

Dilemas éticos y bases filosóficas de la Inteligencia Artificial en educación: una revisión sistemática

Dilemas éticos y bases filosóficas de la Inteligencia Artificial en educación: una revisión sistemática

Adolfo José Apaza Condori

adol3jose@gmail.com

<https://orcid.org/0009-0003-4912-467X>

**Universidad Pública de El Alto. La Paz,
Bolivia**

Richard Jorge Torrez Juaniquina

richardt34@hotmail.com

<https://orcid.org/0009-0001-0170-4718>

**Universidad Pública de El Alto. La Paz,
Bolivia**

Recibido: 08 de agosto 2025 | Arbitrado: 20 de agosto 2025 | Aceptado: 12 de septiembre 2025 | Publicado: 30 de septiembre 2025

Palabras claves:

Bases filosóficas; Dilemas éticos; Epistemología educativa; Gobernanza ética; Inteligencia Artificial

Resumen

El uso adecuado de la tecnología en educación es crucial para potenciar el aprendizaje sin comprometer los valores humanistas esenciales. La presente investigación tiene como objetivo analizar de manera crítica la evidencia científica disponible entre 2020 y 2025 sobre los dilemas éticos y las bases filosóficas de la Inteligencia Artificial en la educación. Se implementó una revisión sistemática cualitativa con 16 estudios seleccionados mediante la guía PRISMA actualizada. El análisis riguroso de las investigaciones seleccionadas posibilita establecer que la integración de esta disciplina en el proceso de enseñanza-aprendizaje trasciende su función instrumental para convertirse en un fenómeno filosófico que transforma los fundamentos del conocimiento y la enseñanza. Frente a los cuestionamientos posthumanistas, surgen tensiones éticas entre privacidad y personalización, eficiencia y equidad, autonomía humana y automatización. Esta dualidad demanda equilibrar innovación con valores humanistas, redefiniendo el rol docente y preservando su juicio. Se recomienda priorizar la equidad mediante gobernanza ética, transparencia algorítmica y auditorías continuas.

Keywords:

Philosophical Foundations; Ethical Dilemmas; Educational Epistemology; Ethical Governance; Artificial Intelligence

Abstract

The appropriate use of technology in education is crucial for enhancing learning without compromising humanistic values. This systematic review critically analyzed evidence published between 2020 and 2025 on the ethical dilemmas and philosophical foundations of Artificial Intelligence (AI) in education. A qualitative review was conducted following updated PRISMA guidelines, with 16 studies included. The analysis highlights that integrating AI into teaching and learning not only supports instructional goals but also fundamentally alters educational philosophy by challenging established concepts of knowledge, teaching, and learning. Ethical tensions arise between privacy and personalization, efficiency and equity, and human autonomy and automation. This duality requires balancing innovation with humanistic principles, redefining the role of teachers while preserving professional judgment. The review recommends prioritizing equity through ethical governance, algorithmic transparency, and ongoing audits.

INTRODUCCIÓN

La Inteligencia Artificial (IA) ha irrumpido en el panorama educativo contemporáneo como una fuerza transformadora de alcance global. Lo que comenzó como herramientas complementarias para la enseñanza ha evolucionado hacia sistemas capaces de mediar procesos cognitivos, personalizar trayectorias de aprendizaje y redefinir las interacciones pedagógicas fundamentales. Esta penetración tecnológica trasciende el ámbito meramente instrumental para convertirse en un fenómeno cultural y filosófico que cuestiona los cimientos mismos sobre los que se ha construido la educación moderna (Tapalova y Zhiyenbayeva, 2022).

En las últimas décadas, la integración de tecnologías inteligentes en entornos educativos ha experimentado una aceleración sin precedentes. Plataformas adaptativas, sistemas de tutoría inteligente, herramientas de evaluación automatizada y asistentes virtuales han dejado de ser prototipos experimentales para convertirse en realidades operativas en instituciones educativas de todos los niveles. Esta implementación masiva responde a promesas de optimización, personalización y escalabilidad, pero simultáneamente genera interrogantes fundamentales sobre el futuro de la práctica educativa y los valores que la sustentan (Gomes, 2025).

El debate sobre la IA en educación ha transitado desde cuestiones técnicas iniciales hacia problemáticas cada vez más profundas de naturaleza ética y filosófica. La literatura especializada evidencia un desplazamiento del enfoque, mientras las primeras investigaciones se centraban en la efectividad pedagógica y la usabilidad tecnológica, los estudios más recientes abordan dilemas relacionados con la autonomía humana, la justicia educativa, la privacidad estudiantil y la reconfiguración de la agencia en entornos algorítmicos (Nopas, 2025). Este cambio de perspectiva revela la creciente conciencia de que la IA no es neutral, sino que incorpora visiones particulares sobre el conocimiento, el aprendizaje y el ser humano (Korteling et al., 2021).

La reflexión filosófica se ha revelado como un componente indispensable para comprender el impacto real de estas tecnologías. Conceptos como Zwischenreich o reino intermedio emergen para describir el espacio liminal donde se negocia la convivencia entre inteligencias humanas y artificiales. Las tensiones entre humanismo y posthumanismo, entre eficiencia algorítmica y desarrollo humano integral, entre personalización educativa y protección de la privacidad, definen un territorio conceptual complejo que demanda aproximaciones interdisciplinares y marcos de análisis sofisticados (Rahm y Rahm, 2023).

La urgencia de investigar sistemáticamente estas dimensiones ético-filosóficas deviene de la constatación de que las decisiones que se tomen en la actualidad respecto a la integración educativa de la IA configurarán el futuro de las sociedades venideras. Se hace necesario, por tanto, examinar críticamente ¿cuáles son los fundamentos conceptuales que subyacen a estas tecnologías?, ¿cuáles son las tensiones valorativas que generan su implementación? y ¿qué propuestas existen para los marcos de actuación que aseguren que el desarrollo tecnológico sirva a los fines más elevados de la educación?. Por lo tanto, la presente investigación tiene como objetivo analizar de manera crítica la evidencia científica disponible entre 2020 y 2025 sobre los dilemas éticos y las bases filosóficas de la Inteligencia Artificial en la educación.

MÉTODO

El abordaje metodológico se sustentó en una revisión sistemática de carácter cualitativo. Para ello, se ejecutó una búsqueda metódica en bases de datos especializadas, con el objetivo de localizar estudios primarios que examinaran los desafíos éticos y filosóficos asociados a la integración de la Inteligencia Artificial en el ámbito educativo. El procedimiento se rigió por los criterios establecidos en la guía PRISMA, la cual permitió organizar la selección de la literatura en las fases de identificación, cribado, evaluación de elegibilidad e inclusión final.

