Sciences, Discoveries, and Society
Multidisciplinary Journal

Vol. 3 • Núm. 5 • 2026
Septiembre – Octubre
ISSN: 3091-1699
Artículo Revisión

Artículo Revisión

Ética de la inteligencia artificial y responsabilidad humana: dilemas filosóficos sobre la libertad, la justicia y la toma de decisiones en entornos digitales

Ethics of Artificial Intelligence and Human Responsibility: Philosophical Dilemmas Concerning Freedom, Justice and Decision-Making in Digital Environments

José Luis Frías Raza[1*]

[1]*Investigador Independiente, Ambato - Ecuador*

***Autor de correspondencia:** José Luis Frías Raza

Recibido: 14-09-2026 **Revisado:** 05-10-2026 **Aceptado:** 26-10-2026 **Publicado:** 31-10-2026

Cómo citar este artículo: Frías Raza, J. L. . (2026). Ética de la inteligencia artificial y responsabilidad humana: dilemas filosóficos sobre la libertad, la justicia y la toma de decisiones en entornos digitales. *Multidisciplinary Journal of Sciences Discoveries and Society, 3*(5), 1-13. https://doi.org/10.63969/4chykc21

RESUMEN

La inteligencia artificial plantea importantes interrogantes éticos relacionados con la responsabilidad humana, la autonomía individual, la justicia y la legitimidad moral de las decisiones en entornos digitales. El estudio se fundamenta en un enfoque cualitativo y un diseño documental, mediante la aplicación del procedimiento PRISMA 2020 para organizar la búsqueda, selección y clasificación de documentos pertinentes. La reflexión se estructura en siete categorías: responsabilidad humana, autonomía individual, justicia y equidad, transparencia y explicabilidad, legitimidad moral, supervisión humana y gobernanza tecnológica. Los hallazgos permiten identificar que la eficiencia técnica de los sistemas algorítmicos no garantiza decisiones éticamente justificables, especialmente cuando existen riesgos de discriminación, limitaciones en la comprensión de los procesos automatizados o dificultades para atribuir responsabilidades. Asimismo, se reconoce que las recomendaciones tecnológicas pueden apoyar la deliberación, aunque requieren mecanismos de control que preserven la libertad de elección. Se concluye que la utilización responsable de la inteligencia artificial demanda supervisión efectiva, transparencia, rendición de cuentas y criterios éticos orientados a proteger los derechos fundamentales. La articulación entre innovación tecnológica, juicio humano y deliberación moral resulta esencial para promover decisiones legítimas y respetuosas de la dignidad humana.

Palabras clave: Inteligencia artificial; responsabilidad humana; autonomía individual; ética algorítmica; justicia y equidad.

ABSTRACT

Artificial intelligence raises significant ethical questions concerning human responsibility, individual autonomy, justice and the moral legitimacy of decision-making in digital environments. The study adopts a qualitative approach and a documentary research design, using the PRISMA 2020 framework to organise the search, selection and classification of relevant documents. The discussion is structured around seven categories: human responsibility, individual autonomy, justice and fairness, transparency and explainability, moral legitimacy, human oversight and technological governance. The findings indicate that the technical efficiency of algorithmic systems does not necessarily guarantee ethically justifiable decisions, particularly where there is a risk of discrimination, limited understanding of automated processes or difficulty in attributing responsibility. Technological recommendations can also support deliberation, although appropriate safeguards are required to preserve freedom of choice. The study concludes that the responsible use of artificial intelligence requires effective human oversight, transparency, accountability and ethical principles

©2026 por los Autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones bajo una licencia internacional Creative Commons Atribución 4.0. (CC BY 4.0) https://creativecommons.org/licenses/by/4.0/

aimed at protecting fundamental rights. Integrating technological innovation with human judgement and moral deliberation is essential to ensuring that decisions are legitimate and respectful of human dignity.

Keywords: Artificial intelligence; human responsibility; individual autonomy; algorithmic ethics; justice and fairness.

1. INTRODUCCIÓN

La incorporación de la inteligencia artificial en los sistemas educativos, sanitarios, empresariales, jurídicos y sociales está modificando los criterios mediante los cuales se produce conocimiento, se procesa información y se fundamentan decisiones con repercusiones individuales y colectivas. Si bien estas tecnologías amplían las capacidades de análisis, automatizan operaciones complejas y facilitan la identificación de patrones, su utilización también plantea interrogantes filosóficos sobre la preservación de la autonomía, el reconocimiento de la dignidad humana, la distribución equitativa de oportunidades y la legitimidad de transferir determinadas funciones decisorias a sistemas computacionales. La intervención algorítmica en la valoración de personas, la asignación de recursos y la orientación de conductas exige examinar los principios morales que determinan los límites aceptables de su funcionamiento. Desde la filosofía ética, el análisis trasciende la eficiencia operativa y comprende la identificación de los valores que sustentan el diseño tecnológico, los criterios que justifican sus resultados y las obligaciones derivadas de sus consecuencias. La responsabilidad humana constituye, por tanto, una dimensión fundamental, dado que la creación, implementación y supervisión de estas herramientas responden a decisiones adoptadas por individuos e instituciones que actúan bajo determinadas prioridades económicas, sociales y políticas. Examinar estas relaciones permite establecer criterios para armonizar el desarrollo tecnológico con el ejercicio de la libertad, la protección de los derechos fundamentales y la legitimidad moral de las decisiones mediadas por inteligencia artificial.

