MULTIVERSO JOURNAL | ISSN: 2792-3681
Volumen 6, Número 10, Edición Enero-junio de 2026
https://doi.org/10.46502/issn.2792-3681/2026.10.4
Cómo citar:
Barragán Ramírez, C.A., Muela Santamaria, X.G., Florez Vasquez, C., & Moreira Arteaga, S.P. (2026). Aspectos éticos del uso general de la inteligencia artificial en el marco de la digitalización social. *Multiverso Journal, 6*(10), 43-52. https://doi.org/10.46502/issn.2792-3681/2026.10.4

Aspectos éticos del uso general de la inteligencia artificial en el marco de la digitalización social

Ethical considerations regarding the widespread use of artificial intelligence in the context of social digitalization

Christian Andrés Barragán Ramírez *
Xavier Geovanny Muela Santamaria **
Carlos Flórez Vasquez ***
Silvia Patricia Moreira Arteaga ****

Recibido el 15/02/2026 - Aceptado el 23/04/2026

Resumen

Este artículo examina los aspectos éticos del uso general de la inteligencia artificial (IA) en el contexto de la digitalización social. Se parte de la idea de que los problemas éticos asociados a la IA no se reducen a prevenir abusos puntuales o errores técnicos, sino que exigen repensar cómo estas tecnologías reconfiguran la dignidad, la autonomía, la equidad y la confianza en las relaciones sociales. El objetivo general fue analizar, de forma acotada, los principales dilemas éticos que surgen del uso extendido de la IA en distintos ámbitos de la vida social digital, con el fin de precisar criterios filosóficos básicos para su valoración responsable. Metodológicamente, se adopta un enfoque interpretativo de base documental, sustentado en la revisión crítica de literatura especializada: obras clásicas de filosofía de la tecnología y ética, junto con artículos científicos de alto impacto indexados en las principales bases de datos internacionales. Los resultados de la investigación muestran que los dilemas éticos de la IA no son simples "fallas" corregibles mediante ajustes técnicos, sino síntomas de una transformación profunda en la distribución del poder de ver, clasificar y decidir sobre las vidas humanas en entornos atravesados, ontológicamente, por datos y algoritmos.

Palabras clave: aspectos éticos, inteligencia artificial, digitalización social, reflexión filosófica.

* Doctor en Ciencias Sociales mención Gerencia, por la Universidad del Zulia, Venezuela. Pontificia Universidad Católica de Ecuador Sede Ambato, Ecuador. https://orcid.org/0000-0001-8027-7883 - Email: cbarragan@pucesa.edu.ec
** Licenciado en Gastronomía por la Universidad Hemisferios, Ecuador, Docente en la Pontificia Universidad Católica del Ecuador Sede Ambato, Ecuador. https://orcid.org/0009-0005-1771-3124 - Email: xmuel@pucesa.edu.ec
*** Doctor en Ciencias Sociales mención Gerencia, por la Universidad del Zulia, Venezuela. Economista, Universidad del Magdalena. Docente de la Universidad nacional abierta y a distancia UNAD, Colombia. https://orcid.org/0009-0000-9172-6138 - Email: florez@unad.edu.co
**** Universidad de Valladolid: Valladolid, Castilla y León, España. https://orcid.org/0009-0008-6783-5500 - Email: patty_ma06@hotmail.com

Abstract

This article examines the ethical aspects of the widespread use of artificial intelligence (AI) in the context of social digitalization. It begins with the premise that the ethical issues associated with AI go beyond merely preventing isolated abuses or technical errors; rather, they require us to rethink how these technologies reshape dignity, autonomy, equity, and trust in social relationships. The overall objective was to analyze, in a focused manner, the main ethical dilemmas arising from the widespread use of AI in various spheres of digital social life, with the aim of identifying basic philosophical criteria for its responsible assessment. Methodologically, an interpretive, document-based approach is adopted, grounded in a critical review of specialized literature: classic works in the philosophy of technology and ethics, along with high-impact scientific articles indexed in major international databases. The research findings show that the ethical dilemmas of AI are not mere "flaws" that can be corrected through technical adjustments, but rather symptoms of a profound transformation in the distribution of power to observe, classify, and decide on human lives in environments ontologically permeated by data and algorithms.

