Año VIII. Julio - Diciembre 2026

e-ISSN:2683-2143

QVADRATA

ESTUDIOS SOBRE EDUCACIÓN, ARTES Y HUMANIDADES

Tensiones éticas en la regulación de IA educativa: Análisis comparativo de marcos internacionales

[en] Ethical Tensions in Educational AI Regulation: Comparative Analysis of International Frameworks

Jesús Humberto Pérez Domínguez

Maricarmen Pérez Carbajal

Claudia Isela Gutiérrez Quezada

Luis Horacio Álvarez Peregrino

Rigoberto Marín Trejo

Centro de Investigación y Docencia (México)

Recibido: 2026/04/26 **Aceptado para su publicación**: 2026/05/12 **Publicado**: 2026/06/26

RESUMEN

Resumen: Los marcos normativos internacionales sobre inteligencia artificial en educación prometen orientar el desarrollo tecnológico hacia principios éticos fundamentales. Sin embargo, persiste una brecha significativa entre estos principios declarados y su implementación efectiva en contextos educativos reales. Este artículo examina esa desconexión mediante análisis documental de cuatro marcos normativos internacionales (UNESCO, Unión Europea, Consenso de Beijing y OCDE) y de literatura académica publicada entre 2018 y 2024. Empleando codificación temática, el estudio identifica cinco dimensiones éticas recurrentes: equidad, privacidad, transparencia, autonomía pedagógica y sesgo algorítmico. Todas aparecen en los marcos analizados, pero arrastran tensiones conceptuales y operativas no resueltas; el análisis profundiza en las tres más agudas. Primero, la equidad se entiende predominantemente como acceso tecnológico, omitiendo cuestiones de justicia epistémica sobre qué conocimientos privilegian los algoritmos y quién participa en su diseño. Segundo, la protección de datos personales enfrenta contradicciones específicas en contextos educativos donde el "consentimiento libre" resulta problemático ante la obligatoriedad escolar. Tercero, la autonomía docente frecuentemente invocada en los marcos convive con sistemas que reconfiguran decisiones pedagógicas sin participación genuina de quienes enseñan. Los hallazgos muestran que los marcos normativos actuales adoptan concepciones limitadas de estas dimensiones éticas, concentrándose en cumplimiento formal más que en transformación sustantiva de dinámicas de poder en el desarrollo e implementación de IA educativa. El estudio

ABSTRACT

Abstract: International regulatory frameworks on artificial intelligence in education promise to steer technological development toward sound ethical principles. Yet a clear gap remains between these stated principles and how they are implemented in real educational settings. This article examines that gap through a documentary analysis of four international frameworks (UNESCO, the European Union, the Beijing Consensus, and the OECD) and of academic literature published between 2018 and 2024. Using thematic coding, the study identifies five recurring ethical dimensions: equity, privacy, transparency, pedagogical autonomy, and algorithmic bias. All five appear across the frameworks, but each carries unresolved conceptual and practical tensions, and the analysis focuses on the three most pronounced. First, equity is mostly understood as access to technology, which sets aside questions of epistemic justice about whose knowledge algorithms privilege and who takes part in their design. Second, data protection meets specific contradictions in schools, where the idea of free consent is hard to sustain given compulsory attendance. Third, teacher autonomy, although frequently invoked, coexists with systems that reshape pedagogical decisions without the genuine involvement of teachers. The frameworks tend to adopt narrow versions of these dimensions and to stress formal compliance rather than a real shift in power. The study also finds a central pattern: the broader a framework's geographic scope, the vaguer its implementation mechanisms. We argue that closing this gap requires not only stronger mechanisms for monitoring and accountability, but also participatory governance and a rethinking of the priorities that guide educational technology.

identifica, además, un patrón central: a mayor alcance geográfico de un marco, menor es la especificidad de sus mecanismos de implementación. Los instrumentos globales enuncian principios amplios con mecanismos vagos, mientras que el único marco regional vinculante fija requisitos detallados, pero choca con límites de capacidad institucional.

PALABRAS CLAVE

inteligencia artificial educativa, regulación algorítmica, ética de la inteligencia artificial, gobernanza educativa, autonomía pedagógica.

KEYWORDS

educational artificial intelligence, algorithmic regulation, artificial intelligence ethics, educational governance, pedagogical autonomy.

Como citar (APA 7ª Edición):

Pérez, J. H., Pérez, M., Gutíerrez, C. I., Álvarez, L. H. y Marín, R. (2026). Tensiones éticas en la regulación de IA educativa: Análisis comparativo de marcos internacionales. *Qvadrata, estudios sobre educación, artes y humanidades*, *8*(16), 41-56. https://doi.org/10.54167/qvadrata.v8i16.2279

La adopción de inteligencia artificial en educación avanza rápidamente. Los sistemas actuales califican exámenes, recomiendan contenido y predicen resultados académicos. Sin embargo, esta expansión tecnológica genera interrogantes éticos sobre control, equidad y privacidad. Aunque organismos internacionales como UNESCO y la Unión Europea han establecido marcos regulatorios, existe una diferencia entre las normas declaradas y su implementación efectiva. Los estudios que comparan estos marcos confirman esa coincidencia. Fjeld et al. (2020), revisaron treinta y seis documentos internacionales de principios para la IA y encontraron acuerdo en ocho temas: privacidad, rendición de cuentas, seguridad, transparencia y explicabilidad, equidad y no discriminación, control humano de la tecnología, responsabilidad profesional y promoción de valores humanos. Pero coincidir en los principios no asegura que estos se cumplan en la práctica.

Mittelstadt (2019), explica por qué ese acuerdo no es suficiente. A diferencia de la medicina, de donde la ética de la IA tomó su modelo de principios, el desarrollo de IA no tiene fines comunes ni deberes claros hacia los usuarios, carece de una tradición profesional establecida, no cuenta con métodos probados para llevar los principios a la práctica y tampoco dispone de formas firmes de exigir responsabilidad legal o profesional. Por eso advierte que el acuerdo en principios generales puede ocultar desacuerdos de fondo.

El presente estudio analiza esta desconexión y sus implicaciones para las instituciones educativas. Estas tecnologías plantean algunas preguntas éticas importantes: ¿Quién controla estos sistemas? ¿Son justos para todos los estudiantes? ¿Qué datos recopilan?, entre otras. Giannini (2023), señala que mientras aún intentamos comprender las implicaciones de revoluciones digitales anteriores, hemos despertado abruptamente para encontrarnos entrando en otra revolución digital, una que puede hacer que las otras parezcan menores en comparación.

