

Dilemas éticos contemporáneos en los sistemas de inteligencia artificial: perspectivas sobre la responsabilidad y los límites humanos

Contemporary ethical dilemmas in artificial intelligence systems: a perspective on responsibility and human limits

Dilemas dilemas éticos contemporâneos nos sistemas de inteligência artificial: perspectivas sobre a responsabilidade e as limitações humanas

Kirenia Chaveco Asin, <https://orcid.org/0009-0007-8765-4321>

Melissa Miniet Alberja, <https://orcid.org/0009-0008-1234-5678>

Universidad de Oriente, Facultad de Ciencias Sociales, Departamento de Filosofía, Santiago de Cuba, Cuba

*Autor para correspondencia: kirenia0404@gmail.com

RESUMEN

La expansión vertiginosa de las Tecnologías de la Información y la Comunicación (TIC) plantean dilemas éticos que trascienden el ámbito tecnológico y se inscriben en la reflexión sobre la condición humana. Este artículo analiza las tensiones que emergen de la relación entre el desarrollo tecnológico responsable y la autonomía del sujeto social. A través de una metodología cualitativa donde se combina el método hermenéutico y el método de revisión documental, se examina cómo la delegación del juicio en sistemas automatizados puede limitar la capacidad deliberativa de los seres humanos. El resultado final muestra cómo los sistemas de inteligencia artificial han provocado transformaciones significativas en diversos ámbitos de la sociedad; sin embargo, la falta de mecanismos regulatorios en su implementación ha dado lugar a una crisis de valores que compromete sus potenciales beneficios, generando impactos adversos en la autonomía de quienes la utilizan y en la responsabilidad de quienes la diseñan. Esto exige repensar los criterios éticos de su funcionamiento y proponer un enfoque que vaya más allá de la eficiencia técnica, procurando alternativas que promuevan tecnologías sostenibles.

Palabras clave: Inteligencia artificial, dilemas éticos, condición humana, autonomía, tecnologías sostenibles.

ABSTRACT

The rapid expansion of information and communication technologies raises ethical dilemmas that transcend the technological sphere and are rooted in reflections on the human condition. This article analyses the tensions that arise from the relationship between responsible technological development and individual autonomy. Using a qualitative methodology that combines hermeneutics and documentary review, it examines how delegating judgement to automated systems can limit the deliberative capacity of humans. The final result shows how artificial intelligence systems have caused significant transformations in various areas of society. However, the lack of regulatory mechanisms in their implementation has given rise to a crisis of values that compromises their potential benefits and generates adverse impacts on the autonomy of those who use them and the social responsibility of those who design them. This requires us to reconsider the ethical criteria for their operation and propose an approach that goes beyond technical efficiency and promotes sustainable technologies.

Keywords: Artificial intelligence, ethical dilemmas, human condition, autonomy, sustainable technologies.

RESUMO

A expansão vertiginosa das Tecnologias da Informação e Comunicação (TIC) levanta dilemas éticos que transcendem o âmbito tecnológico e se enquadram na reflexão sobre a condição humana. Este artigo analisa as tensões que emergem da relação entre o desenvolvimento tecnológico responsável e a autonomia do sujeito social. Através de uma metodologia qualitativa que combina o método hermenêutico e o método de revisão documental, é examinado como a delegação do julgamento em sistemas automatizados pode limitar a capacidade deliberativa dos seres humanos. Os resultados finais mostram como os sistemas de inteligência artificial provocaram transformações significativas em diversos setores da sociedade; no entanto, a falta de mecanismos regulatórios na sua implementação deu origem a uma crise de valores que compromete os seus potenciais benefícios, gerando impactos adversos na autonomia dos utilizadores e na responsabilidade social dos designers. Tal exige repensar os critérios éticos do seu funcionamento e propor um enfoque que vá além da eficiência técnica, procurando alternativas que promovam tecnologias sustentáveis.

Palavras-chave: Inteligência artificial, dilemas éticos, condição humana, autonomia, tecnologias sustentáveis.

