culturales. La IA está en desarrollo. Apenas florecen sus amplias y diversas aplicaciones, su utilidad y sus riesgos.” (Guzmán Gómez, 2024, p.17). Por ello, resulta esencial analizar cómo la IA afecta a la sociedad y por qué debe abordarse su equidad desde múltiples perspectivas.

Garantizar la equidad en los sistemas de IA significa crear programas informáticos que impidan decisiones discriminatorias y promuevan la transparencia hacia el público. Como cuando un algoritmo evalúa solicitudes de crédito o sugiere diagnósticos médicos “está transformando el ámbito de la atención sanitaria al incrementar la precisión diagnóstica, personalizar los tratamientos y optimizar los resultados clínicos” (Zesergio MELO, 2024, 49), no basta con únicamente garantizar precisión: debe ser justo en su trato a todas las poblaciones.

La falta de criterios claros y transparencia en los mismos en la detección y corrección de sesgos puede llevar a el rechazo injustificado de préstamos a personas de comunidades vulnerables o en diagnósticos menos precisos para minorías, lo que

nos da la alerta de establecer reglas y métodos efectivos para evitar casos de este estilo y reducir al máximo el sesgo a diferentes grupos.

La justificación de este enfoque es clara:

1. La IA se usa en áreas importantes como salud, justicia y finanzas, donde un fallo puede ocasionar perjuicios graves. “La falta de transparencia en los algoritmos de IA puede generar desconfianza y preocupación entre los estudiantes, profesores y otras partes interesadas” (Maldonado Zuñiga et all. 2024, p.73).
2. Es necesario definir marcos regulatorios, protocolos y prácticas para asegurar su uso responsable. “Falta de infraestructura tecnológica, la escasez de personal capacitado y la necesidad de políticas claras que regulen el uso de la IA” (Maldonado Zuñiga et all, 2024. p.72).

Bajo este contexto, analizar con mayor profundidad tres aspectos clave que permiten comprender el impacto social de la inteligencia artificial es fundamental. Primero, identificar causas detrás de estos sesgos en los algoritmos, ya que no son errores técnicos, sino de reflejos de prejuicios y desigualdades existentes en los datos que los alimentan. También, se debe examinar cómo la falta de equidad en estos sistemas influye directamente en decisiones que afectan la vida de las personas, aumentando las desigualdades. Y, por último, se deben considerar las posibles soluciones para avanzar hacia una inteligencia artificial más justa, abarcando tanto enfoques técnicos como marcos legales y éticos que garanticen su desarrollo responsable.

Aunque prometen objetividad, los algoritmos pueden reproducir desigualdades cuando se entrenan con datos que las reflejan. En muchos casos, los datos incluyen patrones de exclusión social o económica, por lo que un modelo de IA termina reforzando las mismas barreras que pretendía superar.

Además, los equipos que crean estos sistemas no siempre son diversos. Cuando quienes desarrollan la IA comparten una misma visión sobre género, etnia o formación, es fácil pasar por alto sesgos. La falta de perspectivas diferentes fomenta tecnologías más “opacas” y cerradas, y nadie cuestiona las decisiones de diseño.

Esto puede generar desconfianza, ya que los algoritmos creados no son transparentes. “Un error en un sistema de IA podría reducir la confianza tanto del público como de los profesionales de la salud, afectando negativamente la aceptación de estas tecnologías y frenando su desarrollo.” (Zesergio MELO, 2024, p.52).

La transparencia y la rendición de cuentas son pilares éticos esenciales. Las instituciones deben garantizar que el funcionamiento de los algoritmos sea interpretable y establecer mecanismos de responsabilidad. “Esta claridad no solo facilita decisiones clínicas más fundamentadas, sino que también refuerza la credibilidad de la IA, fomentando la confianza tanto en los profesionales de la salud como en los pacientes.” (Zesergio MELO, 2024).

Los efectos ya se hacen notar: en el sector financiero, algunos bancos rechazan préstamos de forma desproporcionada a solicitantes con ciertos apellidos o de ciertas zonas. En salud, herramientas de diagnóstico muestran menor precisión con pacientes de piel más oscura por falta de representatividad en los datos de entrenamiento.

Por esto, es urgente implementar auditorías continuas, diversificar equipos de desarrollo y fomentar la transparencia. Solo así la IA podrá ser una herramienta justa.

La confianza es vital para que la IA sea aceptada socialmente. “For users to trust and interact effectively with AI systems, they need to understand how these systems make decisions.” (L. Inglada Galiana, 2024, 180).

Uno de los principales problemas de los sistemas sesgados es que el público pierde la confianza en ellos. En el ámbito judicial, algunos algoritmos han mostrado sesgos contra ciertos grupos raciales o socioeconómicos. “...manipulación de la percepción y decisiones” (Patricia Villarreal Novillo, 2023, 9509).