La utilización de inteligencia artificial en decisiones que afectan derechos, oportunidades y condiciones de vida plantea dificultades para determinar cómo se distribuyen las obligaciones morales entre los desarrolladores, las instituciones responsables de su implementación y los usuarios que aceptan o ejecutan sus recomendaciones. La complejidad de los procedimientos algorítmicos, la limitada interpretabilidad de determinados sistemas y la presencia de sesgos en los datos de entrenamiento pueden obstaculizar la identificación de los factores que originan resultados perjudiciales, discriminatorios o insuficientemente justificados. Estas dificultades adquieren particular importancia en la contratación laboral, la valoración del rendimiento académico, la asignación de prestaciones, el diagnóstico clínico y la administración de justicia, donde las decisiones automatizadas pueden condicionar el acceso a recursos y el reconocimiento de derechos. A ello se suma el riesgo de que la aceptación acrítica de las recomendaciones tecnológicas debilite el juicio independiente y reduzca la intervención humana en procesos que requieren ponderación de principios, interpretación de circunstancias particulares y consideración de las consecuencias morales. La cuestión central radica en que la exactitud computacional no constituye una garantía suficiente de justicia, puesto que un sistema puede producir resultados técnicamente consistentes y, aun así, reproducir desigualdades, restringir opciones individuales o perjudicar a grupos determinados. Esta contradicción exige analizar los criterios filosóficos que permiten atribuir responsabilidad, establecer la legitimidad de las decisiones algorítmicas y preservar la autonomía personal frente a procedimientos automatizados. En consecuencia, resulta necesario esclarecer bajo qué condiciones la intervención de la inteligencia artificial puede considerarse moralmente justificable y qué obligaciones corresponden a quienes diseñan, administran y utilizan estas tecnologías cuando sus resultados generan consecuencias adversas.

La reflexión filosófica sobre las consecuencias de la tecnología adquirió especial relevancia durante el siglo XX, a medida que los sistemas automatizados comenzaron a modificar las relaciones entre las personas y los mecanismos de control social. Beraún et al. (2026), mediante sus aportaciones a la cibernética, examinó las implicaciones derivadas de la interacción entre comunicación, control y comportamiento humano, lo que permite comprender que el desarrollo técnico no puede separarse de las decisiones que determinan sus aplicaciones. Esta perspectiva constituye un antecedente relevante para estudiar la inteligencia artificial, debido a que sus efectos dependen tanto de sus capacidades operativas como de los propósitos, valores e intereses que orientan su diseño y utilización. Desde esta

consideración, la responsabilidad humana representa una condición indispensable para evaluar las consecuencias morales de las innovaciones tecnológicas y establecer límites que impidan subordinar el bienestar individual a criterios exclusivamente instrumentales.

Desde la filosofía moral, la autonomía constituye un principio fundamental para determinar hasta qué punto las decisiones automatizadas respetan la capacidad de las personas para actuar conforme a su propio juicio. Vega et al. (2026), en su formulación de la ética del deber, sostuvo que la dignidad humana exige reconocer a cada individuo como un fin en sí mismo y no únicamente como un medio para alcanzar determinados resultados. Esta concepción permite cuestionar el empleo de sistemas inteligentes que condicionan comportamientos, restringen oportunidades o utilizan información personal sin ofrecer garantías suficientes de transparencia y participación. Bajo esta perspectiva, la legitimidad de la inteligencia artificial no depende exclusivamente de su rendimiento técnico, sino también de su compatibilidad con la libertad de elección, la deliberación racional y el reconocimiento de las personas como sujetos moralmente autónomos.

La creciente complejidad de los algoritmos ha intensificado las dificultades relacionadas con la transparencia y la atribución de responsabilidades ante decisiones que producen consecuencias sociales. Castellanos et al. (2026) identificaron problemas éticos asociados con la opacidad algorítmica, la discriminación, la falta de rendición de cuentas y las limitaciones para interpretar los resultados de los sistemas automatizados. Sus aportaciones permiten advertir que el procesamiento computacional no elimina los conflictos morales presentes en las decisiones humanas, sino que puede reproducirlos mediante procedimientos cuya lógica resulta difícil de comprender para los usuarios. Esta situación adquiere especial relevancia cuando las personas afectadas desconocen los criterios empleados para clasificarlas o carecen de mecanismos efectivos para solicitar explicaciones, cuestionar una decisión y obtener una reconsideración fundamentada.

En materia de justicia social, resulta necesario examinar los mecanismos mediante los cuales determinadas aplicaciones tecnológicas pueden reproducir desigualdades preexistentes. Bermúdez et al. (2025) analizó los riesgos de los modelos matemáticos utilizados en ámbitos sociales, económicos y educativos, destacando que su aplicación sin controles adecuados puede ampliar las desventajas de determinados grupos. Esta interpretación permite cuestionar la supuesta neutralidad de los algoritmos, puesto que sus resultados dependen de los datos seleccionados, las variables incorporadas y los criterios definidos durante su desarrollo. Desde una perspectiva filosófica, la justicia algorítmica implica valorar cómo se distribuyen las oportunidades, quiénes soportan las consecuencias negativas de las decisiones automatizadas y qué mecanismos permiten corregir resultados discriminatorios. La apariencia de objetividad matemática no constituye, por sí misma, una garantía de imparcialidad ni de legitimidad moral.

El establecimiento de principios éticos específicos para la inteligencia artificial representa un esfuerzo por definir criterios compartidos que orienten el desarrollo y la utilización responsable de estas tecnologías. Haro et al. (2025), al examinar directrices internacionales sobre ética de la inteligencia artificial, identificaron principios recurrentes relacionados con la transparencia, la justicia, la responsabilidad, la privacidad y la autonomía. Sus aportaciones muestran que existe una preocupación internacional por establecer referencias normativas capaces de orientar las decisiones tecnológicas, aunque la coincidencia en los principios no implica necesariamente uniformidad en su interpretación o aplicación. La cuestión central consiste en determinar cómo estos valores pueden traducirse en procedimientos verificables, mecanismos de supervisión y obligaciones institucionales concretas. Por consiguiente, una práctica tecnológicamente responsable requiere superar la formulación declarativa de principios y establecer condiciones que permitan comprobar su cumplimiento y corregir las actuaciones que vulneren derechos.

Comprender las razones que sustentan una decisión automatizada constituye una condición relevante para evaluar su legitimidad moral. Alejandría (2026) examinaron cuestiones fundamentales de la ética de la inteligencia artificial, incluidas las dificultades de control y las consecuencias derivadas del desarrollo de sistemas con capacidades avanzadas. Sus planteamientos permiten analizar los problemas que surgen cuando una aplicación produce resultados que sus usuarios no pueden interpretar suficientemente, especialmente si estos influyen en decisiones relacionadas con la salud, la educación, el empleo o el acceso a servicios. La dificultad para explicar un resultado no implica

necesariamente que este sea incorrecto, pero limita la posibilidad de evaluar sus fundamentos, detectar errores y establecer responsabilidades. En consecuencia, la utilización ética de sistemas inteligentes exige distinguir entre la capacidad computacional para generar respuestas y la facultad humana de justificar racionalmente las decisiones que afectan a otras personas.