Keywords: ethical issues, artificial intelligence, social digitization, philosophical reflection.

Introducción

El vertiginoso proceso de digitalización social ha convertido a la inteligencia artificial IA en una infraestructura cotidiana de mediación, que organiza información, perfila conductas, automatiza decisiones y reordena formas de interacción que antes dependían casi por completo del juicio humano. En ese tránsito, los aspectos éticos del uso general de la IA no se limitan a impedir abusos puntuales, sino que exigen pensar, con mucha seriedad, cómo estas tecnologías reconfiguran la dignidad, la autonomía, la justicia y la confianza en los vínculos sociales.

Tal como señala la United Nations Educational, Scientific and Cultural Organization (2021), desde una perspectiva filosófica, el problema no es solo qué puede hacer la IA, sino qué tipo de mundo social contribuye a producir cuando se integra de manera amplia en la educación, la administración pública, el trabajo, la salud, la comunicación y el consumo de información. La cuestión ética aparece, entonces, como una reflexión sobre los fines y los límites del progreso digital, que se pregunta cómo evitar que la eficiencia técnica desplace la deliberación moral, cómo prevenir sesgos y exclusiones, y cómo conservar un espacio humano de responsabilidad ante sistemas cada vez más opacos y normalizados.

En este orden de ideas, el objetivo general de esta investigación es analizar, de forma acotada, los principales dilemas éticos que surgen del uso general de la inteligencia artificial en contextos de digitalización social, con el fin de precisar criterios filosóficos básicos para su valoración responsable. A partir de ello, la investigación se guía por dos preguntas clave: ¿qué principios éticos deberían orientar el uso cotidiano de la IA para que no debilite la autonomía y la justicia social? y ¿qué exigencias morales impone la digitalización cuando convierte a la IA en una mediación estable de la vida colectiva?

El artículo se organiza en cinco secciones: después de esta introducción, se desarrolla el marco teórico; luego se expone el diseño metodológico; más adelante, en el análisis y la discusión de los hallazgos, se presentan dos subtítulos que responden a las preguntas orientadoras; y, finalmente, se ofrecen las conclusiones y las recomendaciones del estudio, sin ninguna pretensión de generalidad teórica. Esta estructura permite pasar de una base conceptual sólida a una reflexión crítica y aplicable sobre la ética de la IA en la sociedad digital.

Marco teórico

El marco teórico de esta investigación se articula, en primer lugar, alrededor de un enfoque de ética de la IA basado en derechos humanos, que asume la digitalización social como un espacio donde se juegan, de forma cotidiana, la dignidad, la autonomía y la justicia. En palabras de Morandín-Ahuerma (2023), este modelo normativo parte de la idea de que el despliegue generalizado de sistemas de IA, solo puede considerarse legítimo si respeta límites claros, tales como: proporcionalidad en su uso, prevención del daño, protección robusta de la privacidad y la protección de datos, así como mecanismos efectivos de responsabilidad y rendición de cuentas frente a decisiones automatizadas que afectan biografías concretas. En consecuencia:

> Para garantizar la seguridad de los seres humanos, el medio ambiente y los ecosistemas, es crucial abordar y mitigar los daños no deseados, riesgos de seguridad y las vulnerabilidades que pueden dar lugar a ataques a lo largo de todo el ciclo de vida de los sistemas de IA. Esto implica tomar medidas proactivas para prevenir y eliminar tales riesgos. Para promover eficazmente la seguridad de la IA, es esencial establecer marcos sostenibles para acceder a los datos. (Morandín-Ahuerma, 2023, pp. 87-88)

En el contexto de la digitalización social, este modelo permite interpretar, críticamente, prácticas de vigilancia masiva, perfilado algorítmico y personalización extrema como posibles vulneraciones de libertades fundamentales, más que como simples avances técnicos, y exige, por tanto, una gobernanza de la IA centrada en la persona y no en la eficiencia instrumental.

