

HACIA UN MARCO ÉTICO DE LA INTELIGENCIA ARTIFICIAL EN LA EDUCACIÓN

Towards an Ethical Framework of Artificial Intelligence in Education

Ana María ALONSO-RODRÍGUEZ

Centro de Investigación de Filosofía de la Ciencia y la Tecnología (CIFYT). Universidad de A Coruña. (Campus de Ferrol). España.

a.alonsor@udc.es

<https://orcid.org/0000-0002-2379-657X>

Fecha de recepción: 29/12/2023

Fecha de aceptación: 22/02/2024

Fecha de publicación en línea: 04/06/2024

Cómo citar este artículo / How to cite this article: Alonso-Rodríguez, A. M^a. (2024). Hacia un marco ético de la inteligencia artificial en la educación [Towards an Ethical Framework of Artificial Intelligence in Education]. *Teoría de la Educación. Revista Interuniversitaria*, 36(2), 79-98. <https://doi.org/10.14201/teri.31821>

RESUMEN

Este artículo reflexiona sobre el uso de la Inteligencia Artificial en educación desde una perspectiva ética. Lo hace desde un punto de vista externo, considerando su incidencia en los contextos educativos como vivero de los desafíos éticos y políticos que encara la sociedad. Esto permite dimensionar su alcance y profundidad y proponer medidas para afrontarlos.

Sus *objetivos* se enfocan hacia las bases éticas de la Inteligencia Artificial relacionada con la actividad educativa, buscando identificar: las oportunidades, riesgos asociados y su impacto ético en educación; y b) los principios éticos que puedan guiar el desarrollo, despliegue y uso de estos sistemas inteligentes.

Para ello se realizó un estudio cualitativo, apoyado en una *metodología* de revisión bibliográfica de: (i) trabajos académicos sobre sus usos actuales y riesgos potenciales de la Inteligencia Artificial; y (ii) un análisis comparativo de distintos códigos éticos, explorando la convergencia de principios aplicables a Sistemas de Inteligencia Artificial en los contextos educativos.

Los resultados obtenidos, en primer lugar, sitúan los problemas identificados en la tradición ética, cuestionando la proliferación de subdominios de la disciplina. Se indaga, después, la posibilidad de un marco ético unificado que evite la superposición de principios para cada dominio específico. Se constata la utilidad de un marco ampliamente reconocido e influyente, cuyos principios se adaptan bien a los desafíos de la educación.

Se concluye señalando las líneas en las que se debe avanzar en la investigación: (I) fundamentación ética y regulación normativa para el desarrollo y uso de la Inteligencia Artificial en educación conforme a los principios seleccionados; y (II) definición del nuevo perfil profesional docente y sus implicaciones para la formación inicial del profesorado.

Palabras clave: educación; inteligencia artificial; ética; códigos éticos; marco ético.

ABSTRACT

This article reflects on the use of Artificial Intelligence in education from an ethical perspective. It does so from the external point of view, considering its impact on educational contexts, as a breeding ground for the ethical and political challenges that society faces. This allows us to measure its scope and depth and propose actions to address them.

Its *objectives* focus on the ethical bases of Artificial Intelligence related to educational activity, seeking to identify: a) the opportunities, associated risks and its ethical impact on education; and b) the ethical principles that should guide the development, deployment and use of these intelligent systems.

For this purpose, a qualitative study was carried out, supported by a bibliographic review *methodology*: (i) academic works on current uses and potential risks of Artificial Intelligence and (ii) a comparative analysis of different ethical codes, exploring the convergence of principles applicable to Artificial Intelligence Systems in educational contexts.

The results obtained first place the identified problems in the ethical tradition and question the proliferation of subdomains of the discipline. The possibility of a unified ethical framework that avoids the overlap of principles for each specific domain is then investigated. The usefulness of a widely recognized and influential framework is confirmed, whose principles adapt well to the ethical challenges of education.

It concludes by indicating the lines in which research should advance (I) ethical foundation and normative regulation for the development and use of Artificial Intelligence in education in accordance with the selected principles; and (II) definition of the new teaching professional profile and its implications for initial teacher training.

Keywords: education; artificial intelligence; ethics, ethical codes; ethical framework.

1. CONTEXTO ACTUAL

La digitalización creciente de las tareas habituales en los distintos sectores de la actividad humana ha generado cambios de carácter informacional que modulan nuestra interacción con las Tecnologías de la Información y la Comunicación (TICs), pero también cómo se articulan las relaciones y los procesos sociales. Esto impulsa una “reontologización” del mundo a partir de la infosfera y nos sitúa en una nueva era: la hiperhistoria (Floridi, 2014), donde la información es el recurso fundamental, lo que nos convierte en vitalmente dependientes de las TICs.

La educación es el medio que hace posible esta innovación disruptiva y por eso es ahora “a lo largo de la vida”. Por ser la instancia que facilita la transición al recurso informativo, la Agenda 2030 tiene como uno de los objetivos del Desarrollo Sostenible (ODS 4) “garantizar una educación inclusiva, equitativa y de calidad y promover las oportunidades de aprendizaje durante toda la vida para todos” (United Nations, 2019). La Inteligencia Artificial (IA) puede contribuir de manera eficaz a esos objetivos, pues ya ofrece numerosas oportunidades en los diversos ámbitos de la actividad social (la industria, la comunicación, la salud, etc.).

Su incorporación a la educación impulsada desde el *e-learning* ha sido discreta, pero su uso está en constante crecimiento incluso en los entornos educativos más convencionales. Su potencial para mejorar la educación es inmenso, pero es preciso aprender a gestionar las numerosas implicaciones sociales, éticas y deontológicas. Sobre todo, porque la normalización de la IA en los distintos ámbitos de la vida social conlleva un cambio inevitable al que la educación está obligada a responder.

Este “discurso del cambio imperativo” (Bearman *et al.*, 2023) supone poco realista prescindir de estas herramientas porque entrañan riesgos. Además, sería contrario a los fines de la educación, que incluyen preparar a los estudiantes para incorporarse al mundo en el que tendrán que vivir, que es ya un entorno de digitalización e IA. Hay que pensar entonces en una *Inteligencia Artificial para la Educación* en el contexto de una *Educación para la Inteligencia Artificial*.

El impacto social de la IA en general plantea preocupaciones éticas legítimas respecto de su aplicación en el dominio educativo. Por tratarse de un campo de especial riesgo —y un inmejorable banco de pruebas para la resolución de esta clase de problemas adecuadamente enfocados— es necesaria una reflexión ética. Puede plantearse desde una perspectiva “interna”, como hacen los científicos que crean la IA, los tecnólogos que la implementan o los educadores que la utilizan. Pero aquí se busca profundizar en el enfoque “externo”, que permite considerar la triple dimensión científica, tecnológica y social, de los problemas y aportar criterios para valorar las posibles soluciones, los resultados que se esperan y las posibles consecuencias.