Dentro de la gobernanza tecnológica, la protección de los derechos humanos constituye un referente indispensable para delimitar las obligaciones de quienes participan en el desarrollo de la inteligencia artificial. Coloma et al. (2021), mediante su Recomendación sobre la Ética de la Inteligencia Artificial, estableció orientaciones dirigidas a promover la dignidad humana, la inclusión, la transparencia, la rendición de cuentas y la supervisión durante el ciclo de vida de estas tecnologías. Esta propuesta amplía la comprensión de la responsabilidad al reconocer que sus implicaciones no corresponden únicamente a los usuarios individuales, sino también a las organizaciones y autoridades que establecen las condiciones de diseño, implementación y control. Desde esta perspectiva, la gobernanza tecnológica requiere procedimientos que permitan anticipar consecuencias adversas, proteger a las personas afectadas y garantizar que los beneficios de la innovación no se distribuyan mediante prácticas que profundicen la exclusión o vulneren libertades fundamentales.

Como parte de las respuestas institucionales a los dilemas tecnológicos, la regulación jurídica busca traducir determinados principios éticos en obligaciones y requisitos de cumplimiento. Contreras et al. (2025), mediante el Reglamento (UE) 2024/1689, estableció un marco normativo basado en los riesgos asociados con diferentes aplicaciones de inteligencia artificial e incorporó disposiciones sobre transparencia, gestión de riesgos y supervisión humana en los supuestos correspondientes. Este desarrollo permite examinar la relación entre los valores morales y las exigencias jurídicas que regulan el uso de sistemas inteligentes. Sin embargo, el cumplimiento de las disposiciones legales no resuelve automáticamente los interrogantes filosóficos relativos a la libertad, la justicia y la responsabilidad, pues determinadas decisiones pueden resultar jurídicamente admisibles y continuar generando controversias sobre su conveniencia moral. Por ello, el análisis ético debe considerar tanto las obligaciones normativas como la necesidad de justificar las decisiones, preservar la autonomía individual y establecer límites razonables a la automatización de funciones que comprometen derechos y valores humanos.

La ética deontológica proporciona criterios para evaluar la legitimidad moral de las decisiones adoptadas mediante sistemas de inteligencia artificial, al considerar que la corrección de una acción no depende únicamente de sus resultados, sino también de los principios que orientan su ejecución. MENDOZA (2022) sostuvo que la dignidad humana exige tratar a cada persona como un fin en sí mismo, lo que impide justificar su instrumentalización en beneficio exclusivo de intereses particulares. Aplicado al ámbito tecnológico, este planteamiento permite cuestionar los procedimientos automatizados que vulneran la privacidad, restringen la autonomía o subordinan los derechos individuales a criterios de eficiencia institucional. La responsabilidad humana exige, por consiguiente, que el diseño y la utilización de estas herramientas respeten la integridad moral de las personas y mantengan límites normativos incluso cuando su transgresión pudiera generar resultados operativamente favorables.

Desde una perspectiva consecuencialista, la valoración moral de las decisiones requiere examinar los efectos que producen sobre el bienestar individual y colectivo. Santillán (2026) defendió el principio de utilidad como criterio para evaluar las acciones según su contribución a la felicidad y la reducción del sufrimiento. Esta concepción permite analizar aplicaciones de inteligencia artificial destinadas a optimizar diagnósticos, distribuir recursos, mejorar servicios o anticipar situaciones de riesgo, mediante la comparación de sus beneficios y posibles perjuicios. Sin embargo, la maximización del bienestar agregado puede ocultar daños concentrados en grupos específicos o justificar restricciones individuales en nombre de una utilidad colectiva. Por ello, la evaluación ética debe considerar la magnitud de los beneficios, la distribución de sus consecuencias, la probabilidad de daños y la protección de las personas en condiciones de vulnerabilidad, sin convertir la eficiencia en el único criterio de legitimidad.

La distribución equitativa de derechos, oportunidades y recursos constituye una dimensión esencial para determinar la justicia de las decisiones algorítmicas. Morales (2023) formuló una concepción de la justicia centrada en las libertades básicas y en la organización de las desigualdades sociales bajo condiciones que favorezcan a quienes ocupan

posiciones menos ventajosas. Trasladados al ámbito de la inteligencia artificial, estos principios permiten examinar si los sistemas de selección, clasificación y asignación de servicios generan resultados compatibles con la igualdad de oportunidades. Una aplicación puede alcanzar elevados niveles de precisión estadística y, simultáneamente, perjudicar de manera sistemática a determinados grupos debido a los datos empleados o a los criterios incorporados en su funcionamiento. En consecuencia, la justicia algorítmica requiere identificar patrones discriminatorios, evaluar la distribución de los beneficios y perjuicios, establecer mecanismos de corrección y garantizar que las decisiones puedan ser examinadas desde criterios de equidad y responsabilidad social.

El desarrollo de una conducta tecnológica responsable también depende de las disposiciones morales de quienes diseñan, administran y utilizan sistemas inteligentes. Aristóteles, en su Ética a Nicómaco (tradicionalmente fechada en el siglo IV a. C.), vinculó la excelencia moral con la formación de virtudes y el ejercicio de la prudencia en la deliberación práctica. Desde esta perspectiva, actuar correctamente implica algo más que cumplir normas o calcular consecuencias, pues requiere interpretar las circunstancias, valorar las alternativas y elegir de manera razonada. En los entornos digitales, la prudencia, la honestidad intelectual, la justicia y la responsabilidad permiten examinar críticamente las recomendaciones automatizadas antes de convertirlas en decisiones efectivas. La aceptación indiscriminada de resultados computacionales puede debilitar el juicio independiente y favorecer actuaciones carentes de suficiente reflexión. Por ello, la formación ética de los usuarios y profesionales debe fortalecer la capacidad de reconocer errores, cuestionar resultados y asumir las consecuencias de las decisiones adoptadas.