Entre 2020 y 2024, organismos internacionales como UNESCO, OCDE y Comisión Europea publicaron marcos normativos que intentan orientar el desarrollo y uso de IA en educación. Al mismo tiempo, gobiernos nacionales han comenzado a implementar regulaciones específicas con niveles variables de especificidad y alcance. Estos marcos normativos comparten premisas sobre beneficios potenciales de la IA educativa: personalización de trayectorias de aprendizaje, automatización de tareas administrativas, identificación temprana de dificultades académicas y ampliación de acceso a recursos educativos.

Sin embargo, la literatura académica reciente documenta casos donde la implementación de estos sistemas ha generado consecuencias no previstas. Estas incluyen amplificación de desigualdades existentes, pérdida de control docente sobre sus clases y normalización de vigilancia estudiantil. La distancia entre promesas normativas y resultados documentados plantea interrogantes sobre la suficiencia de los marcos éticos actuales para orientar decisiones técnicas con consecuencias sociales.

Reconociendo la complejidad de este fenómeno, el estudio se propone no solo identificar las brechas existentes, sino comprender sus implicaciones prácticas para quienes toman decisiones en instituciones educativas. Con este fin, se plantean los siguientes objetivos:

Pregunta general

¿Cómo abordan los marcos normativos internacionales las implicaciones éticas del uso de inteligencia artificial en educación y qué coherencia existe entre sus principios declarados y sus mecanismos de implementación?

Preguntas específicas.

- ¿Qué dimensiones éticas priorizan los marcos normativos internacionales sobre IA en educación y qué convergencias o divergencias existen entre ellos?
- ¿Qué contradicciones surgen al intentar aplicar principios como equidad, privacidad o transparencia en escuelas y aulas concretas?
- ¿Qué diferencias existen entre los principios éticos declarados y los mecanismos de supervisión que realmente proponen los marcos normativos?

Objetivo general

Analizar las dimensiones éticas presentes en marcos normativos internacionales sobre inteligencia artificial en educación y sus mecanismos de implementación propuestos.

Objetivos específicos.

- Identificar las dimensiones éticas prioritarias en marcos normativos internacionales sobre IA educativa publicados entre 2018 y 2024.
- Identificar las contradicciones que aparecen al intentar aplicar principios éticos generales en contextos educativos específicos.
- Evaluar las diferencias entre principios éticos declarados y mecanismos de supervisión propuestos en los marcos normativos

Para responder estas interrogantes, el presente estudio adopta un enfoque cualitativo centrado en el análisis documental, cuyas características se describen a continuación.

Metodología

Esta investigación examina dos tipos de fuentes publicadas entre 2018 y 2024. La primera es un conjunto de cuatro marcos normativos internacionales que se analizan en profundidad por representar enfoques regulatorios distintos: la Recomendación de UNESCO sobre Ética de IA (2021), el Reglamento de IA de la Unión Europea, el Consenso de Beijing sobre IA y Educación (2019) y las Directrices de la OCDE sobre IA (2019). De forma complementaria se consultaron otros instrumentos como apoyo contextual: el Reglamento General de Protección de Datos, las guías de la Comisión Europea sobre IA confiable (2019) y la Convención Marco del Consejo de Europa. La segunda fuente es la literatura académica sobre inteligencia artificial aplicada a la educación, identificada mediante búsquedas en bases de datos especializadas.

La búsqueda académica utilizó bases de datos especializadas (ERIC, Google Scholar, Web of Science) con combinaciones de términos en inglés y español: "artificial intelligence + education + ethics", "algorithmic bias + educational technology", "data privacy + students", "teacher autonomy +

e-ISSN: 2683-2143

automation", "algorithmic accountability + schools" y "ética + inteligencia artificial + educación". Se priorizaron estudios empíricos sobre implementación de IA en contextos educativos reales, análisis de marcos normativos existentes y propuestas teóricas sobre gobernanza de sistemas algorítmicos. Se excluyeron textos que abordaran la IA educativa exclusivamente desde perspectivas técnicas o de ingeniería, sin considerar dimensiones sociales o éticas. La selección de la literatura fue intencional y no exhaustiva, acorde con el carácter cualitativo del estudio: se incluyeron los textos que abordaban de manera sustantiva las dimensiones éticas analizadas, hasta que las nuevas lecturas dejaron de aportar categorías, sin buscar cubrir toda la producción disponible.

El análisis siguió un proceso de codificación temática de tipo inductivo, en el que las categorías se construyeron a partir de los documentos y no de un esquema previo. El procedimiento tuvo tres fases. En la primera, una lectura abierta identificó temas recurrentes sobre ética de la IA educativa. En la segunda, esas observaciones se agruparon en categorías provisionales, con el criterio de que cada categoría reuniera un problema ético tratado de forma explícita y reiterada en varios documentos. En la tercera, las categorías se depuraron mediante relectura y comparación constante entre fuentes, hasta estabilizarse en cinco dimensiones: equidad, privacidad, transparencia, autonomía pedagógica y sesgo algorítmico.

Una vez establecidas las categorías, se examinó cómo cada marco normativo las abordaba. Algunos documentos priorizan la protección de datos mediante restricciones técnicas específicas. Otros enfatizan principios generales como la transparencia o la supervisión humana sin definir mecanismos de implementación. Algunos marcos presentan la IA como una herramienta neutral que puede mejorar la educación si se usa correctamente, mientras otros cuestionan si ciertos usos son compatibles con objetivos educativos que van más allá de la eficiencia.

Este enfoque tiene limitaciones reconocibles. La revisión se concentra en documentos escritos en inglés y español, lo que excluye debates importantes en otros idiomas. Los marcos normativos analizados provienen principalmente de contextos con alta capacidad institucional, dejando fuera perspectivas de regiones donde la implementación de IA educativa ocurre sin marcos regulatorios formales. La literatura académica incluida se sesga hacia publicaciones en revistas indexadas, un filtro que puede excluir investigación regional relevante o perspectivas publicadas en formatos alternativos.

Resultados

Dimensiones éticas en marcos normativos sobre IA educativa

Equidad: entre acceso tecnológico y justicia epistémica

Los marcos normativos analizados mencionan la equidad como principio rector, pero sus conceptualizaciones difieren ampliamente. La recomendación de UNESCO sobre ética de la IA (2021) define equidad en términos de acceso universal a tecnologías educativas. Esta definición entiende la inequidad principalmente como problema de distribución de recursos materiales: infraestructura, dispositivos, conectividad.

El marco de la Comisión Europea para IA confiable (2019) amplía esta definición al incluir consideraciones sobre diseño inclusivo de sistemas que consideren diversidad de capacidades, lenguas y contextos culturales. Sin embargo, la literatura académica documenta limitaciones de ambas perspectivas cuando se enfrentan a casos concretos.

