La ética de la Inteligencia Artificial en la educación: hacia un uso responsable e inclusivo[1]

Julio Cabero-Almenara[2]
Orcid: 0000-0002-1133-6031
Julio Barroso-Osuna[2]
Orcid: 0000-0003-0139-9140

Resumen

El artículo aborda la relación entre la inteligencia artificial (IA) y la ética en el ámbito educativo, destacando la necesidad de un uso responsable e inclusivo de esta tecnología. Se analiza cómo la IA, tras la irrupción de herramientas como ChatGPT, está transformando los procesos educativos, desde la personalización del aprendizaje hasta la automatización de tareas docentes. No obstante, este avance tecnológico también conlleva desafíos éticos significativos, como la privacidad de los datos, la equidad en el acceso, los sesgos algorítmicos y el impacto ambiental. El texto enfatiza la importancia de formar tanto a estudiantes como a docentes no solo en el uso técnico de la IA, sino también en su comprensión crítica y en la adopción de principios éticos que guíen su utilización. Para ello, se hace una llamada a establecer normativas internacionales que regulen esta tecnología, protejan los derechos humanos y promuevan una distribución justa de sus beneficios. El artículo subraya que el verdadero desafío radica en integrar la IA en la educación de manera que impulse el pensamiento crítico, fomente la equidad y responda a las necesidades sociales actuales y futuras.

Palabras clave

Inteligencia Artificial – Ética – Brecha digital – Transformación educativa.

1- Disponibilidad de los datos: Todo el conjunto de datos que apoya los resultados de este estudio se ha publicado en el propio artículo.

2- Universidad de Sevilla, Sevilla, España. Contactos: cabero@us.es; jbarroso@us.es

https://doi.org/10.1590/S1678-4634202551293347es
This content is licensed under a Creative Commons attribution-type BY 4.0.

*The ethics of Artificial Intelligence in education: towards responsible and inclusive use**

Abstract

This article addresses the relationship between artificial intelligence (AI) and ethics in education, highlighting the need for responsible and inclusive use of this technology. It analyzes how AI, following the emergence of tools such as ChatGPT , is transforming educational processes, from personalized learning to the automation of teaching tasks. However, this technological advancement also entails significant ethical challenges, such as data privacy, equity of access, algorithmic bias, and environmental impact. The article emphasizes the importance of training both students and teachers not only in the technical use of AI, but also in its critical understanding and the adoption of ethical principles that guide its use. To this end, it calls for the establishment of international regulations that regulate this technology, protect human rights, and promote a fair distribution of its benefits. The article emphasizes that the real challenge lies in integrating AI into education in a way that fosters critical thinking, promotes equity, and responds to current and future social needs.

Keywords

Artificial Intelligence – Ethics – Digital Divide – Educational Transformation.

Referencias iniciales

Hasta la fecha, dos grandes tecnologías han revolucionado las instituciones educativas y las formas de acceder al conocimiento y la información: el libro y la internet. Estas tecnologías no solo han transformado los paradigmas educativos, sino que también han generado impactos significativos en el acceso, la producción y la distribución del conocimiento en la sociedad.

Por lo que se refiere al libro, su incorporación representó un cambio trascendental tanto desde una perspectiva social como educativa. Como destaca Harari (2024), en 1424, la biblioteca de la Universidad de Cambridge contaba únicamente con 122 volúmenes, todos ellos manuscritos, esta cantidad indica las limitaciones en el acceso al conocimiento existente en esa época. Sin embargo, la invención de la imprenta, 30 años después, marcó un antes y un después en la historia de la humanidad, al permitir la producción 50 años después de más de doce millones de libros en Europa.

Este avance tecnológico transformó profundamente la educación, alterando los roles tradicionales que desempeñaban los docentes y estudiantes. Los profesores dejaron de ser los únicos depositarios del conocimiento, ya que los estudiantes podían acceder directamente a la información a través de los libros. Asimismo, el libro de texto emergió como un recurso fundamental que estandarizó los contenidos educativos, configurando así una base común para todos los estudiantes en lo que se refiere a los aprendizajes y los procesos de enseñanza.

CC BY

La aparición de Internet transformó la manera de acceder a la información, deslocalizó la información, y rompió las tres unidades en las cuales se solía apoyar la formación: unidad de tiempo, unidad de espacio y unidad de acción.

Pero esta situación se ha visto transformada con la tecnología de la Inteligencia Artificial (IA), que, aunque existía hace tiempo, desde el lanzamiento el 30 de noviembre de 2022 de ChatGPT, ha adquirido un papel preponderante en todos los sectores de la sociedad, produciendo una revolución no solo tecnológica, sino también social, cultural, económica, política, y por supuesto educativa. Como señalan Miao *et al.* (2021, p. 7):

> Solamente en los últimos cinco años, debido a algunos éxitos destacados y a su potencial disruptivo, la inteligencia artificial (IA) ha pasado de los remansos de la investigación académica a la primera línea de los debates públicos, incluso a nivel de las Naciones Unidas.

Su incorporación ha repercutido, no sólo en la transformación de los procesos seguidos en diferentes acciones y contextos, sino también influyendo fuertemente en el sector laboral, donde se estima que su impacto en el mercado será significativo, ya que esta tecnología generativa podría eliminar aproximadamente 300 millones de empleos a tiempo completo. No obstante, también se prevé una repercusión positiva, pues se piensa que su introducción en diferentes sectores tendrá una repercusión en del 7 % del PIB mundial, impulsado por su integración en múltiples sectores (Hatzius *et al.*, 2023; Greenhouse, 2023).