Así, los *objetivos* de este trabajo se orientan hacia las bases éticas de la Inteligencia Artificial relacionada con la actividad educativa, buscando: a) identificar las oportunidades, riesgos e impacto ético asociado a su incorporación. Esto ayuda

a comprender la necesidad de contar con principios éticos; y b) concretar esos principios para disponer un marco ético que regule el desarrollo, despliegue y uso de los sistemas inteligentes. Debe considerarse también su utilidad en la toma de decisiones pedagógicas y para determinar las capacidades que necesitan docentes y estudiantes, cuando interactúan con la IA, para prevenir sus riesgos.

Con el fin de articular una propuesta, se realizó un estudio cualitativo apoyado en una metodología de revisión bibliográfica de dos tipos de textos: (i) trabajos académicos sobre usos, riesgos y directrices éticas en Inteligencia Artificial —seleccionados en base a la temática específica, la autoridad de los autores y el impacto de los trabajos— que confirmaron la necesidad de vincular las aplicaciones de IA a códigos éticos; y (ii) códigos éticos para sistemas inteligentes buscando consensos en torno a determinados principios. El número y diversidad de los códigos publicados recomendó un muestreo intencionalmente orientado a comparar la perspectiva de distintas partes interesadas: organizaciones internacionales, países implicados en la carrera por la IA y las industrias representativas del sector.

Dentro de este contexto, la investigación sigue varios pasos que definen la estructura del artículo. Primero, se sitúa el problema en el marco filosófico-metodológico de las Ciencias de Diseño. Segundo, se describen usos, oportunidades y riesgos de las herramientas de IA en educación. Tercero, se constatan importantes implicaciones éticas. A este respecto, una vez reconocida su continuidad con problemas de la tradición, se cuestiona la proliferación de subdominios en la ética. Cuarto, como resultado del análisis de distintos códigos éticos, se confirma la convergencia de principios que permiten proponer un marco unificado para la IA en educación. Quinto, se concluye que este planteamiento requiere avanzar en dos direcciones de investigación: (i) fundamentación ética y regulación normativa para el desarrollo y uso de los Sistemas de Inteligencia Artificial (SIA) en educación; y (ii) esbozar el nuevo perfil profesional docente para identificar las necesidades de formación del profesorado.

2. CIENCIAS DE DISEÑO: OBJETIVOS, PROCESOS Y RESULTADOS

La Inteligencia Artificial en Educación (AIED, por sus siglas en inglés) es un quehacer científico en el ámbito de la ciencia de diseño (González, 2017a). Su objetivo es crear herramientas inteligentes que permiten ampliar las posibilidades de diferentes usuarios en contextos específicos de uso educativo. Surge de la confluencia de la Inteligencia Artificial y la Educación, que son también disciplinas científicas de Diseño, pues buscan nuevas metas para ampliar las posibilidades humanas y se articulan en términos de objetivos, procesos y resultados.

A diferencia de las Ciencias de la Naturaleza y las Ciencias Sociales, las Ciencias de Diseño se orientan expresamente a predecir cómo pueden llegar a ser las cosas y a prescribir pautas de actuación que contribuyan a cambiar la realidad existente para que sea “como debe ser” (Simon, 1996; González, 2007). Así, su objetivo principal es modular el futuro de acuerdo con unos fines previamente elegidos.

Sobre esta base científica se pueden diseñar intervenciones en distintos campos de actividad humana y, para hacerlas operativas, pueden incorporar diseños tecnológicos. Es el caso de los desarrollos de IA para resolver problemas y mejorar las posibilidades de la educación en el logro de objetivos. Pero esto tiene profundas implicaciones éticas que han de ser consideradas en relación a varios aspectos:

(i) Los objetivos, que pueden ser discutibles o plantear conflictos entre las distintas partes interesadas. Ello incluye la selección de los problemas cuya resolución encomendamos a la IA. (ii) Los procesos: qué procesos queremos automatizar, acelerar, etc. y qué tecnologías de IA son adecuadas a ese propósito. (iii) Los resultados, que pueden conducir a cuestionar la intervención misma (es decir, toda la secuencia de actividad pedagógica mejorada con IA). (iv) Las consecuencias que se siguen a partir de las acciones realizadas, que pueden no ser las deseadas.

3. USOS Y OPORTUNIDADES DE LA IA EN EDUCACIÓN

Entre la amplia diversidad de usos de la IA en la Educación (Flores Vivar y García Peñalvo, 2023; Unión Europea, 2022; Jara y Ochoa, 2020; Ocaña Fernández *et al.*, 2019; Moreno Padilla, 2019; Sánchez Vila y Lema Penín, 2007), destacan las siguientes:

- (i) Facilitar la gestión, con la programación de horarios, asignación de recursos, etc.
- (ii) Automatizar tareas rutinarias de los profesores, como el seguimiento de estudiantes, la provisión de información sobre los estudiantes a sus familias o tutores, la calificación de ejercicios o pruebas, etc.
- (iii) Apoyar la enseñanza, a través de sistemas de tutoría inteligentes que ofrecen asistencia a los estudiantes, en función de las dificultades que pueda mostrar. Pueden ayudarles de manera efectiva, también fuera del aula, con la preparación de las tareas. En contextos y situaciones en los que el docente humano no esté disponible, podrían sustituirlo *facilitadores virtuales* (personajes virtuales realistas mediante la combinación de tecnologías de IA, juegos en 3-D y animación por ordenador).
- (iv) Personalizar la experiencia educativa, que es quizá la principal contribución de la IA. Los llamados *sistemas de enseñanza adaptativos* son aplicaciones relacionadas con este objetivo. Sitúan a los estudiantes realmente en el centro, ajustando las trayectorias educativas a sus perfiles, características y comportamiento individual. Analizando la información obtenida acerca de su progreso, predicen el rendimiento futuro, lo que permite optimizar los recursos y el contenido que se les ofrece e incluso anticipar medidas correctoras para mejorarlo.

Contribuyen también a una educación más personalizada distintas soluciones basadas en IA, creadas para atender a necesidades educativas especiales, como el subtítulo en directo en caso de discapacidad auditiva o la audiodescripción para quienes presentan discapacidad visual.

Cabe considerar, además, otras muchas aplicaciones algorítmicas que se incorporan al ecosistema de aprendizaje desde la educación informal: a través de las redes sociales, plataformas de juegos *on-line* o aplicaciones móviles, entre otras.