Recibido: 22/3/2025 Aprobado: 12/4/2026

Introducción

En la medida que los sistemas de inteligencia artificial se vuelven más ubicuo en la sociedad, su impacto se percibe de manera directa en las relaciones interpersonales. Estos sistemas tecnológicos capaces de aprender y de resolver problemas, condición que era exclusiva de la humanidad, ahora colocan en tela de juicio los conceptos de autonomía y de responsabilidad. Pudiéramos decir que, la existencia y la esencia humana están siendo moldeadas por algoritmos, que simulan las funciones cognitivas y estimulan el pensamiento utilitarista. Por tanto, su impacto no se limita a la automatización de tareas o al procesamiento de datos, también se expresa en la construcción de la subjetividad. En este sentido, Haraway (1991) hace alusión a la figura del ciborg como símbolo de una subjetividad híbrida que desdibuja las fronteras entre lo humano, lo animal y lo mecánico. Braidotti (2019) continúa esta línea al conceptualizar al sujeto posthumano como un nodo dinámico de afectos, tecnologías y discursos. Cada autor ofrece una perspectiva distinta sobre la condición humana, al pensar la subjetividad desde una mirada relacional y tecnológicamente mediada.

Los sistemas de inteligencia artificial están diseñados para reconfigurarse mediante el procesamiento de datos derivados de la experiencia humana, permitiendo la ejecución de tareas de forma autónoma. Si bien esta arquitectura técnica puede ser conveniente, también crea burbujas de filtros y cámaras de ecos (Pariser, 2017). Estos fenómenos impactan en la alteridad y restringen la exposición del usuario a perspectivas divergentes. En consonancia con Floridi (2023), los agentes artificiales procesan información (sintaxis), pero no producen conocimiento significativo (semántica). Esta limitación ontológica inaugura una crisis en la autodeterminación del sujeto, la cual se agudiza ante la proliferación de interfaces conversacionales que fomentan vínculos afectivos y difuminan la frontera entre lo humano y lo artificial.

Al respecto, la crítica de Virilo (2007) sobre la dromología o aceleración tecnológica, revela cómo el individuo se inserta en un régimen de dependencia que erosiona su autonomía funcional. Dicha perspectiva converge con la tesis de Heidegger (1994) sobre la técnica como Estructura de Emplazamiento (Gestell), un modo de desocultamiento que reduce al ser a una "existencia" o recurso disponible. Por su parte, Han (2012) desplaza el análisis hacia la autoexplotación en la era digital. González y Martínez (2020) expresan que los sistemas inteligentes operan como dispositivos donde los dilemas éticos se intensifican y se tornan más visibles. En última instancia, la convergencia de estos enfoques revela que la tecnología constituye una matriz, cuyas manifestaciones amenazan la emancipación del sujeto contemporáneo. Esta genealogía proporciona el marco teórico necesario para reformular la tríada entre tecnología, autonomía y responsabilidad ética en la sociedad-red.

Las vertiginosas transformaciones tecnológicas de las últimas décadas, han suscitado una amplia reflexión interdisciplinaria en torno a sus implicaciones éticas. Autores como Bostrom (2009) advierten que las innovaciones disruptivas modifican las condiciones materiales de existencia y transforman la naturaleza humana. Por su parte, Beheshti y Kerridge (2025) sostienen que la inteligencia artificial plantea dilemas

vinculados a la delegación de decisiones en sistemas algorítmicos, que generan incertidumbres sobre la responsabilidad ética en caso de error o sesgo. Porcelli (2020) analiza como la creciente digitalización fusiona los mundos físicos, digitales y biológico, generando dilemas éticos sobre la esencia humana. En este sentido, el debate filosófico exige pensar críticamente las condiciones para un desarrollo humano auténtico, donde la tecnología sea medio y no fin.

Los sistemas inteligentes han pasado de ser una herramienta externa para convertirse en la urdimbre de la vida cotidiana. En este punto es necesario volver a Descartes (1637) con su célebre cogito ergo sum, donde sitúa a la razón humana como la fuente primaria de conocimiento y fundamento del ser. Por su parte, Kant (1785) define al sujeto como un ser dotado de la capacidad racional y moral. Esta noción del ser humano permite cuestionar el sentido de la vida en un mundo mediado por el desarrollo tecnológico. Es válido señalar, que los dilemas éticos resultantes de la implementación de la inteligencia artificial, es uno de los problemas filosóficos más complejos de los tiempos contemporáneos que no admiten respuestas simples.