Pero esta fuerte y rápida penetración también ha generado miedos y temores en el sector educativo, llamándose rápidamente la atención, sobre diferentes aspectos como son: la actualización de los métodos de evaluación de los estudiantes, el establecimiento de políticas y reglamentos en las instituciones educativas para su utilización, y en la necesidad de la formación del profesorado y de los estudiantes para un uso de calidad y ético (Lo, 2023). Estos temores rápidamente se tradujeron en la aparición de manuales, guías de buenas prácticas, y normativas de restricción para su utilización elaboradas por las Universidades, como claramente puede observarse en la revisión que de las mismas realiza González (2024).

Las recomendaciones de utilización no solo han sido efectuadas por las instituciones educativas, sino también por instituciones sociales y políticas, como la Unesco (UNESCO, 2021a; 2021b) o la Comisión Europea (Comisión Europea, 2022). En este último caso cabe señalar que la Unión Europea ha sido la primera región del planeta que ha establecido normas armonizadas en materia de inteligencia artificial (Reglamento de la UE 2024/1689 de 13 de junio de 2024), que persiguen el mejorar el funcionamiento del mercado interior mediante el establecimiento de un marco jurídico, en particular para el desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial.

¿Qué se puede entender por la IA?

A la hora de definir que es la IA, lo primero es asumir que no existe una definición única, aunque de acuerdo con la Comisión Mundial de Ética del Conocimiento Científico y la Tecnología de la Unesco (2019), se la puede entender como máquinas que

> [...] son potencialmente capaces de imitar o incluso superar las capacidades cognitivas humanas, incluyendo la detección, la interacción lingüística, el razonamiento y el análisis, la resolución de problemas e incluso la creatividad.

Según Baker, Smith y Anissa (2019), la IA se refiere a sistemas informáticos capaces de realizar tareas cognitivas generalmente asociadas con la mente humana, como el aprendizaje y la resolución de problemas. Esta definición aporta varios aspectos clave: en primer lugar, la IA no se limita a una única tecnología, sino que abarca un conjunto diverso de herramientas, desde algoritmos hasta aplicaciones de aprendizaje automático; en segundo lugar, su evolución futura es incierta, lo que plantea, por una parte, múltiples posibilidades en su desarrollo; y por otra, un gran desconocimiento en comprender sus verdaderas posibilidades y por tanto las consecuencias futuras que pueda tener su implantación.

Pero centrándonos en el ámbito educativo, lo verdaderamente significativo es que está alcanzando a todos los sectores, funciones y personas que participan en las instituciones educativas, desde los directivos, hasta los docentes y discentes, sin olvidarnos del personal de administración y servicios y de las familias. Por otra parte, indicar que también está teniendo fuertes impactos en la forma de organizar el curriculum, generar materiales de enseñanza, evaluar a los estudiantes, o las formas en las cuales pueden ser admitidos los estudiantes en las instituciones.

Tal está siendo su relevancia, que en poco tiempo han aumentado las investigaciones centradas en el análisis de sus posibilidades y usos educativos. Esto se refleja en el aumento de metaanálisis que sobre sus investigaciones se están realizando en la búsqueda de principios sólidos para su incorporación en la enseñanza (Gallent; Zapata; Ortego, 2023; García Peñalvo *et al*., 2024; Mena-Guacas *et al*., 2024; Bond *et al*., 2024; López-Regalado *et al*., 2024). Metaanálisis que indican una serie de cuestiones: hay una fuerte preocupación por las implicaciones éticas del uso de la IA, necesidad de formación del profesorado y del alumnado, facilita la incorporación de prácticas innovadoras de enseñanza, necesidad de realizar investigaciones educativas para conocer su efecto real cuando es incorporada a la enseñanza y que su uso mejora el rendimiento académico.

Por lo que respecta a la temática en la cual se centra el presente artículo, la relación entre la ética y la IA, apuntar que ya han surgido algunos metaanálisis de investigaciones centradas en esta cuestión (Vélez-Rivera *et al*., 2024) que ponen de manifiesto la necesidad de contar con regulaciones y normativas éticas para el uso responsable de la IA en educación superior, destacando la importancia de equilibrar sus beneficios con los desafíos éticos, y de incorporarla desde una perspectiva crítica y no exclusivamente tecnológica-instrumental. Asimismo, estos metaanálisis subrayan la necesidad de desarrollar programas de estudio que permitan la formación de profesionales dotados de competencias digitales y conciencia ética para el diseño de estas herramientas.

La integración de la inteligencia artificial en la educación puede adoptar diversas formas, dependiendo del nivel educativo en el que se aplique. Sin embargo, en términos generales, podemos encuadrarla en las siguientes categorías:

a) Orientar a los estudiantes (enseñar a los estudiantes a utilizar la IA y utilizarla como apoyo a los estudiantes).
b) Orientar al profesorado (apoyo al docente en: investigación, docencia, evaluación, producción de recursos, construcción de instrumentos y pruebas de evaluación, ...).
c) Orientar a la institución (apoyar a la institución en el diagnóstico y planificación).

De forma más específica y centrados en la docencia se concretan en: la utilización para la selección de los estudiantes, la potenciación del aprendizaje personalizado, la tutoría virtual, el análisis de los sentimientos, evaluación automática del proceso de aprendizaje seguido por los estudiantes y el seguimiento del proceso de aprendizaje adquirido (Albarrán, 2023).

Y en todos estos usos, la IA ofrece una serie de ventajas para su utilización e incorporación a las instituciones educativas, que sin pretender agotar el tema se pueden concretar en las siguientes:

a) Suministro de información sobre los antecedentes de los estudiantes, lo que permite adaptar el proceso formativo a sus necesidades, características y debilidades formativas.
b) Ayudar al docente en la toma de decisiones sobre el contenido de aprendizaje.
c) Facilitar la organización de la planificación de actividades por el docente.
d) Monitorizar los aprendizajes adquiridos por los estudiantes.
e) Reducir la carga de trabajo de los profesores.
f) Disminuir la carga administrativa que deben realizar los docentes.
g) Seguimiento del proceso de aprendizaje seguido por el estudiante.
h) Mejor predicción/evaluación del desempeño/resultados de los docentes.
i) Contribuir al logro de los Objetivos de Desarrollo Sostenible.
j) Hacer funcionar una gran mayoría de app, las denominadas IA generativa, mediante el lenguaje natural. Lo que facilita su utilización.
k) Realización de valoraciones y evaluaciones automatizadas. (Celik *et al.*, 2022; Alto, 2023; Cortina, 2024).