En segundo lugar, se incorpora un modelo de teorías éticas clásicas aplicadas a la IA —principalmente de ontologismo y utilitarismo—, que ofrece un andamiaje para evaluar la corrección de las decisiones algorítmicas, tanto por sus principios como por sus consecuencias. Tal como afirman González Arencibia, & Martínez Cardero (2020), desde una perspectiva deontológica, el problema central es que la digitalización social no convierta a las personas en meros medios de extracción de datos o de optimización de procesos, sino que respete su condición de fines en sí mismos, incluso cuando las interacciones se encuentran mediadas por plataformas y sistemas automatizados.

Por su parte, la sensibilidad utilitarista se hace visible en la exigencia de maximizar beneficios sociales y minimizar daños, preguntando hasta qué punto las arquitecturas de IA incrementan el bienestar colectivo, o más bien consolidan asimetrías de poder y nuevas formas de exclusión, en particular sobre los grupos más vulnerables en ecosistemas digitales hiperconectados.

Un tercer modelo teórico procede de la filosofía de la tecnología (Quintanilla, 2017) y de las teorías críticas de la digitalización (Nosthoff & Maschewski, 2022), que no se limitan a preguntar cómo regular la IA, sino qué tipo de forma de vida está siendo producida por su integración generalizada. En términos ontológicos, dicho enfoque entiende la IA, como parte de sistemas sociotécnicos que moldean percepciones, hábitos de atención, formas de reconocimiento recíproco y estructuras de dominación, por lo que los dilemas éticos no pueden aislarse del análisis del poder, la economía de datos y la racionalidad técnica que subyace a la digitalización social (Foucault, 1980).

Para los autores de estas reflexiones, una mirada crítica subraya que la IA no es un simple conjunto de herramientas neutras, sino una cristalización de intereses y visiones del mundo, de modo que la ética debe abrirse a interrogantes sobre justicia social, no dominación algorítmica y participación democrática en el diseño de las infraestructuras digitales.

En diálogo con estos tres modelos (Morandín-Ahuerma, 2023; Nosthoff & Maschewski, 2022), cobran relevancia las propuestas de una ética de la IA "antropocéntrica", que intentan articular principios como transparencia, equidad, supervisión humana significativa y sostenibilidad social como criterios operativos para el despliegue de sistemas algorítmicos en la vida cotidiana.

La literatura reciente sobre ética de la IA (United Nations Educational, Scientific and Cultural Organization, 2021; Quintanilla, 2017), insiste en la necesidad de traducir los grandes valores filosóficos en salvaguardas concretas: explicar las decisiones automatizadas a los ciudadanos, auditar sesgos y discriminaciones, evitar que la opacidad técnica se transforme en una nueva forma de dominación y garantizar que los procesos de digitalización no vacíen de contenido la responsabilidad moral de los agentes humanos. Este modelo "de principios" complementa las teorías previas al mostrar que la cuestión no es solo qué es justo o correcto en abstracto, sino cómo se constituye, en la práctica concreta, una cultura institucional de cuidado y responsabilidad en la gobernanza de la IA.

En una mirada panorámica, estos modelos teóricos ofrecen a la presente investigación un horizonte articulado y, a la vez, crítico: el enfoque basado en derechos humanos proporciona un anclaje normativo fuerte para evaluar el uso general de la IA (Habermas, 2003); las teorías éticas clásicas (Nosthoff & Maschewski, 2022) permiten contrastar principios y consecuencias de las decisiones algorítmicas en distintos ámbitos de la digitalización social; y la filosofía de la tecnología, junto con las perspectivas críticas, visibiliza las dimensiones estructurales de poder, desigualdad y configuración de formas de vida que atraviesan la expansión de la IA (Quintanilla, 2017).

En conjunto, estos marcos permiten abordar los aspectos éticos de la inteligencia artificial, no ya como un listado de riesgos o buenas prácticas, sino como un problema filosófico complejo que exige repensar la relación entre técnica, justicia y autonomía en sociedades crecientemente digitalizada.