Pero, aunque el uso de la IA en la Educación se está normalizando, la investigación empírica sobre su efecto real sigue siendo escasa. Algunos estudios revelan que estas herramientas, con un enfoque basado en pasos para problemas bien definidos, están teniendo éxito en apoyar la enseñanza y el aprendizaje de conocimientos del dominio STEM (Humble & Mozelius, 2022). Se trata, precisamente, de aquellas disciplinas que buscan desarrollar competencias aptas desde el punto de vista tecnológico y científico para los retos del siglo XXI (Moreno Padilla, 2019). También funcionan bien en el aprendizaje de lenguas extranjeras. En cualquier caso, parece conveniente mantener una actitud crítica y supervisada (Unión Europea, 2022, p. 14).

4. RIESGOS POTENCIALES E IMPLICACIONES ÉTICAS

Para una adecuada implementación de la IA en la educación, preocupa la limitada formación de los docentes, porque puede conducir al mal uso o incluso al abuso de esta tecnología. Pero para comprender los riesgos potenciales, conviene mirar sobre todo a los diseños de las herramientas y preguntarse a qué responde su introducción en el “mercado” educativo, quién la impulsa y por qué lo hace.

Los desarrollos de IA en educación pueden interferir en la autonomía y la responsabilidad de las personas y obstaculizar derechos universales (Asamblea General de la ONU, 1948) como la *privacidad* (art. 12), la *igualdad* (art. 1) y la *no discriminación* (art. 2). Esto tiene implicaciones sociales y éticas de carácter general (Crawford, 2023) que se plantean también en otros ámbitos (Linares Salgado, 2022), pero requieren una especial precaución en contextos educativos.

Privacidad: Los sistemas de aprendizaje automático se entrenan con grandes volúmenes de datos. En educación, se refieren a: (i) información personal sobre los estudiantes y sus familias; (ii) datos registrados acerca del rendimiento escolar; y (iii) datos de rastreo generados en el uso digital y las actividades de aprendizaje.

Los datos son la clave del éxito de la IA y constituyen la base de una educación personalizada. Pero la lógica del sector tecnológico según la cual cualquier cosa puede ser un dato que se puede tomar y usar (Crawford, 2023) plantea, en el espacio educativo, una serie de problemas relacionados con: el modo cómo se obtienen esos datos (el consentimiento y la privacidad); cómo se analizan (la transparencia y la confianza); y la posibilidad de que puedan ser usados para fines distintos a los aprobados, de modo que los alumnos y sus familias acaben siendo víctimas de manipulación comercial o de otro tipo (Unesco, 2019).

El riesgo de ciberataques es alto, cuando no existen protocolos de seguridad. Esto es especialmente preocupante cuando se trata de menores de edad. Los centros educativos están obligados a contar con procedimientos adecuados para garantizar la protección y el uso ético de los datos personales (UE, 2022, p. 11), pero la tarea

no es fácil cuando las habilidades y el conocimiento para desarrollar sistemas AIED se encuentra en organizaciones con fines de lucro y no dentro del sector educativo (Humble & Mozelius, 2022).

Igualdad y no discriminación: Los datos con los que se entrenan los algoritmos de aprendizaje automático incorporan sesgos provenientes de ciertos contextos y personas. Esto conduce a que estos sistemas internalicen criterios parciales o discriminatorios (de “raza”, género, edad...) propios de esas fuentes e incluso los amplifiquen.

Históricamente, las distintas formas de desigualdad condicionaron el acceso a los recursos y las oportunidades y esto ahora incide en los datos con los que se entrenan los algoritmos para clasificar y reconocer patrones. Así, “bajo el disfraz de una neutralidad técnica” (Crawford, 2023, p. 200), que a menudo deja atrás a los grupos de usuarios de los que se tienen menos datos disponibles., los sistemas de IA perpetúan las desigualdades sociales. El uso de la IA, también está dilatando, socialmente, la brecha digital, ya que los países y los grupos humanos con más recursos son quienes se benefician de las oportunidades que ofrece. Todo ello compromete la equidad educativa.

Los sesgos son también consustanciales a las prácticas de clasificación de la IA. Los diseñadores quienes tienen la última palabra para decidir cuáles son las variables a considerar en el conjunto de datos de entrenamiento y qué diferencias se tienen en cuenta para *clasificar* correctamente nuevas observaciones (Crawford, 2023). Esto contribuye a mantener y amplificar estereotipos (con claras implicaciones epistemológicas que trascienden el objetivo de este artículo).

Autonomía: La interacción con los SIA puede obstaculizar el desarrollo de la autonomía de los estudiantes y afectar a su capacidad de razonamiento y decisión, lo que plantea pues una dificultad añadida para alcanzar el objetivo de convertirse en aprendices independientes (Humble & Mozelius, 2022). Además, renunciar a la libertad de elección y delegar la toma de decisiones, socava la autonomía como capacidad de autolegislarse y autodeterminarse, ligada al reconocimiento de la dignidad de las personas.

Responsabilidad: Los SIA son agentes – y pueden estar reemplazando la agencia humana–, pero no son agentes morales. Esto genera problemas éticos relacionados con la responsabilidad (e imputabilidad) de sus acciones, incluidas las responsabilidades legales que habría que afrontar en caso de que tuvieran consecuencias negativas para otras personas. Estos problemas se incrementan porque actualmente los algoritmos de aprendizaje automático presentan elementos no predecibles debido a la opacidad con la que funcionan.

4.1. Desafíos éticos específicos del dominio educativo

Junto a estas cuestiones generales, cada uno de los dominios en los que se aplica la IA presenta desafíos éticos específicos que deben ser analizados en su contexto. En educación, hay que tener en cuenta lo siguiente: 1) Los posibles daños

derivados de los diagnósticos y la predicción de los resultados de aprendizaje de los estudiantes que pueden condicionar su desarrollo futuro. 2) Los problemas generados por las decisiones que podría tomar el sistema de IA por su influencia en las decisiones educativas de los docentes, las familias y otras partes interesadas (incluidos los legisladores). 3) Los impactos en el desarrollo y madurez de las personas, especialmente en las primeras etapas educativas, como consecuencia del cambio de roles que modifica la relación entre docentes y discentes.

También hay que considerar: 4) Las pedagogías que orientan los diseños de los sistemas de IA. Generalmente buscan aquello que se monetiza con mayor facilidad y son poco innovadoras. Contribuyen así a mantener el *statu quo* del sistema educativo (Holmes *et al*, 2022), perdiendo su inmenso potencial para cambiar la educación y ampliar realmente las oportunidades de las personas. 5) La imagen del mundo y las concepciones sociales y políticas tácitas a las que responde la IA, que se trasladan al ámbito educativo. Esta tecnología no es inocua y “las clasificaciones que se eligen casualmente para darle forma a un sistema técnico pueden jugar un papel dinámico a la hora de delinear el mundo social y material” (Crawford, 2023, p. 195). 6) Cabe citar también la dimensión ética de algunas malas prácticas de los estudiantes sirviéndose de instrumentos de IA, que conllevan el engaño y el fraude, atentan contra la propiedad y comprometen la honestidad intelectual.