Frente a estas posibilidades, la IA también presenta una serie de limitaciones, riesgos y preocupaciones, que deben ser considerados al momento de incorporarla a las instituciones educativas, tales como:

a) Falta de precisión y validez de la información generada, la aparición de las denominadas "alucinaciones".
b) Problemáticas en torno a la responsabilidad y ética respecto a la adquisición de los datos por parte de la IA.
c) Adaptación curricular y cambio de roles.
d) Limitaciones y sesgos del modelo.
e) Destrucción de puestos de trabajo, con las implicaciones sociales que ello supone.
f) Protección de los derechos de propiedad intelectual.
g) Limitado razonamiento matemático.

h) Privacidad y seguridad de los datos.
i) Problema de asumirla como verdad única.
j) Replanteamiento de las formas de evaluar a los estudiantes, debido a los riesgos de plagio.
k) La falta de comprensión profunda de su funcionamiento.
l) Adopción y aceptación (Cooper, 2023; Jiménez *et al*., 2023; Farrokhnia *et al*., 2023; Mayol, 2023; Cortina, 2024).

Considerar estas limitaciones es de suma importancia, en primer lugar, porque, como se analizará a lo largo de este trabajo, muchas de ellas tienen implicaciones éticas y legales; y, en segundo lugar, porque tanto docentes como estudiantes suelen adoptar actitudes muy positivas hacia la IA, sin desarrollar una visión crítica respecto a sus posibilidades y limitaciones, lo que puede derivar en un uso automático e inconsciente de esta herramienta. Esto es si cabe más peligroso en el caso de los estudiantes, ya que la IA se ha consolidado como una de sus herramientas tecnológicas preferidas para la realización de actividades académicas, que incluso llegan a utilizarlas cuando se les advierte explícitamente que no deben emplearla, como evidenciaron Ramírez y López (2024) en su investigación.

Por otra parte, no se debe obviar que en su utilización subyace cierta asunción de "tecnocentrismo", donde todo lo posible a alcanzar depende de la máquina, como se refleja en el siguiente "post" escrito en una web en el que se le concede a la tecnología un poder mágico para transformar la educación: "Al combinar la inteligencia artificial con enfoques como el aprendizaje autónomo y la colaboración entre pares, podemos crear experiencias educativas más personalizadas y efectivas. La IA no solo optimiza el tiempo y el feedback, sino que también impulsa la participación y el pensamiento crítico. ¡Impulsa tu aula sin fronteras!"

El debate educativo en torno a la IA ha ganado una creciente relevancia en los últimos años, y su análisis puede abordarse desde al menos dos grandes perspectivas complementarias. La primera, de carácter didáctico, que explora cómo las herramientas de IA pueden integrarse en los procesos de enseñanza y aprendizaje, transformando las metodologías pedagógicas, los recursos educativos y las dinámicas entre docentes y estudiantes. La segunda, de naturaleza más filosófica y política, se centra en las implicaciones sociales, culturales y éticas de formar ciudadanos para un mundo profundamente influido por esta tecnología.

Desde un enfoque didáctico, la IA ofrece una amplia gama de posibilidades para personalizar los procesos educativos, mejorar la accesibilidad y enriquecer las experiencias de aprendizaje. Los distemas basados en aprendizaje automático permiten adaptar los contenidos a las necesidades y ritmos de cada estudiante, facilitando una educación más inclusiva y equitativa. Sin embargo, este potencial debe ser equilibrado con una reflexión crítica sobre el papel de la IA en la formación de sujetos autónomos, capaces de pensar críticamente y no depender exclusivamente de esta tecnología.

La segunda perspectiva, que adopta un enfoque filosófico y político, requiere un análisis más profundo de las implicaciones éticas, sociales y políticas de vivir en un mundo transformado por la IA. Como destacan Ivanov (2023), Zawacki-Richter *et al*.

CC BY

(2019) y Cortina (2024), esta reflexión exige un cuestionamiento crítico tanto de los desafíos y riesgos que plantea la IA, así como de los valores y principios que deben guiar su integración en la sociedad.

Como se ha mencionado en esta introducción, el uso de la IA puede conllevar diversos problemas técnicos, tal como han señalado distintos autores e instituciones (UNESCO, 2021a; Comisión Europea, 2022; Alonso-Rodríguez, 2024; Cortina, 2024), cuestión que será abordada a continuación.

IA y ética en su utilización educativa

Desde la aparición de la IA, siempre ha existido una preocupación por los aspectos éticos relacionados con su utilización. Ya la Unesco en su reunión de Beijing (UNESCO, 2018), realizó una serie de recomendaciones (7, 12, 23, 24, 25, 26, 27 y 28) donde llamaba la atención sobre la necesidad de que el desarrollo de la IA debía estar controlado por el ser humano y centrado en las personas. En estas recomendaciones aunque se sugería su utilización en la docencia, también se llamaba la atención respecto una serie aspectos, como por ejemplo: la necesidad de no perder la interacción entre estudiantes y docentes; el velar por que la IA promueva oportunidades formativas sin distinción por razón de género, discapacidad, condición social; la no discriminación de los estudiantes con discapacidades o dificultades de aprendizaje; o el preocuparse porque no se produzcan discriminaciones en función del género de las personas.