Criterios de selección y bases de datos utilizadas

La estrategia de búsqueda se implementó en tres bases de datos multidisciplinares de relevancia en el ámbito de las ciencias sociales y la educación:

SciELO, Scopus y Dialnet. La sintaxis de búsqueda combinó descriptores y operadores booleanos para maximizar la recuperación de literatura pertinente. Las palabras clave utilizadas, adaptadas a cada base de datos, fueron en inglés: "Artificial Intelligence" OR "AI" AND "ethics" OR "ethical dilemmas" OR "challenges" AND "education" OR "learning". Para las búsquedas en español en Dialnet y SciELO, se emplearon los términos equivalentes: "Inteligencia Artificial" OR "IA" AND "ética" OR "dilemas éticos" OR "desafíos" AND "educación" OR "aprendizaje".

Los criterios de exclusión se aplicaron de forma rigurosa para refinar los resultados. Se excluyeron: 1) estudios duplicados entre las bases de datos consultadas; 2) publicaciones que no fueran artículos de investigación originales o revisiones sistemáticas; 3) literatura gris y actas de congresos no revisadas por pares; 4) documentos cuyo texto completo no estuviera disponible; 5) estudios publicados en un idioma distinto al inglés o español; y 6) investigaciones que, a pesar de tratar sobre IA, no abordaran de manera explícita y sustancial las dimensiones éticas, filosóficas o los desafíos asociados en un contexto educativo formal.

Estrategias de búsqueda y proceso de selección de estudios

Dada la heterogeneidad de las plataformas, se diseñaron y aplicaron protocolos de búsqueda individualizados, optimizando la estrategia para cada entorno mediante el uso de su sintaxis específica y los términos de su tesoro correspondiente.

1. Scopus

Se utilizó la interfaz avanzada. La estrategia se aplicó a los campos de Título, Resumen y Palabras clave.

- (TITLE-ABS-KEY ("artificial intelligence" OR "ai" OR "machine learning") AND TITLE-ABS-KEY (ethic* OR "ethical dilemma*" OR "moral philosophy" OR challenge*) AND TITLE-ABS-KEY (educat* OR learn* OR "higher education" OR "K-12"))
- Filtros Aplicados:
 - Límite de fechas: 2020-2025.

- Tipo de documento: Artículo y Revisión.
- Área temática: Ciencias Sociales, Informática.
- Idioma: Inglés, Español.

2. SciELO

La búsqueda se realizó a través de su portal, utilizando el cuadro de búsqueda principal que indexa título, resumen y materias.

- ("Inteligencia Artificial" OR IA OR "Aprendizaje Automático") AND (Ética OR "Dilema Ético" OR "Filosofía Moral" OR Desafío*) AND (Educación OR Aprendizaje OR "Educación Superior")
- Filtros Aplicados:
 - Año: 2020-2025.
 - Formato del documento: Artículo.
 - Idioma: Español, Inglés.

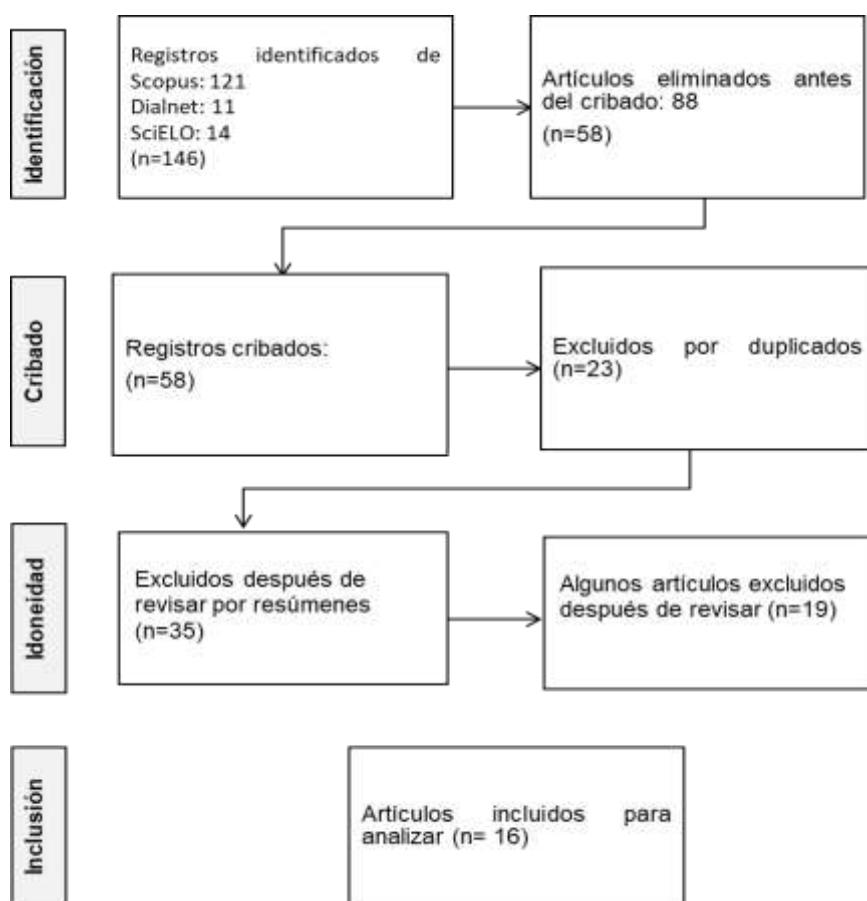
3. Dialnet

S se priorizó la búsqueda en los campos de Título y Palabras clave para asegurar la máxima relevancia, dada la amplitud de su repositorio.

- ("Inteligencia Artificial" OR IA) AND (Ética OR "Dilemas Éticos" OR Desafío*) AND (Educación OR Aprendizaje)
- Filtros Aplicados:
 - Año de publicación: 2020-2025.
 - Tipo de documento: Artículo de revista.
 - Acceso: Texto completo disponible.

El proceso de selección inició con la identificación de 146 referencias. Tras una evaluación inicial basada en los criterios de elegibilidad, se descartaron 88 fuentes. En la etapa de cribado, un análisis más detallado permitió excluir 23 estudios adicionales, resultando en un corpus de 35 trabajos potencialmente relevantes. La fase de idoneidad, que implicó el examen de títulos, resúmenes y contenido completo, condujo a la exclusión de 19 investigaciones. El proceso concluyó con la inclusión de 16 estudios que satisficieron integralmente todos los criterios metodológicos establecidos. (Figura 1).

Figura 1. Diagrama de flujo para la selección de los artículos según PRISMA



Una vez seleccionados los artículos, se procedió a analizar su contenido. Se elaboraron matrices que incluyeron los siguientes indicadores, año de publicación, autores, título, bases filosóficas, dilemas éticos más relevantes del uso de IA en la educación y los hallazgos más significativos encontrados en cada estudio.

RESULTADOS

A continuación, se presentan los hallazgos derivados del análisis cualitativo de los 16 estudios seleccionados. Los resultados se organizan en torno a las principales tensiones ético-filosóficas identificadas en la literatura, revelando un consenso notable sobre la naturaleza dual de la IA en educación como herramienta de potenciación y, simultáneamente, como fuente de desafíos sin precedentes para los valores humanistas fundamentales.

El análisis de procedencia de los estudios incluidos revela una clara hegemonía de Scopus como base de datos indexadora, la cual alberga el 90% de los estudios seleccionados. Esta

predominancia indica que la producción científica más relevante y reconocida sobre los fundamentos filosóficos y dilemas éticos de la IA en educación se concentra en repositorios de alto impacto y alcance internacional. Por su parte, Dialnet y SciELO contribuyeron cada uno con tres publicaciones, lo que sugiere que, si bien existe una producción valiosa en el ámbito iberoamericano, su volumen dentro del corpus de alta relevancia para esta revisión es aún limitado.

Desde una perspectiva regional, la investigación demuestra un carácter marcadamente global, con contribuciones de todos los continentes habitados. Se identifican tres focos geográficos principales: el continente europeo, España, Macedonia del Norte, Turquía, Finlandia; en América Ecuador, Estados Unidos y en Asia Uzbekistán, Indonesia, India, Emiratos Árabes. Es notable la representación de Ecuador, que con cuatro estudios se erige como un contribuyente significativo al debate, particularmente desde una perspectiva humanista y de derecho. Esta

distribución evidencia que la preocupación por las implicaciones ético-filosóficas de la IA en educación es un fenómeno transversal, que no se limita a los países tecnológicamente más avanzados, sino que emerge desde contextos culturales y educativos diversos.

El análisis de la Tabla 1, sobre las bases filosóficas de la IA en educación, demuestra, en primer lugar, que existe una reconfiguración epistemológica y ontológica profunda. La capacidad de los sistemas de IA para generar conocimiento y mediar procesos cognitivos cuestiona, según Abdusattarova (2025), los pilares de la racionalidad científica clásica y redistribuye la autoridad que tradicionalmente residía en el docente y el currículum. Simultáneamente, la noción de agencia se desdibuja, surgiendo una ontología relacional donde la capacidad de acción y decisión se distribuye entre entidades humanas y no humanas. Esto lleva a la educación, como analizan Román et al. (2024), a un Zwischenreich o reino intermedio, un espacio liminal de constante negociación dialéctica entre la potenciación y la alienación.

En segundo término, emerge una tensión irreductible entre visiones humanistas y posthumanistas. Frente a los paradigmas del posthumanismo y transhumanismo que, como analiza Firdaus (2025), promueven la fusión humano-tecnología y priorizan la eficiencia, se erige una postura humanista que, en palabras de Vinicio et al. (2025), aboga por colocar el desarrollo integral del ser humano en el centro. Esta perspectiva alerta sobre el riesgo de que la IA erosione la integridad académica, el pensamiento crítico y los valores fundamentales de la educación. La filosofía de la tecnología, siguiendo a Rajan (2025), aporta aquí una crítica crucial al recordar que la esencia de la IA no es neutral, sino que configura un marco que condiciona nuestra libertad y nuestra comprensión del ser.

Finalmente, el análisis revela que la superación de estas tensiones requiere de marcos éticos y de gobernanza robustos. La propuesta de un diálogo interdisciplinario constante entre pedagogos, tecnólogos y estudiantes, como sugiere Vinicio et al. (2025), se presenta como un camino necesario para guiar el desarrollo de la IA. La conclusión

transversal es que el verdadero desafío no es técnico, sino filosófico: se trata de integrar la tecnología sin abdicar de la responsabilidad humana, asegurando que la IA en la educación sirva, en última instancia, a un proyecto ético y profundamente humanista.

Tabla 1. Investigaciones incluidas en la revisión sistemática sobre las bases filosóficas para reflexionar sobre la IA en educación

Autor / País / Año	Título	Bases filosóficas	Hallazgos
1. Abdusattarova (2025) / Uzbekistán	Artificial Intelligence as a Phenomenon of Contemporary Philosophy of Science and Technology	Filosofía de la Ciencia: La IA desafía los paradigmas tradicionales de objetividad y racionalidad científica en la producción de conocimiento educativo Ontología: Examina la naturaleza del ser y la agencia en contextos donde sistemas no humanos demuestran capacidades de aprendizaje y toma de decisiones Epistemología: Replantea el significado mismo del conocimiento cuando este es generado o mediado por algoritmos Posthumanismo: Investiga la nueva condición humana emergente de la interacción constante entre personas y sistemas inteligentes	El estudio conceptualiza la IA como un evento filosófico que trasciende su función instrumental para desafiar paradigmas fundamentales. Sus principales hallazgos revelan que: 1) Cuestiona nociones clásicas de objetividad, agencia y racionalidad científica; 2) Genera transformaciones ontológicas y epistemológicas al redefinir la inteligencia y autonomía; 3) Capacidades como el aprendizaje automático y la toma de decisiones algorítmica exigen replantear la condición humana y las formas de conocer y actuar en un mundo tecnológicamente mediado.
2. Firdaus (2025) / Indonesia	The Philosophical Construction of Educational Science in Relation to Posthumanism and Transhumanism in Artificial Intelligence	Posthumanismo: Cuestiona la centralidad humana en educación al distribuir la agencia con sistemas no humanos, transformando los objetivos formativos tradicionales Transhumanismo: Examina los riesgos éticos de priorizar la eficiencia tecnológica sobre el desarrollo intelectual y moral humano Hermenéutica Filosófica: Proporciona herramientas para interpretar cómo la IA resignifica las prácticas educativas y afecta valores como el esfuerzo intelectual y la autoría	El estudio, que analizó 8,288 documentos, revela un crecimiento exponencial en la investigación sobre IA educativa. Sus hallazgos principales indican que: 1) La IA potencia la personalización y eficiencia del aprendizaje; 2) Los paradigmas posthumanistas y transhumanistas desafían los valores educativos tradicionales; 3) Existe un riesgo real de erosión de la integridad académica, desarrollo del carácter y pensamiento crítico, amenazando el modelo educativo humanocéntrico.
Rajan (2025) / Estados Unidos	Philosophy of technology for the lost age of freedom: a critical treatise on human essence and uncertain future	Filosofía de la Tecnología (Heidegger): Concibe la tecnología como marco que configura nuestra comprensión del mundo y condiciona la libertad humana Ética de la Información (Floridi): Analiza el impacto de la IA en la construcción de identidad y relaciones sociales dentro de la infosfera	El estudio plantea que la evolución humana, basada en la interacción con entornos naturales y sociales, genera una disonancia ante tecnologías avanzadas como la IA. Esta brecha requiere un análisis filosófico urgente sobre la esencia tecnológica y su impacto en la libertad humana, la naturaleza del ser y las transformaciones sociales. La investigación propone un marco interdisciplinario para

		Fenomenología y Crítica Cultural (Arendt, Huxley): Evalúa cómo la IA transforma la experiencia humana y alerta sobre pérdidas de autonomía en sociedades tecnocráticas	evaluar críticamente estos efectos en los ámbitos público y privado.
4. Vinicio et al. (2025) / Ecuador	Hacia una educación universitaria humanista en la era de la inteligencia Artificial. reflexiones axiológicas contemporáneas.	Humanismo: Coloca al ser humano y su desarrollo integral en el centro del proceso educativo, utilizando la IA como herramienta de potenciación -no de reemplazo- de capacidades humanas Ética de la Justicia: Garantiza que la implementación de IA no profundice brechas sociales, asegurando acceso equitativo y protección de derechos estudiantiles fundamentales	El estudio revela la dualidad de la IA en educación superior: mientras ofrece potencial para innovación pedagógica, también presenta riesgos de deshumanización. Identifica tres desafíos críticos: vulneración de privacidad, inequidad en acceso tecnológico y erosión de valores humanistas. La investigación concluye que se requiere un enfoque inclusivo y diálogo permanente entre todos los actores para alinear la IA con una educación centrada en el ser humano.
5. Román et al. (2024) / Ecuador	Epistemología, Ética, Educación e Inteligencia Artificial	Ontología: Examina cómo la dataficación del estudiante redefine la condición humana y cuestiona la preservación de la autonomía y dignidad Epistemología: Investiga la construcción y validación del conocimiento en la era algorítmica, desafiando las narrativas tradicionales del saber Fenomenología y Hermenéutica: Analiza la experiencia vivida en la relación humano-IA a través del concepto de "Zwischenreich", explorando la tensión entre conciencia humana y potencialidades tecnológicas.	El estudio conceptualiza la IA como una disruptión filosófica que genera una tensión dialéctica en educación entre su promesa liberadora (optimización, personalización) y el riesgo de deshumanización (alienación, desigualdad). Esta tensión crea un "Zwischenreich" o reino intermedio que redefine la relación sujeto-tecnología. Los hallazgos destacan la urgencia de abordar las implicaciones ontológicas y epistemológicas de la IA, examinando cómo los algoritmos reconfiguran el conocimiento, la identidad y la autonomía humana frente a sistemas automatizados.

La Tabla 2 revela que la integración de la IA en la educación genera una constelación de dilemas éticos interconectados que trascienden los problemas técnicos para abordar cuestiones fundamentales sobre derechos humanos, agencia y justicia social. Estos desafíos, documentados globalmente desde España hasta Nigeria y desde Ecuador hasta Finlandia, pueden organizarse en cuatro ejes temáticos principales que demandan una atención urgente por parte de educadores, diseñadores de políticas y desarrolladores tecnológicos.

Privacidad, datos y autonomía: el precio de la personalización

Uno de los dilemas más consistentes identificados en la literatura es la tensión inherente entre la personalización del aprendizaje y la protección de la privacidad estudiantil. Como señalan Salloum (2024) y Espinosa et al. (2025), la recolección exhaustiva de datos necesaria para adaptar los contenidos educativos se enfrenta con el derecho fundamental a la intimidad. Esta tensión se intensifica con tecnologías de vigilancia como el reconocimiento facial, que según Akgun y Greenhow (2022), crean un conflicto entre el monitoreo del bienestar estudiantil y la intrusión en su libertad. El problema se agrava cuando consideramos, con Dave (2025), que esta hiperpersonalización puede limitar la autonomía del aprendiz, creando burbujas algorítmicas que restringen su exposición a contenidos diversos y su capacidad de elección. La paradoja resulta evidente: para potenciar al estudiante, el sistema debe recolectar datos íntimos que simultáneamente pueden vulnerar su privacidad y limitar su agencia.

Sesgo algorítmico y equidad: la amenaza a la justicia educativa

El segundo gran conjunto de dilemas gira en torno a la promesa incumplida de objetividad algorítmica. Múltiples investigaciones, incluyendo las de Salloum (2024) y Gallent et al. (2024), alertan que los sistemas de IA pueden perpetuar e incluso amplificar sesgos sociales existentes, priorizando la eficiencia operativa sobre la equidad sustantiva. Khan (2024), identifica esto como una paradoja fundamental, herramientas diseñadas para ser

objetivas reproducen prejuicios históricos codificados en sus datos de entrenamiento. Esta problemática adquiere dimensiones particularmente graves en contextos jurídicos, donde García (2025), advierte que la automatización puede violar el principio de igualdad ante la ley y el debido proceso. La brecha digital emerge como un multiplicador de estas desigualdades, creando, en palabras de Gallent et al. (2024), un sistema educativo de, donde la implementación de tecnologías avanzadas termina marginando a quienes carecen de acceso o competencias digitales.

Agencia humana vs. automatización: ¿quién decide en educación?

El tercer eje temático explora la redistribución de agencia entre humanos y sistemas automatizados. El estudio comparativo de Karakuş et al. (2025), revela una divergencia crucial en los procesos de toma de decisiones, mientras la IA prioriza un enfoque analítico y utilitarista (resultado vs. principio), los docentes enfatizan la empatía y la contextualización. Esta diferencia no es meramente técnica sino profundamente ética, pues como señala Akgun y Greenhow (2022), la dependencia de sistemas automatizados erosiona progresivamente el juicio profesional docente. El problema de la responsabilidad difusa, identificado por Salloum (2024), complica aún más este panorama, pues cuando una decisión educativa errónea o sesgada emana de un sistema autónomo, resulta extremadamente difícil atribuir accountability. Khan (2024), añade a esto la preocupación por el desplazamiento de roles educativos humanos, planteando la disyuntiva entre ganar eficiencia operativa y deshumanizar procesos educativos fundamentales.

Gobernanza, integridad académica y escalabilidad ética

Finalmente, la literatura identifica desafíos macroéticos relacionados con la gobernanza y la integridad académica. Nguyen et al. (2023), plantean la tensión entre la necesaria estandarización global de principios éticos y la adaptación a contextos locales específicos, mientras que Shalevska (2024), alerta sobre las amenazas a la propiedad intelectual con la generación autónoma de contenido. La integridad académica emerge

como una preocupación transversal, particularmente con herramientas generativas como

ChatGPT que, según Khan (2024), facilitan el plagio y erosionan los principios de honestidad académica. Frente a la imprevisibilidad de escenarios identificada por Casas et al. (2021), el diseño ético previo resulta insuficiente, demandando la integración de sensores éticos en tiempo real. La velocidad del desarrollo tecnológico crea además una tensión crítica, descrita por Nguyen et al. (2023), entre la presión por implementar rápidamente nuevas herramientas y la obligación de realizar evaluaciones éticas exhaustivas.

Convergencia en las soluciones propuestas

Frente a esta compleja constelación de dilemas, la literatura converge en la necesidad de marcos de gobernanza robustos y procesos de diálogo interdisciplinario continuos. La educación misma emerge, en el análisis de Shalevska (2024), como el mecanismo más poderoso para mitigar estos riesgos, posicionando a los educadores como actores clave en la formación de una ciudadanía digital crítica. La implementación exitosa requerirá, como concluyen Espinosa et al. (2025), un enfoque integral que equilibre la innovación tecnológica con la equidad educativa y la responsabilidad ética, asegurando que el desarrollo de la IA en educación sirva finalmente para expandir, rather que restringir, las oportunidades de aprendizaje humano.

Tabla 2. Investigaciones incluidas en la revisión sistemática sobre los dilemas éticos más relevantes del uso de IA en la educación

Autor / País / Año	Título	Dilemas éticos	Hallazgos
1. Casas et al. (2021) / España	AI Perils in Education: Exploring Ethical Concerns	Elección binaria: Sistemas que deben priorizar entre rendimiento académico y bienestar emocional estudiantil Imprevisibilidad: La autonomía de la IA genera escenarios éticos no anticipados en su diseño Contextualización limitada: Dificultad para programar la comprensión de realidades socio-familiares únicas en la toma de decisiones	El diseño ético previo resulta insuficiente para la IA educativa debido a su complejidad y autonomía, que generan escenarios impredecibles. Se propone integrar herramientas de moralidad artificial como "sensores éticos" que permitan a los sistemas autónomos identificar nuevos dilemas en tiempo real. Esta solución requiere incorporar el contexto único de cada estudiante -necesidades, preferencias y metas- en el proceso de razonamiento moral de la IA para una toma de decisiones más adaptativa.
2. Salloum (2024) / Emiratos Árabes Unidos	AI Perils in Education: Exploring Ethical Concerns	Privacidad vs. Personalización: Conflicto entre recolección de datos y protección estudiantil Eficiencia vs. Equidad: Algoritmos que optimizan operaciones perpetuando desigualdades Autonomía vs. Automatización: Tensión entre criterio docente y decisiones de IA Responsabilidad Difusa: Desafío de atribuir culpas en fallos de sistemas autónomos.	El estudio identifica cuatro riesgos éticos principales de la IA en educación: 1) vulneración de la privacidad estudiantil, 2) amplificación de sesgos algorítmicos, 3) erosión de la autonomía educativa, y 4) opacidad en la rendición de cuentas. Estos desafíos alteran la dinámica educativa tradicional, requiriendo con urgencia políticas que equilibren los beneficios tecnológicos con una implementación ética y responsable.
3. Karakuş et al. (2025) / Turquía	Ethical Decision-Making in Education: A Comparative Study of Teachers and Artificial Intelligence in Ethical Dilemmas	Resultado vs. Principio: Conflicto entre maximizar el bienestar general y cumplir normas morales inviolables Racionalidad vs. Empatía: Tensión entre el análisis lógico y la compasión en las decisiones Uniformidad vs. Contextualización: Disyuntiva entre aplicar reglas consistentes y adaptarse a circunstancias específicas	El estudio muestra una coincidencia del 62.5% entre IA y docentes en dilemas éticos, pero con enfoques distintos: la IA prioriza el análisis utilitario de resultados, mientras los docentes enfatizan la empatía y principios morales. Esto confirma el valor complementario de la IA para desarrollar conciencia ética, pero reafirma la indispensabilidad del juicio humano en decisiones educativas sensibles que requieren compasión y contextualización.
4. Shalevska (2024) / Macedonia del Norte	Human Rights in the Age of AI: Understanding the Risks, Ethical Dile	Progreso vs. Privacidad: Innovación tecnológica que requiere datos versus protección de la intimidad personal	El estudio advierte que el desarrollo de la IA, aunque transformador, amenaza derechos humanos fundamentales mediante tres riesgos principales: vulneración masiva de privacidad, perpetuación de sesgos por discriminación algorítmica, y amenaza a la propiedad intelectual. La

	mmas, and the Role of Education in Mitigating Threats	Eficiencia vs. Equidad: Automatización algorítmica que perpetúa sesgos históricos y sociales Creación autónoma vs. Propiedad intelectual: Desafío de atribuir autoría en obras generadas por IA	investigación concluye que la educación constituye el pilar esencial para mitigar estos riesgos, posicionando a los educadores como actores clave en la resolución de estos dilemas éticos.
5. Espinosa et al. (2025) / Ecuador	La inteligencia artificial en la educación: Una revisión sistemática de aplicaciones, beneficios y desafíos éticos	Personalización vs. Privacidad: Conflicto entre la recolección de datos para adaptar la enseñanza y la protección de la intimidad estudiantil Objetividad vs. Sesgo: Tensión entre la promesa de neutralidad algorítmica y la perpetuación de desigualdades existentes en los datos de entrenamiento	La revisión confirma que la IA mejora el aprendizaje mediante personalización y optimización de procesos, pero enfrenta tres retos principales: brecha digital, capacitación docente insuficiente y dilemas éticos sobre privacidad y sesgos algorítmicos. Una implementación exitosa requiere equilibrar innovación tecnológica con equidad educativa y responsabilidad ética para garantizar su sostenibilidad.
6. García (2025) / Ecuador	La Inteligencia Artificial como herramienta para el diseño de estrategias didácticas e intervención jurídica en la educación superior ecuatoriana	Eficiencia vs. Debido Proceso: Agilización judicial versus opacidad que vulnera derechos de defensa Innovación vs. Equidad: Implementación tecnológica versus ampliación de brechas de acceso Automatización vs. Igualdad: Optimización algorítmica versus perpetuación de sesgos históricos en decisiones legales	El estudio reconoce el potencial de la IA para transformar la educación jurídica en Ecuador, pero identifica tres barreras principales: brecha digital, falta de regulación específica y formación docente insuficiente. En el ámbito judicial, la opacidad algorítmica amenaza derechos fundamentales como el debido proceso. Su implementación exitosa requiere un marco regulatorio sólido, capacitación especializada y supervisión ética continua.
7. Gallent et al. (2024) / España	Inteligência Artificial: entre riscos e potencialidades	Personalización vs. Privacidad: Conflicto entre la recolección de datos para adaptar la enseñanza y la protección de la intimidad estudiantil Eficiencia vs. Equidad: Riesgo de que los algoritmos prioricen la optimización operativa perpetuando sesgos existentes Innovación vs. Inclusión: Tensión entre adoptar tecnologías avanzadas y ampliar las desigualdades por falta de acceso o competencias digitales	El estudio reconoce el potencial de la IA para transformar la educación mediante personalización del aprendizaje, tutorías inteligentes y monitoreo del desempeño. Sin embargo, advierte que su implementación sin salvaguardias puede agravar desigualdades, identificando tres riesgos críticos: vulneración de privacidad estudiantil, reforzamiento de sesgos algorítmicos y ampliación de la brecha digital. Concluye que se requieren marcos éticos y regulatorios robustos, con participación activa de educadores y policymakers, para asegurar que la IA promueva equidad en lugar de discriminación.

8. Khan (2024) / Nigeria	Ethical Dilemma Of Artificial Intelligence in education: balancing automation and human teaching roles: a review	Eficiencia vs. Empleo: Automatización que mejora productividad pero amenaza roles educativos tradicionales Innovación vs. Integridad: Herramientas de aprendizaje que pueden facilitar el plagio y fraude académico Personalización vs. Privacidad: Recopilación de datos estudiantiles que vulnera la confidencialidad Objetividad vs. Sesgo: Algoritmos que perpetúan desigualdades sociales existentes	La IA revoluciona la educación mediante la automatización, personalización y optimización de procesos. Sin embargo, plantea cuatro desafíos éticos cruciales: 1) posible desplazamiento de docentes, 2) riesgos a la integridad académica, 3) vulnerabilidad de datos estudiantiles, y 4) sesgos algorítmicos que comprometen la equidad. La investigación subraya la importancia de equilibrar tecnología y factor humano, recomendando marcos de gobernanza, políticas claras y directrices éticas para garantizar una implementación transparente e inclusiva.
9. Akgun y Greenhow (2022) / Estados Unidos	Artificial intelligence in education: Addressing ethical challenges in K-12 settings	Vigilancia vs. Bienestar: Monitoreo estudiantil que vulnera privacidad bajo pretexto educativo Personalización vs. Autonomía: Plataformas que guían aprendizaje restringiendo libertad educativa Eficiencia vs. Criterio docente: Sistemas automatizados que erosionan el juicio profesional del profesor	La IA transforma la educación K-12 mediante plataformas personalizadas y evaluación automatizada, pero sus dilemas éticos suelen subestimarse. Este estudio enfatiza que la comprensión de estos desafíos por parte de docentes y estudiantes es crucial para una implementación responsable. Como solución, propone recursos educativos concretos (MIT Media Lab, Code.org) para enseñar a navegar estos riesgos, equilibrando los beneficios de la IA con la necesaria mitigación de sus impactos negativos.
10. Dave (2025) / India	The ethical implications of AI in education	Eficiencia vs. Transparencia: Sistemas complejos que optimizan procesos pero cuyas decisiones resultan inescrutables Personalización vs. Libertad: Adaptación curricular que limita la autonomía y desarrollo del estudiante Innovación vs. Igualdad: Tecnología avanzada que puede ampliar brechas educativas y sociales existentes	La IA educativa presenta una dualidad: potencia la personalización y eficiencia, pero genera riesgos éticos como opacidad algorítmica, mal uso de datos estudiantiles, amplificación de desigualdades e impacto psicológico. El estudio concluye que se requiere una gobernanza ética con algoritmos transparentes y marcos colaborativos que involucren a la comunidad educativa para garantizar equidad, autonomía estudiantil y bienestar, transformando así la IA en una herramienta de innovación responsable.
11. Nguyen et al. (2023) / Finlandia	Ethical principles for artificial intelligence in education	Estandarización vs. Contexto: Tensión entre principios éticos universales y su adaptación a realidades locales diversas	A pesar del potencial transformador de la IA educativa y la proliferación de directrices éticas, persiste una falta de consenso global sobre sus principios rectores. Este estudio responde mediante un análisis temático de políticas internacionales, proponiendo un marco unificado de

Progreso vs. Autonomía: Conflicto entre usar datos para mejorar el aprendizaje y respetar la soberanía del estudiante

Velocidad vs. Precaución: Disyuntiva entre implementación ágil y evaluaciones éticas exhaustivas

principios éticos. El marco busca guiar integralmente a todos los actores educativos en el desarrollo e implementación de una IA confiable, y servir como base para futuros estudios de impacto en el campo educativo.

La IA se revela, en última instancia, como un espejo tecnológico que devuelve a la educación su pregunta fundamental, el significado del ser humano en un mundo crecientemente mediado por inteligencias no humana. Los marcos filosóficos analizados, del posthumanismo al humanismo, de la ontología a la epistemología, no ofrecen respuestas definitivas, sino que delinean el espacio mismo donde debe librarse esta batalla conceptual. La verdadera disyuntiva no es entre adopción o rechazo, sino entre una implementación acrítica que naturalice la tecnología como destino y una apropiación reflexiva que la subordine a un proyecto educativo deliberado. En este reino intermedio, el imperativo ya no es solo enseñar con IA, sino aprender a pensar filosóficamente sobre ella, reafirmando que toda tecnología educativa es, en esencia, una tecnología del yo y de la convivencia.

Frente a la diversidad de dilemas éticos que la IA introduce, la educación se enfrenta a una prueba de integridad. La paradoja es profunda, para personalizar, debe vigilar; para optimizar, puede discriminar; para asistir, podría suplantar. Superar estas tensiones requiere trascender la búsqueda de soluciones técnicas superficiales o regulaciones reactivas. El desafío es de arquitectura moral, se trata de diseñar sistemas educativos donde la equidad esté codificada en el algoritmo, la transparencia sea una característica no negociable y la privacidad se erija como derecho fundamental.

Esto exige una gobernanza distribuida donde desarrolladores, educadores, estudiantes y políticos compartan la responsabilidad de cultivar un ecosistema en el que la IA no compita con la inteligencia humana, sino que amplifique la capacidad colectiva para el juicio prudencial, la empatía contextual y la sabiduría práctica que constituyen el núcleo irreductible de una educación verdaderamente humana.

DISCUSIÓN

Los hallazgos de esta revisión sistemática permiten establecer un diálogo crítico con la literatura académica no incluida en el corpus de análisis, identificando tanto consensos como matices significativos en la comprensión de los fundamentos filosóficos y dilemas éticos de la IA en

educación. Esta comparación con autores externos a la revisión enriquece la contextualización de los resultados y revela aportes distintivos del presente estudio.

En el ámbito de las bases filosóficas, los resultados coinciden con las posturas de Hosseini y Sakhaei (2025), que reconocen a la IA como un fenómeno que redefine las relaciones de poder epistémico y cuestiona los fundamentos del conocimiento educativo. Igualmente, se observa sintonía con la perspectiva de Adewojo (2025), sobre la necesidad de repensar la IA en entornos tecnológicamente mediados. Sin embargo, mientras autores como Hallström (2022), enfatizan un determinismo tecnológico que subordina lo educativo a lo digital, y Ranyal et al. (2022), conciben la tecnología como estructura condicionante, esta revisión revela una realidad más compleja donde emerge un espacio de negociación dialéctica, el *Zwischenreich*, que matiza visiones puramente deterministas y sugiere una interacción más dialéctica entre agentes humanos y tecnológicos.

Respecto a los dilemas éticos, se corrobora con investigadores como la de Boateng y Boateng (2025), la naturaleza estructural de los sesgos algorítmicos y su impacto en la equidad educativa, así como con los planteamientos de Holzmeyer (2021), sobre modelos y algoritmos que perjudican de manera sistemática y opaca a grupos poblacionales vulnerables, escalando la injusticia social bajo una apariencia de objetividad técnica. La tensión entre personalización y privacidad coincide con las alertas de Jindal y Gouri (2024) y Selvakumar et al. (2025), sobre las burbujas de filtro o cámaras de eco algorítmicas en entornos digitales.

En el plano de la gobernanza y valores educativos, se identifican puntos de contacto con los planteamientos de Mura y Stehlíková (2025), sobre la necesidad de preservar el desarrollo humano integral frente a tendencias utilitaristas. Asimismo, existe coincidencia con Koutsikouri et al. (2024), en cuanto a la importancia de mantener espacios para el juicio contextual y la sabiduría práctica en entornos crecientemente automatizados. Sin embargo, mientras que Islam et al. (2024), enfatiza

la potencialidad de los sistemas colaborativos, los hallazgos de esta revisión sugieren que la implementación actual de la IA en educación tiende más hacia lógicas de optimización individual que de construcción colectiva, marcando una divergencia significativa con las visiones más utópicas de la tecnología educativa.

En síntesis, la comparación con autores ajenos a esta revisión evidencia que, si bien se confirman alertas previas sobre riesgos éticos y transformaciones filosóficas, los hallazgos actuales matizan visiones deterministas y profundizan en la naturaleza relacional de la agencia, avanzando hacia una comprensión más dialéctica de la co-constitución entre humanos y sistemas inteligentes en educación. El concepto de Zwischenreich, dominio liminal, emerge como aporte distintivo para comprender este espacio de negociación permanente, superando dicotomías simplificadoras y proponiendo una visión más matizada de las transformaciones educativas en la era de la inteligencia artificial.

CONCLUSIONES

El análisis evidencia que la integración de la IA en la educación demanda una reflexión profunda que articule sus bases filosóficas con los urgentes desafíos éticos identificados. Trasciende su función instrumental para convertirse en un evento filosófico que cuestiona los fundamentos de la educación. Los enfoques posthumanistas y transhumanistas desafían la centralidad humana, mientras la ontología y epistemología revelan la manera en que los algoritmos reconfiguran el conocimiento y la agencia, exigiendo replantear la condición humana en entornos tecnomediados.

Se identifican cuatro dilemas éticos principales: privacidad vs. personalización, eficiencia vs. equidad, autonomía humana vs. automatización, y responsabilidad difusa. Estos conflictos surgen de la tensión entre optimización técnica y valores humanistas, requiriendo marcos que equilibren innovación con protección de derechos fundamentales en entornos educativos.

La implementación de IA en educación genera una dualidad irreductible, potencia la personalización y eficiencia, pero simultáneamente amenaza con deshumanizar el aprendizaje. Esta

tensión exige redefinir los roles docentes, desarrollar nuevas competencias digitales críticas y mantener el juicio humano como elemento central en la toma de decisiones educativas sensibles. De ahí, la necesidad de implementar marcos de gobernanza ética con algoritmos transparentes y auditoría continua. Desarrollar políticas educativas que prioricen la equidad sobre la eficiencia. Fortalecer la formación docente en competencias digitales críticas. Promover investigación longitudinal sobre impactos psicosociales. Establecer comités éticos multidisciplinares para evaluar implementaciones de IA.

REFERENCIAS

- Abdusattarova, S. (2025). Artificial Intelligence as a Phenomenon of Contemporary Philosophy of Science and Technology. *International Journal of Scientific Trends*, 3(6), 77-83. <https://scientifictrends.org/index.php/ijst/article/view/599/553>
- Adewojo, A. A. (2025). Beyond the Physical: Exploring the Ontological Implications of Future Classrooms in Digital and AI-Mediated Spaces. *Digital Education Review*, 46, 1-12. <https://doi.org/10.1344/der.2025.47.1-12>
- Akgun, S. y Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI Ethics*, 2(3), 431-440. <https://doi.org/10.1007/s43681-021-00096-7>
- Boateng, O. y Boateng, B. (2025). Algorithmic bias in educational systems: Examining the impact of AI-driven decision making in modern education. *World Journal of Advanced Research Reviews*, 25(1), 2012-2017. <https://doi.org/10.30574/wjarr.2025.25.1.0253>
- Casas, J., Conesa, J. y Caballé, S. (2021). Education, ethical dilemmas and AI: from ethical design to artificial morality. *International conference on human-computer interaction*,

- https://doi.org/10.1007/978-3-030-77857-6_11
- Dave, A. (2025). The ethical implications of AI in education. *Research Review Journal of Social Science*, 5(1), 1-8. https://doi.org/10.31305/rrjss.2025.v05.n01.001
- Espinosa, J. G., Moreno, S. R., Arana, R. R., Cadena, S. V., Cadena, O. E. y Pichogagón, F. I. (2025). La inteligencia artificial en la educación: Una revisión sistemática de aplicaciones, beneficios y desafíos éticos. *Revista Latinoamericana de Calidad Educativa*, 2(4), 41-47. https://doi.org/10.70625/rlce/134
- Firdaus, T. (2025). The philosophical construction of educational science in relation to posthumanism and transhumanism in artificial intelligence. *Turkish Academic Research Review*, 10(1), 70-83. https://doi.org/10.30622/tarr.1610935
- Gallent, C., Arenas, B., Vallespir, M. y Foltynek, T. (2024). Inteligencia Artificial en educación: entre riesgos y potencialidades. *Praxis educativa*, 19. https://doi.org/10.5212/praxeduc.v.19.23760.083
- García, J. F. (2025). La Inteligencia Artificial como Herramienta para el Diseño de Estrategias Didácticas e Intervención Jurídica en la Educación Superior Ecuatoriana. *Latam: revista latinoamericana de Ciencias Sociales y Humanidades*, 6(3), 47. https://doi.org/10.56712/latam.v6i3.4192
- Gomes, D. (2025). A Comprehensive Study of Advancements in Intelligent Tutoring Systems Through Artificial Intelligent Education Platforms. In *Improving Student Assessment With Emerging AI Tools* (pp. 213-244). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-6170-2.ch008
- Hallström, J. (2022). Embodying the past, designing the future: technological determinism reconsidered in technology education. *International Journal of Technology Design Education*, 32(1), 17-31. https://doi.org/10.1007/s10798-020-09600-2
- Holzmeyer, C. (2021). Beyond 'AI for Social Good'(AI4SG): social transformations—not tech-fixes—for health equity. *Interdisciplinary Science Reviews*, 46(1-2), 94-125. https://doi.org/10.1080/03080188.2020.1840
- Hosseini, S. H. y Sakhaei, S. (2025). Educating intelligence, producing power: Iranian sociologists on AI, knowledge production, and global hierarchies. *Journal of World Sociopolitical Studies*, 9(4), 887-921. https://doi.org/10.22059/wsps.2025.396408.1529
- Islam, A., Ali, R., Singh, G., Islam, B., Islam, A. y Hossain, S. (2024). An evaluation of AI-enhanced collaborative learning platforms. *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*. https://doi.org/10.1109/IC3SE62002.2024.10593320
- Jindal, P. y Gouri, H. (2024). AI-personalization paradox: Navigating consumer behavior in a filter bubble era. In *AI Impacts in Digital Consumer Behavior* (pp. 82-111). IGI Global Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-1918-5.ch004
- Karakuş, N., Gedik, K. y Kazazoğlu, S. (2025). Ethical Decision-Making in Education: A Comparative Study of Teachers and Artificial Intelligence in Ethical Dilemmas. *Behavioral Sciences*, 15(4), 469. https://doi.org/10.3390/bs15040469
- Khan, W. N. (2024). Ethical challenges of AI in education: Balancing innovation with data privacy. *AI EDIFY Journal*, 1(1), 1-13. https://doi.org/10.56201/ijcsmt.vol.11.no4.2025
- Korteling, J. E., van de Boer, G. C., Blankendaal, R. A., Boonekamp, R. C. y Eikelboom, A. R. (2021). Human-versus artificial

- intelligence. *Frontiers in artificial intelligence*, 4, 622364. <https://doi.org/10.3389/frai.2021.622364>
- Koutsikouri, D., Hylving, L., Bornemark, J. y Lindberg, S. (2024). Human judgment in the age of automated decision-making systems. In *Research Handbook on Artificial Intelligence and Decision Making in Organizations* (pp. 144-159). Edward Elgar Publishing. <https://doi.org/10.4337/9781803926216.00017>
- Mura, L. y Stehlíková, B. (2025). The ethics of artificial intelligence: Safeguarding human dignity, social justice and environmental stability in the age of AI. *Equilibrium. Quarterly Journal of Economics Economic Policy*, 20(2), 479-507. <https://doi.org/10.24136/eq.3743>
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B. y Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education information technologies*, 28(4), 4221-4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Nopas, D.-s. (2025). Algorithmic learning or learner autonomy? Rethinking AI's role in digital education. *Qualitative Research Journal*. <https://doi.org/10.1108/QRJ-11-2024-0282>
- Rahm, L. y Rahm, J. (2023). Imaginaries and problematisations: A heuristic lens in the age of artificial intelligence in education. *British Journal of Educational Technology*, 54(5), 1147-1159. <https://doi.org/10.1111/bjet.13319>
- Rajan, I. (2025). Philosophy of technology for the lost age of freedom: a critical treatise on human essence and uncertain future. *AI Society*, 40(3), 1705-1722. <https://doi.org/10.1007/s00146-024-01942-y>
- Ranyal, E., Sadhu, A. y Jain, K. (2022). Road condition monitoring using smart sensing and artificial intelligence: A review. *Sensors*, 22(8), 3044. <https://doi.org/10.3390/s22083044>
- Román, C. A., Gamarra, J. O. M., Vallejo, M. V., Pisfil, F. y Alegría, C. M. (2024). Epistemología, Ética, Educación e Inteligencia Artificial. Editorial Internacional Alema. <https://editorialalema.org/libros/index.php/alema/article/view/29/32>
- Salloum, S. A. (2024). AI perils in education: Exploring ethical concerns. *Artificial intelligence in education: The power dangers of ChatGPT in the classroom*, 669-675. https://doi.org/10.1007/978-3-031-52280-2_43
- Selvakumar, P., Sudheer, P. y Kannan, N. (2025). Balancing Innovation and Privacy: Understanding AI in the Digital World. In *Digital Citizenship and the Future of AI Engagement, Ethics, and Privacy* (pp. 279-304). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-9015-3.ch011>
- Shalevska, E. (2024). Human Rights in the Age of AI: Understanding the Risks, Ethical Dilemmas, and the Role of Education in Mitigating Threats. *Journal of Legal Political Education*, 1(2), 38-52. <https://doi.org/10.47305/JLPE2412038sh>
- Tapalova, O. y Zhiyenbayeva, N. (2022). Artificial intelligence in education: AIEd for personalised learning pathways. *Electronic Journal of e-Learning*, 20(5), 639-653. <https://files.eric.ed.gov/fulltext/EJ1373006.pdf>
- Vinicio, M., Anneriz, B., Maldonado, P. R. y Carruyo, J. E. (2025). Hacia una educación universitaria humanista en la era de la inteligencia Artificial. reflexiones axiológicas contemporáneas.: Towards a humanist university education in the era of artificial intelligence. contemporary axiological reflections. *Revista Científica Multidisciplinar G-nerando*, 6(1), ág.

3129–3144-ág. 3129–3144.
<https://doi.org/10.60100/rcmg.v6i1.556>