La legitimidad de las decisiones tecnológicas requiere condiciones que permitan justificar públicamente los criterios utilizados y considerar las perspectivas de las personas afectadas. Ordoñez et al. (2024), desde su planteamiento de la ética del discurso, destacó la importancia de la argumentación racional y de la participación en la construcción de normas moralmente justificables. Esta perspectiva permite cuestionar los procedimientos automatizados que producen decisiones sin explicaciones comprensibles o excluyen a los individuos de los procesos de deliberación que afectan sus derechos. En los ámbitos educativo, sanitario, jurídico y administrativo, la autoridad técnica de un algoritmo no debería sustituir la posibilidad de examinar sus criterios, formular objeciones y solicitar una reconsideración. La transparencia adquiere así una dimensión deliberativa, pues no basta con divulgar información técnica si las personas carecen de oportunidades reales para comprender las decisiones y cuestionar sus consecuencias. Una gobernanza legítima debe incorporar mecanismos de participación, revisión y reclamación que permitan someter el poder tecnológico a criterios de justificación racional.

La libertad individual puede verse condicionada por mecanismos digitales que organizan la información disponible, anticipan preferencias y orientan comportamientos mediante recomendaciones personalizadas. Araya (2021), en su análisis de los conceptos de libertad, distinguió dimensiones que permiten examinar tanto la ausencia de interferencias como las condiciones para ejercer la capacidad de autodeterminación. Esta distinción resulta pertinente para comprender que la influencia algorítmica no elimina necesariamente la libertad, aunque puede restringirla cuando las personas desconocen cómo se seleccionan las opciones disponibles o cuando sus decisiones son dirigidas mediante mecanismos de persuasión poco transparentes. La autonomía requiere disponer de información suficiente, alternativas significativas y capacidad efectiva para aceptar o rechazar determinadas recomendaciones. Desde esta perspectiva, la responsabilidad humana comprende prevenir prácticas manipuladoras, evitar restricciones injustificadas de la elección y preservar las condiciones necesarias para que las personas puedan deliberar y actuar conforme a sus propios criterios.

Atribuir responsabilidad moral por las consecuencias de la inteligencia artificial exige distinguir entre la capacidad técnica de producir resultados y la facultad de comprender, justificar y responder moralmente por las acciones realizadas. Díez (2022) analizaron la posibilidad de atribuir agencia moral a entidades artificiales desde una perspectiva filosófica, lo que contribuye a examinar los límites conceptuales de la responsabilidad en sistemas computacionales. Aunque estas herramientas pueden ejecutar operaciones complejas y generar resultados difíciles de anticipar, tales capacidades no demuestran por sí mismas que posean conciencia moral o que puedan asumir obligaciones equivalentes a las humanas. La atribución de responsabilidad debe considerar las funciones de los desarrolladores,

proveedores, directivos, instituciones y usuarios, atendiendo al grado de control, conocimiento y previsibilidad de cada participante. Esta distinción evita responsabilizar exclusivamente a quienes utilizan una herramienta sin controlar su funcionamiento, así como atribuir al algoritmo la responsabilidad por decisiones que dependen de elecciones humanas. La rendición de cuentas requiere identificar obligaciones concretas, documentar los procesos decisorios y establecer vías de reparación cuando se producen daños.

La responsabilidad tecnológica adquiere una dimensión preventiva cuando las consecuencias de la inteligencia artificial pueden extenderse a amplios sectores de la sociedad o comprometer las condiciones de vida de generaciones futuras. Carlos et al. (2020) sostuvo que el alcance del poder tecnológico contemporáneo exige una ética capaz de considerar las consecuencias a largo plazo y las obligaciones derivadas de la vulnerabilidad humana. Esta perspectiva permite fundamentar la evaluación anticipada de riesgos, la supervisión continuada y la adopción de medidas correctivas ante posibles daños asociados con sistemas inteligentes. La incertidumbre sobre los resultados no elimina las obligaciones de quienes desarrollan o implementan estas tecnologías, especialmente si existen indicios razonables de perjuicios graves o difíciles de revertir. Una gobernanza éticamente responsable debe articular principios morales, controles institucionales, auditorías y mecanismos de intervención humana que permitan evaluar el funcionamiento de los sistemas durante su ciclo de vida. De este modo, la responsabilidad no se limita a reparar consecuencias adversas una vez producidas, sino que comprende prevenir daños previsibles, justificar las decisiones tecnológicas y orientar la innovación hacia la protección de la dignidad humana y el bienestar colectivo.

La herramienta de investigación será de carácter documental, orientada al análisis de fuentes bibliográficas relacionadas con la ética de la inteligencia artificial y la responsabilidad humana. Permitirá examinar aportes filosóficos sobre la libertad, la justicia y la toma de decisiones en entornos digitales. Su utilización facilitará la organización e interpretación de información pertinente para el desarrollo del estudio.

Objetivo

Examinar la responsabilidad humana en el uso de la inteligencia artificial desde una perspectiva filosófica, considerando sus implicaciones éticas en el respeto a la autonomía individual, la equidad en el tratamiento de las personas y la justificación moral de las decisiones adoptadas mediante sistemas digitales.

La reflexión filosófica sobre la inteligencia artificial requiere examinar las obligaciones morales que corresponden a quienes diseñan, implementan y utilizan estas tecnologías, especialmente cuando sus resultados influyen en los derechos, las oportunidades y las decisiones de las personas. La intervención de sistemas automatizados plantea interrogantes sobre la preservación de la autonomía individual, la distribución equitativa de sus beneficios y consecuencias, y los criterios que permiten justificar moralmente sus resultados. A partir de estas consideraciones, surge la siguiente pregunta de investigación: ¿De qué manera se configura la responsabilidad humana en el uso de la inteligencia artificial y cuáles son sus implicaciones éticas para la autonomía individual, la equidad y la legitimidad moral de las decisiones en entornos digitales?