La opacidad agrava el problema. “La naturaleza a menudo opaca de los sistemas de IA puede hacer difícil entender cómo se llega a ciertas decisiones” (Patricia Villarreal Novillo, 2023, 9498). Esta falta de claridad impide cuestionar decisiones o auditar su validez, especialmente grave en contextos judiciales, donde la equidad y la rendición de cuentas son esenciales, este nivel de opacidad no debería ser aceptable.

En el ámbito empresarial también hay discriminación. Algoritmos usados en procesos de contratación han favorecido a hombres sobre mujeres, o han excluido personas según su lugar de residencia o apellido. “...IA pueden perpetuar o amplificar los sesgos existentes en los datos sobre los que se entrenan” (Patricia Villarreal Novillo, 2023, 9513). Esta situación es preocupante, niega oportunidades laborales a sectores ya vulnerables y, daña la credibilidad de las empresas que las usan.

Aun así, la IA puede ser muy beneficiosa. Por ejemplo, “...herramientas de gestión y análisis de datos permiten a las empresas tomar decisiones más informadas, prever

tendencias de mercado o incluso automatizar tareas repetitivas...” (Patricia Villarreal Novillo, 2023, 9497) y también: “AI is being used to detect diseases in early stages... AI algorithms in radiology can identify subtle signs of diseases such as cancer...” (L. Inglada Galiana, 2024, 182). o en otros sectores como el transporte donde también se ven sus beneficios como “vehículos autónomos, que hace unas décadas parecían cosa de ciencia ficción, hoy son una realidad en el desarrollo gracias a la IA. Estos sistemas pueden interpretar el entorno, tomar decisiones y desplazarse con una eficiencia y seguridad cada vez mayores” (Patricia Villarreal Novillo, 2023, 9497)

Pero si la gente percibe que las decisiones de la IA son injustas o incomprensibles, se genera fricción con su uso, incluso en áreas donde podría traer grandes beneficios. En resumen, podemos decir que la confianza hacia las nuevas tecnologías es fundamental para su aceptación, pero la falta de equidad y transparencia daña esta misma.

¿Es posible diseñar una IA completamente justa y ética?

The possibility that future developments in AI will lead to systems that can display forms of consciousness or advanced autonomy has profound ethical and philosophical implications. Reflecting on these possibilities involves exploring fundamental questions about the nature of consciousness, identity, and ethics in relation to non-human entities. (L. Inglada Galiana, et al 2024, 18)

Para avanzar hacia una IA más justa, se requieren acciones desde lo técnico, ético y legal. “A legal framework must be developed that specifically addresses the unique challenges posed by AI...” (L. Inglada Galiana, 2024, 182).

Como soluciones técnicas algunas pueden ser, auditar los algoritmos de forma constante. Medir sesgos y corregirlos antes de que causen daño. "...transparency and liability. Explainability can help in compliance with these legal requirements and standards" (L. Inglada Galiana, 2024, 180).

También se deben usar datos diversos y equilibrados: "Diversificación de datos de entrenamiento: asegurarse de que los datos utilizados para entrenar los modelos de IA sean diversos y representativos..." (López Martínez, Heriberto García Peña, 2024, 118). De no ser así, "la falta de diversidad en los datos de entrenamiento... puede llevar a resultados sesgados y poco representativos" (López Martínez, Heriberto García Peña, 2024, 114). Estas técnicas, si se llegan a implementar con responsabilidad, podrán ayudar a reducir los sesgos algorítmicos y mejorar la equidad en los resultados.

Sin embargo, estas soluciones técnicas por sí solas no bastan. Es igual de necesario establecer marcos legales que regulen el uso de la inteligencia artificial.

Es imprescindible establecer leyes claras que regulen el uso de la IA. Estas deben asegurar transparencia, equidad y la posibilidad de apelar decisiones.

"Auditorías algorítmicas: realizar auditorías regulares de los algoritmos para detectar y corregir sesgos..." (López Martínez, Heriberto García Peña, 2024, 114).

Además, el diseño de IA no puede estar solo en manos de ingenieros. Es fundamental incluir expertos de otras disciplinas:

Enfoques interdisciplinarios: colaborar con expertos de diferentes disciplinas, como la sociología, la psicología y el derecho, para desarrollar enfoques más holísticos para la detección y corrección de sesgos. La integración de

perspectivas diversas puede ayudar a identificar y abordar problemas que pueden no ser evidentes desde una única disciplina (Mullane citado en López Martínez, Heriberto García Peña, 2021)

En resumen, llegar a tener una inteligencia artificial equitativa requiere no solo avances técnicos: si no también es necesario un enfoque integral. Diversificar los datos de entrenamiento para que representen a toda la población, realizar auditorías que detecten sesgos y promover la transparencia en los algoritmos son pasos fundamentales. Además, es crucial educar a los desarrolladores en ética y fomentar la participación de expertos de distintas disciplinas como el derecho y la sociología. Solo con soluciones combinadas y una mirada crítica y colaborativa podremos construir tecnologías inclusivas y garantizar que sus decisiones respeten los derechos de todas las personas. (López Martínez, Heriberto García Peña, 2024).

La equidad en la inteligencia artificial no es un ideal opcional, sino una necesidad urgente para evitar decisiones que perjudiquen a grupos vulnerables y para garantizar que esta tecnología beneficie a todos por igual. Como se ha demostrado, los algoritmos pueden reproducir o incluso amplificar desigualdades si no se controlan adecuadamente. Esto se debe, en parte, a que los sistemas de IA aprenden de datos que ya contienen sesgos históricos, y a que muchas veces quienes los desarrollan no representan a toda la diversidad social existente.

Además, cuando las decisiones de una IA no son comprensibles ni auditables, la confianza del público se erosiona. “La naturaleza a menudo opaca de los sistemas de IA puede hacer difícil entender cómo se llega a ciertas decisiones” (Patricia Villarreal

Novillo, 2023, p. 9498). Sin confianza, incluso los avances más prometedores pueden generar rechazo o temor.

Frente a este panorama, es esencial combinar soluciones técnicas —como auditorías, diversificación de datos y transparencia algorítmica— con marcos legales claros. “A legal framework must be developed that specifically addresses the unique challenges posed by AI...” (L. Inglada Galiana, 2024, p. 182). No basta con programar mejor: es imprescindible regular, supervisar y garantizar que los principios éticos se mantengan en el diseño y la implementación de la tecnología.

Este desafío se vuelve aún más complejo en países en desarrollo, donde “falta infraestructura tecnológica, la escasez de personal capacitado y la necesidad de políticas claras que regulen el uso de la IA” (Maldonado Zuñiga et al., 2024, p. 72). Por eso, además de establecer normas, es fundamental evaluar con el tiempo si estas soluciones están funcionando y adaptarlas según los contextos específicos.

Solo con un enfoque interdisciplinario, que incluya a técnicos, sociólogos, juristas y representantes de comunidades diversas, se podrá construir una inteligencia artificial verdaderamente justa. “La integración de perspectivas diversas puede ayudar a identificar y abordar problemas que pueden no ser evidentes desde una única disciplina” (Mullane citado en López Martínez, Heriberto García Peña, 2021). La equidad no es solo una condición ética: es la base para una tecnología que respete los derechos y necesidades de todas las personas.

Bibliografía:

Guzmán Gómez, C. (2024). Inteligencia artificial y equidad de género: Una perspectiva histórica de los sesgos culturales y su impacto en la relación humana con las tecnologías de la información y comunicación. *Sintaxis: Revista Científica del Centro de la Investigación para la Comunicación Aplicada*, 8, 14–30.

<https://revistas.anahuac.mx/index.php/sintaxis/article/view/2839/2854>

Inglada Galiana, L., et al. (2024). Ethics and artificial intelligence. *Revista Clínica Española*, 224(3), 178–186.

<https://sciencedirect.tec.elogim.com/science/article/pii/S2254887424000213/pdf?md5=8104cc0184366e08b8bf19c9031e5b38&pid=1-s2.0-S2254887424000213-main.pdf>

López Martínez, C., & García Peña, H. (2024). IA y sesgos: Una visión alternativa expresada desde la ética y el derecho. *Revista Iberoamericana de Derecho Informático*, 1(15), 109–121.

<https://dialnet.unirioja.es/descarga/articulo/9870496.pdf>

Maldonado Zuñiga, L., et al. (2024). Desafíos universitarios: Equidad, transparencia, responsabilidad y oportunidades en el uso de la inteligencia artificial. *Serie*

Científica de la Universidad de las Ciencias Informáticas, 17(5), 70–83.

<https://publicaciones.uci.cu/index.php/serie/article/view/1597>

Melo, Z. (2024). Inteligencia artificial en salud: Desafíos éticos para lograr la aplicación de las tecnologías a la salud del paciente. *Dignidad Humana e Inteligencia Artificial: Aplicaciones Éticas para una Armonía Futura*, 18, 49–62.

<https://www.unilim.fr/trahs/6349>

Villarreal Novillo, P. (2023). La justicia en tiempos de inteligencia artificial. *Ciencia Latina Revista Científica Multidisciplinar*, 7(5), 9494–9519.

<https://dialnet.unirioja.es/descarga/articulo/9481847.pdf>