A estas preocupaciones hay que incorporarle la visión acertada que muestra Cortina (2024, p. 205) cuando afirma

> [...] que no es menos cierto que la tecnoeducación se ha convertido en una poderosa arma económica en manos de las empresas y los países capaces de liderarla, de modo que quienes queden atrás en la competición por el primer puesto tendrán que pagar caro el coste de oportunidad, perderán poder económico y político.

Para O'Neil (2017), el uso de la IA está generando diversos problemas. Por un lado, la concentración de poder en manos de un pequeño grupo de corporaciones y países, principalmente Estados Unidos y China, está incrementando la brecha digital a nivel global. Por otro, el impacto ambiental de la IA es significativo, ya que el funcionamiento de los centros de datos y la formación de modelos requieren grandes cantidades de energía y recursos, contribuyendo de manera considerable a la crisis climática. Además, surgen brechas en la ética y la transparencia, dado que los sistemas de IA a menudo operan de manera opaca, dificultando comprender las razones existentes detrás de sus decisiones. Asimismo, su implementación ha derivado en una vigilancia que amenaza derechos fundamentales como la privacidad y la libertad de expresión, al tiempo que sus algoritmos pueden discriminar a ciertos colectivos. En la IA hay dos modelos de control de los datos, en un caso lo hacen las empresas, caso de EE. UU., y en otro el estado, caso de China. A lo anterior se suma su impacto en el ámbito laboral y su potencial uso con fines militares.

Ante este panorama, la autora subraya la necesidad de desarrollar políticas internacionales que regulen la IA, protejan los derechos humanos y garanticen una distribución equitativa de sus beneficios.

Lo apuntado lleva a una necesaria reflexión, puesto que el verdadero desafío no radica simplemente en aprender a manejar las apps de la IA, ya que este aprendizaje será un proceso natural e inevitable que la mayoría de las personas irán adquiriendo con el tiempo y la práctica, puesto que la IA estará incorporada en una diversidad de tecnologías, y funcionan casi sin darnos cuentas. Más bien, el reto principal está en comprender y gestionar los profundos cambios que la IA traerá consigo a medida que evoluciona y se vuelve cada vez más potente y sofisticada, tendiendo a la realización de habilidades cognitivas claramente humanas, como, por ejemplo: el cálculo mental, la navegación espacial, la resolución de problemas rutinarios, o el pensamiento analítico básico (Tuomi, 2018; Loján *et al*., 2024). Estos cambios afectarán no solo la forma en que interactuamos con la tecnología, sino también nuestras estructuras sociales, los valores éticos que guían nuestras decisiones, los modelos económicos predominantes y hasta nuestra percepción del mundo y de nosotros mismos como seres humanos en una sociedad digitalizada. Por lo tanto, el enfoque debe estar en cómo anticipar, adaptar y orientar esos impactos hacia un futuro más equitativo, inclusivo y consciente.

Entrando ya directamente en el tema de las relaciones entre la IA y la ética, y siguiendo a Cortina (2024), se pueden señalar dos perspectivas fundamentales en esta relación, or un lado, establecer principios y, por otro, que las personas adquieran principios éticos para su utilización responsable. Ambas perspectivas serán abordadas a continuación.

Establecer principios éticos en el diseño y construcción de modelos de IA

Una de las críticas realizadas a la IA se refiere a los sesgos que presenta la información que ofrece. De forma concreta podemos hablar de lo que Sabzalieva & Valentini (2023, p. 11) denominan sesgo cognitivo que sugiere que la IA

> [...] no se rige por principios éticos y no puede distinguir entre lo correcto y lo incorrecto, lo verdadero y lo falso. Esta herramienta sólo recoge información de las bases de datos y textos que procesa de Internet, por lo que también aprende cualquier sesgo cognitivo que se encuentre en dicha información.

A esta perspectiva se le puede incorporar la dimensión de género y diversidad, como consecuencia de la "falta de participación femenina en temas relacionados con la IA, en la investigación y desarrollo de la IA y, por otro, el poder de la IA generativa para producir y difundir contenidos que discriminan o refuerzan estereotipos de género y de otro tipo". (Sabzalieva; Valentini, 2023, p. 11). Todo ello es resultado de los datos con los que se entrenan a estos sistemas y que pueden llevar a ofrecer sesgos en sus respuestas, y de este modo afectar la calidad, veracidad y precisión de las respuestas ofrecidas. En un informe elaborado por IBM (2023), se presentan diversos ejemplos que evidencian la existencia de sesgos en los sistemas de inteligencia artificial, los cuales pueden derivar en desigualdades

significativas en distintos ámbitos. En el sector de la salud, por ejemplo, se ha identificado que los sistemas de diagnóstico asistido por ordenador (CAD) presentan una menor precisión en la evaluación de pacientes pertenecientes a grupos minoritarios, como la población negra, en comparación con los pacientes blancos, debido a la infrarrepresentación de ciertos datos en los conjuntos de entrenamiento. De manera similar, en el ámbito de la publicidad digital, se ha observado que los algoritmos de segmentación utilizados por plataformas como Google tienden a mostrar con mayor frecuencia ofertas laborales mejor remuneradas a los hombres que a las mujeres, lo que perpetúa las brechas de género en el acceso a oportunidades económicas. Por otro lado, en el ámbito de la seguridad, el uso de herramientas de vigilancia y predicción del crimen basadas en IA ha revelado sesgos derivados del entrenamiento con datos históricos de detenciones, lo que ha llevado a una sobrerrepresentación de perfiles raciales y a una focalización desproporcionada en comunidades marginalizadas.

Estos sesgos producidos por la IA se observan con claridad en relación con el género, la edad y el color de piel de la persona en los programas de generación de imágenes mediante IA, donde suele privilegiarse la edad joven, los hombres y las personas blancas, en la generación de imágenes referidas a puestos de trabajo (Thomson; Thomas, 2023). Como señala Ortiz de Zárate (2023, p. 9): "Durante años, distintos estudios, algunos de los más recientes del año 2018, demostraban que, mientras que casi el 100% de los hombres blancos eran reconocidos de manera exitosa por sistemas de reconocimiento facial, la tasa de éxito disminuía hasta un 35% en el caso de las mujeres racializadas."