El estudio de Baker y Hawn (2022), sobre sistemas de tutoría inteligente en escuelas estadounidenses reveló que, incluso garantizando la paridad en el acceso tecnológico, los algoritmos reproducen supuestos sobre cómo deben aprender los estudiantes. Penalizando a quienes no siguen trayectorias "normales", estos sistemas terminan excluyendo precisamente a los estudiantes con patrones de progreso menos convencionales, aquellos cuyos datos no coinciden con los modelos de entrenamiento. La inequidad, en este caso, no es producto de la falta de dispositivos sino del diseño algorítmico mismo.

e-ISSN: 2683-2143

D'Ignazio y Klein (2020), plantean preguntas sobre quién decide qué conocimientos son válidos, qué formas de aprender se consideran legítimas y qué voces participan en el diseño de sistemas que configuran experiencias educativas. Esta perspectiva cambia la atención desde la distribución de acceso hacia la estructura de poder en el desarrollo tecnológico. Los marcos normativos analizados incorporan esta dimensión de manera limitada, sin abordarla como cuestión central de diseño tecnológico.

Las diferencias entre estas conceptualizaciones tienen consecuencias prácticas. Como argumenta Giannini (2023), a pesar de las promesas de la IA y otras tecnologías digitales de diversificar sistemas de conocimiento, podemos estar moviéndonos en la dirección opuesta. Si la equidad se entiende solo como acceso, las políticas se concentran en adquisición de tecnología. Si se entiende como inclusión en el diseño, las políticas exigen participación de diversos usuarios en procesos de desarrollo. Si se trata de justicia en el conocimiento, las políticas deben preguntarse qué tipo de aprendizaje promueven los algoritmos. Los marcos actuales no explicitan cuál concepción adoptan, lo que dificulta evaluar si las medidas propuestas son coherentes con sus objetivos declarados.

Privacidad y protección de datos: Entre cumplimiento normativo y vigilancia normalizada

La protección de datos personales de estudiantes aparece en todos los marcos normativos revisados como preocupación central. El Reglamento General de Protección de Datos de la Unión Europea (2016), establece estándares sobre consentimiento informado, minimización de recolección de datos, limitación de propósito y derecho al olvido. Estos estándares han influido en legislaciones de otros países. Sin embargo, la aplicación de estos principios en contextos educativos enfrenta complicaciones específicas que los marcos generales no anticipan.

Zuboff (2019), documenta cómo plataformas educativas recolectan datos que exceden lo necesario para sus funciones declaradas. Esta práctica genera un desequilibrio donde instituciones educativas y familias carecen de información completa sobre qué datos se recolectan, cómo se procesan y con qué fines secundarios se utilizan.

Selwyn et al. (2021), documentan que tanto estudiantes como docentes carecen de conocimiento sobre los mecanismos mediante los cuales se recopilan, procesan y utilizan los datos que producen al interactuar con plataformas educativas digitales. Los términos de servicio de plataformas educativas populares contienen cláusulas que autorizan usos amplios de información estudiantil, pero estos documentos están redactados en lenguaje técnico-legal que dificulta su comprensión por parte de usuarios no especializados.

La cuestión del consentimiento presenta problemas particulares en educación. Cuando una institución educativa adopta una plataforma de IA como parte de su operación regular, ¿qué significa que estudiantes o familias consientan? ¿Existe posibilidad real de rechazar el uso si esto implica exclusión de actividades académicas regulares? Williamson (2017), argumenta que hablar de consentimiento libre en contextos donde existe obligatoriedad educativa resulta conceptualmente problemático.

Adicionalmente, las plataformas de aprendizaje adaptativo registran cada interacción: tiempo dedicado a cada ejercicio, secuencia de respuestas, patrones de error, momentos de abandono, entre otros. Este nivel de detalle permite conjeturas sobre estados emocionales, capacidad de atención y dificultades cognitivas. Los datos no solo describen el rendimiento académico inmediato, sino que construyen registros que pueden influir en cómo se percibe al estudiante a largo plazo.

Los marcos normativos actuales recomiendan limitar recolección de datos a lo estrictamente necesario, pero no definen qué cuenta como necesario en contextos educativos. Giannini (2023), advierte que uno de los riesgos primarios y más evidentes de la IA es su potencial para manipular a usuarios humanos, y que los niños y jóvenes son altamente susceptibles a la manipulación.

Los marcos tampoco abordan el problema de inferencias derivadas de esto. Aunque una institución no recolecte directamente información sobre condición socioeconómica, algoritmos pueden

e-ISSN: 2683-2143

inferirla por los patrones de búsqueda, dispositivos utilizados o tipos de errores cometidos. La protección de privacidad requiere considerar no solo qué datos se recolectan explícitamente, sino qué información puede extraerse mediante procesamiento algorítmico.

Transparencia algorítmica: límites técnicos y políticos de la explicabilidad

La demanda de transparencia en sistemas de IA educativa aparece consistentemente en documentos normativos, pero su operacionalización enfrenta obstáculos técnicos y políticos. El Reglamento de IA de la Unión Europea (2024), clasifica aplicaciones educativas como alto riesgo y exige que sus decisiones sean explicables. Sin embargo, qué constituye una explicación adecuada permanece en disputa.

Burrell (2016), identifica tres tipos de opacidad en sistemas algorítmicos. La opacidad corporativa surge del secreto comercial que impide acceso a códigos. La opacidad técnica resulta del analfabetismo especializado que dificulta comprensión por no expertos. La opacidad epistémica deriva de la complejidad intrínseca de modelos de aprendizaje profundo que impide explicaciones causales simples. Los marcos normativos suelen asumir que la transparencia puede resolverse mediante documentación técnica o interfaces que muestren factores considerados en decisiones algorítmicas. Esta aproximación ignora que incluso desarrolladores de sistemas complejos de aprendizaje automático no siempre pueden explicar por qué un modelo produce cierto resultado.

La investigación de Holstein et al. (2019), con docentes que utilizan sistemas de IA educativa muestra que explicaciones técnicas sobre funcionamiento algorítmico no necesariamente les permiten tomar mejores decisiones pedagógicas. Los docentes necesitan entender no cómo funciona matemáticamente un algoritmo, sino qué supuestos sobre aprendizaje incorpora, qué evidencia respalda esos supuestos y en qué contextos el sistema ha mostrado limitaciones. Este tipo de transparencia requiere documentación centrada en implicaciones pedagógicas más que en especificaciones técnicas, lo cual resulta determinante para la toma de decisiones y el diseño de estrategias pedagógicas.

La opacidad de los sistemas de IA educativa tiene implicaciones específicas sobre el conocimiento mismo. Como señala Giannini (2023), las respuestas proporcionadas por chatbots de IA no se remontan a mentes humanas. Más bien, provienen de un laberinto de cálculos tan complejos que no son completamente comprensibles incluso para las personas que desarrollan la tecnología. Esta característica genera un problema de conocimiento donde las respuestas carecen de humanidad y pueden llevar a un estado donde el conocimiento generado por máquinas se vuelve dominante.