5. LA IMPORTANCIA DE LA ÉTICA

Los aspectos positivos y los riesgos asociados a la IA están íntimamente relacionados y no es posible acceder a las enormes ventajas que ofrece sin enfrentarnos a sus consecuencias negativas. Se hace necesario analizar y comprender estas ambivalencias para (i) anticipar y calibrar los impactos de determinados desarrollos y (ii) seleccionar los valores y principios que debemos proteger. Por eso ha surgido una ética de la IA como área de investigación multidisciplinaria en cuestiones de ética aplicada relacionadas con los problemas normativos que plantea el desarrollo, despliegue y utilización de sistemas inteligentes. No es una ética para sistemas inteligentes, sino para personas que diseñan, desarrollan o usan esos sistemas inteligentes, que son quienes pueden causar, aunque sea de manera involuntaria, los problemas morales.

El análisis normativo de los impactos de la IA, presenta paralelismos con otros subdominios específicos en la ética de la tecnología (de internet, de datos, de robots...) conformados en torno a distintas innovaciones. Con frecuencia abordan las mismas cuestiones (privacidad, sesgos...) con planteamientos similares, pero se desarrollan sin interacción entre ellas y no llegan a conectar sus discusiones ni a ponerlas en relación con los enfoques históricos (Sætra & Danaher, 2022). Esto conduce a duplicar esfuerzos sin aumentar el conocimiento real e incluso a obviar ideas consolidadas en la tradición ética. Así, muchas veces, las dificultades para afrontar los desafíos éticos de la IA, proceden más del olvido de cuestiones

fundamentales de la ética –y del conocimiento históricamente adquirido–, que de nuestra comprensión de la IA.

La ética de la IA ha recibido mucha atención por parte de los investigadores (Boddington, 2017; Coeckelberg, 2021; Floridi, 2022) y también de diversos organismos, instituciones y corporaciones. Actualmente, está centrada en la elaboración de códigos éticos para identificar los valores y principios que se deben poner en práctica. Son múltiples las iniciativas y propuestas que surgieron en los últimos años y distintos países han intentado también establecer sugerencias y regulaciones. Con frecuencia, la proliferación de códigos con sus pautas y principios es vista como un problema (Sætra & Danaher, 2022) ya que puede confundir (Floridi & Cowls, 2021). Sin embargo, muchos de estos códigos contemplan principios comunes, con ciertos énfasis y matices afectados por valores culturales y sociales diferentes.

El análisis comparativo de seis códigos de alto perfil para la IA –publicados entre 2017 y 2018¹–, ha permitido a Floridi & Cowls (2021) establecer un marco general con cinco principios básicos, cuatro de los cuales convergen con los utilizados en Bioética: *beneficencia*, *no maleficencia*, *autonomía* y *justicia* (Beauchamp & Childress 2019). No obstante, la transposición a los desafíos de la IA no es perfecta y requiere, en algunos casos, una traducción adecuada:

Beneficencia supone promover el bienestar, preservar la dignidad y el sostenimiento del planeta. *No maleficencia* remite a la privacidad, seguridad y precaución. *Autonomía* se entiende como el poder de decidir. *Justicia* implica promover la prosperidad, preservar la solidaridad y evitar la injusticia. Cuestiones derivadas del desarrollo mismo de la IA hacen necesaria la incorporación de un nuevo principio: la *explicabilidad*, que permite habilitar los otros principios a través de la inteligibilidad y la rendición de cuentas (Floridi & Cowls, 2021).

Este marco se presenta como la arquitectura dentro de la cual se puedan desarrollar las regulaciones, los estándares técnicos y las mejores prácticas para sectores específicos. Puede desempeñar tanto un papel habilitador (p. ej. el uso de la IA para alcanzar los Objetivos de Desarrollo Sostenible de las Naciones Unidas), como restrictivo (Floridi & Cowls, 2021, p. 14). Es un marco ético general y realista para el seguimiento y evaluación del diseño, desarrollo y uso de los SIA, porque dispone de un número manejable de principios compatibles con valores fundamentales universalizables e irrenunciables. Estos principios ofrecen una base para las regulaciones específicas que se propongan en una amplia variedad de contextos, para las leyes y acuerdos internacionales, las instancias de supervisión y evaluación que los estados establezcan e incluso los observatorios ciudadanos... Proporcionan, de ese modo, legitimidad moral, jurídica y política.

¹ Principios de Asilomar sobre IA (2017); Declaración de Montreal para una IA Responsable (2017); Segunda versión de Diseño Éticamente Alineado (2017); Declaración sobre Inteligencia Artificial, Robótica y Sistemas 'Autónomos', del Grupo Europeo de Ética en la Ciencia y las Nuevas Tecnologías de la Comisión Europea (2018); Informe del Comité de Inteligencia Artificial de la Cámara de los Lores del Reino Unido (2018); Asociación sobre IA de múltiples partes interesadas (2018).

6. UN MARCO ÉTICO UNIFICADO PARA LA INTELIGENCIA ARTIFICIAL EN EL CONTEXTO EDUCATIVO

Desde las ideas expuestas, se puede avanzar de lo general a lo particular. Esto requiere adoptar marcos éticos para el uso y desarrollo de la IA en la educación. Su importancia ha sido señalada en investigaciones anteriores (Holmes *et al.*, 2022; Holmes y Porayska-Pomsta, 2023). Pero desde las *Directrices para una Ética de la IA en Educación* que Aiken y Epstein publicaron hace más de dos décadas (Aiken & Epstein 2000), ha habido una sorprendente escasez de trabajos publicados que se centren explícitamente en la ética (Holmes, *et al.*, 2022, pp. 505-506). Recientemente, esta cuestión volvió a suscitar el interés de la comunidad investigadora en educación (Chu *et al.*, 2022; Holmes y Porayska-Pomsta, 2023; Boddington, 2023; Nguyen *et al.*, 2023) y, en los últimos años, surgieron también algunas iniciativas relevantes de organizaciones internacionales (Unesco, 2019; Unión Europea, 2022).

La Unión Europea, dentro de su Plan de Acción para la Educación Digital (2021-2027) -cuyo fin es apoyar una adaptación sostenible y eficaz de los sistemas de educación y formación de los Estados miembros a la era digital- estableció una serie de *directrices éticas sobre el uso de la Inteligencia Artificial (IA) y los datos en la educación y formación para los educadores* (Unión Europea, 2022). En esas directrices, basadas en los principios y requisitos para una IA fiable (Unión Europea, 2019), hay cuatro consideraciones clave que deben orientar a los agentes educativos en sus decisiones sobre el uso de la IA: la capacidad de acción humana, la equidad, la humanidad y la elección justificada.