Esta situación está repercutiendo en diversos ámbitos, lo que ha llevado a que se reclame (Yu; Guo, 2023) establecer medidas adecuadas para desarrollar algoritmos explicables y justos, actualizar la tecnología de cifrado y formular leyes y regulaciones relevantes para proteger los datos, así como mejorar la calidad y cantidad de los conjuntos de datos con los cuales se entrenan los modelos. Al mismo tiempo sugerir nuevos modelos de entrenamiento de la IA, y en este sentido la Unesco (2023) propone cambiar los modelos utilizados para el entrenamiento de la IA, y adoptar nuevos formatos y propuestas:

> [...] es esencial que se realicen esfuerzos para refinar los modelos básicos, no solo mediante la adición de conocimientos sobre la materia y la eliminación de prejuicios, sino también mediante la adición de conocimientos sobre métodos de aprendizaje relevantes, y cómo esto puede reflejarse en el diseño de algoritmos y modelos. El reto consiste en determinar hasta qué punto los modelos EdGPT pueden ir más allá del conocimiento de la asignatura para dirigirse también a la pedagogía centrada en el alumno y a las interacciones positivas entre profesores (UNESCO, 2023, p. 13).

Además de los riesgos derivados de los sesgos, surge otro problema de tipo ético, referido a la dificultad con que nos podemos encontrar respecto a proteger la privacidad de la ciudadanía, ya que cada vez que interaccionamos con la IA, le estamos ofreciendo información sobre nuestra persona, gustos y fobias; y por tanto facilitando información, que permita que nos conozcan en profundidad.

Este aspecto de la privacidad es especialmente preocupante cuando se analiza una de las posibilidades que la IA ofrece a la enseñanza con el denominado "aprendizaje personalizado"; es decir, la adaptación de la acción formativa a las características cognitivas y necesidades del estudiante. Para la aplicación de este modelo es necesario que la máquina disponga de cierta información del estudiante, y por ello se pueden hacer una serie de preguntas: ¿quién garantiza que dicha información no pudiera ser utilizada en un futuro con otros propósitos y fines diferentes para los que inicialmente fueron solicitados? O ¿quién asegura que en un futuro esos datos que se posean sobre el estudiante no serán utilizados en procesos selectivos, como por ejemplo para selección de puestos laborales, o para ofrecerles productos comerciales a él o a su familia?

En este último aspecto, Alonso-Rodríguez (2024, p. 84) llama la atención respecto a que 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.

Sin olvidarnos que en el terreno educativo hay que hacer una serie de reflexiones por las consecuencias futuras que pueden tener para las personas algunas de las decisiones que actualmente se adoptan en el uso de la IA:

> 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 (Alonso-Rodríguez, 2024, p. 85).

Las palabras que apunta Cortina (2024, p. 31) pueden servir para esclarecer lo que se está comentado, para esta autora: "entre hacer uso de sistemas inteligentes (sean máquinas, algoritmos o robots) a la hora de tomar decisiones y delegar en esos sistemas inteligentes decisiones significativas para la vida de las personas y de la naturaleza."

Pero hablar de establecer principios éticos en el uso de la IA implica no olvidarque su incorporación de la IA al sistema educativo está ampliando la brecha digital entre los países y los colectivos de personas, pues aquellos con mayores recursos la están incorporando con mayor facilidad y de manera masiva, y en consecuencia tienen más posibilidades de beneficiarse de las ventajas que ofrece, y ello repercutirá en la ampliación de la brecha digital ya existente, como se puso de manifiesto durante la Covid-19, afectando directamente a la equidad social y educativa. Esta situación se ve agravada por el hecho de que no todos los servicios ni todas las versiones de las apps de IA son gratuitas. En consecuencia, el tipo de uso que puede hacer la persona y la calidad del producto que le

CC BY

ofrezca la herramienta, estará condicionada por las características sociales y económicas de la persona, y ello se convertirá en un elemento de ampliación de la brecha digital entre las personas. Por ejemplo, en el caso del ChatGPT, la versión gratuita 3.5, tiene frente a la 4.0, que es de pago, una capacidad limitada de mantener conversaciones largas solo funciona con texto, comete más errores e información sesgada, y revisa la información hasta una fecha concreta.

Adquisición de valores éticos por la persona a la hora de la utilización de la IA

No se trata solo de adoptar una serie de medidas para que en el diseño de los modelos se deban tener en cuenta una serie de medidas éticas. Sino también de formar a las personas para que hagan un uso ético y responsable de la IA.

Nos es ninguna novedad afirmar que, desde la aparición de ChatGPT, las principales preocupaciones de los docentes y administradores educativos, se han centrado en la posibilidad que ofrece para facilitar el plagio de los trabajos que realizan los estudiantes, y por tanto favorecer la falta de honestidad de ciertos estudiantes para la realización de trabajos académicos.

Esta situación se ha visto agravada por la limitada eficacia de las herramientas tradicionales de detección de plagio disponibles en el mercado, o por la ineficacia mostrada por las herramientas de detección de trabajos generados por ChatGPT o herramientas similares de generación de contenidos para garantizar al cien por cien cómo se ha generado el texto (GPTZero, Plagscan, ...) así como por la falta de opciones más avanzadas. Además, la idea de que los docentes podrían identificar el uso indebido de estas tecnologías basándose en el lenguaje empleado tampoco ha sido respaldada por las investigaciones realizadas.