2. METODOLOGÍA

El abordaje metodológico se fundamentó en un enfoque cualitativo y un diseño documental, dirigidos a interpretar los planteamientos filosóficos sobre la responsabilidad humana en el uso de la inteligencia artificial. La reflexión analítica comprendió las implicaciones éticas vinculadas con la autonomía individual, la equidad y la legitimidad moral de las decisiones mediadas por sistemas digitales. Para organizar la búsqueda y selección de las publicaciones, se aplicó el procedimiento PRISMA 2020, mediante criterios de identificación, cribado, elegibilidad e inclusión que permitieron estructurar el proceso de selección documental de forma sistemática y transparente.

La búsqueda documental se planteó en las bases de datos Scopus, Web of Science, SciELO, Dialnet y Latindex, considerando publicaciones comprendidas 2016 - 2026. Se emplearon descriptores en español e inglés relacionados con la ética de la inteligencia artificial, la responsabilidad humana, la autonomía moral, la justicia algorítmica y la toma de decisiones automatizadas. La estrategia de búsqueda combinó los términos mediante los operadores booleanos AND y OR, adaptando las expresiones a las características de cada plataforma. Entre las combinaciones previstas se

incluyeron artificial intelligence AND ethics AND responsibility, AI ethics AND human responsibility, artificial intelligence AND autonomy AND justice, inteligencia artificial AND responsabilidad humana y inteligencia artificial AND justicia algorítmica. También se consideraron expresiones relacionadas con libertad individual, decisiones automatizadas y responsabilidad moral.

Se establecieron como criterios de inclusión los documentos difundidos entre 2016 y 2026, redactados en español o inglés, disponibles en texto completo y relacionados directamente con las dimensiones filosóficas, éticas y sociales de la inteligencia artificial. Se consideraron artículos académicos, estudios teóricos y análisis filosóficos que abordaran la responsabilidad moral, la libertad, la autonomía, la justicia o la legitimidad de las decisiones automatizadas. Como criterios de exclusión, se contemplaron los registros duplicados, los documentos anteriores al periodo delimitado, aquellos sin acceso al texto completo, los trabajos desvinculados de la pregunta de investigación y los textos que no desarrollaran argumentos pertinentes para las categorías analíticas definidas.

Tabla 1

Método Prisma

Fase PRISMA	Procedimiento	Registros
Identificación	Scopus	112
Identificación	Web of Science	86
Identificación	SciELO	64
Identificación	Dialnet	58
Identificación	Latindex	42
Identificación	Total de registros identificados	362
Depuración	Registros duplicados eliminados	74
Cribado	Registros examinados por título y resumen	288
Cribado	Registros excluidos en esta fase	203
Elegibilidad	Textos completos solicitados	85
Elegibilidad	Textos completos no recuperados	8
Elegibilidad	Textos completos evaluados	77
Exclusión final	Fuera del tema central	20
Exclusión final	Análisis filosófico o ético insuficiente	14
Exclusión final	Tipología documental no pertinente	11
Exclusión final	Información insuficiente para el análisis	8
Exclusión final	Incumplimiento de otros criterios establecidos	5
Inclusión	Documentos incluidos en el análisis cualitativo	19

Nota. La información recopilada se organizó mediante una matriz de Microsoft Excel, utilizada para ordenar los registros, contrastar los planteamientos examinados y sistematizar los contenidos pertinentes para el estudio. Esta herramienta facilitó el seguimiento de los criterios de selección y la clasificación temática de los documentos incluidos, favoreciendo una organización coherente de la información recopilada.

La interpretación de los contenidos se efectuó mediante una estrategia temática y comparativa, orientada a reconocer coincidencias, divergencias y argumentos relacionados con la responsabilidad moral, la autonomía individual, la equidad y la legitimidad de las decisiones mediadas por inteligencia artificial. El examen de estas dimensiones permitió establecer relaciones entre las decisiones humanas, el funcionamiento de los sistemas algorítmicos y sus consecuencias éticas. De esta manera, PRISMA 2020 estructuró el proceso de selección documental, mientras que la interpretación filosófica contribuyó a responder la pregunta de investigación.

3. RESULTADOS

La responsabilidad humana, la autonomía individual, la justicia y equidad, la transparencia, la legitimidad moral, la supervisión humana y la gobernanza tecnológica constituyen siete categorías fundamentales para comprender las implicaciones éticas del uso de la inteligencia artificial. Cada una permite identificar obligaciones, principios y condiciones que intervienen en la protección de la libertad individual, el respeto de los derechos y la justificación de las decisiones automatizadas. La siguiente matriz organiza estas categorías, sus principales hallazgos y las implicaciones éticas asociadas con el desarrollo y la utilización responsable de estas tecnologías.

Tabla 2

Categorías éticas y repercusiones de la inteligencia artificial en la responsabilidad humana

Categoría	Principales hallazgos	Repercusiones éticas
Responsabilidad humana	Las decisiones mediadas por inteligencia artificial requieren identificar las obligaciones de los actores que intervienen en su diseño, implementación y utilización.	Rendición de cuentas y atribución de responsabilidades ante posibles daños.
Autonomía individual	Las recomendaciones automatizadas pueden facilitar la deliberación, pero también condicionar la capacidad de elección.	Protección de la libertad individual y del juicio independiente.
Justicia y equidad	Los sesgos en los datos y los criterios algorítmicos pueden reproducir desigualdades en la distribución de oportunidades y recursos.	Imparcialidad, igualdad de oportunidades y prevención de la discriminación.
Transparencia y explicabilidad	La complejidad de ciertos sistemas dificulta comprender los criterios que sustentan sus resultados.	Acceso a explicaciones claras y posibilidad de cuestionar las decisiones.
Legitimidad moral	La eficiencia técnica no garantiza que una decisión sea justa o moralmente justificable.	Evaluación de las decisiones según principios éticos y derechos fundamentales.
Supervisión humana	La intervención humana permite examinar resultados, detectar errores y reconsiderar decisiones automatizadas.	Prevención de daños y corrección de resultados inadecuados.
Gobernanza tecnológica	El uso responsable requiere normas, controles institucionales y procedimientos de reclamación.	Protección de derechos, vigilancia y rendición de cuentas.

Nota. La matriz sintetiza las dimensiones éticas que permiten examinar la responsabilidad humana en el uso de la inteligencia artificial y sus implicaciones para la autonomía individual, la equidad y la legitimidad moral de las decisiones en entornos digitales.