Metodología

Tomando en consideración la naturaleza del tema abordado, la metodología se concibe aquí como una revisión documental de enfoque cualitativo-interpretativo, informada por lineamientos de transparencia propios de revisiones sistemáticas (como PRISMA) (Rethlefsen et al., 2021), pero aplicada a un objeto eminentemente filosófico: los dilemas éticos de la IA en la digitalización social.

Como es lógico suponer, este diseño metodológico parte de una base hermenéutica. En la lógica de Ricoeur (2008), se asume que los textos no solo contienen datos, sino interpretaciones situadas de conflictos morales, y que el trabajo del investigador consiste, por lo tanto, en reconstruir, comparar y evaluar críticamente esos horizontes de sentido. Así, se combina una lógica sistemática en la búsqueda y selección de estudios con una lógica interpretativa en el análisis, lo que permite mantener cierto rigor metodológico sin renunciar a la densidad filosófica del objeto de estudio.

En lo concreto, los criterios de selección de fuentes se definen *ex ante* y se documentan de forma explícita. Como criterios de inclusión se consideraron: a) publicaciones de acceso abierto indexadas, preferiblemente pero no exclusivamente, en WOS, Scopus u otras bases reconocidas, b) periodo temporal acotado al último lustro, c) abordaje explícito de dilemas éticos vinculados al uso general de la IA en contextos de digitalización social (no solo aplicaciones clínicas o técnicas aisladas) y d) presencia de un marco filosófico, normativo o crítico relevante (ética de la tecnología, derechos humanos, justicia social, etc.).

Al mismo tiempo, se excluyeron trabajos puramente técnicos sin reflexión ética, literatura gris y documentos sin revisión por pares. Además, se justificó la inclusión de algunas obras filosóficas clásicas y manuales contemporáneos sobre ética de la tecnología, siempre que ofrecieran categorías conceptuales clave para interpretar los hallazgos de la literatura reciente sobre ética de la IA.

El procedimiento siguió cuatro fases claramente diferenciadas, inspiradas en la lógica de los diagramas de flujo PRISMA (identificación, cribado, elegibilidad e inclusión) tal como lo explica Rethlefsen et al. (2021), pero adaptadas al carácter cualitativo-interpretativo de este estudio. En la fase de identificación se aplicaron cadenas de búsqueda combinando descriptores en inglés y español (por ejemplo: "ethical dilemmas" AND "artificial intelligence" AND "digitalization/social media/public services"), registrando bases, fechas y filtros utilizados.

En el cribado se revisaron títulos y resúmenes para descartar duplicados y trabajos manifiestamente irrelevantes. La fase de elegibilidad, por su parte, implicó la lectura completa de los textos preseleccionados, aplicando los criterios definidos e identificando el tipo de enfoque filosófico o normativo. Finalmente, en la fase de inclusión, se conformó el corpus documental definitivo y se procedió a un análisis de contenido cualitativo, con codificación temática orientada a extraer tipos de dilemas éticos y principios invocados, seguido de una etapa interpretativa donde se pusieron en diálogo estos resultados con los modelos teóricos reseñados anteriormente.

Pese a su coherencia, esta estructura metodológica presenta límites que condicionan el alcance de las conclusiones. En primer lugar, toda revisión documental depende de lo que ha sido publicado y catalogado en las bases consultadas, por lo que ciertas perspectivas —sobre todo de contextos no hegemónicos o tradiciones filosóficas minoritarias— pueden quedar infrarrepresentadas. En segundo lugar, el componente interpretativo introduce un margen inevitable de subjetividad, tal como sostiene Gadamer (1993), en consecuencia, el investigador selecciona, codifica y articula los sentidos a partir de su propio horizonte teórico, de modo que la reflexividad y la transparencia en las decisiones analíticas resultan cruciales para sostener la credibilidad del estudio.

Finalmente, el énfasis en la dimensión textual hace que esta metodología no capture, al menos no directamente, experiencias vividas de usuarios y comunidades afectadas por la IA, de modo que sus resultados deben entenderse como un mapa normativo y crítico que invita, más que clausura, a futuros trabajos empíricos y mixtos sobre la ética de la inteligencia artificial, en sociedades digitalizadas del Norte y del Sur Global.