Pero la situación no solo debe enfocarse desde lo ya comentado, sino desde aspectos mucho más transcendentales para el futuro de las personas y de la sociedad. Autores como Desmurget (2020) o Luri (2020), vienen llamando la atención, por una parte, sobre el empobrecimiento conceptual y científico de las personas que se están formando en nuestras instituciones educativas, y por otra, sobre la disminución de los niveles de exigencia y requerimientos solicitado a los estudiantes para poder superar los cursos y ciclos formativos. Exigencias que progresivamente van disminuyendo, debido, por una parte, por la presión de las instituciones, que en sus estándares deben alcanzar unos niveles exitosos de estudiantes que supera el curso, y por otra, por un cierto "progresismo pedagógico" ilustrado de muchos pedagogos y pedagogas "YouTube", donde el esfuerzo a realizar por los estudiantes está en el ver, y en el "buenismo" y no en el comprender y la realización de esfuerzo cognitivo.

No debe caber la menor duda que esta situación, de manera progresiva, no solo traerá un empobrecimiento conceptual de los estudiantes, sino también situaciones de calado para el futuro y para las posibilidades de promoción social de los estudiantes de clases sociales populares. Siempre hemos defendido que instituciones formativas públicas deben requerir un esfuerzo elevado, ya que ello es la mejor política social para facilitar

la promoción y el avance social de las clases populares. En este sentido, el razonamiento expuesto por Cortina es claro e irrefutable:

> [...] las gentes con poder adquisitivo se cuidan de dar a sus hijos una educación personalizada, que exige esfuerzo, competencias lingüísticas y culturales de todo tipo para que estén bien preparados, mientras en la educación pública, a menudo, en vez de pertrechar de una preparación semejante al alumnado, se dan todo tipo de facilidades para que vaya obteniendo los diplomas sin esfuerzo alguno. Es éste un proceder claramente demagógico, que refuerza lo que convendría llamar «la trampa de la pobreza cultural», porque es un callejón sin salida. No ayuda a formar en la excelencia, sino en la mediocridad (Cortina, 2024, p. 218).

Pero frente a este aspecto, un uso laxo de la IA en los procesos formativos, que se limite a facilitar la realización de los trabajos y actividades encomendadas por los docentes, existe un problema más complejo y es la falta de la visión y reflexión crítica por parte del estudiante de la información con la que trabaja, aspecto clave para la formación de una personalidad profesional que pueda llevar a cabo una transformación de la realidad social. Precisamente, esta es una de las grandes limitaciones que la IA puede traer a la educación, ya que, al ofrecer una herramienta cómoda y sencilla para la realización de las tareas escolares, puede llevar a que los estudiantes releguen el esfuerzo cognitivo necesario para construir un pensamiento crítico sobre la realidad que analizan, lo que termina simplificando y empobreciendo su madurez intelectua.

En cierta medida, y en relación con lo comentado anteriormente, surge el problema de la posible falta de autonomía que puedan ir adquiriendo los estudiantes, al ir teniendo una excesiva dependencia de la herramienta para su desenvolvimiento educativo y delegar en ella la realización de diferentes acciones formativas y asumir con displicencia los resultados que esta ofrece.

Como ha señalado recientemente Morozov (2024) podemos incorporar la IA desde dos perspectivas, una sería utilizarla para aumentar a la persona otra para mejorarla.

> La diferencia entre estos dos paradigmas es sutil pero crucial. Aumentar es cuando utilizamos el GPS del teléfono móvil para orientarnos en un lugar desconocido: permite llegar antes y más fácilmente al destino. El aspecto beneficioso es efímero. Si nos quitaran esta ayuda tecnológica nos sentiríamos todavía más desamparados. Mejorar consiste en utilizar la tecnología para desarrollar competencias nuevas -en este caso, refinar el sentido innato de la orientación recurriendo a técnicas avanzadas de memorización o aprendiendo a comprender las señales de la naturaleza. En definitiva, la aumentación nos desprovee de ciertas capacidades en nombre de la eficacia, mientras que la mejora nos hace adquirir otras y enriquece nuestras interacciones con el mundo (Morozov, 2024, p. 26).

Por otra parte, es fundamental adoptar una visión crítica de la IA, y ello pasa por reconocer que las tecnologías, incluida la IA, no son neutras; están impregnadas de las ideologías, intereses y valores de quienes las diseñan y controlan. Este hecho pone de relieve la necesidad de formar ciudadanos con competencias digitales avanzadas, que les

permitan no solo utilizar estas tecnologías, sino también comprender cómo funcionan, cómo son creadas y qué intereses subyacen a su desarrollo y aplicación.

Ello lleva a señalar que el verdadero desafío educativo no reside únicamente en enseñar a utilizar las herramientas de IA, sino en preparar a los estudiantes para un mundo en el que esta tecnología desempeña un papel central. Esto implica fomentar una educación crítica y ética, que trascienda el mero dominio técnico y promueva una comprensión profunda de las implicaciones sociales, económicas y culturales de la IA.

Formar ciudadanos críticos y conscientes en un mundo con IA implica cuestionar las narrativas que presentan a esta tecnología como inevitable o inherentemente beneficiosa, promoviendo, en su lugar, un diálogo que lleve a conocer sus limitaciones, sesgos y riesgos. Solo a través de esta reflexión crítica será posible construir una sociedad en la que la IA se utilice de manera responsable y equitativa, respetando los derechos y la dignidad de todas las personas.

Este enfoque no solo redefine las prioridades educativas, sino que también plantea una cuestión urgente: ¿cómo educar en una era donde la IA no es solo una herramienta, sino un actor transformador del tejido social? o ¿cómocrear una conciencia crítica en la ciudadanía que permita comprender y adaptarse a las situaciones caóticas que surgirán en los primeros momentos de la penetración de la IA en todos los sectores? Como ha señalado recientemente Ferrarelli (2024) en un informe para la OEI respecto a cómo abordar la IA desde Latinoamérica, en su utilización se van a dar no solo una serie de dilemas éticos, algunos de los cuales se han analizado en el presente artículo, sino también dilemas de ordenes ambientales, laborales y pedagógicas. Además, como advierte la autora del informe, hay una serie de malentendidos que conviene desentramar, como son que la IA no es: “un buscador”, “neutral”, “confiable” y “creativa”.

Referencias

ALBARRÁN, Eloy. **Hacia una educación personalizada y adaptativa**: la disrupción de la inteligencia artificial. Caracas: Centro Internacional de Educación Continua: Universidad Pedagógica Experimental Libertador, 2023.

ALONSO-RODRÍGUEZ, Ana María. Hacia un marco ético de la inteligencia artificial en la educación. **Teoría de la Educación: Revista Interuniversitaria**, v. 36, n. 2, p. 79-98, 2024. https://doi.org/10.14201/teri.31821

ALTO, Valentina. **Inteligencia artificial generativa como modelos de ChatGPT y OpenAI**. Madrid: Anaya, 2023.

BAKER, Toby; SMITH, Laurie; ANISSA, Nandra. Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. **Nesta**, 2019. Disponible en: https://www.nesta.org.uk/report/education-rebooted/. Acesso en: 20 Mayo 2025.

BOND, Melissa; KHOSRAVI, Hassan; DE LAAT, Maarten; BERGDAHL, Nina; NEGREA, Violeta; OXLEY, Emily; PHAM, Phuong; CHONG, Sin Wang; SIEMENS, George. A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. **International Journal of Educational Technology in Higher**, v. 21, 2024. https://doi.org/10.1186/s41239-023-00436-z

CELIK, Ismail; DINDAR, Muhterem; MUUKKONEN, Hanni; JÄRVELÄ, Sanna The promises and challenges of artificial intelligence for teachers: a systematic review of research. **TechTrends**, v. 66, p. 616-630, 2022. https://doi.org/10.1007/s11528-022-00715-y

COMISIÓN EUROPEA. **Directrices éticas sobre el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores**. Luxemburgo: Unión Europea, 2022.

COOPER, Grant. Examining science education in ChatGPT: an exploratory study of generative artificial intelligence. **Journal of Science Education and Technology**, v. 32, p. 444-452, 2023. https://doi.org/10.1007/s10956-023-10039-y

CORTINA, Adela. **¿Ética o ideología de la inteligencia artificial?** Barcelona: Paidós, 2024.

DESMURGET, Michel. **La fábrica de cretinos digitales**. Barcelona: Paidós, 2020.

FARROKHNIA, Mohammadreza; BANIHASHEM, Seyyed Kazem; NOROOZI, Omid; WALS, Arjen. A SWOT analysis of ChatGPT: implications for educational practice and research. **Innovations in Education and Teaching International**, v. 61, n. 3, p. 460-474, 2023. https://doi.org/10.1080/14703297.2023.2195846

FERRARELLI, Mariana. **Inteligencia artificial y educación**: insumos para su abordaje desde Iberoamérica. Buenos Aires: OEI, 2024.

GALLENT, Cinta; ZAPATA, Alfredo; ORTEGO, José Luis. El impacto de la inteligencia artificial generativa en educación superior: una mirada desde la ética y la integridad académica. **Relieve**: Revista Electrónica De Investigación y Evaluación Educativa, v. 29, n. 2, 2023. https://doi.org/10.30827/relieve.v29i2.29134

GARCÍA PEÑALVO, Francisco; LLORENS-LARGO, Faraón; VIDAL, Javier. La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. **Ried:** Revista Iberoamericana de Educación a Distancia, v. 27, n. 1, p. 9-39, 2024. https://doi.org/10.5944/ried.27.1.37716

GONZÁLEZ, Gregorio. **1 d.C. (Después de ChatGPT):** inteligencia artificial en la educación superior. Valencia: PUV Universitat de València, 2024.

GREENHOUSE. US experts warn AI likely to kill off jobs and widen wealth inequality. **The Guardian**, 2023. Disponible en: https://is.gd/n38xqn. Acceso en: 12 ene. 2025.

GRUPO DE TRABAJO DE INTELIGENCIA ARTIFICIAL GENERATIVA DE LA UNAM. **Recomendaciones para el uso de inteligencia artificial generativa en la docencia**. México, DC: UNAM, 2023.

HARARI, Yuvak. **Nexus**: una breve historia de las redes de información desde la Edad de Piedra hasta la IA. Barcelona: Debate, 2024.

HATZIUS, Jan; BRIGGS, Joseph; KODNANI, Devest; PIERDOMENICO, Giovanni. **The potentially large effects of artificial intelligence on economic growth**. Goldman Sachs, 2023. Diponible en: https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html. Acesso en: 20 Mayo 2025.

CC BY

IVANOV, Stanislav. The dark side of artificial intelligence in higher education. **The Service Industries Journal**, v. 43, p. 1055-1082, 2023. https://doi.org/10.1080/02642069.2023.2258799

INTERNATIONAL BUSINESS MACHINES. El sesgo de la IA, a la luz de ejemplos reales. **IBM**, 2023. Disponible en: https://www.ibm.com/es-es/think/topics/shedding-light-on-ai-bias-with-real-world-examples. Acesso em: 19 maio 2025.

JIMÉNEZ, Luis; LÓPEZ-GÓMEZ, Julio; MARTÍN-BAOS, José Angel; ROMERO, Francisco; SERRANO-GUERRERO, Jesús. ChatGPT: reflexiones sobre la irrupción de la inteligencia artificial generativa en la docencia universitaria. **Actas de las Jenui**, n. 8, p. 113-120, 2023.

LO, Chung. What is the impact of ChatGPT on education? A rapid review of the literature. **Education Sciences**, v. 13, n. 4, p. 410, 2023. https://doi.org/10.3390/educsci13040410

LOJÁN, María del Cisne.; ROMERO, J. Antonio; SANCHO AGUILERA, David; ROMERO, A. Yajaira. Consecuencias de la dependencia de la inteligencia artificial en habilidades críticas y aprendizaje autónomo en los estudiantes. **Ciencia Latina Revista Científica Multidisciplinar,** v. 8, n. 2, p. 2368-2382, 2024. https://doi.org/10.37811/cl_rcm.v8i2.10678

LÓPEZ REGALADO, Oscar; NÚÑEZ-ROJAS, Nemecio; RAFAEL LÓPEZ GIL, Oscar; SÁNCHEZ-RODRÍGUEZ, José. El análisis del uso de la inteligencia artificial en la educación universitaria: una revisión sistemática. **Pixel-Bit**: Revista de Medios y Educación, n. 70, p. 97-122, 2024. https://doi.org/10.12795/pixelbit.106336

LURI, Gregori. **La escuela no es un parque de atracciones**. Barcelona: Paidós, 2020.

MAYOL, Julio. Inteligencia artificial generativa y educación médica. **Educación Médica**, v. 24, n. 4, p. 1-3, 2023. https://doi.org/10.1016/j.edumed.2023.100851

MENA-GUACAS, Andrés; VÁZQUEZ-CANO, Estevan; FERNÁNDEZ-MÁRQUEZ, Esther; LÓPEZ-MENESES, Eloy. La inteligencia artificial y su producción científica en el campo de la educación. **Formación Universitaria**, v. 17, n. 1, p. 155-164, 2024. http://dx.doi.org/10.4067/s0718-50062024000100155

MIAO, Fengchun; HOLMES, Wayne; HUANG, Ronghuai; HUI, Zhang. **Inteligencia artificial y educación:** promesas e implicancias. París: Unesco, 2021.

MOROZOV, Evgeny. Otra inteligencia artificial es posible. **Le Monde Diplomatique en Español**, n. 346, p. 25-28, 2024.

O'NEIL, Cathy. **Armas de destrucción matemática**. Madrid: Capitán Swing, 2017.

ORTIZ DE ZÁRATE, Lucía. Sesgos de género en la inteligencia artificial. **Revista de Occidente**, n. 502, p. 5-20, 2023.

RAMÍREZ, Gina; LÓPEZ, Alison. Reflexiones éticas sobre la evaluación y la inteligencia. *In*: MUÑOZ, Juan Miguel; LORENZO, Neus; SUÑÉ, Xavier (org.). **Inteligencia artificial en la microeducación**: transformando el aula del futuro. [*S. l.*]: Odite, 2024. p. 123-134.

SABZALIEVA, Emma; VALENTINI, Arianna. **ChatGPT e inteligencia artificial en la educación superior**. Santiago de Chile: Unesco-Iesalc, 2023.

THOMSON, Travis Jason; THOMAS, Ryan. Ageism, sexism, classism and more: 7 examples of bias in AI-generated images. **The Conversation**, 2023. Disponible en: https://theconversation.com/ageism-sexism-classism-and-more-7-examples-of-bias-in-ai-generated-images-208748. Acesso em: 19 maio 2025.

TUOMI, Ilkka. The impact of artificial intelligence on learning, teaching, and education. Policies for the future. *In*: CABRERA, Marcelino.; VUORIKARI, Riina.; PUNIE, Yves. (ed.). **EUR 29442 EN**. Luxembourg: European Union, 2018. p. 1-48. Disponível em: 10.2760/12297. JRC113226. Acesso em: 19 maio 2025.

UNESCO. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. **Beijing consensus on artificial intelligence and education**. Beijing: Unesco, 2019. Disponívle em: https://unesdoc.unesco.org/ark:/48223/pf0000368303. Acesso em: 19 maio 2025.

UNESCO. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. **Guidance for generative AI in education and research**. París: Unesco, 2023.

UNESCO. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. **Inteligencia artificial y educación**: guía para las personas a cargo de formular políticas. París: Unesco, 2021a.

UNESCO. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. **Recomendación sobre la ética de la inteligencia artificial**. París: Unesco, 2021b.

VÉLEZ, Rocio; MUÑOZ ÁLVAREZ, Diana; LEAL ORELLANA, Priscila; RUIZ GARRIDO, Alejandra. Uso de Inteligencia Artificial en educación superior y sus implicancias éticas: mapeo sistemático de literatura. **Hachetetepé: Revista científica de Educación y Comunicación**, v. 28, p. 1105, 2024. https://doi.org/10.25267/Hachetetepe.2024.i28.1105

YU, Hao; GUO, Yunyun. Generative artificial intelligence empowers educational reform: current status, issues, and prospects. **Frontiers in Education**, v. 8, 1183162, 2023. https://doi.org/10.3389/feduc.2023.1183162

ZAWACKI-RICHTER, Olaf; MARÍN, Victoria; BOND, Melissa; GOUVERNEUR, Franziska. Systematic review of research on artificial intelligence applications in higher education – where are the educators? International **Journal of Educational Technology in Higher Education**, v. 16, n. 1, p. 1-27, 2019. https://doi.org/10.1186/s41239-019-0171-0

Recibido en: 12.01.2025
Revisado en: 11.02.2025
Aprobado en: 24.02.2025

Editor: Prof. Dr. Hugo Heredia Ponce

Julio Cabero Almenara es catedrático de la Universidad de Sevilla. Director del Grupo de Investigación Didáctica (GID- HUM-0390) y presidente de EDUTEC. Su área de conocimiento está relacionada con las Tecnologías Aplicadas a la Educación.

Julio Barroso Osuna es catedrático de la Universidad de Sevilla. Miembro del Grupo de Investigación Didáctica (GID- HUM-0390) y de EDUTEC. Su área de conocimiento está relacionada con las Tecnologías Aplicadas a la Educación.