En el contexto de América Latina, los sistemas de inteligencia artificial configuran un horizonte de posibilidades y tensiones que no pueden ser ignoradas. Por un lado, su impacto positivo se evidencia en la capacidad de contribuir al desarrollo social mediante la optimización de procesos y la generación de soluciones innovadoras. No obstante, los riesgos emergen cuando la tecnología no se concibe como un medio orientado a la liberación, tal como lo advierte Dussel (1998). En esta dirección, Tello (2023) introduce el concepto de colonialismo digital para referirse a la tecnología como un instrumento de dominación del poder y del saber. Se reconoce que el verdadero dilema no es la fascinación por el avance de la tecnología, sino la orientación ética de sus fines.

Este artículo tiene como propósito examinar las tensiones que surgen en la relación entre el desarrollo tecnológico responsable y la autonomía del sujeto social, destacando cómo los avances en innovación pueden coexistir con la necesidad de preservar la capacidad crítica, la participación activa y la autodeterminación de las personas en los distintos ámbitos de la vida social.

Metodología

El artículo que se presenta es de investigación, se adopta un enfoque cualitativo fundamentado en la convergencia de tres métodos principales que permiten abordar la complejidad del objeto de estudio. El hermenéutico, se empleó para la interpretación de los textos clásicos y contemporáneos, su pertinencia radica en la necesidad de trasladar los principios éticos fundamentales al contexto digital actual, permitiendo comprender cómo la técnica redefine categorías esenciales como la dignidad y la autonomía. El analítico-sintético, permitió en una primera fase, descomponer la complejidad de los sistemas de inteligencia artificial en sus dimensiones constitutivas; posteriormente, se procedió a la síntesis de estos elementos para articular una visión integral sobre la crisis de responsabilidad y la reconfiguración de las relaciones intersubjetivas.

La revisión documental constituyó la base empírica del estudio. Para ello, se seleccionaron y analizaron fuentes bibliográficas junto con evidencias de funcionamiento en plataformas digitales. La búsqueda se realizó en bases de datos académicas de amplio reconocimiento, como Scopus, SciELO y otras, lo que garantizó la calidad, pertinencia y actualidad de la información recopilada. La inclusión de estos materiales permitió contrastar los dilemas teóricos con fenómenos verificables de la praxis tecnológica contemporánea. Esta estrategia metodológica se complementó con un enfoque ético, orientado a asegurar la transparencia en la selección de fuentes, el respeto a la propiedad intelectual y la responsabilidad en la interpretación de los hallazgos.

El apartado ético de la investigación se centra en la construcción de una prospectiva que no solo atienda los desafíos tecnológicos, sino que también promueva un marco de acción integral comprometido con la equidad, la inclusión y la autonomía del sujeto social, evitando sesgos y garantizando que las propuestas derivadas contribuyan al bienestar colectivo.

Resultados y discusión

Dado el carácter reflexivo de la investigación, los hallazgos se expresan como aportaciones conceptuales y argumentativas. Se evidencia que los sistemas de inteligencia artificial (IA) han reconfigurado el campo de la ética, situando a la filosofía ante escenarios donde los principios tradicionales son tensionados por la

complejidad tecnológica. En este punto, resulta imperativo establecer una distinción epistemológica entre la autonomía técnica y la autonomía moral. Mientras que la primera, refiere a la capacidad operativa de un sistema para ejecutar tareas complejas sin intervención humana directa, la segunda, fundamentada en la tradición kantiana, alude a la facultad de un sujeto para dictar sus propias leyes bajo el imperativo de la razón (Kant, 1785).

Bajo esta premisa, el dilema ético plantea una contradicción que se expresa entre la autonomía operativa y la autonomía moral. Ahora bien, los sistemas de inteligencia artificial no tienen la capacidad de reconocer la alteridad, ni de asumir la dimensión ética de sus resultados. Al respecto, Bustamante-Cabrera (2024) sostiene que atribuir autonomía plena a la IA, induce a una confusión conceptual que diluye la responsabilidad humana. Es precisamente esta distinción entre la capacidad de procesamiento y la sensibilidad moral, lo que define el límite infranqueable entre la decisión automatizada frente al juicio humano.