A modo de análisis y discusión de los hallazgos:

Principios éticos para un uso emancipador y justo de la inteligencia artificial

Ante la legítima pregunta ¿qué principios éticos deberían orientar el uso cotidiano de la IA para que no debilite la autonomía y la justicia social? Todo indica que, los principios éticos que deberían orientar el uso cotidiano de la IA solo tienen sentido si se entienden como condiciones materiales y morales para preservar la autonomía del ser humano y la justicia social como base de la convivencia democrática.

Por lo tanto, para autores como González Arencibia, & Martínez Cardero (2020), no basta con prohibir ciertos usos extremos; se trata de asegurar que ninguna persona quede reducida a un expediente de datos ni a un perfil probabilístico que otros explotan sin su conocimiento ni control. De ahí que principios como el respeto a la autonomía, la no maleficencia, la justicia y la explicabilidad deban leerse en clave política: son barreras frente a la tentación de una gobernanza algorítmica que "administre" la vida social sin participación ciudadana y sin rendición de cuentas clara.

Desde esta perspectiva, orientar éticamente la IA en la vida cotidiana exige articular varios principios complementarios. El respeto de la autonomía humana implica que los sistemas no manipulen, coaccionen ni sustituyan el juicio de las personas, sino que ofrezcan información comprensible y espacios reales de consentimiento, revocación y desacuerdo. La justicia y la no discriminación obligan a detectar y corregir

sesgos, evitando que los algoritmos reproduzcan o profundicen desigualdades históricas, especialmente sobre grupos vulnerables (Morandín-Ahuerma, 2023).

La transparencia y la explicabilidad demandan, por su parte, que las decisiones automatizadas sean auditables y entendibles, mientras que la responsabilidad y la rendición de cuentas garantizan que siempre haya agentes humanos e instituciones identificables que respondan por los efectos de la IA en la vida social (United Nations Educational, Scientific and Cultural Organization, 2021).

Tabla 1.
Principios éticos para un uso cotidiano de la IA compatible con autonomía y justicia social

Principio	Foco ético central	Riesgo que busca evitar	Exigencia práctica en el uso cotidiano de IA
Respeto de la autonomía.	Capacidad de decidir y comprender.	Manipulación, tutela algorítmica y opacidad.	Información clara, posibilidad de optar, revocar y cuestionar decisiones.
Justicia y no discriminación.	Igual dignidad y trato justo.	Sesgos, exclusiones y reproducción de desigualdades.	Evaluaciones de impacto, auditorías de sesgos, inclusión de grupos vulnerables.
Transparencia y explicabilidad.	Comprensibilidad y control democrático.	Cajas negras incontrolables.	Modelos auditables, explicaciones accesibles a usuarios y autoridades.
Responsabilidad y rendición de cuentas.	Asignación clara de responsabilidades.	Dilución de culpas, impunidad técnica.	Mecanismos legales, institucionales y éticos para reparar daños.

Nota metodológica: la tabla sintetiza, en clave analítica, los principios convergentes en marcos internacionales recientes y en la literatura filosófica y política sobre ética de la IA, adaptándolos al contexto de la digitalización social. Fuente: elaboración propia (2026).

La información de la Tabla 1 permite comprender que estos principios no son eslóganes abstractos, sino criterios operativos que reordenan la relación entre tecnología, poder y ciudadanía. Al exigir respeto efectivo de la autonomía, la justicia y la explicabilidad, lo que se está reclamando es que el diseño y uso cotidiano de la IA se sometan a la lógica de los derechos humanos (Asamblea General de las Naciones, 1948) y de la deliberación democrática de la que habla Habermas (1999), en su teoría de la acción comunicativa, y no solo a la optimización técnica o al interés económico. De este modo, la IA puede convertirse en una herramienta que amplía capacidades y oportunidades, siempre que permanezca anclada en instituciones que supervisen sus impactos, escuchen a quienes resultan afectados y estén dispuestas a corregir o, incluso desactivar sistemas, que socaven la autonomía personal o la justicia social.