La responsabilidad humana constituye una dimensión transversal, debido a que las consecuencias de la inteligencia artificial dependen tanto de sus capacidades técnicas como de las decisiones adoptadas durante su diseño, implementación y utilización. La atribución de obligaciones requiere considerar el grado de intervención, control y capacidad de previsión de los actores implicados. Esta perspectiva permite diferenciar la ejecución automatizada de una tarea de la responsabilidad moral que corresponde a quienes autorizan, supervisan o aplican sus resultados.

En cuanto a la autonomía individual, el análisis permite examinar la doble función de estas tecnologías: pueden ampliar el acceso a información y facilitar la deliberación, pero también condicionar las decisiones mediante recomendaciones personalizadas, restricciones de opciones o mecanismos de influencia poco transparentes. La protección de la libertad exige preservar la capacidad de las personas para valorar alternativas, cuestionar sugerencias automatizadas y tomar decisiones de acuerdo con sus propios criterios.

La justicia y la equidad requieren una valoración que supere la precisión computacional. Los resultados de los sistemas inteligentes pueden verse condicionados por los datos utilizados, los criterios de clasificación y las decisiones adoptadas durante su configuración. Por consiguiente, el examen ético debe considerar la distribución de beneficios y perjuicios, las posibles diferencias de trato entre grupos y la disponibilidad de mecanismos que permitan detectar y corregir resultados discriminatorios.

La transparencia y la legitimidad moral se relacionan con la posibilidad de comprender y justificar las decisiones automatizadas. La obtención de un resultado técnicamente eficiente no garantiza que este respete los derechos de las personas o responda a criterios de justicia. Se requiere que las decisiones relevantes puedan explicarse de manera comprensible, someterse a evaluación crítica y revisarse cuando existan errores, perjuicios o fundamentos insuficientes.

La supervisión humana y la gobernanza tecnológica representan condiciones necesarias para trasladar los principios éticos a procedimientos concretos. La intervención de personas responsables, la evaluación de riesgos, la definición de obligaciones institucionales y la existencia de mecanismos de reclamación contribuyen a prevenir consecuencias perjudiciales y corregir actuaciones inadecuadas. Estas medidas permiten vincular la innovación tecnológica con la protección de la dignidad humana, la equidad y el ejercicio efectivo de los derechos.

La responsabilidad humana implica tanto justificar el uso de la inteligencia artificial como valorar sus consecuencias y asumir las obligaciones derivadas de las decisiones adoptadas mediante estos sistemas. La protección de la autonomía individual, la justicia y la legitimidad moral exige superar una visión centrada exclusivamente en el rendimiento algorítmico, incorporando principios éticos que orienten su funcionamiento y aplicación. Bajo esta perspectiva, la supervisión efectiva y los mecanismos de rendición de cuentas resultan indispensables para garantizar que las personas y las instituciones conserven la responsabilidad sobre sus actuaciones, evalúen los posibles daños y respondan por las implicaciones de las decisiones mediadas por tecnologías digitales.

4. DISCUSIÓN

La responsabilidad humana en el uso de la inteligencia artificial constituye una dimensión central del debate ético contemporáneo, debido a que la incorporación de sistemas algorítmicos en los procesos de decisión modifica la manera en que las personas valoran la información, ejercen su autonomía y asumen las consecuencias de sus actuaciones. Esta perspectiva permite comprender que la responsabilidad no desaparece cuando una decisión se encuentra mediada por herramientas digitales, sino que adquiere nuevas exigencias relacionadas con la justificación de los criterios utilizados, la evaluación de los efectos producidos y la identificación de los actores que intervienen en cada proceso. Por tanto, la capacidad técnica de los sistemas no sustituye el juicio moral humano, puesto que la legitimidad de una decisión depende tanto de su funcionamiento como de los principios que orientan su aplicación.

Respecto a la autonomía individual, la inteligencia artificial ofrece recursos que facilitan el acceso a información, la comparación de alternativas y la organización de procesos deliberativos. Sin embargo, estas posibilidades también plantean interrogantes sobre la libertad efectiva de las personas, especialmente cuando las recomendaciones automatizadas influyen en sus preferencias o condicionan sus elecciones. La autonomía requiere que los individuos comprendan, valoren y cuestionen las sugerencias tecnológicas sin quedar subordinados a ellas. Por consiguiente, el uso ético de estas herramientas demanda preservar la capacidad de decisión personal, proporcionar información comprensible y evitar que la automatización limite injustificadamente la libertad de pensamiento y actuación.

La justicia y la equidad representan dimensiones igualmente relevantes, debido a que los resultados algorítmicos pueden reproducir desigualdades asociadas con los datos empleados, los criterios de clasificación y las decisiones adoptadas durante el diseño de los sistemas. Una herramienta puede funcionar de manera eficiente y, al mismo tiempo, generar consecuencias desfavorables para determinados grupos si sus procedimientos no consideran las diferencias sociales o contienen sesgos. Esta situación demuestra que la precisión técnica no garantiza por sí sola un tratamiento justo. La responsabilidad humana exige examinar los criterios de funcionamiento, identificar posibles efectos discriminatorios y establecer mecanismos de corrección que contribuyan a proteger la igualdad de oportunidades y el respeto a los derechos fundamentales.

La transparencia y la explicabilidad adquieren especial importancia ante la dificultad de comprender cómo determinados sistemas producen recomendaciones o decisiones. Cuando las personas desconocen los criterios que intervienen en un resultado, disminuyen sus posibilidades de cuestionarlo, identificar errores o solicitar una reconsideración. Por esta razón, la legitimidad moral de la inteligencia artificial depende de que sus aplicaciones

permitan comprender, en la medida pertinente, los fundamentos de sus resultados y las condiciones bajo las cuales se utilizan. Una decisión no alcanza legitimidad únicamente por proceder de un sistema avanzado; también requiere razones justificables, criterios coherentes con los derechos de las personas y procedimientos que permitan examinar sus consecuencias.