Algunos autores argumentan que el énfasis en transparencia algorítmica puede desviar atención de cuestiones más sustantivas. Ananny y Crawford (2018), sostienen que la transparencia no garantiza rendición de cuentas si no existen mecanismos para cuestionar o modificar decisiones algorítmicas. Un sistema completamente transparente puede seguir generando consecuencias injustas si los afectados carecen de poder para impugnar sus resultados. La transparencia es condición necesaria pero insuficiente para la gobernanza democrática de IA educativa.

Autonomía pedagógica: Redefinición del trabajo docente

La introducción de sistemas de IA en prácticas educativas modifica la distribución de decisiones entre docentes, algoritmos y administradores. Los marcos normativos revisados afirman que la IA debe apoyar o asistir a docentes sin reemplazarlos. El informe del Departamento de Educación de Estados Unidos (2023), ilustra bien esta postura: recomienda mantener siempre a las personas en el centro de las decisiones (humans in the loop) y sostiene que la IA debe apoyar a los docentes, no sustituirlos, dejando en sus manos las decisiones de enseñanza y de evaluación.

Sin embargo, esta distinción resulta difícil de mantener en implementaciones concretas. Cuando un sistema de tutoría inteligente recomienda actividades específicas para cada estudiante basándose en

e-ISSN: 2683-2143

análisis de datos, ¿el docente conserva autonomía pedagógica si seguir esas recomendaciones se convierte en expectativa institucional o un mandato administrativo?

El estudio etnográfico de Nemorin (2017), en escuelas británicas que implementaron sistemas de análisis predictivo documenta cómo docentes experimentan tensión entre su juicio profesional y las predicciones algorítmicas. Cuando un sistema identifica estudiantes en riesgo de bajo desempeño, docentes sienten presión para intervenir según lo indicado, incluso cuando su experiencia sugiere que la predicción puede ser incorrecta. Esta presión se intensifica cuando administradores utilizan concordancia entre decisiones docentes y recomendaciones algorítmicas como criterio de evaluación de desempeño.

La autonomía pedagógica no se refiere solo a libertad para tomar decisiones, sino a capacidad para ejercer juicio profesional basado en conocimiento especializado sobre contextos específicos de aprendizaje. Los sistemas de IA educativa suelen operar bajo supuestos universales. Este supuesto entra en conflicto con tradiciones pedagógicas que enfatizan el trato particular de cada situación educativa y la imposibilidad de reducir el aprendizaje a variables medibles.

Algunos marcos normativos recomiendan que docentes reciban formación en competencias digitales para trabajar efectivamente con IA. Sin embargo, esta formulación asume que la adaptación debe ocurrir del lado docente, sin cuestionar si los sistemas están diseñados de manera que respeten formas establecidas de expertise pedagógico. Selwyn et al. (2021), plantean la necesidad de cuestionar procesos de automatización en educación, particularmente aquellos relacionados con evaluación de aprendizajes, gestión de espacios educativos y plataformas que individualizan trayectorias formativas. Esta automatización frecuentemente oculta procesos de descualificación profesional donde decisiones que anteriormente requerían juicio experto se delegan a algoritmos.

La cuestión de la autonomía se complica cuando sistemas de IA median la relación entre docentes y estudiantes. Plataformas de aprendizaje adaptativo crean experiencias personalizadas donde cada estudiante recorre trayectorias distintas. Esto dificulta que docentes mantengan conocimiento compartido sobre lo que el grupo está aprendiendo, erosionando posibilidades de discusión colectiva y construcción social de conocimiento. La personalización algorítmica puede resultar en atomización de experiencias educativas.

Giannini (2023), observa que sería ingenuo pensar que las utilidades futuras de IA no fortalecerán los llamados a una mayor automatización de la educación: escuelas sin maestros, educación sin escuelas y otras visiones distópicas. Esta preocupación se intensifica cuando consideramos que la automatización digital de la educación históricamente se ha propuesto como solución para comunidades donde los desafíos educativos son más severos, impactando primero a los estudiantes más desfavorecidos.

Sesgo algorítmico: Cuando los datos reproducen desigualdades

Los marcos normativos advierten sobre riesgos de sesgo en sistemas de IA educativa, pero sus recomendaciones suelen limitarse a sugerencias genéricas sobre diversificar datos de entrenamiento o auditar resultados. La literatura académica documenta casos específicos donde sesgos algorítmicos han generado consecuencias discriminatorias, mostrando que el problema es más complejo que malas prácticas técnicas corregibles mediante mejores datos.

El análisis de Obermeyer y otros (2019), mostró que un algoritmo ampliamente utilizado en contextos de salud exhibía sesgo racial porque utilizaba gasto médico como indicador de necesidad de atención. Este sesgo surgía sin considerar que disparidades en acceso a servicios de salud hacían que pacientes negros con igual necesidad clínica tuvieran menores gastos registrados. Este caso ilustra cómo sesgos no derivan necesariamente de prejuicios codificados explícitamente, sino de la elección de variables que reflejan desigualdades estructurales.

e-ISSN: 2683-2143

Barocas y Selbst (2016), ya habían descrito este mecanismo. La discriminación que producen los sistemas de minería de datos no suele ser una decisión consciente de quien los programa, sino un resultado que emerge del proceso mismo. Decisiones como qué se quiere predecir, qué datos se usan para entrenar el modelo o qué rasgos se toman en cuenta pueden afectar de forma desigual a ciertos grupos, incluso cuando nadie tuvo la intención de discriminar.

En educación, sistemas de predicción de deserción escolar frecuentemente utilizan indicadores como ausentismo, calificaciones previas y participación en actividades extracurriculares. Estos indicadores correlacionan con nivel socioeconómico, lo que hace que algoritmos identifiquen estudiantes de menores recursos como en riesgo con mayor frecuencia, incluso cuando controlan por desempeño académico. Esta sobreidentificación puede generar profecías autocumplidas: estudiantes categorizados como en riesgo reciben intervenciones estigmatizantes que afectan su motivación y sentido de pertenencia escolar.

El trabajo de Noble (2018), sobre algoritmos de búsqueda documenta cómo sesgos de género y raza se reproducen en sistemas de información porque los datos que los entrenan reflejan estereotipos sociales existentes. Si un sistema de recomendación de carreras se entrena con datos históricos sobre qué estudiantes ingresaron a qué campos, reproducirá patrones de segregación ocupacional por género. Corregir este sesgo requiere decisiones normativas sobre qué constituye una recomendación justa, no solo mejoras técnicas.