Responsabilidades morales ante una vida social mediada por la inteligencia artificial

Cuando la digitalización convierte a la IA en una mediación estable de la vida colectiva, nos obliga a asumir una primera exigencia moral: repensar la distribución del poder en sociedades gobernadas, en parte, por infraestructuras algorítmicas. Siguiendo las reflexiones de Foucault (1980), en su microfísica del poder, no se trata solo de "usar bien" las herramientas, sino de preguntarnos quién diseña los sistemas, con qué fines, con qué datos y bajo qué formas de control democrático.

Allí donde la IA filtra la información, organiza la conversación pública o condiciona el acceso a derechos, se hace moralmente ineludible garantizar que los valores de dignidad, justicia y pluralismo, que son la base de los derechos humanos (Asamblea General de las Naciones, 1948), no queden subordinados a la sola eficiencia o al interés económico de unos pocos actores hegemónicos (Estados imperialistas, Megacorporaciones, grupos de poder).

Una segunda exigencia moral tiene que ver con la responsabilidad y el cuidado en contextos donde la mediación algorítmica se vuelve invisible, pero omnipresente. La digitalización introduce sistemas que observan, clasifican y anticipan nuestras conductas, y esto genera obligaciones de transparencia sustantiva, de prevención del daño y de inclusión de las voces más vulnerables en el diseño y la evaluación de la IA. En la educación, la salud, la justicia o la comunicación, la pregunta ya no es definitivamente si debemos usar IA, sino cómo hacerlo sin erosionar la confianza, la equidad y la posibilidad de deliberación compartida sobre los fines que perseguimos como comunidad política.

Tabla 2.
Exigencias morales en una vida colectiva mediada por IA

Exigencia moral central	Pregunta ética que formula	Tipo de responsabilidad implicada
Redistribuir el poder algorítmico.	¿Quién decide qué ve, recibe o puede hacer cada ciudadano?	Responsabilidad política y de gobernanza digital.
Proteger la vulnerabilidad y la dignidad humana y de todas las formas de vida.	¿Cómo se evita que la IA agrave desigualdades y estigmas?	Responsabilidad de cuidado e inclusión.
Garantizar transparencia y contestabilidad.	¿Podemos comprender, cuestionar y corregir decisiones algorítmicas?	Responsabilidad epistémica y procedimental.
Mantener espacios de deliberación humana.	¿En qué momentos la decisión debe seguir siendo propiamente humana?	Responsabilidad moral y prudencial.

Nota metodológica: el cuadro sintetiza, de forma analítica, las exigencias morales identificadas en la literatura reciente sobre ética de la IA y digitalización, organizándolas según el tipo de responsabilidad que introducen en la vida colectiva.
Fuentes: elaboración propia con base a los aportes de González Arencibia, & Martínez Cardero (2020) y United Nations Educational, Scientific and Cultural Organization (2021).

La Tabla 2 muestra que las exigencias morales no son simples "principios decorativos", sino obligaciones concretas que atraviesan el diseño, la regulación y el uso cotidiano de la IA. Redistribuir el poder algorítmico exige instituciones democráticas capaces de someter los sistemas a evaluación pública y a controles democráticos; proteger la vulnerabilidad supone reconocer que no todos los grupos están expuestos por igual a los daños de la digitalización; garantizar la transparencia y la contestabilidad implica que los ciudadanos puedan desafiar decisiones automatizadas; y preservar espacios de deliberación humana recuerda que no toda decisión que puede automatizarse debe automatizarse. En conjunto, estas exigencias dibujan una moral de la mediación digital que no renuncia a la IA, pero la subordina a la tarea siempre inacabada de construir una vida colectiva justa y verdaderamente humana.

Conclusiones y recomendaciones

Los dilemas éticos del uso general de la inteligencia artificial en la digitalización social no son simples "fallas" corregibles mediante ajustes técnicos, se trata de síntomas de una transformación profunda de la forma en que se distribuye el poder de ver, clasificar y decidir sobre las vidas humanas en entornos cada vez más mediados por datos.