Dentro de la comunidad académica, se han analizado de manera sistemática directrices e informes éticos de organizaciones internacionales y se han seleccionado principios aplicables para garantizar que los sistemas de IA desarrollados para la educación, sean esencialmente éticos por diseño: Principio de gobernanza y administración, Principio de transparencia y responsabilidad, Principio de sostenibilidad y proporcionalidad, Principio de privacidad, Principio de seguridad y protección, Principio de inclusión y Principio de AIED centrado en el ser humano (Nguyen *et al.* 2023).

Consideradas estas propuestas, hace falta contar con un marco unificado para una IA responsable, ética y confiable. Con ello se busca evitar la superposición de principios y recomendaciones para cada dominio específico. Existe un marco de prestigio, que ha desempeñado un importante papel en los debates del Grupo de expertos de alto nivel en IA de la Comisión Europea e influido en distintas recomendaciones (Floridi & Colws, 2021). Sus principios se adaptan sorprendentemente bien a una amplia gama de contextos (Linares Salgado, 2022). La cuestión es si pueden regular también problemas relativos al cuándo, cómo y con qué fin se utiliza en educación un sistema inteligente, para situar decisiones, acciones y prácticas pedagógicas concretas que aseguren un uso adecuado de la IA.

Esta cuestión llevó a un nuevo análisis en tres fases:

- 1) Revisión de guías y recomendaciones que contienen códigos éticos, relevantes por su impacto, publicados posteriormente a los examinados por Floridi & Cowls (2021). El objetivo es comprobar la vigencia y prevalencia de los principios éticos que proponen. En el muestro se buscó que estuvieran representadas las diversas perspectivas influyentes del sector: a) recomendaciones de Organizaciones internacionales: UE (2019), UNESCO (2022) y OCDE (2019); b) documentos de países implicados en la carrera por la IA: EE UU (Executive Office of the President, 2020), China (Ministerio de Ciencia y Tecnología de la República Popular China, 2017), UE (2019) y España (Gobierno de España, 2020); y c) iniciativas de varias empresas relevantes en el sector: Microsoft (2022), Google (2022) y Meta (Meta Platforms, 2021).
- 2) Comprobación de la convergencia, en el marco ético anterior, de: (i) los principios seleccionados por Nguyen *et al.* (2023) y (ii) las directrices éticas de la Comisión Europea para guiar el uso de la Inteligencia Artificial en educación.
- 3) Constatación de que los valores que orientan estos principios están alineados con los valores propios de la educación (Alonso, 2022 y 2023).

6.1. Resultados

El análisis comparativo, como puede verse en la Tabla 1, muestra un alto grado de superposición entre los principios de los nueve códigos, que convergen claramente en el marco general propuesto por Floridi & Cowls (2021). Aunque solo la *explicabilidad* aparece con idéntica formulación –o como *transparencia*– en la totalidad de los códigos, la “traducción” realizada para transponer los de la bioética a la IA, permite identificarlos.

- I) El principio de *no-maledicencia* (que remite al *non-laedere* del derecho romano)² figura expresamente como *prevención del daño* en las directrices de la UE, como *seguridad y protección* en seis códigos y centrada en la protección de la *privacidad* en otros dos. En la *Estrategia Nacional para la IA* española, se asocia la prevención al principio de Gobierno de datos y sistemas: “los datos no pueden usarse para perjudicar a las personas o violar sus derechos fundamentales” (Gobierno de España, 2020, p. 67).
- II) La *justicia* (*suum cuique tribuere*) se cita de forma explícita en uno de los códigos (China), siendo más frecuente la *equidad* como criterio de justicia (en siete de ellos). Es reconocible en la *inclusión* (España) y en la prevención de amenazas para la justicia como *evitar crear y reforzar sesgos* (Google) y la *no discriminación*.

² “Turis praecepta sunt haec: honeste vivere, alterum non laedere, suum cuique tribuere” (Ulpiano. Digesto 1, 1, 10, 1).

Tabla 1
 VISIÓN SINÓPTICA 1

	Beneficencia	No Maledicencia	Autonomía	Justicia	Explicabilidad
UNESCO	Promover el bienestar, preservar la Dignidad y el Sosteenimiento del Planeta	Privacidad, Seguridad y Precaución de Capacidad	Poder de decidir (Decidir)	Promover la prosperidad, preservar la solidaridad, evitar la injusticia.	Habilitar los otros principios a través de la inteligibilidad y la rendición de cuentas
OCDE	Crecimiento inclusivo, desarrollo sostenible y bienestar	Robustez, seguridad y protección	Respetar los valores humanos y la justicia (incluye la autonomía)	Respetar los valores humanos y la justicia	Transparencia y explicabilidad Responsabilidad
UE		Prevención del daño	Respeto por la autonomía humana	Equidad	Explicabilidad
EE. UU.		Seguridad y protección		Equidad y no discriminación	Divulgación y transparencia
CHINA	Armonía y amistad: promover el bienestar común	Privacidad: respetar y proteger la privacidad	Privacidad: respetar y proteger el derecho de las personas de conocer y elegir	Equidad y Justicia	Seguridad: transparencia, explicabilidad, trazabilidad, confiabilidad, auditabilidad y seguridad

(continúa)

TABLA 1
 VISIÓN SINÓPTICA 1

	Beneficencia	No Maledicencia	Autonomía	Justicia	Explicabilidad
ESPAÑA	Bienestar social Sostenibilidad	Gobierno de datos y sistemas: los datos no pueden usarse para perjudicar a las personas o violar sus derechos fundamentales	Bienestar social: no reducir, limitar o desviar la autonomía de las personas	Inclusión	Transparencia
GOOGLE	Beneficio a la sociedad	Garantizar la seguridad de los sistemas inteligentes Incorporar principios de diseño de privacidad	Ser responsables ante las personas: los sistemas deben estar sujetos a un adecuado control y dirección humana	Evitar crear y reforzar sesgos	Ser responsables ante las personas: proveer oportunidades adecuadas de respuesta, explicación y apelación
MICRO-SOFT		Privacidad y Seguridad		Equidad	Transparencia Responsabilidad
META		Solidez y seguridad		Equidad e inclusión	Transparencia y control

Fuente: elaboración propia

- III) El principio de *autonomía* está asociado a la toma de decisiones dirigida por humanos, que puede ponerse en riesgo a medida que aumenta la agencia de las máquinas. Es llamativo que este principio no esté recogido en la “Guía para la regulación de aplicaciones de IA” de los EEUU y en dos de las tres iniciativas de la industria privada (Microsoft y Meta). Aparece en los demás códigos bajo la formulación expresa de *autonomía*, como *supervisión*, *control* o *dirección humanas* o como *derecho a decidir*.
- IV) Los mismos códigos que no incluyen la autonomía entre sus principios, tampoco incorporan el de *beneficencia*, que limita la creación de SIA a aquellos que han de ser beneficiosos para la humanidad. Tampoco aparece expresamente en las directrices éticas de la UE, sin embargo, el *bienestar social* y *medioambiental* sí figura entre los requisitos que derivan del principio de *equidad*. En el resto de los códigos presenta formulaciones diversas, vinculado al *bienestar* y la *sostenibilidad*.

Cabe concluir que el marco presentado por Floridi & Cowsls (2021) ofrece una relación suficiente y completa de los principios éticos clave para la IA. Puede resultar útil –como sugieren los autores– para orientar regulaciones y prácticas en dominios específicos. Debería recoger también los principios que se proponen en el ámbito educativo, encaminados en dos direcciones: (i) dirigir el desarrollo, despliegue y uso de las herramientas y (ii) orientar a los agentes educativos en sus prácticas profesionales.

Es claro el paralelismo entre los principios de *no maleficencia* y *beneficencia* y los “metaprincipios” que propusieron Aiken y Epstein como fundamento filosófico básico para cualquier discusión sobre los sistemas AIED. Según el *meta-principio negativo* “la tecnología AIED no debe disminuir al estudiante en ninguna de las dimensiones fundamentales del ser humano”. El *meta-principio positivo* establece que “la tecnología AIED debe potenciar al estudiante en al menos una de las dimensiones fundamentales del ser humano” (Aiken & Epstein, 2000, p. 170).

También, como se observa en la Tabla 2, la propuesta de Nguyen y colaboradores (2023) presenta un alto grado de coherencia con el marco unificado: (i) El principio de *no maleficencia*, implica *privacidad*, *seguridad* y *precaución de capacidad*. Incorpora, por lo tanto, los *principios de privacidad y de seguridad y protección*. (ii) *Beneficencia* supone *promover el bienestar*, *preservar la dignidad* y el *sostenimiento del planeta*, donde se reconoce el *principio de sostenibilidad y proporcionalidad*. (iii) *Autonomía* significa *poder de decidir*. Ligada al reconocimiento de la dignidad de las personas conduce al principio básico que debe presidir nuestra relación con la IA: la *irremplazabilidad de lo humano*, que exige el *principio de IA educativa centrado en el ser humano*. (iv) *Justicia* es *promover la prosperidad*, *preservar la solidaridad*, *evitar la injusticia*, lo que persigue el *principio de inclusión*. (v) La *explicabilidad* permite habilitar los otros principios a través de la *inteligibilidad* y la *rendición de cuentas*. La *inteligibilidad* requiere

transparencia y la *rendición de cuentas* supone atribución de *responsabilidades*, que es lo que exige el *principio de transparencia y responsabilidad*.

Asimismo, están en sintonía con estos principios las consideraciones éticas identificadas por la Unión Europea para orientar a los agentes educativos en sus decisiones sobre la IA (Unión Europea, 2022, p. 18). Se reconocen en ellas: (i) los valores que constituyen los fines de la Educación (autonomía, autodeterminación, responsabilidad, cohesión social, bienestar, etc.); (ii) los valores indispensables para su desarrollo (igualdad de oportunidades, no discriminación, transparencia, explicabilidad...); y (iii): los valores que sirven a la educación como criterio de juicio, entre ellos la inclusión, la distribución justa de derechos y responsabilidades, o una conexión humana significativa (Alonso, 2022 y 2023).

Los principios éticos de *beneficencia*, *no maleficencia*, *autonomía*, *justicia* y *explicabilidad* aplicados a los SIA, pueden prevenir los riesgos asociados a estos sistemas para usarlos en beneficio del Derecho a la Educación (ODS 4), como fin en sí mismo y medio para alcanzar todos los demás derechos. Los instrumentos que —en su diseño, despliegue y uso— pudieran conculcar alguno de estos principios, no serían legítimos al poner en peligro el respeto a la dignidad humana que los derechos humanos preservan.

No maleficencia. La IA debe diseñarse de modo que no agrave daños existentes ni cree otros nuevos. Así, no sería ético diseñar instrumentos que obstaculicen el desarrollo de alguna de las capacidades propias de la inteligencia humana en sus tres dominios principales: la percepción —conocimiento—, la volición y los valores (González, 2017b), limitando el uso de los sentidos, la imaginación, el pensamiento y el razonamiento de un modo “verdaderamente humano” (Nussbaum, 2012, p. 53).

Beneficencia. Además de no dañar, los SIA educativos deben diseñarse para que los estudiantes puedan desarrollar su potencial y usar todas sus capacidades para elegir la vida que desean vivir de forma plena y creativa, conforme a la dignidad humana.

Autonomía. En el ámbito de la ética se entiende por *autonomía* la capacidad de autolegislarse y autodeterminarse, que solo es aplicable a los seres humanos. Esto impone y justifica el control humano sobre la “máquina”. Por eso: (a) No pueden delegarse en máquinas inteligentes, sin supervisión humana, decisiones que afectan a la vida de las personas. En consecuencia, las decisiones educativas no pueden estar basadas únicamente en el proceso automático de datos. (b) Es exigible que un humano sepa siempre si está interactuando con otro humano o con una máquina. Así, cuando se usen instrumentos de IA, las precauciones especiales con personas vulnerables, como los niños, hacen necesario el consentimiento informado de los padres o tutores legales.

La educación constituye un ámbito de especial riesgo por su incidencia en el desarrollo de las personas, en la configuración del pensamiento y la acción a lo largo de la vida. Habría que enfatizar la precaución, que aparece ahora como exigencia de la no maleficencia, e incorporar el *Principio Precautorio* en aquellas intervenciones educativas donde la IA represente una amenaza para la autonomía y la dignidad humana (ampliando la Declaración de Wingspread).