Algunos autores argumentan que la noción misma de sesgo es problemática porque asume la existencia de un estándar neutral contra el cual medir desviaciones. Hoffmann (2019), propone que, en lugar de buscar algoritmos libres de sesgo, deberíamos preguntarnos qué valores queremos que los sistemas reflejen y cómo asegurar que esos valores sean decididos democráticamente más que impuestos por diseñadores técnicos o fuerzas de mercado.

Análisis comparativo de mecanismos de implementación en marcos normativos seleccionados

Para evaluar la correspondencia entre principios éticos declarados y mecanismos de implementación propuestos, se analizaron cuatro marcos normativos que representan diferentes aproximaciones regulatorias. Estos marcos son la Recomendación de UNESCO sobre Ética de IA (2021), el Reglamento de Inteligencia Artificial de la Unión Europea (2024), el Consenso de Beijing sobre IA y Educación (2019), y las Directrices de la OCDE sobre IA (2019). Este análisis comparativo destaca diferencias sustanciales en especificidad, obligatoriedad y viabilidad de los mecanismos propuestos.

Tabla 1. *Análisis comparativo de mecanismos de implementación en marcos normativos seleccionados*

Marco (año y naturaleza)	Principios éticos	Mecanismos de implementación	Principales limitaciones
UNESCO. Recomendación sobre Ética de IA (2021). Instrumento global, no vinculante (soft law).	Dignidad humana, justicia, protección de datos, transparencia y explicabilidad, supervisión humana.	Evaluaciones de impacto ético durante el ciclo de vida; el resto, en términos generales.	No fija estándares mínimos, autoridades responsables ni consecuencias por incumplir.
Unión Europea. Reglamento de IA, 2024 (propuesta de	IA educativa como alto riesgo: gestión de riesgos,	Requisitos obligatorios y autoridades	Depende de capacidades de supervisión que casi

e-ISSN: 2683-2143

Marco (año y naturaleza)	Principios éticos	Mecanismos de implementación	Principales limitaciones
2021). Instrumento regional, vinculante.	gobernanza de datos, documentación, transparencia, supervisión humana, datos representativos y control de sesgos.	nacionales con poder de investigación y sanción (hasta 15 millones de euros o 3 % en alto riesgo; 35 millones o 7 % en prácticas prohibidas).	ningún país tiene aún; su nivel de detalle puede volverse rígido.
UNESCO y gobierno de China. Consenso de Beijing sobre IA y Educación (2019). Declaración internacional, no vinculante.	Uso inclusivo y equitativo, privacidad y seguridad de datos, transparencia y auditabilidad, IA que apoya a los docentes.	Casi inexistentes; pide a los gobiernos crear marcos regulatorios apropiados.	Más débil que UNESCO 2021; tono optimista; no evalúa si la IA amplía o reduce brechas.
OCDE. Directrices sobre IA (2019). Principios intergubernamentales, no vinculantes (soft law).	Crecimiento inclusivo, valores centrados en las personas, transparencia y explicabilidad, robustez y seguridad, rendición de cuentas.	Deliberadamente débiles; remiten a las leyes nacionales; sin supervisión ni auditoría.	Formulaciones amplias y ambiguas; su influencia depende de decisiones voluntarias.

Nota. Elaboración propia a partir del análisis documental de los cuatro marcos.

Recomendación de UNESCO sobre Ética de IA (2021)

La Recomendación de UNESCO articula principios éticos comprehensivos organizados en torno a valores como dignidad humana, justicia y protección de datos. Sin embargo, sus mecanismos de implementación presentan ambigüedad sistemática. Para la dimensión de equidad, el documento recomienda que los Estados miembros deberían esforzarse por garantizar la inclusión de diversos actores en todas las etapas del ciclo de vida de los sistemas de IA. Esta formulación no especifica qué constituye inclusión, quiénes son los diversos actores en contextos educativos específicos, o qué consecuencias se derivan de no cumplir esta recomendación.

En materia de privacidad, la Recomendación establece que los sistemas de IA no deberían comprometer la privacidad de datos y deben garantizar protección adecuada. Sin embargo, el documento no define estándares mínimos de protección adecuada para datos educativos, ni propone autoridades responsables de verificar cumplimiento. La transparencia se aborda mediante el principio de que debe existir transparencia y explicabilidad de los sistemas de IA, pero el documento no establece niveles requeridos de explicabilidad según tipo de aplicación educativa.

El único mecanismo de implementación relativamente específico que propone UNESCO es la creación de evaluaciones de impacto ético que deberían realizarse a lo largo del ciclo de vida de los

e-ISSN: 2683-2143

sistemas de IA. No obstante, el documento no especifica quién debe conducir estas evaluaciones, con qué metodología, en qué momentos precisos del ciclo de vida, ni qué sucede cuando una evaluación identifica problemas éticos graves. Esta falta de especificidad operativa caracteriza el enfoque de UNESCO: principios robustos con mecanismos de implementación débiles que dependen de voluntad política de Estados miembros sin crear obligaciones verificables.

Al tratarse de una recomendación, este instrumento no obliga jurídicamente: orienta la formulación de políticas nacionales, pero su cumplimiento queda a criterio de cada Estado y no puede exigirse por vía legal. Su influencia real depende de que los gobiernos decidan convertirlo en normas internas.

Reglamento de Inteligencia Artificial de la Unión Europea (2024)

El reglamento de la Unión Europea representa el extremo opuesto en términos de especificidad y obligatoriedad legal. El documento clasifica sistemas de IA educativa como alto riesgo y establece requisitos vinculantes. Estos requisitos incluyen sistemas de gestión de riesgos, gobernanza de datos, documentación técnica, registro automático de eventos, transparencia hacia usuarios y supervisión humana. A diferencia de las recomendaciones de UNESCO, estos requisitos son legalmente obligatorios. En su versión finalmente adoptada, Reglamento (UE) 2024/1689, en vigor desde agosto de 2024, el incumplimiento de las obligaciones aplicables a los sistemas de alto riesgo, categoría en la que se ubican los sistemas de IA educativa, puede sancionarse con multas de hasta 15 millones de euros o el 3 % de la facturación mundial anual total, lo que sea mayor; monto que asciende a 35 millones de euros o el 7 % en el caso de las prácticas prohibidas.

Para la dimensión de equidad, el Reglamento exige que los conjuntos de datos de entrenamiento sean relevantes, representativos, libres de errores y completos considerando el propósito previsto del sistema. Además, requiere que los proveedores examinen los datos de entrenamiento con respecto a posibles sesgos y implementen medidas apropiadas de detección, prevención y mitigación de sesgos. Este nivel de especificidad supera ampliamente lo propuesto por UNESCO, aunque persisten ambigüedades sobre qué constituye un conjunto de datos suficientemente representativo en contextos educativos diversos.