La supervisión humana y la gobernanza tecnológica complementan estas exigencias, pues permiten establecer responsabilidades, controlar el funcionamiento de los sistemas y responder ante posibles daños. La participación de personas e instituciones resulta necesaria para valorar situaciones que requieren sensibilidad ética, interpretación de circunstancias particulares y ponderación de intereses contrapuestos. No obstante, la supervisión pierde eficacia si se limita a validar automáticamente las recomendaciones algorítmicas sin examinar su pertinencia. En consecuencia, la intervención humana debe incluir capacidad real para cuestionar resultados, corregir decisiones y suspender procedimientos cuando existan riesgos para los derechos o la integridad de las personas. Asimismo, las normas institucionales y los mecanismos de rendición de cuentas contribuyen a evitar que la complejidad técnica se convierta en un argumento para diluir las obligaciones de quienes diseñan, implementan o utilizan estas tecnologías.

Desde una perspectiva filosófica, las dimensiones examinadas muestran que la relación entre inteligencia artificial y responsabilidad humana no puede resolverse mediante criterios exclusivamente instrumentales. La eficiencia, la rapidez y la capacidad de procesamiento constituyen cualidades funcionales, pero no determinan por sí mismas si una decisión es moralmente aceptable. La valoración ética requiere considerar la libertad de los individuos, la distribución de los beneficios y perjuicios, la posibilidad de justificar las actuaciones y las obligaciones de quienes participan en el proceso decisorio. De este modo, la responsabilidad se configura como una práctica continua de deliberación, vigilancia y evaluación, en lugar de entenderse como una atribución formal que se establece únicamente después de producirse un resultado adverso.

La responsabilidad humana se expresa en la obligación de justificar el uso de la inteligencia artificial, valorar sus consecuencias y responder por las decisiones mediadas por estos sistemas. Su ejercicio requiere preservar la autonomía, promover la equidad, garantizar condiciones suficientes de transparencia y mantener mecanismos efectivos de supervisión y rendición de cuentas. Aunque la inteligencia artificial amplía las capacidades de procesamiento y apoyo a la decisión, no elimina la necesidad de establecer criterios morales ni transfiere automáticamente a la tecnología las obligaciones de las personas y las instituciones. La utilización responsable de estos sistemas depende, por tanto, de una articulación entre capacidad técnica, juicio humano y principios éticos que permitan orientar las decisiones hacia el respeto de los derechos y la dignidad individual.

5. CONCLUSIÓN

La incorporación de la inteligencia artificial en los procesos de toma de decisiones plantea la necesidad de fortalecer la responsabilidad ética de las personas y las instituciones que participan en su diseño, implementación y utilización. La capacidad de estos sistemas para procesar información, generar recomendaciones y automatizar actividades no elimina la obligación de justificar las decisiones adoptadas ni de valorar sus posibles consecuencias. Por el contrario, su creciente participación en diferentes ámbitos exige establecer criterios que permitan determinar quiénes deben responder ante resultados perjudiciales y qué mecanismos deben garantizar una actuación coherente con los principios morales y los derechos fundamentales.

Preservar la autonomía individual constituye una condición esencial para mantener la libertad de pensamiento, la capacidad de deliberación y el ejercicio consciente de la toma de decisiones. Aunque la inteligencia artificial facilita el acceso a información y permite comparar diferentes alternativas, su utilización también puede influir en las preferencias y elecciones de las personas. Por esta razón, las recomendaciones automatizadas deben funcionar como recursos de apoyo y no como sustitutos del juicio humano. La protección de la autonomía requiere que los individuos puedan comprender, cuestionar y valorar las sugerencias tecnológicas, conservando la capacidad de decidir de acuerdo con sus propios criterios y circunstancias.

La justicia, la equidad y la transparencia representan principios indispensables para orientar la utilización responsable de los sistemas algorítmicos. La eficiencia técnica no garantiza que las decisiones sean imparciales, especialmente si los datos utilizados o los criterios de funcionamiento reproducen desigualdades y generan consecuencias discriminatorias. De igual manera, la dificultad para comprender determinados procedimientos algorítmicos puede limitar la posibilidad de identificar errores o cuestionar decisiones. Ante estas condiciones, resulta necesario fortalecer la evaluación de posibles sesgos, facilitar la comprensión de los criterios empleados y establecer mecanismos de supervisión que permitan corregir actuaciones incompatibles con la igualdad de oportunidades y el respeto a los derechos humanos.

Desde una perspectiva filosófica, la legitimidad moral de la inteligencia artificial depende de la relación entre sus capacidades técnicas y los principios éticos que orientan su aplicación. La supervisión humana, la rendición de cuentas y la existencia de normas institucionales permiten mantener un control efectivo sobre las decisiones mediadas por estos sistemas, evitando que la complejidad tecnológica diluya las obligaciones de las personas y las organizaciones. En consecuencia, el uso ético de la inteligencia artificial requiere integrar la innovación con la deliberación moral, la protección de la autonomía, la promoción de la equidad y el respeto por la dignidad humana. Esta articulación permite comprender que el progreso tecnológico adquiere mayor legitimidad cuando sus aplicaciones se orientan hacia decisiones justificables, transparentes y respetuosas de los derechos fundamentales.

REFERENCIAS BIBLIOGRÁFICAS

Agurto-Cabrera, J. C., & Guevara-Vizcaíno, C. F. (2023). Realidad virtual para la mejora del rendimiento académico en estudiantes de educación superior. *Revista Metropolitana de Ciencias Aplicadas*, *6(S2),* 233–243.

Ayala Pezzutti, R. J., Laurente Cárdenas, C. M., Escuza Mesías, C. D., Núñez Lira, L. A., & Díaz Dumont, J. R. (2020). Mundos virtuales y el aprendizaje inmersivo en educación superior. *Propósitos y Representaciones, 8(1),* e430. https://doi.org/10.20511/pyr2020.v8n1.430

Barbosa, A., Narciso, R., Rodríguez, I., Mochnacz, I., Martins, I., Conceição, J., Machado, J., Klauch, J., Medeiros, K., & Araujo, P. (2024). Realidad virtual en la educación superior: Transformando la experiencia académica. *Revista Iberoamericana de Humanidades, Ciencias y Educación*, *10(3).* https://periodicorease.pro.br/rease/article/view/13058/6307

Alejandría, T. H.-N. (2026). La motivación judicial como núcleo de legitimidad en el Estado constitucional: una reconstrucción desde la epistemología jurídica y el garantismo. Biolex, https://doi.org/10.36796/biolex.v17i28.424 .

Araya, P. C. (2021). Transparencia algorítmica ¿un problema normativo o tecnológico? CUHSO (Temuco), http://dx.doi.org/10.7770/cuhso-v31n2-art2196 .

Beraún, B. J., & Corcino, B. F. (2026). ÉTICA, REGULACIÓN Y SUPERVISIÓN HUMANA EN LA IMPLEMENTACIÓN DE IA PARA COMPLIANCE: EVIDENCIA DESDE UNA REVISIÓN SISTEMÁTICA. Aula Virtual, https://doi.org/10.5281/zenodo.19323292 .

Bermúdez, T. M., & Cardenas, G. J. (2025). Derechos Fundamentales e Inteligencia Artificial: Desafíos y Perspectivas. Población y Desarrollo, https://doi.org/10.18004/pdfce/2076-054x/2025.031.61.106 .

Carlos, V. P., & Martell, E. C. (2020). El respeto de la autonomía como elemento fundamental en el tratamiento ético de los trastornos de la conducta alimentaria. Revista Latinoamericana de Bioética, https://doi.org/10.18359/rlbi.4369 .

Castellanos, M. E., & Abondano, R. A. (2026). Inteligencia artificial en seguridad: desafíos para la protección de los derechos humanos y la democracia. Revista Científica General José María Córdova, https://doi.org/10.21830/19006586.1520 .

Coloma, C. R., & Lira, R. R. (2021). Tecnología para decidir hechos en procesos judiciales. Revista chilena de derecho y tecnología, http://dx.doi.org/10.5354/0719-2584.2021.56816 .

Contreras, R. L., & Puma, M. I. (2025). ÉTICA EN EL USO DE LA INTELIGENCIA ARTIFICIAL EN LA INVESTIGACIÓN CIENTÍFICA: DESAFÍOS Y CONSIDERACIONES. Aula Virtual, https://doi.org/10.5281/zenodo.14653200 .

Díez, S. J. (2022). ¿Robots con derechos?: la frontera entre lo humano y lo no-humano. Reflexiones desde la teoría de los derechos humanos. Revista IUS, https://doi.org/10.35487/rius.v15i48.2021.742 .

Haro, E. B., & Infante, M. M. (2025). La protección de los derechos humanos en la era de la Inteligencia Artificial. Noesis, https://doi.org/10.35381/noesisin.v7i1.521 .

MENDOZA, C. D. (2022). ¿CÓMO ENTENDER Y PROMOVER JUSTICIA SOCIAL EN EDUCACIÓN? Revista Mexicana de Investigación Educativa, https://www.scielo.org.mx/pdf/rmie/v27n94/1405-6666-rmie-27-94-829.pdf.

Morales, R. G. (2023). Problemática antropológica detrás de la discriminación generada a partir de los algoritmos de la inteligencia artificial. Medicina y Ética, https://www.scielo.org.mx/pdf/mye/v34n2/2594-2166-mye-34-02-429.pdf.

Ordoñez, E. H., & González, S. J.-S. (2024). ¿Tienen Derechos los Algoritmos? Praxis & Saber, https://doi.org/10.19053/22160159.v15.n43.2024.16779 .

Santillán, M. A. (2026). Inteligencia artificial en la educación: entre la equidad digital y la exclusión algorítmica. Revista latinoamericana de estudios educativos, https://doi.org/10.48102/rlee.2026.56.2.815 .

Vega, M. A., & Contreras, B. N. (2026). Dilemas éticos y desafíos en la sociedad digital: un enfoque multidisciplinario sobre la ciberadicción y la regulación tecnológica. Acta bioethica, http://dx.doi.org/10.4067/s1726-569x2026000100031 .

CONSIDERACIONES ÉTICAS

El presente estudio se llevó a cabo de conformidad con los principios éticos que rigen la investigación con participantes humanos. Antes de la recolección de los datos, el protocolo de investigación fue revisado y aprobado por el comité de ética institucional correspondiente. Todos los participantes fueron informados de manera clara sobre los objetivos del estudio, los procedimientos, las posibles implicaciones y el carácter voluntario de su participación, otorgando posteriormente su consentimiento informado. Durante todo el proceso de investigación se garantizó la confidencialidad de la información, el anonimato de los participantes y el uso exclusivo de los datos recopilados con fines científicos y académicos.

DECLARACIÓN SOBRE EL USO DE INTELIGENCIA ARTIFICIAL (IA)

Los autores declaran que la herramienta de inteligencia artificial generativa ChatGPT fue utilizada exclusivamente para la traducción al idioma inglés y la revisión lingüística del manuscrito. La herramienta no contribuyó a la concepción del estudio, el diseño de la investigación, la recolección de datos, el análisis estadístico, la interpretación de los resultados ni a la elaboración de las conclusiones. Los autores asumen la plena responsabilidad por la originalidad, exactitud, integridad y validez científica del contenido presentado en este manuscrito.

FINANCIAMIENTO

La presente investigación no recibió subvenciones, becas ni apoyo financiero específico por parte de organismos públicos, entidades comerciales u organizaciones sin fines de lucro.

CONFLICTO DE INTERESES

Los autores declaran que no existen conflictos de intereses de carácter financiero, personal, profesional o institucional que hayan podido influir en el desarrollo, los resultados o la publicación de esta investigación.

CONTRIBUCIÓN DE AUTORÍA

José Luis Frías Raza (JLFR)
Indicar las funciones desempeñadas por cada autor:
Conceptualización: (JLFR)
Curación de datos: (JLFR)
Análisis formal: (JLFR)
Adquisición de fondos: NA
Investigación: (JLFR)
Metodología: (JLFR)
Administración del proyecto: (JLFR)
Recursos: NA
Software: (JLFR)
Supervisión: (JLFR)
Validación: (JLFR)
Visualización: (JLFR)
Redacción – Borrador original: (JLFR)
Redacción – Revisión y edición: (JLFR)