TABLA 2
VISIÓN SINÓPTICA 2

Floridi & Cows, 2021		AIED		
		Nguyen <i>et al.</i> , 2023	Valores de la Educación (UE, 2022)	
Principios		Principios	Consideraciones éticas	Valores
No maleficencia:	Privacidad	Privacidad		
	Seguridad	Seguridad y protección		
	'Precaución de Capacidad'			
Beneficencia:	Promover el Bienestar		La humanidad (consideración para las personas, identidad y dignidad)	Bienestar Seguridad
	Preservar la Dignidad			Cohesión social Contacto valioso
	Sostenimiento del Planeta	Sostenibilidad y proporcionalidad		Conexión humana significativa
Autonomía:	Poder de decidir (decidir) (Conduce a la irreemplazabilidad de lo humano)	IA educativa centrada en el ser humano	La capacidad de acción humana	Autonomía Autodeterminación Responsabilidad
Justicia:	Promover la prosperidad	Principio de inclusión	La equidad (trato justo)	Igualdad de oportunidades Inclusión
	Preservar la solidaridad			No discriminación
	Evitar la injusticia			Distribución justa de derechos y responsabilidades
Explicabilidad:	Inteligibilidad	Principio de transparencia y responsabilidad	La elección justificada (uso de conocimientos, hechos y datos para justificar las elecciones de las distintas partes interesadas en el entorno escolar)	Transparencia Explicabilidad
	Rendición de cuentas (La inteligibilidad requiere transparencia y la rendición de cuentas supone atribución de responsabilidades)			

Fuente: elaboración propia

7. CONCLUSIÓN Y DISCUSIÓN

La perspectiva externa sobre IA y su uso en educación, atiende a la repercusión social y la responsabilidad en las prácticas mediadas por estos sistemas (Bearman *et al.*, 2023) y destacan la importancia de la ética en el debate contemporáneo de la educación (Holmes *et al.*, 2022). Aquí se ha buscado un marco unificado para este problema actual y con recorrido futuro, donde Ciencia, Tecnología y Sociedad se han tenido en cuenta.

Este marco constituye un punto de partida útil, ya que (i) puede orientar una regulación coherente y bien elaborada, (ii) permite ubicar las buenas prácticas, (iii) indica las competencias que necesitan desarrollar los docentes, y (iv) ofrece criterios a los estudiantes para interactuar con la IA y aprovechar su potencial en los contextos formativos. Pero es preciso seguir avanzando. La investigación en este ámbito está experimentando un incremento notable y conviene, en futuros esfuerzos, poner estos consensos en relación con los resultados del metaanálisis de una muestra mucho mayor de documentos. A su vez:

- 1) Los principios representan los valores que se han de realizar a través de los fines (p. ej., no sería legítimo un fin que de ser alcanzado pueda dañar). Pero además de procurar fines éticos para la IA, es preciso determinar los medios para desarrollarla de manera ética, porque una cosa es hacer cosas éticas y otra hacerlas éticamente (Holmes *et al.*, 2022, p. 504) y aunque lo que se busque sea algo beneficioso, no todo puede estar permitido para alcanzarlo (piénsese en el uso indebido de los datos).
- 2) Los valores a los que remiten los principios son valores “confiables” y constituyen algunas de nuestras certezas como garantía de la dignidad humana y la convivencia democrática. Pero (a) pueden conceptualizarse de maneras distintas, como se observa si se atiende a los distintos criterios de *justicia* que aparecen en los diferentes códigos y (b) son valores que surgieron dentro de la tradición occidental y, para garantizar un alcance global, es preciso incorporar la perspectiva de otras regiones y culturas. Esto nos sitúa de nuevo ante el viejo debate del multiculturalismo. Como señala Adela Cortina, “ante retos universales no cabe sino la respuesta de una actitud ética universal, que tiene como horizonte para la toma de decisiones el bien universal” (Cortina, 1997, p. 261).
- 3) Pocas veces se concretan los mecanismos efectivos para que todas las partes interesadas asuman esos principios o la responsabilidad de su transgresión. Esto atañe de modo especial a las empresas tecnológicas que desarrollan los productos de IA para la educación y que además controlan la investigación actual (Ahmed *et al.*, 2023). Como sujetos sociales, pueden y tienen que asumir la responsabilidad de sus acciones en la medida que: a) muestran intencionalidad en la actividad que despliegan y b) la toma de decisiones es controlada por la propia organización (González, 2020).

Las preocupaciones suscitadas por la IA entroncan con cuestiones tradicionales de la ética. En general, no son novedosas. Cambia el grado de relevancia, puesto

que la IA conforma en buena medida cómo interaccionamos con el mundo, modula las relaciones sociales e, incluso, incide en la configuración de nuestra propia identidad. Esto tiene consecuencias a la hora de considerar las nociones tradicionales de moralidad y responsabilidad individual, que quizá necesiten ser revisadas a la luz de las nuevas condiciones.

El debate ha de plantearse en dos direcciones principales: Primera, la fundamentación ética que dote a los principios de legitimidad moral. Conviene evitar simplificaciones excesivas de cuestiones especialmente complejas, como sucede cuando son abordadas por expertos en IA. Decidir qué es lo que no puede sacrificarse o qué se debe preferir, requiere poder ofrecer razones sólidas acerca de lo que se considera “bueno” o de lo que se “debe hacer”. Y esto es tarea de la ética.

Para que el desarrollo y uso de los SIA en educación se realice conforme a los principios, hace falta también una regulación jurídica y política. Pero las leyes han de conectar con los valores y estar en sintonía con la razón y los deseos de las personas. Porque una cosa es una ley (que puede ser impuesta o asumida en el proceso de socialización) y otra las razones que un ser humano tenga para tenerla por suya (Cortina, 1997). Entonces hay que considerar una educación en valores.

Segunda, la determinación del nuevo perfil profesional docente, que requiere precisar el conocimiento experto que ahora es necesario para el ejercicio de la profesión. Esto ha de incorporarse a los planes de formación inicial del profesorado. Sobre la base de lo aquí expuesto, deberá contemplar al menos: (a) formación investigadora científicamente sólida que combine el rigor con la resolución práctica de problemas; (b) formación ética para abordar los desafíos de la IA; y (c) formación en IA que permita a los educadores ajustar su enfoque y diseñar estrategias pedagógicas más efectivas para avanzar hacia el objetivo de una “educación inclusiva, equitativa y de calidad y promover las oportunidades de aprendizaje durante toda la vida para todos”.

REFERENCIAS BIBLIOGRÁFICAS

- Ahmed, N., Wahed, M., & Thompson, N. C. (2023). The growing influence of industry in AI research. *Science*, 379 (6635), 884-886. <https://www.science.org/doi/10.1126/science.ade2420>
- Aiken, R. M., & Epstein, R. G. (2000). Ethical Guidelines for AI in Education: Starting a Conversation. *International Journal of Artificial Intelligence in Education*, 11, 163-176.
- Alonso, A. M. (2022). Las Ciencias de Diseño como Ciencias Aplicadas y el papel de los valores: Análisis del caso de la Educación. *Revista de Investigación Filosófica*, 9(1), 3-27. https://doi.org/10.26754/ojs_arif/arif.202216381
- Alonso, A. M. (2023). *Filosofía de la Ciencia de la Educación: Análisis como Ciencia Aplicada de Diseño*. Tirant lo Blanch.
- Asamblea General de la ONU. (1948). Declaración Universal de los Derechos Humanos. Paris. <http://www.un.org/en/universal-declaration-human-rights/>
- Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: a critical literature review. *Higher Education*, 86, 369-385. <https://doi.org/10.1007/s10734-022-00937-2>

- Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics*. Oxford University Press.
- Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Springer.
- Boddington, P. (2023). Introduction, Why AI Ethics? In P. Boddington, *AI Ethics. A Textbook* (pp. 1-34). Springer.
- Chu, J., Xi, L., Zhang, Q., & Lin, R. (2022). Research on Ethical Issues of Artificial Intelligence in Education. In J. Yang, *et al.* (Eds.), *Resilience and Future of Smart Learning* (pp. 101-108). Springer. https://doi.org/10.1007/978-981-19-5967-7_12
- Coeckelberg, M. (2021). *Ética de la Inteligencia Artificial*. Cátedra.
- Comisión Europea. (2019). Grupo de Expertos de Alto Nivel sobre IA. *Directrices Éticas para una IA fiable*. <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>
- Cortina, A. (1997). *Ciudadanos del mundo. Hacia una teoría de la ciudadanía*. Alianza editorial.
- Crawford, K. (2023). *Atlas de I. Poder, Política y Costes Planetarios de la Inteligencia Artificial*. Ned ediciones.
- Declaración de Wingspread sobre el Principio Precautorio (1998). <http://www.linea-e.com/cuadernos/pdfs/numero11/elprincipiodeprecaucion.pdf>
- Executive Office of the President. (2020). *Guidance for regulation of Artificial Intelligence applications*. <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>
- Flores Vivar, J. M. y García-Peñalvo, F. J. (2023). Reflexiones sobre la ética, potencialidades y retos de la Inteligencia Artificial en el marco de la Educación de calidad (ODS4). *Comunicar*, 31(74), 37-47. <https://doi.org/10.3916/C74-2023-03>
- Floridi, L. (2014). *The Fourth Revolution - How the Infosphere is Reshaping Human Reality*. Oxford University Press.
- Floridi, L. (2022). *Etica dell'intelligenza artificiale. Sviluppo, opportunità, sfide*. Raffaello Cortina.
- Floridi L., & Cowls, J. (2021). A Unified Framework of Five Principles for AI in Society. In L. Floridi (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence* (5-17). Springer.
- Gobierno de España. (2020). *Estrategia Nacional de Inteligencia Artificial*. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/021220-ENIA.pdf>
- González, W. J. (2007). Configuración de las Ciencias de Diseño como Ciencias de lo Artificial: Papel de la Inteligencia Artificial y de la racionalidad limitada. En W. J. González (Ed.), *Las Ciencias de Diseño: Racionalidad limitada, predicción y prescripción* (41-69). Netbiblo.
- González, W. J. (2017a). Artificial Intelligence in a New Context: "Internal" and "External" Factors. *Minds & Machines*, 27(3), 393-396. <https://doi.org/10.1007/s11023-017-9444-3>
- González, W. J. (2017b). From Intelligence to Rationality of Minds and Machines in Contemporary Society: The Sciences of Design and the Role of Information. *Minds & Machines*, 27(3), 397-424. <https://doi.org/10.1007/s11023-017-9439-0>
- González, W. J. (2020). The Internet at the Service of Society: Business Ethics, Rationality, and Responsibility. *Éndoxa*, 46, 383-412. <https://doi.org/10.5944/endoxa.46.2020.28029>
- Google LLC. (2022). *Artificial Intelligence at Google: Our Principles*. <https://ai.google/principles/>
- Holmes, W., & Porayska-Pomsta, K. (2023). *The Ethics of Artificial Intelligence in Education*. Routledge Taylor.
- Holmes, W., Porayska-Pomsta, K., Holstein, K. et al. (2022). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 32, 504-526. <https://doi.org/10.1007/s40593-021-00239-1>

- Humble, N., & Mozellius, P. (2022). The threat, hype, and promise of artificial intelligence in education. *Discover Artificial Intelligence*, 2(22). <https://doi.org/10.1007/s44163-022-00039-z>
- Jara, I., & Ochoa, J. M. (2020). *Usos y efectos de la inteligencia artificial en educación*. Banco Interamericano de Desarrollo. <http://dx.doi.org/10.18235/0002380>
- Linares Salgado, J. E. (2022). Principios éticos para el desarrollo de la Inteligencia Artificial y su aplicación en los sistemas de salud. *ArtefaCToS. Revista de Estudios de la Ciencia y la Tecnología*, 11(2), 137-161.
- Meta Platforms. (2021). *Facebook five pillars of responsible AI*. <https://ai.facebook.com/blog/facebook-five-pillars-of-responsible-ai/>
- Microsoft Corporation (2022). *Responsible IA*. https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimar_yr6
- Ministerio de Ciencia y Tecnología de la República Popular China. (2017). *Principios de Gobernanza para una Nueva Generación de IA: Desarrollo Responsable de la Inteligencia Artificial*. <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/>
- Moreno Padilla, R. (2019). La llegada de la inteligencia artificial a la educación. *RITI Journal*, 7(14), 260-270. <https://doi.org/10.36825/RITI.07.14.022>
- Nguyen A, Ngo H. N, Hong, Y., Dang, B., & Nguyen, B.T. (2023). Ethical principles for artificial intelligence in education. *Educ Inf Technol*, 28(4), 4221-4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Nussbaum, M. (2012). *Crear capacidades. Propuesta para el desarrollo humano*. Paidós.
- Ocaña-Fernandez, Y., Valenzuela-Fernandez, L., & Garro-Aburto, L. (2019). Inteligencia artificial y sus implicaciones en la educación superior. *Propósitos y Representaciones*, 7(2), 536-552. <http://dx.doi.org/10.20511/pyr2019.v7n2.274>
- OECD (2019). *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- Sætra, H.S., & Danaher, J. (2022). To Each Technology Its Own Ethics: The Problem of Ethical Proliferation. *Philosophy & Technology*, 35, 93. <https://doi.org/10.1007/s13347-022-00591-7>
- Sánchez Vila, E. M., & Lama Penín, M. (2007). Monografía: Técnicas de la Inteligencia Artificial Aplicadas a la Educación. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 11(33), 7-12.
- Simon, H. A. (1996). *The Sciences of the Artificial*. The MIT Press. 3ª ed.
- Ulpiano, E. D. (1836). Digesto. En *Academicun parisiense, Corpus Iuris civilis*. <https://gallica.bnf.fr/ark:/12148/bpt6k65768518/f13.item.r=digestum>
- Unesco. (2019). *Beijing consensus on artificial intelligence and education*. <https://unesdoc.unesco.org/ask/48223/pf0000368303>
- Unesco. (2022). *Recomendación sobre la ética de la inteligencia artificial*. https://unesdoc.unesco.org/ark:/48223/pf0000379920_spa.page=15.
- Unión Europea. (2022). *Directrices éticas sobre el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2766/898>
- United Nations. (2019). *The Sustainable Development Goals Report 2019*. <https://bit.ly/34nbq60>