En privacidad, el Reglamento se articula con el GDPR existente, estableciendo que los sistemas de IA deben cumplir requisitos de protección de datos personales ya codificados en legislación europea. Para transparencia, el documento exige que sistemas de alto riesgo proporcionen a usuarios información concisa, completa, correcta y clara sobre capacidades y limitaciones, en formato apropiado y comprensible. Sin embargo, el Reglamento no define estándares específicos sobre qué constituye explicabilidad apropiada para diferentes audiencias educativas: estudiantes, docentes, familias o administradores.

El mecanismo de supervisión propuesto incluye autoridades nacionales de vigilancia con poderes de investigación, acceso a datos y capacidad sancionatoria. No obstante, la efectividad de este mecanismo depende críticamente de capacidades técnicas que actualmente no existen en la mayoría de estados miembros. El Reglamento requiere que autoridades nacionales competentes tendrán recursos técnicos, financieros y humanos suficientes, pero no especifica niveles mínimos ni cómo Estados con presupuestos limitados desarrollarán expertise necesario para auditar sistemas algorítmicos complejos.

Consenso de Beijing sobre IA y Educación (2019)

El Consenso de Beijing, producido por la Conferencia Internacional sobre IA y Educación organizada por UNESCO y el gobierno chino, adopta una aproximación centrada en oportunidades más que en riesgos. El documento articula principios sobre IA al servicio de la educación, pero sus mecanismos de implementación son incluso más débiles que los de la Recomendación de UNESCO de 2021.

e-ISSN: 2683-2143

Para equidad, el Consenso recomienda garantizar uso inclusivo y equitativo de IA en educación mediante desarrollo de políticas que aborden brechas digitales. Esta formulación asume que inequidad deriva principalmente de acceso diferenciado a tecnología, sin considerar cómo sistemas de IA pueden perpetuar desigualdades incluso cuando existe paridad de acceso. El documento no propone mecanismos para evaluar si implementaciones específicas de IA educativa amplían o reducen brechas existentes.

En privacidad, el Consenso establece que debe protegerse la privacidad individual y la seguridad de datos sin especificar estándares mínimos, autoridades responsables o procedimientos de verificación. La transparencia se menciona mediante el principio de que algoritmos deben ser transparentes, auditables y comprensibles, pero nuevamente sin operacionalización concreta. Respecto a autonomía docente, el documento afirma que IA debe empoderar docentes, una formulación que no reconoce dificultades documentadas donde sistemas algorítmicos erosionan más que empoderan autonomía profesional.

El único mecanismo de implementación que el Consenso propone es que gobiernos deben desarrollar marcos regulatorios apropiados, una recomendación circular que delega responsabilidad sin proporcionar orientación sobre qué constituyen marcos apropiados. El Consenso de Beijing refleja optimismo tecnológico característico de documentos producidos en contextos donde existe fuerte impulso gubernamental y corporativo hacia adopción acelerada de IA, con menor énfasis en gobernanza robusta y protección de derechos.

Directrices de la OCDE sobre IA (2019)

Las Directrices de la OCDE sobre IA representan un intento de articular principios que puedan adoptarse por países con diferentes sistemas legales y niveles de desarrollo económico. El documento establece cinco principios: crecimiento inclusivo y desarrollo sostenible, valores centrados en humanos, transparencia y explicabilidad, robustez y seguridad, y rendición de cuentas. Aunque no son específicas al sector educativo, varios países han utilizado estos principios como base para políticas nacionales sobre IA educativa.

Para equidad, las Directrices recomiendan que actores de IA deben respetar estado de derecho, derechos humanos y valores democráticos y promover inclusión social. Esta formulación amplia permite interpretaciones divergentes: ¿qué significa concretamente promover inclusión social al diseñar un sistema de tutoría inteligente? El documento no lo especifica. En privacidad, las Directrices establecen que sistemas de IA deben ser robustos, seguros y protegidos a lo largo de su ciclo de vida, pero remiten a legislaciones nacionales existentes sobre protección de datos sin proponer estándares mínimos globales.

La transparencia se aborda mediante el principio de que debe haber divulgación significativa sobre sistemas de IA para fomentar comprensión general. El calificativo significativo introduce flexibilidad interpretativa sin criterios claros sobre qué nivel de divulgación resulta suficiente. Para rendición de cuentas, las Directrices establecen que actores de IA deben ser responsables del funcionamiento apropiado de sistemas, pero no definen qué constituye funcionamiento apropiado ni quién determina cuando un sistema falla en cumplir este estándar.

Los mecanismos de implementación de las Directrices de OCDE son deliberadamente débiles debido a su naturaleza como instrumento de soft law diseñado para consenso amplio. El documento recomienda que países inviertan en investigación sobre IA y fomenten ecosistemas de IA confiable, formulaciones aspiracionales que no crean obligaciones verificables. A diferencia del Reglamento de la UE, las Directrices no establecen autoridades de supervisión, procedimientos de auditoría o consecuencias por incumplimiento. Su influencia depende enteramente de decisiones voluntarias de gobiernos nacionales sobre cómo traducir principios generales en regulaciones específicas.

e-ISSN: 2683-2143

Análisis comparativo de especificidad y viabilidad

La comparación entre estos cuatro marcos muestra un patrón consistente: existe relación inversa entre alcance geográfico de un marco normativo y especificidad de sus mecanismos de implementación. Los marcos con aspiración global articulan principios éticos amplios con mecanismos de implementación vagos, mientras que el marco regional establece requisitos específicos y vinculantes con mecanismos de supervisión detallados, pero enfrenta desafíos de capacidad institucional para implementación efectiva.

Este patrón refleja diferencias entre diferentes objetivos regulatorios. Marcos globales priorizan consenso amplio entre países con contextos políticos, económicos y culturales diversos, lo que requiere formulaciones generales para acomodar variación contextual. Algo similar observa Hagendorff (2020): tras comparar veintidós directrices éticas sobre IA, encontró que la rendición de cuentas, la privacidad y la equidad aparecen en cerca del 80 % de ellas, pero que el campo no tiene forma de hacer cumplir lo que propone; apartarse de los códigos éticos no trae consecuencias, así que muchas veces la ética funciona más como una carta de presentación que como un límite real.

El Reglamento de la UE evita esta trampa mediante requisitos específicos y vinculantes, pero enfrenta el desafío opuesto: alta especificidad puede resultar en rigidez que dificulta adaptación a contextos educativos diversos dentro de Europa. Además, la efectividad del Reglamento depende de capacidades de supervisión que deben construirse desde cero en la mayoría de los países. Como documenta el AI Index Report 2023 citado por Giannini (2023), menos del 1% de expertos en IA trabajan en gobierno, creando asimetría de expertise donde desarrolladores privados poseen conocimiento técnico que reguladores carecen.

Vacíos en marcos normativos actuales.