En este contexto, la vigilancia algorítmica, la extracción masiva de información personal, la opacidad de los modelos y la concentración del control tecnológico, en pocos actores, configuran un escenario en el que la autonomía se ve tensionada, la dignidad corre el riesgo de reducirse a un patrón de datos y la justicia se ve comprometida por sesgos que se incrustan en infraestructuras digitales, supuestamente neutrales.

- En este orden de ideas, un primer criterio filosófico ineludible, para quienes suscriben esta investigación, es el de la primacía de la persona sobre la lógica de la eficiencia; por lo tanto, ningún

diseño o despliegue de IA puede considerarse, éticamente aceptable, si sacrifica la posibilidad de contestación, de explicación y de reparación de quienes son afectados por sus decisiones.

- Un segundo foco de dilemas éticos emerge allí donde la IA, integrada en plataformas, servicios públicos y dispositivos cotidianos, reconfigura silenciosamente las condiciones de ejercicio de la democracia y la justicia social. La personalización extrema de contenidos, la segmentación política, la automatización de decisiones en el ámbito laboral o en el acceso a prestaciones pueden amplificar desinformación, polarización y exclusiones estructurales, al mismo tiempo que diluyen la responsabilidad entre diseñadores, empresas y autoridades (United Nations Educational, Scientific and Cultural Organization, 2021).

De aquí se derivan los criterios filosóficos centrados en la justicia y la no dominación. En la lógica de la filosofía de la liberación de Dussel (2013), se trata de materializar la exigencia de que ningún grupo sea sistemáticamente desfavorecido por sistemas algorítmicos, el deber de prevenir nuevas formas de control invisible sobre las conciencias, y la responsabilidad de asegurar que las instituciones políticas conserven la capacidad final de deliberar, corregir y responder por decisiones en las que interviene la IA.

- Un tercer conjunto de dilemas se vincula con la tendencia a naturalizar la presencia de la IA como si fuera una "segunda naturaleza" digital, con el riesgo de que el juicio crítico se subordine a la lógica de la automatización (Maphosa, 2024). Así, cuando delegamos en sistemas de IA tareas de evaluación moral implícita —desde recomendar información hasta priorizar pacientes o asignar recursos— corremos el riesgo de habituarnos a decisiones que ya vienen "pre-interpretadas" por arquitecturas técnicas invisibles.

Por eso, un criterio filosófico clave es el de la responsabilidad irrenunciable: incluso cuando la IA participa en la cadena de decisión, el sentido último de la acción debe permanecer abierto a la deliberación humana, de modo que los sistemas se diseñen como soportes de la prudencia y el cuidado, y no como sustitutos de la conciencia moral. A ello se suma la exigencia de transparencia, no únicamente formal, sino comprensible y accesible, que permita en cada momento a ciudadanos y comunidades disputar, reinterpretar y reorientar los usos de la IA.

Desde una perspectiva interdisciplinaria en el campo de las ciencias humanas, estas tensiones sugieren varias líneas de investigación que vale la pena consolidar y recomendar. Se necesitan estudios que articulen filosofía moral, teoría política y sociología digital para analizar cómo los sistemas de IA reconfiguran la esfera pública, la confianza institucional y las desigualdades de poder, así como investigaciones que crucen filosofía del derecho, estudios críticos de datos y antropología para examinar los modos en que normas abstractas se traducen en prácticas algorítmicas concretas. Resulta igualmente urgente promover proyectos de investigación-acción en los que la reflexión filosófica acompañe procesos participativos de diseño y evaluación de sistemas de IA en salud, educación, administración pública o comunicación, con una atención especial a las voces de colectivos históricamente marginados.