Esta incapacidad de los sistemas para asumir una autonomía moral no impide que operen como dispositivos de poder (Foucault, 1998). Por el contrario, se manifiesta una crisis estructural en la que, las herramientas digitales actúan como dispositivos de subjetivación, interviniendo activamente en la identidad y la percepción del sujeto. Por consiguiente, pudiéremos referirnos a un gobierno algorítmico (Ritzer & Degli Esposti, 2020), que se ha consolidado en el ecosistema de las relaciones sociales configurando la experiencia de los usuarios de manera compleja y, a menudo invisible.

La delegación sistemática del juicio en asistentes virtuales genera una erosión de la capacidad deliberativa. Al predecir comportamientos humanos, estos sistemas fomentan un régimen de dependencia donde la autonomía es desplazada por una heteronomía técnica; en consecuencia, el individuo deja de operar como un sujeto soberano para constituirse en un sujeto producido algorítmicamente. Este escenario es conceptualizado por Zuboff (2019) como capitalismo de vigilancia, un modelo donde la interacción cotidiana en plataformas como Facebook, Instagram, TikTok, es extraída y procesada para generar modelos predictivos de comportamiento con fines comerciales y de control.

Paralelamente, la mediación de asistentes virtuales como Gemini, Bixby, Siri y Alexa y el auge de los chatbots de acompañamiento emocional no solo optimizan tareas, sino que están diseñados para modificar los estilos comunicativos, naturalizando la simbiosis entre humano y máquina. Se advierte, por tanto, una crisis de valores donde la técnica trasciende su función instrumental para consolidarse como una tecnología de subjetivación. En este sentido, la tecnología pasa de ser una herramienta, para convertirse en una fuerza capaz de moldear las estructuras internas del pensamiento y la voluntad, redefiniendo la esencia misma de la agencia humana en la era digital.

Sobre esta base, se observa que múltiples sistemas de IA incorporan en su diseño modelaciones algorítmicas para bloquear contenidos que contravengan sus políticas. No obstante, cabe cuestionar si estas restricciones responden genuinamente a un código de ética universal o a mecanismos de gobernanza privada. Esta insuficiencia se agrava con la persistencia de sesgos raciales y sociales, como se corroboró en el estudio de Angwin *et al.*, (2016) sobre el sistema COMPAS. Weizenbaum (1976) ya advertía que existen decisiones que no deberían delegarse a sistemas que carecen de la capacidad de sentir y valorar.

Llegado a este punto, la autonomía se descubre compartida con sistemas que intervienen en toma de decisiones, lo que exige recuperar la ética de la responsabilidad de Jonas (1995). El debate se amplía hacia la distribución de la responsabilidad, cuando un algoritmo incurre en un acto de discriminación, ¿recae la falta en el programador, la empresa o la sociedad? Según Rivera *et al.* (2025), esta fragmentación genera un vacío ético que dificulta la imputación moral. Desde el utilitarismo (Bentham, 1789; Mill, 1863), se privilegia la optimización de recursos para el bienestar colectivo, pero este enfoque falla al medir dimensiones cualitativas como la justicia.

Finalmente, la crítica decolonial (Quijano, 2000; Mignolo, 2011) resulta esencial para visibilizar cómo los algoritmos reproducen hegemonías epistémicas. Al ser desarrollados mayoritariamente, en centros de poder global, imponen una visión del mundo unidimensional. Ricourte (2019) expone que la arquitectura algorítmica actual, impone una racionalidad única que invisibiliza los saberes locales y las realidades del Sur Global, ejerciendo una violencia simbólica que estandariza la existencia bajo parámetros noratlántico. Por

consiguiente, el debate ético no debe limitarse a la corrección técnica de los sesgos, debe orientarse hacia una decolonización de la matriz del poder tecnológico. Esto refuerza la necesidad de construir tecnologías inclusivas y situadas, que promuevan una soberanía digital, capaz de conocer la pluralidad de saberes y responder a las urgencias de los territorios donde se despliegan.

Ante la urgencia de trascender los dilemas éticos que emergen en la implementación de los sistemas de inteligencia artificial, y orientar la praxis tecnológica a marcos axiológicos efectivos, se proponen líneas estratégicas para el debate y acción ética, concebidas como esquemas de acción para el desarrollo humano:

☒ Transitar de la opacidad de los sistemas de inteligencia artificial que figuran como cajas negras hacia modelos donde la trazabilidad sea la norma. Esto implica establecer protocolos que permitan la fiscalización de los criterios de ponderación, garantizando que toda la decisión automatizada sea auditable y sujeta a impugnación humana.

☒ La dimensión ética debe integrarse como un requisito técnico desde la fase de arquitectura del sistema. Esto supone la creación de entornos de desarrollo transdisciplinarios donde la identificación de sesgos potenciales y la evaluación del impacto sean variables determinantes en la validación del software.

☒ En respuesta al colonialismo de datos, se requiere el impulso de algoritmos situados. Esta línea propone el uso de conjuntos de datos locales y representativos que eviten la transferencia acrítica de sesgos globales y promuevan una gobernanza alineada con las realidades del contexto regional.

☒ Se debe ratificar que la responsabilidad es una cualidad humana indelegable. Ninguna decisión crítica sobre la integridad o los derechos de los sujetos debe ser definitiva sin la mediación de un agente moral capaz de asumir las consecuencias de tales actos.

☒ La acción ética demanda la formación de una ciudadanía capaz de decodificar las estructuras del gobierno algorítmico. Solo mediante la comprensión crítica de estos dispositivos es posible mitigar la dependencia ontológica y restaurar la capacidad de juicio deliberativo en la sociedad contemporánea.

CONCLUSIONES

El análisis permite concluir que los sistemas de inteligencia artificial, representan una innovación tecnológica que nos sitúan ante un desafío ontológico y ético, que tensionan los fundamentos de la modernidad. Además, se ratifica la irreductibilidad de la autonomía moral frente a la autonomía técnica. La investigación demuestra que, aunque los sistemas algorítmicos poseen capacidad de procesamiento y ejecución superiores a las humanas en tareas específicas, carecen de sensibilidad axiológica necesaria para asumir la responsabilidad de sus actos. Por tanto, delegar decisiones de orden ético o social a dispositivos de caja negra no constituye un avance, sino una abdicación del juicio humano que diluye la imputabilidad moral en una red de procesos automatizados.

Se evidencia, que la proliferación de asistentes virtuales y algoritmos de recomendación ha trascendido la esfera de la utilidad, para convertirse en dispositivos de subjetivación. En este sentido, la normalización de la mediación técnica en la afectividad y la cognición está gestando una dependencia ontológica, que erosiona la capacidad deliberativa del sujeto. Desde esta percepción, el riesgo inminente no es la rebelión de las máquinas, sino la atrofia del criterio ético humano ante la comodidad de la predicción algorítmica. Asimismo, desde una perspectiva situada, se considera que los sistemas de IA bajo las lógicas del capitalismo de vigilancia y el colonialismo de datos, tiende a reproducir hegemonías epistémicas del Norte Global.

La persistencia de sesgos y la extracción masiva de experiencia humana, exige un cambio de paradigma hacia una soberanía tecnológica que reconozca la pluralidad del Sur Global, y subordine el desarrollo tecnológico a las necesidades democráticas y culturales de los territorios. La presente reflexión sostiene que el diseño ético y la alfabetización crítica de los sistemas de IA no deben ser meros requisitos técnicos, sino imperativos de supervivencia social. En última instancia, la técnica debe permanecer como una herramienta al servicio de la vida, y nunca como juez de la dignidad humana.

Agradecimiento

Las autoras expresan su agradecimiento a la Universidad de Oriente por el respaldo institucional brindado. Asimismo, se reconoce el valioso aporte de los revisores y colegas cuyas observaciones críticas permitieron fortalecer el rigor conceptual de esta investigación.

Referencias bibliográficas

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica.
- Beheshti, A., & Kerridge, I. (2025). Artificial intelligence and ethics: Autonomy, privacy, and social responsibility. *Journal of Bioethical Inquiry*, 22(1), 45-62
- Bentham, J. (1789). *An introduction to the principles of morals and legislation*. T. Payne and Son.
- Bostrom, N. (2009). The future of humanity. In J-K. Berg Olsen, E. Selinger & S. Riis (Eds.), *New waves in philosophy of technology* (pp.186-216). New York: Palgrave Macmillan.
- Braidotti, R. (2019). *Posthuman knowledge*. Cambridge: Polity Press
- Bustamante-Cabrera, G. (2024). *La confusión de la autonomía en los sistemas de inteligencia artificial*. Editorial Universitaria.
- Descartes, R. (1637/2009). *Discurso del Método* (Trad. M. García. Morente). Madrid: Espalsa-Calpe.
- Dussel, E. (1998). *Ética de la liberación en la edad de la globalización y de la exclusión*. Editorial Trotta, SA, 9-659.
- Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press. 243 pp.
- Foucault, M. (1998). *Vigilar y castigar: Nacimiento de la prisión*. Siglo XXI.
- González, R., & Martínez, L. (2020). Dilemas éticos en el escenario de la Inteligencia Artificial. *Economía y Sociedad*, 25(57), 93-109.
- Han, B-C. (2012). *La sociedad del cansancio*. Barcelona: Herder
- Haraway, D. (1991). *Simians, Cyborgs, and Women: The reinvention of nature*. Routledge.
- Heidegger, M. (1994). *La pregunta por la técnica*. Barcelona: Ediciones del Serbal. (Obra original publicada en 1954)
- Jonas, H. (1995). *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica*. Herder.
- Kant, I. (1785/2006). *Crítica de la Razón Pura* (Trad. M. García Morente). México: Fondo de Cultura Económica.
- Mignolo, W. (2011). *The darker side of Western modernity: Global futures, decolonial options*. Duke University Press.
- Mill, J. S. (1863). *Utilitarianism*. Parker, Son, and Bourn.
- Pariser, E. (2017). *El filtro burbuja: Cómo la red decide lo que leemos y lo que pensamos* (J. C. Guix Viaplana, Trad.). Taurus. (Obra original publicada en 2011).
- Porcelli, A. (2020). La Inteligencia Artificial y la robótica: sus dilemas sociales, éticos y jurídicos. *Derecho global, Estudios sobre derecho y justicia*, 6(16), 49-105.
- Quijano, A. (2000). Colonialidad del poder, eurocentrismo y América Latina. En E. Lander (Ed.), *La colonialidad del saber: Eurocentrismo y ciencias sociales* (pp. 201-246). CLACSO.
- Ricourte, P. (2019). Data epistemologies, the coloniality of power, and resistance. *Television & New Media*, 20(4), 350-365.
- Ritzer, G., & Degli Esposti, P. (2020). Creative Destruction and Cultural Lag in the Digital Age, 1-10
- Rivera, J., et al. (2025). *Responsabilidad y algoritmos: Vacíos éticos en la era digital*. Alianza Editorial.
- Tello, A. (2023). Sobre el colonialismo digital: Datos, algoritmos y colonialidad tecnológica del poder en el sur global. *InMediaciones de la Comunicación*, 18(2), 89-110.
- Virilo, P. (2007). Velocidad e información. *Contraste: Revista Cultural*, (47), 15-28-
- Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Declaración de conflicto de intereses: Los autores no presentan ningún conflicto de interés.

Declaración de contribución de los autores/as utilizando la Taxonomía CRediT:

Kirenia Chaveco Asín (Autora principal): Idea de investigación, introducción, métodos, discusión y conclusiones.

Melissa Miniet Alberja (Coautora): Redacción del apartado de resumen, corrección de las referencias.

Declaración de aprobación por el Comité de Ética: Los autores declaran que la investigación fue aprobada por el Comité de Ética de la institución responsable, en tanto la misma implicó a seres humanos.

Declaración de originalidad del manuscrito:

Los autores confirman que este texto no ha sido publicado con anterioridad, ni ha sido enviado a otra revista para su publicación.