El análisis de estos documentos normativos indica ausencias significativas que limitan su capacidad para orientar gobernanza efectiva de IA educativa. Tres vacíos resultan particularmente problemáticos: la falta de atención a dimensiones de poder en el desarrollo tecnológico, la ausencia de mecanismos de participación de comunidades educativas en decisiones sobre adopción de sistemas, y la insuficiente consideración de alternativas no tecnológicas para problemas educativos.

Los marcos actuales tratan la IA educativa como herramienta neutral cuyas consecuencias dependen del uso que se le dé. Esta perspectiva ignora que tecnologías incorporan valores y supuestos desde su diseño. La decisión de desarrollar un sistema de predicción de deserción escolar en lugar de invertir recursos en reducir tamaños de grupo refleja prioridades sobre cómo abordar problemas educativos. La concentración del desarrollo de IA educativa en corporaciones tecnológicas privadas genera estructuras de dependencia donde instituciones educativas públicas pierden capacidad de determinar sus propias trayectorias tecnológicas.

Bietti (2020), documenta cómo grandes corporaciones tecnológicas han establecido laboratorios de innovación educativa en escuelas públicas que funcionan como espacios de experimentación para tecnologías posteriormente comercializadas. Estas asociaciones generan asimetrías: corporaciones acceden a datos y contextos reales de implementación, mientras que escuelas dependen de donaciones tecnológicas sin desarrollar capacidades propias. Los marcos normativos no cuestionan estas estructuras de poder, limitándose a recomendar que las asociaciones público-privadas respeten principios éticos sin examinar si esas asociaciones son deseables en sí mismas.

Giannini (2023), plantea un dilema de inversión que los marcos normativos no abordan: ¿Hasta qué punto deberíamos dirigir inversiones, incluyendo inversiones públicas, hacia la construcción de capacidades de máquinas que actúan como humanos inteligentes, o hacia la construcción de capacidades de personas vivas? Esta pregunta cobra urgencia cuando se observa que miles de millones de dólares ahora se están invirtiendo en compañías de IA generativa, cuando podrían dirigirse hacia el desarrollo docente y hacer mejoras necesarias a las escuelas. La autora advierte sobre la posibilidad de que las inversiones dirigidas a hacer la IA más inteligente y capaz podrían algún día superar las inversiones

e-ISSN: 2683-2143

dirigidas a educar niños y otras personas. Esta reconfiguración de prioridades de inversión representa un cambio estructural en cómo las sociedades valoran el desarrollo humano frente al desarrollo tecnológico, tema que los marcos normativos actuales no contemplan.

La participación de comunidades educativas en decisiones sobre adopción de IA es mencionada superficialmente en algunos documentos, pero ninguno especifica mecanismos concretos para hacerla efectiva. ¿Cómo pueden docentes, estudiantes y familias participar significativamente en decisiones técnicas complejas? ¿Qué información necesitan para evaluar implicaciones de diferentes sistemas? ¿En qué momento del proceso de adopción ocurre la participación? Las recomendaciones genéricas sobre consulta a partes interesadas son insuficientes sin estructuras que redistribuyan poder de decisión. Giannini (2023), propone que los sistemas educativos necesitan devolver agencia a los estudiantes y recordarles a los jóvenes que permanecemos al timón de la tecnología. No hay un curso predeterminado.

Los marcos normativos asumen que los problemas educativos que la IA busca resolver existen y requieren soluciones tecnológicas. No consideran que algunos problemas pueden ser construcciones que benefician a ciertos actores más que responder a necesidades educativas reales. Por ejemplo, la demanda de personalización masiva del aprendizaje puede reflejar más intereses de desarrolladores tecnológicos buscando mercados que evidencia sobre insuficiencias de enfoques pedagógicos existentes. La falta de evaluación crítica sobre qué problemas educativos justifican intervenciones tecnológicas complejas representa un vacío conceptual en la gobernanza actual.

Conclusiones

El hallazgo más claro de este análisis es que el alcance y la especificidad de los marcos se mueven en direcciones opuestas: cuanto más global es la aspiración de un instrumento, más vagos resultan sus mecanismos de implementación. Los marcos normativos internacionales han logrado poner sobre la mesa los principales riesgos éticos de la IA educativa. Equidad, privacidad, transparencia, autonomía docente y sesgo algorítmico aparecen ahora de manera consistente en documentos de UNESCO, OCDE y la Unión Europea.

Este consenso básico es valioso porque establece un lenguaje común para discutir problemas que hace pocos años ni siquiera se reconocían. Sin embargo, esta convergencia en principios abstractos contrasta con la ausencia de medios concretos para implementarlos. Los marcos declaran objetivos sin especificar cómo lograrlos, quién debe hacerlo con qué recursos, o qué sucede cuando no se cumplen.

Las cinco dimensiones éticas analizadas revelan que los problemas no son solo de implementación sino conceptuales. La equidad, por ejemplo, requiere decisiones políticas sobre qué diferencias educativas son legítimas y cuáles reproducen injusticias que los sistemas algorítmicos no deberían perpetuar. Pero ningún marco operacionaliza estos criterios ni explica quién debe tomarlos. De manera similar, la privacidad enfrenta una contradicción estructural cuando las escuelas dependen de plataformas comerciales cuyo modelo de negocio es precisamente la extracción de datos.

Los marcos recomiendan minimizar la recolección de información sin reconocer que esta dependencia hace imposible proteger la privacidad de manera sustantiva. La transparencia algorítmica, por su parte, choca tanto con límites técnicos reales como con intereses corporativos. Los modelos de aprendizaje profundo son opacos incluso para quienes los desarrollan, y las empresas protegen sus algoritmos como secretos comerciales. Exigir explicabilidad sin poder obligar a las corporaciones a cumplirla resulta ineficaz.

Estas tensiones se agravan cuando consideramos que la autonomía pedagógica se erosiona precisamente cuando los algoritmos toman decisiones que antes requerían juicio docente. Los sistemas de recomendación automática presionan a los docentes a seguir indicaciones algorítmicas incluso cuando contradicen su experiencia profesional, y esta presión se intensifica cuando los administradores utilizan la concordancia entre decisiones humanas y algorítmicas como criterio de evaluación.

e-ISSN: 2683-2143

El sesgo, finalmente, no puede eliminarse con mejores datos porque los datos necesariamente reflejan desigualdades estructurales existentes. Un algoritmo entrenado con información histórica sobre ingreso a carreras universitarias reproducirá la segregación ocupacional por género. Corregir esto requiere decisiones normativas sobre qué constituye justicia educativa, no solo mejoras técnicas en la calidad de los datos.

Las regulaciones actuales enfrentan problemas prácticos que limitan su efectividad independientemente de qué tan bien articulen principios éticos. La mayoría son recomendaciones voluntarias sin consecuencias legales, lo que permite a las instituciones declarar adherencia a principios sin modificar prácticas reales. Cuando existen obligaciones legales, como en el Reglamento de IA de la Unión Europea, las autoridades educativas carecen de capacidad técnica para supervisar cumplimiento. Auditar algoritmos complejos requiere expertise que los gobiernos no poseen: según el AI Index Report 2023 citado por Giannini (2023), menos del 1% de expertos en IA trabajan en el sector público.

Los documentos analizados tampoco abordan las estructuras de poder que determinan qué tecnologías se desarrollan y para qué propósitos. Corporaciones privadas concentran el diseño de IA educativa, lo que genera dependencia institucional donde escuelas públicas pierden capacidad de decidir sus propias trayectorias tecnológicas. Esta concentración tiene consecuencias: los problemas que la IA supuestamente resuelve reflejan las prioridades de quienes desarrollan los sistemas más que necesidades identificadas por comunidades educativas. La personalización masiva del aprendizaje, por ejemplo, puede responder más a intereses comerciales de empresas tecnológicas que a evidencia sobre limitaciones de enfoques pedagógicos existentes. Las recomendaciones genéricas sobre "consulta a partes interesadas" no cambian esta dinámica porque no especifican cómo pueden docentes y familias participar significativamente en decisiones técnicas complejas, qué información necesitan para evaluar implicaciones, o si tienen poder real para rechazar sistemas o solo pueden opinar después de que las decisiones ya se tomaron.

Las decisiones de inversión pública reflejan esta reconfiguración de prioridades. Cuando gobiernos destinan fondos a IA generativa en lugar de reducir tamaño de clase o mejorar salarios docentes, están eligiendo un modelo específico de educación. Giannini (2023), plantea este dilema con claridad: miles de millones de dólares se invierten ahora en compañías de IA generativa cuando podrían dirigirse hacia el desarrollo docente y mejoras necesarias en infraestructura escolar. Esta redistribución de recursos representa un cambio estructural en cómo las sociedades valoran el desarrollo humano frente al desarrollo tecnológico, pero los marcos normativos actuales no lo cuestionan ni siquiera lo reconocen como decisión política que requiere justificación.

En este sentido, la investigación futura debe documentar consecuencias reales de implementación de IA en contextos educativos diversos mediante estudios longitudinales que superen los experimentos controlados de corto plazo que dominan la literatura actual. Se necesita análisis comparativo de efectividad de diferentes modelos regulatorios para entender qué aproximaciones funcionan en qué condiciones institucionales y políticas. Es necesario desarrollar marcos conceptuales que integren perspectivas técnicas, pedagógicas y políticas, reconociendo que la gobernanza de IA educativa no puede ser responsabilidad exclusiva ni de expertos en ciencias de la computación ni de pedagogos que desconocen funcionamiento algorítmico.

La ética de IA educativa debe construirse mediante procesos participativos que reconozcan la pluralidad de valores legítimos en educación y establezcan estructuras democráticas para negociar conflictos entre ellos. La pregunta central no es cómo la IA cambiará la educación, sino cómo la educación moldeará nuestra recepción de estas tecnologías. Las escuelas no deben ser receptoras pasivas de innovaciones desarrolladas externamente, sino espacios de deliberación democrática sobre qué tipo de futuro queremos construir. Este momento requiere decisiones sobre el papel que la tecnología debe jugar en ese futuro, decisiones que deben tomarse con la urgencia que la situación demanda, pero sin sacrificar la deliberación cuidadosa sobre el tipo de educación que necesitamos.

e-ISSN: 2683-2143

Referencias

Ananny, M. y Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973-989. https://doi.org/10.1177/1461444816676645

Baker, R. S. y Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*(4), 1052-1092. https://doi.org/10.1007/s40593-021-00285-9

Barocas, S. y Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671-732. https://doi.org/10.15779/Z38BG31

Bietti, E. (2020). *From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy*. En Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 210-219). https://doi.org/10.1145/3351095.3372860

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1-12. https://doi.org/10.1177/2053951715622512

Comisión Europea. (2019). *Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Consejo de Europa. (2024). *Convención marco sobre la inteligencia artificial y los derechos humanos, la democracia y el Estado de derecho* (Serie de Tratados del Consejo de Europa n.º 225). Consejo de Europa. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

D'Ignazio, C. y Klein, L. F. (2020). *Data feminism*. MIT Press. https://doi.org/10.7551/mitpress/11805.001.0001

Departamento de Educación de EE.UU. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. Office of Educational Technology. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. y Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center for Internet & Society. https://doi.org/10.2139/ssrn.3518482

Giannini, S. (2023). *Generative AI and the future of education*. UNESCO. https://doi.org/10.54675/HOXG8740

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*(1), 99-120. https://doi.org/10.1007/s11023-020-09517-8

Hoffmann, A. L. (2019). Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. Information, *Communication & Society, 22*(7), 900-915. https://doi.org/10.1080/1369118X.2019.1573912

Holstein, K., McLaren, B. M. y Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher-AI complementarity. *Journal of Learning Analytics, 6*(2), 27-52. https://doi.org/10.18608/jla.2019.62.3

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11), 501-507. https://doi.org/10.1038/s42256-019-0114-4

Nemorin, S. (2017). Post-panoptic pedagogies: The changing nature of school surveillance in the digital age. *Surveillance & Society, 15*(2), 239-253. https://doi.org/10.24908/ss.v15i2.6033

e-ISSN: 2683-2143

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press. https://doi.org/10.18574/nyu/9781479833641.001.0001

Obermeyer, Z., Powers, B., Vogeli, C. y Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447-453. https://doi.org/10.1126/science.aax2342

OCDE. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development*. OECD Publishing. https://doi.org/10.1787/db14d5e9-en

Selwyn, N., Rivera-Vargas, P., Passeron, E., y Miño-Puigcercós, R. (2021). *¿Por qué no todo es (ni debe ser) digital? Interrogantes para pensar sobre digitalización, datificación e inteligencia artificial en educación*. https://doi.org/10.31235/osf.io/vx4zr

UNESCO. (2019). *Beijing consensus on artificial intelligence and education*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000368303

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

Unión Europea. (2016). *Reglamento general de protección de datos (GDPR)*. Reglamento (UE) 2016/679. http://data.europa.eu/eli/reg/2016/679/oj

Unión Europea. (2024). *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial* (Reglamento de Inteligencia Artificial). Diario Oficial de la Unión Europea. http://data.europa.eu/eli/reg/2024/1689/oj

Williamson, B. (2017). *Big Data in Education: The digital future of learning, policy and practice*. SAGE Publications. https://doi.org/10.4135/9781529714920

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs. https://doi.org/10.1007/s00146-020-01100-0

e-ISSN: 2683-2143