Finalmente, una agenda de investigación futura debería plantearse preguntas que solo pueden abordarse desde la colaboración estrecha entre la reflexión filosófica y la investigación empírica. Algunas de ellas son: ¿cómo viven y resignifican los ciudadanos, en distintos contextos culturales, la presencia ubicua de sistemas de IA en su vida cotidiana? ¿qué condiciones institucionales y materiales son necesarias para que los principios éticos —como justicia, autonomía, no discriminación y responsabilidad— se traduzcan efectivamente en prácticas de diseño, regulación y uso de la IA? ¿de qué manera las métricas con las que evaluamos el "éxito" de un sistema de IA incorporan (o silencian) valores morales y sociales? y ¿qué modelos de gobernanza democrática pueden sostenerse cuando parte de las decisiones relevantes para la vida colectiva pasan por infraestructuras técnicas opacas? Conviene recordar que, la respuesta a estas preguntas no puede provenir, ni solo de la filosofía ni solo de las ciencias empíricas: exige un ejercicio compartido de imaginación crítica sobre el tipo de sociedad digital que consideramos digna de ser habitada.

Declaración de Conflictos de Interés:

No declaran conflictos de interés.

Financiamiento:

Ninguno

Declaración de contribución de los autores (CRediT)

Conceptualización: Christian Andrés Barragán Ramírez, Xavier Geovanny Muela Santamaria.
Metodología: Christian Andrés Barragán Ramírez, Xavier Geovanny Muela Santamaria, Carlos Florez Vasquez y Silvia Patricia Moreira Arteaga.

Investigación y recopilación de datos: Christian Andrés Barragán Ramírez, Carlos Florez Vasquez y Silvia Patricia Moreira Arteaga.

Análisis formal: Christian Andrés Barragán Ramírez y Xavier Geovanny Muela Santamaria.
Redacción – borrador original: Christian Andrés Barragán Ramírez.

Redacción – revisión y edición: Xavier Geovanny Muela Santamaria, Carlos Florez Vasquez y Silvia Patricia Moreira Arteaga.

Referencias Bibliográficas

Asamblea General de las Naciones. (1948). *Declaración Universal de Derechos Humanos*. ONU. https://www.un.org/es/about-us/universal-declaration-of-human-rights

Dussel, E. (2013). *Ética de la liberación en la edad de la globalización y la exclusión*. Buenos Aires, Argentina: Editorial Docencia.

Foucault, M. (1980). *Microfísica del poder*. Madrid, España: Ediciones de La Piqueta.

Gadamer, H.-G. (1993). *Verdad y método*. Salamanca, España: Ediciones Sígueme.

González Arencibia, M., & Martínez Cardero, D. (2020). Dilemas éticos en el escenario de la inteligencia artificial. *Economía y Sociedad, 25*(57), 93–109. https://www.scielo.sa.cr/pdf/eys/v25n57/2215-3403-eys-25-57-93.pdf

Habermas, J. (1999). *Teoría de la acción comunicativa I*. Madrid, España: Taurus.

Habermas, J. (2003). *The future of human nature*. Cambridge, United Kingdom: Polity Press.

Maphosa, V. (2024). The rise of artificial intelligence and emerging ethical and social concerns. *Artificial Intelligence, 2*(1), 1–24. https://doi.org/10.5772/acrt.20240020

Morandín-Ahuerma, F. (2023). *Principios normativos para una ética de la inteligencia artificial.* Secretaría de Educación de Puebla. https://philarchive.org/archive/MORDRD-2

Nosthoff, A. V., & Maschewski, F. (2022). Hacia una teoría crítica de la digitalidad. Günther Anders en la era del capitalismo de plataformas y las tecnocracias inteligentes. Constelaciones. *Revista de Teoría Crítica, 14*, 322–346. https://constelaciones-rtc.net/article/view/4978/5464

Quintanilla, M. (2017). *Tecnología: un enfoque filosófico y otros ensayos de filosofía de la tecnología*. Ciudad de México, México: Fondo de Cultura Económica. https://perio.unlp.edu.ar/catedras/wp-content/uploads/sites/210/2023/03/Tecnologia_un_enfoque_filosofico_y_otros.pdf

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews, 10*(1), 39. https://doi.org/10.1186/s13643-020-01542-z

Ricoeur, P. (2008). *Hermenéutica y acción: De la hermenéutica del texto a la hermenéutica de la acción*. Buenos Aires, Argentina: Prometeo Libros.
United Nations Educational, Scientific and Cultural Organization. (2021). *Recomendación sobre la ética de la inteligencia artificial*. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa