

Inteligencia Artificial en educación: entre riesgos y potencialidades

Inteligência Artificial: entre riscos e potencialidades

Artificial Intelligence: between risks and potentialities

Cinta Gallent-Torres*

 <https://orcid.org/0000-0002-4260-7594>

Begoña Arenas Romero**

 <https://orcid.org/0009-0000-6561-4183>

María Vallespir Adillón***

 <https://orcid.org/0009-0007-0218-7162>

Tomáš Foltýnek****

 <https://orcid.org/0000-0001-8412-5553>

Resumen: Este artículo profundiza en el impacto de la Inteligencia Artificial (IA) en la educación, explorando tanto su potencial transformador como los riesgos asociados. A medida que las tecnologías de IA se integran cada vez más en los sistemas educativos, ofrecen oportunidades sin precedentes para mejorar las experiencias de aprendizaje a través de la educación personalizada, sistemas de tutoría inteligentes y monitoreo del rendimiento. Estas aplicaciones no solo facilitan un enfoque educativo a medida, sino que también introducen desafíos significativos y riesgos, como preocupaciones sobre la privacidad de los datos, sesgo algorítmico y el potencial de ampliar las desigualdades educativas. El artículo describe de manera exhaustiva los riesgos involucrados con la implementación de la IA en contextos educativos y propone medidas estratégicas para mitigarlos. Enfatiza la importancia de establecer directrices éticas robustas y marcos regulatorios para proteger contra el mal uso de los datos y asegurar la equidad y transparencia en las aplicaciones de IA. Además, el artículo aboga por un papel proactivo de los educadores y los formuladores de políticas en la configuración de la integración de la IA en la educación para alinear con los estándares éticos y los objetivos educativos. A través de un análisis detallado de las capacidades y peligros de la IA, el artículo ofrece una perspectiva equilibrada sobre cómo navegar las complejidades de la IA en la educación.

* Autor de correspondencia: Cinta Gallent-Torres, Universitat de València, Universitat de les Illes Balears, Stockholm University. E-mail: <cinta.gallent@uv.es>.

** Universidad Internacional de La Rioja (España). E-mail: <begona.arenas@unir.net>.

*** Universitat de les Illes Balears (España). E-mail: <m.vallespir@uib.es>.

**** Masaryk University (República Checa). E-mail: <foltynnek@fi.muni.cz>.

Invoca a un esfuerzo concertado para aprovechar los beneficios de la IA mientras se abordan los desafíos éticos, asegurando que la IA sirva como una herramienta para el enriquecimiento educativo equitativo.

Palabras clave: Inteligencia artificial. Desafíos éticos. Educación Superior.

Resumo: Este artigo investiga o impacto da Inteligência Artificial (IA) na educação, explorando tanto o seu potencial transformador quanto os riscos associados. À medida que as tecnologias de IA se tornam cada vez mais integradas aos sistemas educacionais, elas oferecem oportunidades sem precedentes para melhorar as experiências de aprendizado por meio de educação personalizada, sistemas de tutoria inteligentes e monitoramento de desempenho. Essas aplicações não só facilitam uma abordagem educacional sob medida, mas também introduzem desafios significativos e riscos, como preocupações com a privacidade dos dados, viés algorítmico e o potencial de ampliar as desigualdades educacionais. O artigo delinea de forma abrangente os riscos envolvidos com a implantação da IA em contextos educacionais e propõe medidas estratégicas para mitigar esses riscos. Ele enfatiza a importância de estabelecer diretrizes éticas robustas e frameworks regulatórios para proteger contra o uso indevido de dados e para garantir justiça e transparência nas aplicações de IA. Além disso, o artigo defende um papel proativo dos educadores e formuladores de políticas na modelagem da integração da IA na educação para alinhar com padrões éticos e objetivos educacionais. Por meio de uma análise detalhada das capacidades e armadilhas da IA, o artigo oferece uma perspectiva equilibrada sobre como navegar as complexidades da IA na educação. Ele convoca um esforço concertado para aproveitar os benefícios da IA enquanto se abordam os desafios éticos, garantindo que a IA sirva como uma ferramenta para o enriquecimento educacional equitativo.

Palavras-chave: Inteligência Artificial. Desafios éticos. Ensino Superior.

Abstract: This article focuses into the impact of Artificial Intelligence (AI) on education, exploring both its transformative potential and associated risks. As AI technologies become increasingly integrated into educational systems, they offer unprecedented opportunities for enhancing learning experiences through personalized education, intelligent tutoring systems, and performance monitoring. These applications not only facilitate a tailored educational approach but also introduce significant challenges and risks, such as data privacy concerns, algorithmic bias, and the potential for widening educational inequalities. The paper comprehensively outlines the risks involved with AI deployment in educational contexts and proposes strategic measures to mitigate these risks. It emphasizes the importance of establishing robust ethical guidelines and regulatory frameworks to safeguard against data misuse and to ensure fairness and transparency in AI applications. Furthermore, the article advocates for proactive roles by educators and policymakers in shaping the integration of AI in education to align with ethical standards and educational goals. Through a detailed analysis of AI's capabilities and pitfalls, the article provides a balanced perspective on navigating the complexities of AI in education. It calls for a concerted effort to harness AI's benefits while addressing the ethical challenges, ensuring that AI serves as a tool for equitable educational enhancement.

Keywords: Artificial intelligence. Ethical challenges. Higher Education.

1 INTRODUCCIÓN: LA IA EN LA EDUCACIÓN, UNA RELACIÓN NO TAN RECIENTE

El debate sobre el papel de la inteligencia artificial (IA) en la educación es tan antiguo como el propio concepto de IA. La primera conferencia dedicada exclusivamente a la IA en la educación se llevó a cabo en 1983, y en 1989 se creó la revista *International Journal of Artificial Intelligence in Education* (IJAED), según Self (2016). Desde ese momento, la Educación Asistida por Inteligencia Artificial (AIED) ha evolucionado y se ha implementado en niveles educativos que van desde la primaria hasta la educación superior. Un análisis exhaustivo realizado por Xu y Ouyang (2022) identificó diversas aplicaciones de la IA en la educación STEM (Ciencia, Tecnología, Ingeniería y Matemáticas), incluyendo la predicción del rendimiento estudiantil, sistemas de tutoría inteligentes, monitoreo del comportamiento estudiantil y la implementación de robots educativos, manteniéndose todas ellas pertinentes hasta la actualidad.

El uso de sistemas que alguna vez se consideraron IA en la década de 1980 es objeto de debate, destacando la importancia de una definición clara de IA, un desafío complicado por la falta de consenso sobre la conceptualización de "inteligencia". Alan Turing, en su artículo seminal "Computing Machinery and Intelligence" (1950), exploró el rendimiento de los sistemas computacionales y su indistinguibilidad del comportamiento humano. Esta perspectiva de la IA persiste, con la Comisión Europea (2018, p. 4), describiéndola como sistemas que demuestran "comportamiento inteligente" a través de la interacción autónoma con su entorno para alcanzar metas específicas, y la UNESCO (Miao et al., 2021, p. 3) que destaca su capacidad para procesar información e integrar algoritmos que facilitan el aprendizaje y la toma de decisiones.

La personalización del aprendizaje es uno de los principales beneficios atribuidos a la IA y a la tecnología en educación. Según Pelánek (2022), existen tres niveles de personalización que van desde sugerencias personalizadas simples hasta la adaptabilidad del contenido completo de un curso, pasando por ajustes dentro de actividades específicas. Los sistemas de aprendizaje personalizado buscan presentar desafíos adecuados, evitando tanto la desmotivación por dificultad excesiva como el aburrimiento por tareas demasiado simples. Los sistemas de tutoría inteligente, por ejemplo, ofrecen contenido instructivo personalizado y recomendaciones adaptadas a las necesidades individuales en diversas áreas, como la física, las matemáticas o la programación (Xu y Ouyang, 2022).

Otra vertiente de aplicación incluye sistemas que analizan comportamientos, patrones y características de aprendizaje, posibilitando mejoras en las estrategias docentes e incrementando la motivación estudiantil. La predicción del rendimiento académico y la identificación de estudiantes en riesgo son, por ejemplo, componentes fundamentales en la prevención del fracaso y abandono escolar (Takaki et al., 2022). Además, existen sistemas que ofrecen diversas formas de automatización en los procesos de evaluación, desde la creación de preguntas y retroalimentación hasta la evaluación automatizada, especialmente útil en áreas muy estructuradas como la programación (Gómez, 2023).

El aprendizaje colaborativo también se ha beneficiado de la IA, con sistemas diseñados para recolectar y estructurar datos de interacción y ofrecer asesoramiento, usando técnicas que hoy son pilares de la IA moderna, como árboles de decisión y redes neuronales (Soller et al., 2005). Según Xu y Ouyang (2022), la AIED ha demostrado tener un impacto positivo en el rendimiento estudiantil, mejorando patrones de aprendizaje y motivación.

El interés en AIED se intensificó tras la irrupción de ChatGPT en noviembre de 2022, seguido por la rápida evolución de aplicaciones similares basadas en grandes modelos de lenguaje (LLM), que han mostrado avances significativos en el procesamiento del lenguaje natural (PLN) desde las primeras investigaciones en las décadas de 1950 y 1960. La introducción de la arquitectura Transformer por Vaswani et al. (2017) marcó un avance crucial, facilitando tareas complejas de PLN mediante mecanismos de auto-atención. Los modelos GPT, como GPT-3 y GPT-4 ejemplifican la capacidad de estos sistemas para generar contenido (textual y audiovisual) coherente y relevante, abriendo nuevas posibilidades en numerosos campos.

Sin embargo, la integración de LLM en entornos educativos presenta retos significativos, incluyendo, entre otros, la protección de la privacidad, la mitigación de sesgos y la superación de la brecha digital. En este contexto de rápida evolución tecnológica, es imperativo que investigadores, educadores y políticos aborden estos desafíos de forma consciente para asegurar que los beneficios de la IA en la educación sean accesibles de manera justa y responsable.

2 OBJETIVOS

El objetivo de este trabajo es analizar el impacto de la Inteligencia Artificial (IA) en la educación, identificando tanto su potencial transformador como los riesgos asociados. Se propone explorar cómo la IA puede mejorar las experiencias de aprendizaje mediante la educación personalizada y los sistemas de tutoría inteligentes, así como abordar los desafíos significativos, tales como la privacidad de los datos, el sesgo algorítmico y la ampliación de las desigualdades educativas. Además, se busca proponer medidas estratégicas para mitigar estos riesgos, estableciendo directrices éticas y marcos regulatorios que aseguren la equidad y la transparencia en las aplicaciones de IA en la educación.

3 METODOLOGÍA

En este artículo se realiza una revisión narrativa como metodología principal para explorar el uso de la inteligencia artificial (IA) en la educación. El objetivo de esta metodología es ofrecer una visión comprensiva de las principales potencialidades y riesgos asociados con esta tecnología. Para llevar a cabo esta revisión, se han recopilado y analizado estudios previos y bibliografía relevante en este campo, utilizando bases de datos de referencia como Scopus, Web of Science y Dialnet. Se identificaron las tendencias emergentes y las áreas de aplicación predominantes. Los trabajos seleccionados abordaban directamente el impacto de la IA en la educación, excluyéndose aquellos centrados exclusivamente en aplicaciones no educativas de la IA o que carecían de rigor metodológico. Se realizó un análisis temático de los estudios seleccionados para identificar las tendencias emergentes y las áreas de aplicación predominantes de la IA en la educación. Este análisis permitió categorizar la información en temas clave como la personalización del aprendizaje, los sistemas de tutoría inteligentes, el monitoreo del rendimiento, la privacidad de los datos, el sesgo algorítmico, y la equidad en el acceso. La revisión permitió evaluar los principales retos de la integración de la IA en contextos educativos. Basándose en los hallazgos, se proponen posibles líneas de intervención para enfrentar estos retos como el desarrollo de políticas reguladoras éticas y la formación de educadores en el uso de tecnologías emergentes.

4 RIESGOS ASOCIADOS A LA IA

Tal y como hemos señalado en la introducción del artículo, la IA ha avanzado a pasos agigantados en las últimas décadas, transformando industrias, penetrando fuertemente en los procesos educativos, mejorando la eficiencia operativa en numerosos sectores y ofreciendo nuevas soluciones a problemas complejos (Gallent-Torres y Comas-Forgas, 2024). Sin embargo, este progreso no está exento de riesgos y desafíos significativos que podrían tener implicaciones profundas tanto a nivel social como individual. A medida que integramos más estrechamente la IA en nuestro quehacer diario, se hace imperativo comprender y abordar estos riesgos para asegurar un futuro en el que la tecnología y la humanidad puedan coexistir en armonía (Floridi, 2023). En la presente sección se abordarán algunos de estos riesgos.

4.1 Los sesgos algorítmicos

Existe un principio básico en cualquier sistema basado en el uso de algoritmos: si los datos de entrada presentan sesgos, es probable que los resultados hereden, reproduzcan y, en ciertas ocasiones y bajo ciertas circunstancias, amplifiquen esos mismos sesgos (Huang y Rust, 2021). El sesgo en la IA es una anomalía procedente de los algoritmos de aprendizaje automático causada

por suposiciones preconcebidas hechas durante la fase de desarrollo del algoritmo o por los conjuntos de datos de entrenamiento seleccionados (Dilmegani, 2024). A modo de ejemplo, se ha evidenciado que Amazon aplicaba herramientas de IA para evaluar a los candidatos a puestos de trabajo para su empresa, proceso durante el cual se discriminaba a las mujeres candidatas (Van Giffen et al., 2022). Otro caso muy llamativo y con claras consecuencias perniciosas lo encontramos en el ámbito de las aseguradoras: se han identificado errores de la IA empleada al calcular automáticamente las primas en base en la religión de las personas tomadoras del seguro (Akter et al., 2021). Estos errores llevan a que los sistemas automatizados muestren predisposiciones y respuestas sesgadas y discriminatorias en muchas ocasiones (Shams et al., 2023). En cierta manera podríamos decir que los prejuicios pueden infiltrarse en los algoritmos, y ello suele producirse a partir del conjunto de datos empleados en su entrenamiento (Varsha, 2023).

Estas limitaciones y sesgos también se encuentran muy presentes en los algoritmos educativos, como se evidencia en un número creciente de publicaciones sobre el tema (Li et al., 2023). El problema del sesgo en los procesos de enseñanza-aprendizaje ha sido documentado desde la década de 1960 anticipando muchos aspectos de la literatura moderna sobre el sesgo algorítmico y la equidad (Hutchinson y Mitchell, 2019). En años recientes, los algoritmos se han aplicado a gran escala en prácticas educativas para una gama de aplicaciones, a menudo de mucha importancia, incluyendo la predicción de abandono escolar (Del Bonifro et al., 2020; Melo et al., 2022), la calificación automatizada de trabajos y actividades académicas (Huawei y Aryadoust, 2023), procesos de admisión (Holmes y Tuomi, 2022) y procesos de evaluación institucional o de las asignaturas y docentes (Gervet et al., 2020). Los expertos en el tema han advertido y evidenciado sobre potenciales riesgos de desigualdad y falta de generalización en algoritmos educativos durante varios años (Kizilcec y Lee, 2022). Esta preocupación se volvió muy mediática y despertó gran interés para el público general a raíz de la controversia en las calificaciones de los exámenes de acceso a bachillerato y universidad (GCSE y A-Level) del Reino Unido en 2020 (Smith, 2020), donde un conjunto de fórmulas fue desarrollado por el regulador nacional de calificaciones para asignar calificaciones de exámenes basadas en predicciones de los profesores; el algoritmo asignó calificaciones más bajas a los estudiantes en escuelas financiadas por el estado y mejores calificaciones (incluso mejores que las predicciones de los docentes) a los estudiantes en escuelas independientes más pequeñas.

La investigación sobre los sesgos algorítmicos en la educación se ha enfocado principalmente en tres categorías: raza/etnicidad, nacionalidad y género. En cuanto a raza y etnicidad, Kai et al. (2017) mostraron diferencias de previsión de rendimiento entre estudiantes afroamericanos y blancos en un modelo que predecía la retención estudiantil en un programa de universidad en línea dependiendo del modelo algorítmico aplicado. Hu y Rangwala (2020) investigaron varios algoritmos para predecir si un estudiante está en riesgo de suspender y repetir un curso, encontrando que los modelos generalmente rinden peor para estudiantes afroamericanos. Anderson et al. (2019) diferenciaron el rendimiento entre estudiantes de distintos grupos raciales/étnicos en un modelo que predecía la graduación universitaria en seis años, usando cinco algoritmos diferentes. Observaron que los algoritmos generalmente tenían tasas de falsos positivos más altas para estudiantes blancos y tasas de falsos negativos más altas para estudiantes latinos.

En términos de nacionalidad, existen también interesantes evidencias sobre fallos en la aplicación de la IA en la educación; por ejemplo, el sistema E-Rater fue acusado de dar puntuaciones significativamente más altas a estudiantes chinos y coreanos en una prueba de competencia en lengua extranjera, mientras que era más preciso para estudiantes de otras nacionalidades (Uto y Okano, 2021).

En cuanto al género, Kai et al. (2017) diferenciaron el rendimiento entre estudiantes masculinos y femeninos en un modelo que predecía la retención de estudiantes en un programa de universidad en línea, encontrando que ciertos algoritmos conferían un rendimiento moderadamente mejor para estudiantes femeninas que masculinos. En el mismo trabajo citado anteriormente, Anderson et al. (2019) diferenciaron el rendimiento entre estudiantes masculinos y femeninos, y hallaron que los algoritmos generalmente tenían tasas de falsos negativos más altas para estudiantes masculinos.

Todos estos problemas sugieren que los desafíos planteados por el sesgo algorítmico no son triviales a menos que se eduque a los creadores de los algoritmos (factor input) y a los usuarios últimos de la IA (factor output) (López-Meneses et al., 2022). Por un lado, los desarrolladores y científicos de datos deben estar equipados con el conocimiento y las herramientas necesarias para identificar y mitigar los prejuicios inconscientes que pueden infiltrarse en los sistemas de IA durante el proceso de desarrollo. Esto implica no solo una formación técnica avanzada en técnicas de aprendizaje automático y modelado de datos, sino también una educación robusta en ética de la IA y conciencia sociocultural. Al entender mejor las complejidades y las consecuencias de los sesgos algorítmicos, los creadores de algoritmos pueden adoptar prácticas de desarrollo más inclusivas y representativas, garantizando que sus productos sirvan equitativamente a una amplia gama de usuarios (Riedl, 2019).

Por otro lado, educar a los usuarios finales sobre los sesgos inherentes en la IA es crucial para fomentar una comprensión más crítica y una adopción más informada de estas tecnologías. Los usuarios deben ser capaces de reconocer cuando los resultados de un sistema de IA podrían estar sesgados y saber cómo cuestionar o contrarrestar dichos resultados. Esto es especialmente importante en sectores sensibles como la educación, la salud, las finanzas y el ámbito laboral, donde los sesgos algorítmicos pueden tener impactos significativos en la vida de las personas. Una mayor conciencia de los usuarios puede impulsar la demanda de sistemas de IA más transparentes y justos, incentivando a las empresas y desarrolladores a priorizar la equidad en el diseño de sus soluciones (Southworth, 2023).

Para lograr estos objetivos, es fundamental que la educación en sesgos algorítmicos y ética de la IA se integre tanto en la formación académica de los futuros profesionales del campo como en los programas de capacitación continua para los usuarios de IA en diversos sectores (Charlie Project, 2023). Instituciones educativas, organizaciones profesionales y empresas de tecnología deben colaborar para desarrollar currículos y recursos educativos que aborden estos temas de manera comprensiva. Al hacerlo, se puede avanzar hacia un ecosistema de IA más consciente y responsable, donde tanto los creadores como los usuarios jueguen un papel activo en la mitigación de los sesgos y la promoción de una tecnología verdaderamente inclusiva y equitativa.

4.2 Privacidad y derechos de autor

La IA ha suscitado también importantes preocupaciones en torno a la privacidad y los derechos de autor, cuestiones que se han manifestado en diversas denuncias y controversias (Herrera de las Heras, 2022). Un análisis de algunas de estas recientes problemáticas y disputa ofrece una perspectiva valiosa sobre los desafíos éticos y legales emergentes en la era de la IA.

Uno de los casos más sonados es la denuncia del *New York Times* hacia ChatGPT de OpenAI, en el cual se plantea una problemática significativa sobre los derechos de autor (Grynbaum y Mac, 2023). El uso extensivo de información y contenido protegido para entrenar modelos de IA sin el consentimiento explícito de los autores originales pone en tela de juicio la legalidad de estas prácticas y sus implicaciones para los creadores de contenido. Este caso resalta

la tensión entre el desarrollo tecnológico y la protección de la propiedad intelectual, un equilibrio delicado que entendemos está lejos de alcanzarse.

En una línea similar, la denuncia de George R.R. Martin a ChatGPT pone de relieve las preocupaciones individuales de los creadores sobre el uso no autorizado de sus obras (Desbrières, 2023). La capacidad de los modelos de IA para generar texto que imita el estilo de escritores específicos sin su permiso plantea interrogantes sobre la originalidad, la autenticidad y, en última instancia, sobre quién posee los derechos sobre los contenidos generados por IA que se basan en obras protegidas por derechos de autor.

Otro ejemplo de esta problemática lo representa el caso de Reddit, sitio web de marcadores sociales y agregador de noticias donde los usuarios pueden añadir textos, imágenes, videos o enlaces, que acusó a grandes empresas de tecnología de utilizar sus datos de usuario para entrenar modelos de IA, lo que podría constituir una violación de la privacidad y un uso indebido de contenido protegido (Choi et al., 2023). Este incidente subraya la importancia de establecer límites claros sobre cómo se pueden utilizar los datos recopilados en plataformas en línea y quién tiene el derecho de beneficiarse de ellos.

Finalmente, destaca la demanda colectiva presentada por el bufete de abogados Clarkson Law Firm contra OpenAI que pone evidencia las preocupaciones legales y éticas en torno al uso de datos personales sin el consentimiento de los usuarios. Esta demanda argumenta que OpenAI utilizó datos de usuarios de manera inapropiada para entrenar sus modelos de IA, lo que podría representar una infracción de las leyes de privacidad y derechos de autor (Frenkel y Thompson, 2023). Este caso ilustra el creciente escrutinio público y legal hacia las compañías de IA y sus prácticas de recopilación de datos.

Estos incidentes resaltan un desafío fundamental en la era de la IA: la necesidad de un marco legal y ético robusto que equilibre el avance tecnológico con la protección de los derechos individuales y de propiedad. La legislación actual en muchos países aún no se ha adaptado completamente a las complejidades introducidas por la IA.

Si centramos la atención en el terreno educativo se pueden identificar varios casos en los que la IA puede poner en riesgo aspectos relacionados con la privacidad, la protección de datos y los derechos de autoría, siendo:

- Recopilación de datos sensibles: herramientas de IA utilizadas para personalizar la experiencia de aprendizaje pueden recopilar datos personales sensibles de los estudiantes sin su consentimiento explícito. Esto incluye información sobre rendimiento académico, hábitos de estudio e incluso datos biométricos en casos de plataformas que monitorean la atención mediante cámaras. La recopilación y almacenamiento de estos datos sin medidas de protección adecuadas podrían exponer a los estudiantes a riesgos de privacidad (Sánchez et al., 2023).
- Vigilancia en exámenes online: con el auge del aprendizaje a distancia durante la pandemia, muchas instituciones adoptaron sistemas de *proctoring* basados en IA para supervisar a los estudiantes durante los exámenes. Estos sistemas, que a menudo requieren acceso a la cámara y micrófono del dispositivo del estudiante, han suscitado preocupaciones sobre la intrusión en la privacidad personal, ya que podrían monitorear a los estudiantes y familias en sus propios hogares (Dawson, 2024).
- Contenidos generados por IA: la creación de material didáctico y trabajos académicos mediante el uso de herramientas de IA que generan texto, como ChatGPT, plantea

interrogantes sobre los derechos de autor (Galli y Kanobel, 2023). Sirva un ejemplo: si un docente utiliza una herramienta de IA para crear contenido de curso basándose en una amplia gama de fuentes, determinar la originalidad y titularidad del material resultante puede ser complicado, especialmente si la herramienta no cita sus fuentes o incorpora texto de obras protegidas sin permiso.

- Uso de material protegido para entrenamiento de IA: algunas plataformas educativas emplean modelos de IA que han sido entrenados con grandes conjuntos de datos, incluyendo material educativo con derechos de autor. Esto puede llevar a situaciones en las que el material protegido se utiliza sin autorización, infringiendo potencialmente los derechos de los autores originales (Crawford, 2023).
- Plagio y originalidad: El uso de herramientas de IA por parte de los estudiantes para la realización de tareas y trabajos puede complicar la detección de plagio y la evaluación de la originalidad. Esto no solo afecta la integridad académica, sino que también plantea preguntas sobre los derechos de autor del contenido generado por IA, especialmente cuando los estudiantes presentan como propio el trabajo generado por estas herramientas (Gallent-Torres y Comas-Forgas, 2024).

La solución a estos problemas no es simple y requerirá un esfuerzo concertado de legisladores, tecnólogos, juristas, creadores de IA y la comunidad educativa. La creación de políticas que regulen el entrenamiento de modelos de IA, el establecimiento de directrices claras sobre el uso de datos y la implementación de mecanismos para el consentimiento y compensación justa son pasos cruciales hacia la mitigación de estos desafíos. Además, la transparencia en las prácticas de recopilación y uso de datos por parte de las compañías de IA es fundamental para ganar la confianza del público y asegurar que el desarrollo de la IA se alinee con los valores éticos y legales de la sociedad (Crawford, 2023).

4.3 IA-dependencia e IA-fobia: dos caras de la misma moneda

Recientemente, dos fenómenos emergentes relacionados con la IA han captado la atención de psicólogos, educadores y tecnólogos: el "síndrome de la IA-dependencia" y la "IA-fobia". Estas tendencias opuestas destacan como peligros inherentes al creciente papel de la IA en nuestra vida cotidiana y en los entornos educativos.

El "síndrome de la IA-dependencia" se refiere a la creciente tendencia de confiar en herramientas de IA para realizar tareas cotidianas, decisiones y, en el ámbito educativo, para facilitar el aprendizaje y la enseñanza (Crawford et al., 2024). Aunque el uso de la IA puede aumentar la eficiencia y proporcionar soluciones personalizadas, existe una preocupación creciente sobre cómo esta dependencia afecta nuestras habilidades cognitivas y emocionales.

Una de las consecuencias más significativas de esta dependencia es la pérdida de habilidades de pensamiento crítico y resolución de problemas. A medida que las herramientas de IA asumen tareas complejas, los usuarios pueden encontrarse menos motivados para desarrollar o mantener habilidades analíticas, ya que prefieren confiar en soluciones generadas por IA. Esta tendencia no solo reduce nuestra autonomía, sino que también nos deja vulnerables en situaciones donde la tecnología no está disponible o falla (Mahmut y Sen, 2023). Además, la dependencia de la IA puede tener efectos psicológicos profundos. La constante interacción con sistemas de IA, que ofrecen respuestas instantáneas y soluciones a medida, puede generar ansiedad y depresión cuando enfrentamos desafíos reales sin su ayuda (Liu et al., 2023). En algunos casos, el uso desmedido de la IA para la interacción social puede conducir al aislamiento, ya que los usuarios pueden preferir

la compañía de asistentes virtuales a las relaciones humanas complejas y, a veces, desafiantes (Mucci y Marazziti, 2023).

Otro de los riesgos a los que se enfrenta la educación superior, y que afecta tanto a docentes como alumnado, es la IA-fobia, que se define como “el sentimiento al miedo o ansiedad hacia la adopción de la inteligencia artificial” (Wang et al., 2020, p. 356). Este trastorno, relacionado con la “tecnofobia”, por provocar en los usuarios los mismos sentimientos irracionales, se manifiesta al evitar el uso de tecnología avanzada y resistirse, además, a su implementación. Se produce cuando se experimenta una amenaza o se percibe que algo puede producir un daño físico o psicológico (Ekman 1992; Adolphs, 2013). En este contexto, se vincula con una falta de comprensión sobre el funcionamiento y comportamiento de la IA en un entorno académico, lo que conlleva, en paralelo, la resistencia a la interacción con esta tecnología y la negación de su potencial.

Actualmente existe una preocupación real en torno al uso de la IA en el ámbito educativo, fundamentada en discursos apocalípticos sobre la suplantación humana (Bostrom, 2014), el debilitamiento del pensamiento crítico o la reducción en la calidad de la educación (Ouyang et al., 2022). Si bien a muchos asombran sus capacidades (realiza tareas rutinarias en segundos, analiza grandes cantidades de datos, crea contenido educativo personalizado, retroalimenta al alumnado, etc.), también suscita en algunos incomodidad, desconfianza y dudas sobre su integración, lo que produce que se alimenten ciertas fobias hacia esta tecnología.

En base a lo anterior, autores como Lee y Park (2023) clasifican las emociones vinculadas a la IA en cuatro tipos de fobias, a los que denominan: IA-alfabetización, IA-sustituibilidad, IA-responsabilidad y IA-implementabilidad. El primero de ellos hace referencia al temor de uso por desconocimiento, es decir, a la ansiedad que proviene de “la ignorancia y asimetría de la información” (p. 2) sobre el funcionamiento y los efectos de la IA en el ámbito profesional; en otras palabras, al miedo a explorar esa “caja negra” que genera una brecha digital entre aquellos que la conocen bien y los que no. El segundo tipo de fobia, denominada IA-sustituibilidad, alude al miedo al reemplazo, a verse sustituido por la tecnología, poniendo en tela de juicio la capacidad del ser humano. El tercer tipo alude a la ansiedad por la falta de claridad en la rendición de cuentas, es decir, a la responsabilidad que se deriva de las decisiones incorrectas o las informaciones inexactas proporcionadas por esta tecnología. Incluso a la falsa atribución de méritos o autoría, al considerarla de facto una entidad intelectual. Por último, el cuarto temor se define como IA-implementabilidad o “fiebre de la IA”, y se refiere a la obligación de adoptarla que algunos sienten solo porque tienen que hacerlo. Esa sensación de verse arrastrado por las tendencias tecnológicas genera una percepción de pérdida de control y autonomía en la toma de decisiones (Slater, 2024).

Por su parte, Li y Huang (2020) analizan también la ansiedad asociada a la IA con el objetivo de comprender mejor este fenómeno, sus dimensiones e implicaciones sociales. Utilizan como marco teórico la teoría de adquisición del miedo de Rachman (1977) y la del miedo no correlacionado de Menzies (1995), vinculando, de manera novedosa, la ansiedad a la tecnología. A través de dichas teorías analizan su origen y las causas directas, indirectas e innatas que la producen. Inciden en que la ansiedad no solo surge de experiencias directas condicionadas, sino también de la transmisión de información o instrucción por parte de terceros (indirectas) y de experiencias innatas frente a lo desconocido (Poulton et al., 2001). Tras la revisión de la literatura, identifican ocho factores que contribuyen a la ansiedad ante la IA: i) violación de privacidad (fuga de datos personales, uso generalizado de biometría, etc.); ii) comportamiento sesgado o tendencioso; iii) sustitución del puesto de trabajo; iv) falta de confianza para aprender a utilizar la IA; v) riesgo existencial (creencia en la extinción de la raza humana); vi) violación de las reglas de la ética en la interacción humana; vii) conciencia artificial (IA entendida como una nueva especie humana), y viii) falta de transparencia en sus procesos de toma de decisiones (opacidad del funcionamiento

interno de los algoritmos y datos de entrenamiento, responsabilidad, etc.). Algunos de estos factores están relacionados con el potencial de la IA y las actitudes individuales hacia esta tecnología, pero otros escapan al control de su diseño.

A estos factores, en los que coinciden también autores como Dahlin (2019) y Hopcan et al. (2023), cabe añadir también la preocupación por las desigualdades sociales (Zajko, 2022) y la vulneración al derecho a la intimidad (Elliott y Soifer, 2022). Desde una perspectiva académica, la falta de comprensión sobre la integración de la IA en la práctica docente y la incertidumbre sobre si los sistemas de IA pueden captar los matices de la enseñanza –como la importancia de construir relaciones con los alumnos y adaptar la enseñanza a sus necesidades (Ouyang et al., 2022)–, podrían fomentar el desarrollo de fobias en el alumnado y el profesorado. Que esto suceda dependerá de la experiencia personal de cada uno, la formación académica, el nivel de comprensión de la tecnología, la posibilidad de acceso (Hopcan et al., 2023), y la actitud hacia la misma.

Mitigar la IA-fobia es un desafío complejo que requiere acciones diversas. Entre las propuestas por algunos autores destacan las medidas relacionadas con la formación y capacitación de la comunidad académica (Li y Huang, 2020). Por ejemplo, ofertar cursos diseñados específicamente para disipar los miedos asociados con esta tecnología y promover una comprensión informada de sus riesgos y beneficios. En esta línea, se sugiere facilitar el acceso a recursos educativos como tutoriales y foros virtuales que mejoren el conocimiento técnico de los sistemas de IA y cultiven actitudes positivas hacia su adopción. También es importante fomentar la curiosidad por la innovación, explorar el potencial de estas herramientas en un entorno práctico y colaborativo, y concienciar sobre las implicaciones sociales y éticas de su uso, evitando caer así en lo que algunos autores denominan la “ceguera sociotécnica” (sociotechnical blindness) (Hopcan et al., 2023).

4.4 Alucinaciones, desinformación y postverdad

Otro de los riesgos inherentes a esta tecnología son las alucinaciones, es decir, interpretaciones erróneas o distorsionadas generadas por sistemas de IA que no tienen una correspondencia directa con la realidad (Athaluri et al., 2023). Las alucinaciones presentan contenido aparentemente auténtico, convincente y bien formulado, pero, en verdad, son construcciones completamente producidas por modelos de lenguaje, lo que dificulta su detección. Además, pueden manifestarse de diversas formas, desde textos, referencias incorrectas y perfiles falsos, hasta videos o imágenes manipuladas, voces clonadas y conversaciones simuladas, lo cual plantea problemas de carácter ético y social.

Las respuestas que generan los sistemas de IA son tan realistas y elocuentes que pueden conllevar a una percepción errónea de la realidad, influir en la opinión pública y distorsionar la toma de decisiones. A veces, las alucinaciones se producen por falta de supervisión y actualización de los modelos, utilización de patrones de entrenamiento incorrectos, errores de software o codificación (Beutel et al., 2023). También se producen cuando la información proporcionada por los usuarios es insuficiente, por lo que los sistemas de IA tienden a hacer suposiciones sobre lo que el usuario quiere oír en lugar de hacer preguntas que le permitan comprender mejor el contexto de su consulta. De ahí que autores como Shen et al. (2023) afirmen que estos sistemas se limitan a seguir instrucciones en lugar de participar en una verdadera interacción con el usuario.

Identificar si la información generada por estos sistemas es veraz o no, es todo un desafío hoy en día. Reproducir contenidos fabricados o consumir información no contrastada conlleva consecuencias en todos los ámbitos. Entre ellas, la propagación de la desinformación, la diseminación de sesgos y prejuicios, la adquisición de conocimientos erróneos, la falta de confianza

en el conocimiento adquirido (Amaro et al., 2023), el aumento de la confusión y desorientación, y la pérdida de credibilidad en fuentes de referencia, entre otras.

La confianza ciega de los usuarios en la IA podría llevar a un uso inapropiado de esta tecnología (Glikson y Woolley, 2020), así como a la propagación de noticias falsas y ultrafalsos (deepfakes), y la manipulación de información, acciones que nos llevan a pensar directamente en el concepto de posverdad. La posverdad alude a la prevalencia de las creencias personales sobre los hechos objetivos; considera lo incierto como verdadero, y privilegia lo subjetivo, lo instintivo e irracional, aun cuando contradiga abiertamente lo confirmado por la ciencia. Se apoya en sesgos cognitivos y se aprovecha de las redes sociales para exponer al individuo a noticias y opiniones manipuladas, presentándole aquello que quiere ver y escuchar. Esto influye en su capacidad de razonamiento, en la manera de interpretar la realidad, en la noción de objetividad, y en la habilidad para discernir lo veraz de lo fabricado (Athaluri et al., 2023).

Trasladando esta reflexión al ámbito académico, preocupa la falta de formación y concienciación entre el alumnado universitario respecto a este fenómeno, sobre todo, teniendo en cuenta su impacto en el entorno académico y personal. Es importante, pues, que los estudiantes comprendan qué es la posverdad y cómo se manifiesta en distintos contextos; que aprendan a cuestionar la información y a buscar evidencias sólidas que la respalden, y participen de manera crítica en la sociedad. A menudo, la información que reciben tiende a ser aceptada como verdadera, lo que lleva a pensar en el sesgo de confirmación (Wason, 1960). Es decir, se trata de la tendencia a consumir información que confirme las propias creencias, rechazando cualquier contenido que las contradiga. Por lo tanto, al acceder a una información novedosa, directamente se acepta como veraz, pero si se analiza con detenimiento es probable que se descarte. Cybenko y Cybenko (2018) señalan los cuatro criterios por los cuales esto podría suceder: i) cuando la información no se alinea con la visión del mundo actual; ii) carece de coherencia; iii) proviene de una fuente poco confiable, o iv) se percibe que otras personas no la respaldan. Dependerá, pues, de la capacidad crítica de cada uno para discernir entre verdad y manipulación. En un contexto en el que las herramientas para descubrir la verdad y las estrategias para subvertirla están en constante evolución, es importante reconocer cualquier dinámica que socave la integridad de la información y comprometa la comprensión del mundo que nos rodea.

4.5 Fraude y deshonestidad académica

Otro de los riesgos que suscita preocupación es el fraude académico que se produce cuando se lleva a cabo una conducta deshonesta en este ámbito. Autores como Martínez (2022) señalan que la deshonestidad académica se debe principalmente a dos variables: contextuales (escenarios institucionales, y culturales), y psicológicas (actitudes, creencias, valores, motivaciones, etc.). A dichas variables debería añadirse la tecnológica, que alude a la influencia de las herramientas digitales, la accesibilidad a Internet y el desarrollo de tecnologías emergentes como la IA, las cuales pueden facilitar el plagio y otras formas de deshonestidad académica en el contexto educativo actual. En este sentido, Gendron et al. (2022) consideran que la facilidad con la que la IA genera contenido de forma automática y personalizada contribuye a que el alumnado utilice estas herramientas de manera indebida para elaborar sus trabajos académicos. Esta circunstancia les podría hacer caer en la tentación de plagiar conscientemente, utilizar cualquier material sin indicar la fuente original y evitar el esfuerzo para crear nuevo conocimiento (Vera, 2023). El estudio de Tapia et al. (2023) sigue esta misma línea, y arroja luz sobre cómo la tecnología puede burlar los principios fundamentales de la educación superior (pensamiento crítico, originalidad, responsabilidad, etc.), y alentar a algunos a tomar el camino más corto en lugar de comprometerse con el propio proceso de aprendizaje.

Además del plagio académico, la IA podría fomentar en el alumnado el desarrollo de otras conductas deshonestas como la manipulación de resultados en pruebas o exámenes, la creación de perfiles falsos en entornos virtuales de aprendizaje o redes sociales (robo de identidad académica), la falsificación de títulos académicos, etc. Estas prácticas, aunque no son nuevas, adquieren mayor sofisticación con las funcionalidades de las herramientas digitales actuales. En cualquier caso, estas prácticas socavan la integridad académica y generan desigualdad entre aquellos que respetan los principios éticos de la educación y los que recurren a métodos fraudulentos para obtener ventajas injustas; y entre aquellos que adquieren conocimiento de forma honesta y los que lo hacen de manera artificial. Ciertamente, la prevalencia de estas conductas podría conllevar consecuencias negativas a largo plazo como la disminución de determinadas competencias fundamentales en el alumnado (capacidad de análisis de documentación, síntesis, y búsqueda bibliográfica, etc.), la confiabilidad de las calificaciones y certificaciones, e incluso la credibilidad de las instituciones educativas, lo que comprometería la calidad de la educación.

Autores como Susnjak (2022) también apoyan la idea que estas herramientas tecnológicas facilitan nuevas formas de fraude académico. Algunas de ellas podrían incluir la manipulación de algoritmos de evaluación automatizada, la invención de citas y referencias, el uso de avatares para engañar en entornos virtuales, la fabricación de datos experimentales, el plagio mediante herramientas de traducción automática, la creación de contenido académico falso para obtener becas o premios, la edición de material audiovisual falso, etc. Estos ejemplos ilustran cómo la tecnología amplía el espectro de posibilidades para quienes buscan aprovecharse del sistema.

Paradójicamente, y a pesar de las inquietudes que genera el uso de la IA en el ámbito académico, es importante reconocer su potencial para combatir el fraude académico. En este sentido, los algoritmos de detección del plagio pueden mejorar su precisión y eficacia gracias a la implementación de técnicas avanzadas de aprendizaje automático y procesamiento de lenguaje natural (González-González, 2023). Estas técnicas permiten a los algoritmos analizar grandes cantidades de datos y patrones de texto, lo que les permite identificar similitudes y discrepancias entre los trabajos académicos presentados y las fuentes originales (Khalil y Erkan, 2023). Además, pueden detectar características específicas de la escritura generada por IA y distinguirlas de la escritura humana.

4.6 Impacto medioambiental

Finalmente, otro de los riesgos asociados a esta tecnología es su impacto medioambiental. La creciente demanda de recursos para entrenar modelos de IA, junto con la infraestructura necesaria para sostener esta tecnología, contribuyen a un aumento considerable en la huella de carbono. Estos modelos emiten cientos de miles de kilogramos de dióxido de carbono, algo que resulta difícil de imaginar. Como señala Dhar (2020), urge detenerse a cuantificar su impacto para comprender mejor su magnitud porque “entrenar un solo modelo [de IA] equivale a unos 300.000 kg de emisiones de dióxido de carbono. Esto es del orden de 125 vuelos de ida y vuelta entre Nueva York y Pekín, una cuantificación que los profanos pueden visualizar” (p. 423). Esta comparación visualmente impactante puede ayudar a sensibilizar a un público que, en su imaginario o conciencia social, tenga dificultades para comprender la realidad del problema.

Reducir el impacto climático de la IA pasa, pues, por ser transparentes con la divulgación de esta información. En esta línea, Crawford y Joler (s.f., citados en Dhar, 2020) destacan que la falta de comunicación clara sobre estos costos puede llevar a una percepción errónea de la facilidad de construir un sistema de aprendizaje automático y, por ende, de ser consciente de sus consecuencias. Asimismo, subrayan la necesidad de establecer un estándar de medición que permita

a los usuarios e investigadores cuantificar estos costos. Uno de los recursos disponibles actualmente para evaluar dicho impacto es el desarrollado por Lacoste et al. (2019), conocido como Machine Learning CO2 Impact (<https://mlco2.github.io/impact/#about>). Este instrumento permite examinar las emisiones producidas durante el entrenamiento de un modelo de red neuronal, las cuales dependen de factores como la ubicación del servidor de entrenamiento, la red energética utilizada, la duración del proceso de entrenamiento y el hardware empleado.

Además de desarrollar este tipo de herramientas, se exploran alternativas que promuevan un enfoque más sostenible de la IA. Una de ellas es el uso de redes de energía renovable para el entrenamiento de modelos de IA. En paralelo, es importante implementar políticas y estándares que impulsen el uso eficiente de la energía y la reducción de la huella de carbono en todas las etapas del ciclo de vida de esta tecnología. Esto requiere una colaboración activa entre investigadores, industria y gobiernos para abordar estos desafíos de manera integral y encontrar soluciones innovadoras. En última instancia, lo que se pretende es construir un futuro en el que la tecnología pueda coexistir en armonía con el medio ambiente, garantizando un desarrollo sostenible y responsable en todos los aspectos.

5. ESTRATEGIAS PARA LIMITAR LOS RIESGOS DE LA IA EN EL ENTORNO EDUCATIVO

5.1 Medidas legales y políticas

Durante 2023, la evolución de la política y regulación de la IA ha experimentado una acelerada transición desde un campo de interés residual hasta erigirse en un tema central de debate público y académico (Guenduez y Mettler, 2023). Según Ryan-Mosley (2024), este periodo ha sido testigo de una serie de desarrollos críticos en el ámbito de la política de IA, destacándose la implementación de legislaciones pioneras en la Unión Europea, iniciativas legislativas y órdenes ejecutivas en Estados Unidos, así como la introducción de regulaciones específicas en China dirigidas a los algoritmos de recomendación.

La dinámica global en torno a la regulación de la IA ha mostrado una amplia gama de enfoques y marcos regulatorios que varían significativamente entre distintas jurisdicciones. La Unión Europea, por ejemplo, ha establecido un hito al promulgar la primera legislación específica a nivel mundial que aborda de manera directa los usos permitidos y restringidos de la IA, sentando un precedente importante en términos de gobernanza tecnológica. Esta legislación, que contempla aspectos clave como la prohibición de ciertos usos de la IA y establece límites y sanciones claras por incumplimientos, refleja el esfuerzo de la UE por liderar el desarrollo y uso ético de la IA dentro de un marco legal bien definido (Ersoy, 2024).

En contraste, Estados Unidos ha adoptado una postura que integra la IA dentro del discurso político, llevando a la ejecución de acciones concretas como la emisión de una "orden ejecutiva sobre IA" por parte de la administración Biden. Esta medida, que exige mayor transparencia en el funcionamiento de los modelos de IA y establece nuevas regulaciones, indica un enfoque que busca equilibrar el fomento de la innovación tecnológica con la necesidad de una regulación que garantice prácticas responsables y transparentes en la industria de la IA (Parinandi et al., 2024).

China, por su parte, ha reforzado su marco regulatorio con la implementación de normativas específicas, como aquellas dirigidas a regular los algoritmos de recomendación. Este enfoque subraya la intención del país de supervisar estrechamente la integración de la IA en

diversos aspectos de la sociedad, buscando equilibrar los beneficios de la tecnología con la protección de los derechos y la seguridad de los ciudadanos (Huang et al., 2024).

A nivel global, la regulación de la IA se presenta como un paisaje diverso, donde distintos países y regiones adoptan enfoques variados que reflejan sus prioridades y preocupaciones específicas. Mientras que algunos estados optan por modelos regulatorios similares al de la Unión Europea, otros exploran vías propias, adaptando sus marcos legales y políticos a las necesidades y desafíos particulares de su contexto (Walter, 2024).

En este panorama en evolución destaca la importancia de un diálogo internacional y colaboración entre jurisdicciones para forjar un entendimiento común y aproximaciones coordinadas hacia la regulación de la IA. A medida que estas políticas comienzan a materializarse en acciones concretas, se vislumbra un futuro donde la regulación de la IA se configura como un elemento clave para guiar el desarrollo tecnológico hacia prácticas seguras, éticas y beneficiosas para la sociedad en su conjunto. La convergencia hacia principios y estándares compartidos a nivel internacional podría facilitar este proceso, promoviendo un necesario equilibrio entre innovación tecnológica y responsabilidad social (Walter, 2024).

5.2 Normativas y regulaciones a nivel centro educativo

En este marco cambiante, y de forma progresiva, los centros educativos son cada vez más conscientes tanto de los beneficios como los riesgos de la implantación de la IA. Riesgos y desafíos que deben ser abordados mediante normativas y políticas a nivel de centros educativos y aulas. La UNESCO elaboró ya en 2021 una, *Inteligencia artificial y educación: guía para las personas a cargo de formular políticas* (Miao et al., 2021), en donde, desde un enfoque generalista, propone garantizar que la tecnología se diseñe para servir a las personas de acuerdo con los marcos de derechos humanos acordados a escala internacional, y que se saque provecho de las tecnologías digitales como un bien común para apoyar la consecución del ODS 4 - Educación 2030 y para construir futuros compartidos de la educación más allá de 2030. Algunas de las consideraciones más importantes a las que se enfrentarán en este ámbito los centros educativos en los próximos años son (Miao et al., 2021):

1. **Ética y uso responsable:** las normativas deberán establecer principios éticos para el uso de la IA en la educación, incluyendo la transparencia en los algoritmos, la equidad en el acceso y la protección de los derechos de los estudiantes.
2. **Privacidad de los estudiantes:** los protocolos deberán garantizar la privacidad de los datos de los estudiantes, limitando el acceso a información personal y asegurando que los datos se utilicen de manera segura y confidencial de conformidad con el Reglamento general de protección de datos (GDPR).
3. **Formación docente:** los profesores deberán recibir capacitación sobre cómo integrar la IA en el aula de manera efectiva y ética, incluyendo una comprensión de las implicaciones y las limitaciones de la tecnología con actividades prácticas y cercanas a su tarea docente.
4. **Evaluación y seguimiento:** las normativas de los centros deberán establecer claramente cómo se evaluará el impacto de la IA en el aprendizaje de los estudiantes y cómo se monitoreará su uso.

5. Inclusión y equidad: se deberá garantizar que la IA no amplíe las brechas existentes en el acceso a la educación: todos los estudiantes deben beneficiarse por igual.

El diseño, creación y aplicación de normativas claras en los centros educativos para regular el uso de la Inteligencia Artificial (IA) en la educación comparte muchas similitudes con los enfoques globales adoptados para prevenir el plagio y la deshonestidad académica en general. A lo largo de los años, las instituciones educativas de todo el mundo han implementado diversas estrategias para combatir el plagio, obteniendo resultados significativos cuando estas estrategias se basan en el consenso de toda la comunidad educativa, establecen reglamentos claros y comprensibles, y se aplican efectivamente más allá del papel (Sureda et al., 2016). Estas experiencias previas ofrecen valiosas lecciones que pueden ser aplicadas al desafío de integrar de manera ética y segura la IA en los entornos educativos.

Consenso de la comunidad educativa

Al igual que las políticas antiplagio más efectivas, las normativas sobre el uso de la IA deben ser el resultado de un proceso colaborativo y participativo que involucre a educadores, estudiantes, padres, administradores y expertos en tecnología. El consenso no solo garantiza que las políticas sean equitativas y relevantes para las necesidades y preocupaciones de todos los interesados, sino que también facilita su aceptación y adherencia. Por ejemplo, las políticas de integridad académica que son diseñadas con aportes de estudiantes y profesores tienden a ser más efectivas, ya que estos grupos entienden claramente las expectativas y las razones detrás de las reglas (Sureda et al., 2016). De manera similar, las políticas de IA fruto de un consenso amplio pueden abordar de manera efectiva los riesgos potenciales, al mismo tiempo que maximizan los beneficios de la tecnología para el proceso educativo.

Diseño y establecimiento de reglamentos y políticas claras y entendibles

La claridad y comprensibilidad son cruciales tanto para las políticas de prevención del plagio e integridad académica en general (Sureda et al., 2016) como para las regulaciones sobre la IA. Las normativas deben ser accesibles y fácilmente comprensibles para todos los miembros de la comunidad educativa, independientemente de su experiencia tecnológica. Esto implica evitar el jergón técnico y proporcionar ejemplos concretos de lo que se permite y lo que no. Al igual que las políticas antiplagio efectivas que definen claramente lo que constituye el plagio y las consecuencias de infringir las reglas, las normativas sobre la IA deben detallar explícitamente cómo se debe usar la tecnología de manera ética y segura, así como las repercusiones de su mal uso.

Aplicación efectiva

Para que cualquier política sea efectiva, debe ir acompañada de una implementación y cumplimiento rigurosos (Sureda et al., 2016). En el caso del plagio, las políticas solo han reducido la incidencia de esta mala práctica cuando se aplican consistentemente y se acompañan de herramientas efectivas de detección y sanción. Del mismo modo, las políticas sobre el uso de la IA en la educación necesitan mecanismos claros de monitoreo y evaluación para garantizar su cumplimiento. Esto podría incluir la formación regular de estudiantes y personal sobre los principios éticos de la IA, la realización de auditorías periódicas de las herramientas de IA utilizadas en la educación para detectar sesgos y vulnerabilidades, y la implementación de sanciones para quienes violen las normativas.

5.3 Sistemas y dispositivos de detección

Ahondando en el tema de la ética y el uso responsable, y teniendo en cuenta el uso cada vez más generalizado de herramientas de redacción de textos usando IA, se está volviendo un imperativo que los centros escolares tomen medidas al respecto. La incapacidad de controlar lo que hacen los estudiantes en sus dispositivos personales evidencia que simplemente prohibir este tipo de herramientas en el entorno escolar/educativo/académico puede no ser una medida acertada. En el futuro, lo más probable es que la detección de IA forme parte del software de evaluación y seguimiento, tal y como en su momento fueron las herramientas anti-plagio, que permitan a los profesores discernir qué tareas están escritas con IA y cuáles no.

Si bien existen numerosos detectores de contenido generado por IA, su eficacia ha sido objeto de críticas debido a la falta de consistencia y fiabilidad. Mujezinovic (2023) puso en duda la fiabilidad y prestancia de los detectores de contenido de IA como Writer, GPTZero, entre otros. Las pruebas y experimentos realizados con algunos de estos detectores de contenido de IA han arrojado resultados desalentadores, siendo algunos motivos de burla. Por ejemplo, una prueba realizada sobre la eficacia de GPTZero por Barsee (2023) concluyó que la Constitución de EE.UU. fue escrita por una IA si nos atenemos a los resultados del detector empleado. Otro experimento de Islam (2023) reveló discrepancias en la detección de contenido de IA usando Writer y el clasificador de ChatGPT. Por lo tanto, existen severos cuestionamientos acerca del uso de detectores de IA en su estado actual para determinar la originalidad de los trabajos y actividades entregadas por el alumnado. Además, tal y como sucedía con los sistemas de detección de plagio estamos nuevamente ante el juego del gato y el ratón: a medida que los sistemas de detección de texto elaborado por IA mejoran, salen nuevas formas de saltarse o evitar estos programas (los lectores interesados simplemente deben acceder al servidor de videos de YouTube y buscar por maneras para evadir el control de los detectores de IA). Por lo tanto, es imperativo que las universidades inviertan en investigación sobre la eficacia de los detectores de IA antes de penalizar injustamente a los estudiantes utilizando herramientas defectuosas. Las universidades, y centros educativos en general, cometerían un grave error si se basan en detectores de contenido de IA defectuosos para controlar el uso de la IA en las evaluaciones.

5.4 Alfabetización en IA

La formación en IA se ha enfocado tradicionalmente en el nivel universitario; sin embargo, recientemente ha surgido un notable interés por incorporarla desde la educación básica hasta el nivel de bachillerato (Steinbauer et al., 2021). Este ámbito, aún considerado innovador en estos niveles educativos, se reconoce como fundamental desde la infancia hasta la juventud (Ayanwale et al., 2022). Yang (2022) resalta la importancia de iniciar la educación en IA desde la primera infancia por tres motivos esenciales: primero, la comprensión y utilización de aplicaciones de IA constituyen un componente esencial de la alfabetización digital moderna; en segundo lugar, es imprescindible que los niños adquieran la capacidad de entender, utilizar y valorar la IA; y, por último, aunque existan diversas inquietudes, los menores tienen la capacidad de asimilar las funciones básicas de la IA, y es factible educarlos a través de actividades adaptadas a su edad.

El debate actual se centra en la alfabetización en IA, considerada una especialización dentro de la alfabetización digital, que emerge como un conjunto de habilidades imprescindibles ante el rol cada vez más preponderante de la IA en nuestro día a día (Yang, 2022). La alfabetización en IA abarca las competencias para emplear tecnologías de IA de forma efectiva y ética, definida como "la comprensión de las funciones básicas de la IA y el uso ético de aplicaciones de IA en la cotidianidad" (Yang, 2022, p. 2). Esta se ha establecido como una competencia clave para todos

(Su et al., 2023), con el objetivo de brindar a las personas nuevas capacidades para su integración en la sociedad digital (Kong et al., 2021). La alfabetización en IA incluye cuatro dimensiones: el conocimiento y comprensión de la IA, su uso, evaluación y creación y la ética en IA (Ng et al., 2021). No obstante, no se requiere dominar las teorías y desarrollos esenciales de la IA; basta con el uso competente y razonable de las herramientas de IA para considerarse alfabetizado en esta área (Wang et al., 2022).

Kong et al. (2022) destacan dos ventajas principales de impartir esta alfabetización desde la educación básica: fomentar un entendimiento de la IA en los estudiantes desde una edad temprana y estimular su interés en esta tecnología para inspirar su aplicación en futuros estudios y carreras profesionales, permitiéndoles no solo ser usuarios, sino también creadores de IA. La relevancia de la alfabetización en IA para la adaptación social ha sido ampliamente debatida, representando un desafío crítico en la educación contemporánea al preparar a las nuevas generaciones para un mundo en el que la IA es omnipresente (Carvalho et al., 2022). Las oportunidades de aprendizaje en IA empoderan a los estudiantes al proporcionarles los conocimientos y habilidades necesarios para prosperar en un mundo cada vez más definido por la IA (Southworth et al., 2023). Comprender los principios y conceptos de la IA, junto con la habilidad de usar sus herramientas y analizar sus beneficios y aspectos sociales y éticos a largo plazo, son destrezas cruciales para el siglo XXI (Steinbauer et al., 2021). Los estudiantes deben ser capaces de emplear tecnologías de IA distinguiendo entre prácticas éticas y no éticas (Ng et al., 2021), mientras que los ciudadanos deben utilizar la IA protegiendo sus intereses y privacidad (Kong et al., 2021). La expansión de las tecnologías de IA a diversos campos ha hecho necesario desarrollar una fuerza laboral con el conocimiento y la capacidad para trabajar con IA (Lee et al., 2021). Responsables empresariales y gubernamentales enfatizan la importancia de fomentar la alfabetización en IA para el éxito personal y el bienestar común (Kong et al., 2021). Para integrarse de manera productiva en la era de la IA, todos los jóvenes necesitan entender cómo funciona la IA y cómo impactará en sus vidas (Lee y Perret, 2022).

5.5 Pensamiento crítico

El desarrollo de competencias de pensamiento crítico en el alumnado emerge como un imperativo para fomentar un uso ético y efectivo de la IA. El pensamiento crítico, definido como la capacidad de analizar hechos de manera objetiva para formar un juicio (Facione, 2007), es fundamental para transitar en el complejo paisaje de la IA. Al equipar a los estudiantes con habilidades de pensamiento crítico, los preparamos no solo para comprender mejor los mecanismos subyacentes y las aplicaciones de la IA, sino también para cuestionar y evaluar las implicaciones éticas y sociales de su uso (Darwin et al., 2024).

Una de las razones por las que el pensamiento crítico es tan crucial en el contexto de la IA es que, como hemos relatado en secciones anteriores, esta tecnología no está exenta de sesgos. Los algoritmos de IA, aunque parezcan objetivos, están diseñados por seres humanos y pueden perpetuar o incluso exacerbar prejuicios existentes. Un alumnado con competencias sólidas de pensamiento crítico estará mejor equipado para identificar y cuestionar estos sesgos, promoviendo así el desarrollo y la implementación de sistemas de IA más justos y equitativos (Rusandi et al., 2023).

Además, el pensamiento crítico permite a los estudiantes entender la diferencia entre el uso efectivo y el uso indebido de la IA. En un mundo donde la información puede ser manipulada y donde las noticias falsas pueden propagarse rápidamente a través de las redes sociales impulsadas por algoritmos, la capacidad de evaluar críticamente las fuentes de información y los datos

presentados es más importante que nunca. Esto es esencial no solo para la toma de decisiones informadas a nivel personal, sino también para el mantenimiento de una sociedad informada y democrática (Rusandi et al., 2023).

El fomento de un uso ético de la IA también depende en gran medida del pensamiento crítico. Los dilemas éticos, como la privacidad de los datos, el consentimiento informado y la autonomía en el uso de sistemas de IA, requieren que los individuos puedan analizar situaciones complejas y considerar las consecuencias a largo plazo de las decisiones tecnológicas. Por lo tanto, enseñar a los estudiantes a abordar problemas desde múltiples perspectivas y a cuestionar las normas existentes es fundamental para garantizar que la IA se desarrolle y utilice de manera que respete los derechos y la dignidad de todas las personas (Haejoong y Sangmin, 2021).

Para cultivar efectivamente el pensamiento crítico en relación con la IA, la educación debe ir más allá de la mera transmisión de conocimientos técnicos. Es esencial adoptar un enfoque interdisciplinario que combine las ciencias de la computación con la ética, la filosofía, la sociología y otras disciplinas. Esto no solo amplía la comprensión de los estudiantes sobre lo que la IA puede hacer, sino que también los sensibiliza sobre lo que debería hacer (Haejoong y Sangmin, 2021). Además, es importante promover un ambiente de aprendizaje que incentive la curiosidad, la pregunta y el debate. Los estudiantes deben sentirse libres de expresar sus opiniones, explorar diferentes puntos de vista y aprender de sus errores. Los ejercicios prácticos, como el análisis de casos de estudio y la participación en debates éticos sobre el uso de la IA, pueden ser particularmente efectivos para desarrollar habilidades de pensamiento crítico (Alam, 2023).

Por último, pero no menos importante, es crucial que los educadores mismos sean modelos a seguir en el uso ético y crítico de la IA. Esto implica mantenerse informados sobre los avances tecnológicos, reflexionar sobre su propia práctica pedagógica y buscar constantemente maneras de incorporar principios éticos en la enseñanza de la IA (Crawford et al., 2024).

En definitiva, el desarrollo de competencias de pensamiento crítico entre el alumnado es indispensable para promover un uso ético y efectivo de la IA. Al cultivar estas habilidades, preparamos a los estudiantes no solo para ser usuarios conscientes y responsables de la tecnología, sino también para ser ciudadanos activos y éticos en una sociedad cada vez más digital. La educación en pensamiento crítico sobre la IA no solo beneficia a los individuos, sino que también contribuye a una sociedad más justa, equitativa y democrática. Por lo tanto, es una inversión esencial en nuestro futuro colectivo.

5.6 Adoptar procesos basados en la transparencia en la comunicación del uso de la IA en las actividades académicas/formativas

La integración de la inteligencia artificial en el ámbito educativo presenta un horizonte repleto de posibilidades para transformar y mejorar los procesos de enseñanza y aprendizaje. Sin embargo, para que esta integración sea exitosa y éticamente responsable, es fundamental adoptar procesos basados en la transparencia en la comunicación sobre el uso de la IA en las actividades académicas. La transparencia no solo fomenta un ambiente de confianza y apertura entre los estudiantes, docentes y la comunidad educativa en general, sino que también asegura que todos los implicados comprendan cómo, por qué y con qué propósito se utiliza la IA (Chaudhry et al., 2022).

En primer lugar, la transparencia implica una comunicación clara y detallada sobre los tipos de tecnologías de IA que se están implementando dentro de las aulas y plataformas de aprendizaje. Esto significa explicar las funcionalidades específicas de las herramientas de IA, como los sistemas de tutoría personalizada, las plataformas de evaluación automatizada o los asistentes virtuales, y

cómo estas tecnologías pueden contribuir al proceso educativo. La claridad en la presentación de estas herramientas ayudará a disipar miedos y malentendidos sobre la IA, promoviendo una actitud más receptiva entre los estudiantes y docentes.

Además, la transparencia debe abordar las preocupaciones éticas y de privacidad relacionadas con el uso de datos en aplicaciones de IA. Es esencial que los centros educativos establezcan políticas claras sobre la recolección, análisis y almacenamiento de datos, asegurando que todas las partes interesadas entiendan sus derechos y las medidas de protección implementadas para salvaguardar su información personal. Esto incluye informar a los estudiantes sobre cómo sus datos son utilizados para personalizar su experiencia de aprendizaje y los mecanismos que tienen a su disposición para controlar esta información.

La participación activa de estudiantes y docentes en el proceso de decisión sobre la implementación de la IA es otro aspecto crucial de la transparencia. Facilitar espacios para el diálogo, la retroalimentación y la colaboración permitirá que la comunidad educativa exprese sus inquietudes, sugerencias y expectativas respecto al uso de la IA. Esta colaboración puede conducir al desarrollo de prácticas de IA más inclusivas y efectivas, ajustadas a las necesidades y objetivos educativos específicos.

La evaluación y revisión continua de las tecnologías de IA en uso es también parte integral de un enfoque transparente. Los centros educativos deben comprometerse a monitorear el impacto de la IA en el aprendizaje y el bienestar de los estudiantes, compartiendo los hallazgos con la comunidad educativa y ajustando las estrategias según sea necesario. Este proceso de evaluación abierta y continua garantiza que las aplicaciones de IA permanezcan alineadas con los principios éticos y los objetivos educativos.

Finalmente, la adopción de procesos transparentes en el uso de la IA en educación requiere de un esfuerzo conjunto para educar a toda la comunidad educativa sobre los principios fundamentales de la IA. Desarrollar programas de formación y recursos informativos que expliquen de manera accesible cómo funciona la IA y sus aplicaciones potenciales en educación, empoderará a estudiantes y docentes para interactuar de manera crítica y consciente con estas tecnologías.

En resumen, la transparencia en la comunicación del uso de la IA es esencial para construir un entorno educativo inclusivo, ético y efectivo. Al adoptar un enfoque basado en la claridad, la participación y la evaluación continua, los centros educativos pueden navegar los desafíos y aprovechar al máximo las oportunidades que la IA ofrece para enriquecer la experiencia educativa.

5.7 Auditorias de IA

Para garantizar el uso de la IA sea ético y responsable en el entorno educativo, es crucial realizar auditorías regulares en los centros de educación. Este proceso no solo ayuda a identificar y mitigar posibles riesgos, sino que también asegura que las herramientas de IA se alineen con los valores éticos y los objetivos educativos de la institución (Sabzalieva y Valentini, 2023).

Una auditoría ética y responsable de las herramientas de IA en la educación implica una evaluación sistemática que considera tanto la eficacia técnica como el impacto ético y social de estas tecnologías. El objetivo es garantizar que las herramientas de IA se utilicen de manera que respeten los derechos y la dignidad de todos los miembros de la comunidad educativa, incluidos estudiantes, docentes y personal administrativo (Simbeck, 2023). Este proceso debe asentarse sobre el análisis y evaluación continua de diversos elementos entre los que, basándonos en Mokander (2023), destacamos:

- **Transparencia y explicabilidad:** uno de los principios fundamentales para el uso ético de la IA en la educación superior es la transparencia. Las auditorías deben evaluar si las herramientas de IA son transparentes en términos de sus algoritmos, los datos que utilizan y las decisiones que toman. Esto es esencial para generar confianza entre los usuarios y para permitir una supervisión efectiva por parte de las instituciones. Además, la explicabilidad de las decisiones tomadas por las IA es crucial, especialmente cuando estas afectan aspectos significativos de la vida académica, como la evaluación de los estudiantes o la asignación de recursos.
- **Privacidad y seguridad de los datos:** la protección de la privacidad y la seguridad de los datos es otro aspecto crítico que las auditorías deben abordar. Las herramientas de IA a menudo requieren grandes volúmenes de datos para funcionar eficazmente, lo que puede incluir información personal sensible de estudiantes y personal. Las auditorías deben asegurar que se cumplan las leyes de protección de datos aplicables y que existan medidas robustas para proteger contra la pérdida, el mal uso o el acceso no autorizado a estos datos.
- **Equidad y no discriminación:** las herramientas de IA tienen el potencial de perpetuar o incluso exacerbar las desigualdades existentes en el sistema educativo. Las auditorías deben examinar críticamente si estas tecnologías introducen sesgos en la toma de decisiones, lo que podría resultar en discriminación contra ciertos grupos. Esto implica revisar los conjuntos de datos utilizados para entrenar los algoritmos de IA, así como analizar los resultados de su implementación para detectar cualquier patrón de desigualdad.
- **Impacto en el aprendizaje y el bienestar:** además de los aspectos técnicos y éticos, las auditorías deben considerar el impacto de las herramientas de IA en los procesos de enseñanza y aprendizaje. Esto incluye evaluar cómo estas tecnologías afectan la experiencia educativa de los estudiantes, su bienestar y su capacidad para alcanzar los resultados de aprendizaje deseados. Es fundamental que las herramientas de IA se utilicen para complementar y enriquecer la enseñanza, en lugar de reemplazar las interacciones humanas esenciales para el proceso educativo.
- **Participación de la comunidad educativa:** para que una auditoría sea realmente efectiva, debe involucrar a toda la comunidad educativa en el proceso. Esto incluye recoger opiniones y preocupaciones de estudiantes, docentes y personal administrativo sobre el uso de herramientas de IA. Su participación asegura que la auditoría tenga en cuenta una amplia gama de perspectivas y contribuye a crear un entorno de aprendizaje inclusivo y democrático.

Implementación de mejoras: finalmente, la realización de una auditoría es solo el primer paso. Es crucial que las instituciones de educación superior estén comprometidas con implementar las mejoras necesarias identificadas durante el proceso de auditoría. Esto puede requerir actualizar las políticas de privacidad, modificar los algoritmos para reducir los sesgos o implementar nuevas prácticas de seguridad de datos.

6. Conclusiones

La IA en el ámbito educativo no es solo una herramienta para mejorar la eficiencia o automatizar tareas; representa una oportunidad fundamental para repensar y enriquecer el proceso educativo en su conjunto. A través de sus características y potencial, la IA puede resultar sumamente positiva en los procesos de enseñanza-aprendizaje. Sin embargo, este potencial solo

puede realizarse plenamente si abordamos de manera proactiva los desafíos éticos y prácticos que surgen con su implementación.

Una de las claves para maximizar los beneficios de la IA en la educación es el compromiso con la transparencia y la equidad. Esto significa no solo informar a los estudiantes y docentes sobre cómo se utilizan los datos y los algoritmos, sino también asegurar que estas herramientas no perpetúen sesgos o desigualdades existentes. La participación activa de la comunidad educativa en la evaluación y el desarrollo de tecnologías de IA puede ayudar a garantizar que estos sistemas sean inclusivos y representativos de la diversidad de la población estudiantil.

Además, las auditorías regulares y exhaustivas de las herramientas de IA que se usen en los centros educativos son esenciales para mantener altos estándares éticos y prácticos. Estas auditorías deben considerar no solo la eficacia de la tecnología, sino también su impacto en el bienestar y el aprendizaje de los estudiantes. Al evaluar críticamente el uso de la IA desde una perspectiva ética y práctica, las instituciones educativas pueden identificar áreas de mejora y adaptar sus prácticas para servir mejor a sus comunidades.

La corresponsabilidad entre todos los miembros de la comunidad educativa es otro fundamento crucial para abordar estos desafíos. Esto incluye no solo a los estudiantes y profesores, sino también a los administradores y responsables políticos. Todos deben colaborar para establecer principios y prácticas que guíen el uso ético y efectivo de la IA en la educación. La educación y formación en competencias digitales y éticas para estudiantes y docentes es vital para fomentar un entendimiento profundo de los beneficios y riesgos asociados con la IA.

Para aprovechar al máximo el potencial transformador de la IA en la educación, es esencial adoptar un enfoque holístico y colaborativo. Esto significa no solo implementar tecnologías avanzadas, sino también repensar los enfoques pedagógicos, las estructuras curriculares y las políticas educativas para asegurar que la IA se integre de manera que enriquezca el aprendizaje y promueva la equidad y la inclusión. Al hacerlo, las instituciones educativas pueden liderar el camino hacia un futuro donde la IA no solo mejore la eficiencia, sino que también profundice y enriquezca la experiencia educativa para todos.

En resumen, la integración de la IA en la educación representa una oportunidad sin precedentes para transformar la enseñanza y el aprendizaje. Para lograrlo, las instituciones educativas deben comprometerse con la transparencia, la equidad y la corresponsabilidad, asegurando que la tecnología se utilice de manera que beneficie a todos los estudiantes y fomente un entorno de aprendizaje más inclusivo, personalizado y enriquecedor.

Agradecimientos:

Artículo elaborado en el marco de los proyectos: a) 2022-1-ES01-KA220-HED-000085257 cofinanciado por la UE, b) PID2022-141031NB-I00, financiado por MICIU/AEI /10.13039/501100011033/ y FEDER “Una manera de hacer Europa”, y c) Red-IA financiada por la AUIP.

Referencias

Adolphs, R. (2013). The biology of fear. *Current Biology*, 23(2), R79–R93. <http://dx.doi.org/10.1016/j.cub.2012.11.055>

- Akter, S., Dwivedi, Y. K., Biswas, K., Michael, K., Bandara, R.J., y Sajib, S. (2021). Addressing algorithmic bias in AI-driven customer management. *Journal of Global Information Management*, 29(6), 1-27. <https://www.doi.org/10.4018/JGIM.20211101.0a3>
- Alam, A. (2023). Developing a Curriculum for Ethical and Responsible AI: A University Course on Safety, Fairness, Privacy, and Ethics to Prepare Next Generation of AI Professionals. *Intelligent Communication Technologies and Virtual Mobile Networks* (pp. 879-894), Springer Nature Singapore. https://doi.org/10.1007/978-981-99-1767-9_64
- Amaro, I., Barra, P., Della Greca, A., Francese, R., y Tucci, C. (2023). Believe in Artificial Intelligence? A User Study on the ChatGPT's fake information impact. *IEEE Transactions on Computational Social Systems*, 1-10. <https://www.doi.org/10.1109/TCSS.2023.3291539>
- Anderson, H., Boodhwani, A., y Baker, R. S. (2019). Assessing the Fairness of Graduation Predictions. EDM. <https://tinyurl.com/39mypts8>
- Athaluri, S., Manthena, S., Kesapragada, V., Yarlagadda, V., Dave, T., y Duddumpudi, R.T. (2023). Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*, 15(4), e37432. <https://www.doi.org/10.7759/cureus.37432>
- Ayanwale, M. A., Sanusi, I.T., Adelana, O.P., Aruleba, K. D., y Oyelere, S. S. (2022). Teachers' readiness and intention to teach artificial intelligence in schools. *Computers and Education: Artificial Intelligence*, 3, 100099. <https://doi.org/10.1016/j.caeai.2022.100099>
- Barsee (2023). GPTZero, An AI Detector, thinks the US Constitution was written by AI. <https://twitter.com/heyBarsee/status/1631920236046827520?lang=es>
- Beutel, G., Geerits, E., y Kielstein, J.T. (2023). Artificial hallucination: GPT on LSD? *Critical Care*, 27, 148. <https://doi.org/10.1186/s13054-023-04425-6>
- Bostrom, N. (2014). *Superintelligence, Paths, Dangers, Strategies*. Oxford University Press. <https://tinyurl.com/5n92ath5>
- Carvalho, L., Martínez-Maldonado, R., Tsai, Y., Markauskaite, L., y De Laat, M. (2022). How can we design for learning in an AI world? *Computers and Education: Artificial Intelligence*, 3, 100053. <https://doi.org/10.1016/j.caeai.2022.100053>
- Charlie Project Competency Matrices for Algorithmic Bias (2023). https://charlie-project.uib.es/lib/files/competencymatrix_EN.pdf
- Chaudhry, M.A., Cukurova, M., y Luckin, R. (2022). A transparency index framework for AI in education. *International Conference on Artificial Intelligence in Education* (pp. 195-198), Springer International Publishing. https://doi.org/10.1007/978-3-031-11647-6_33
- Choi, W., Zhang, Y., y Stvilia, B. (2023). Exploring Applications and User Experience with Generative AI Tools: A Content Analysis of Reddit Posts on ChatGPT. *Proceedings of the Association for Information Science and Technology*, 60(1), 543-546. <https://doi.org/10.1002/pra2.823>
- Crawford, J., Allen, K.A., Pani, B., y Cowling, M. (2024). When artificial intelligence substitutes humans in higher education: the cost of loneliness, student success, and retention. *Studies in Higher Education*, 1-15. <https://doi.org/10.1080/03075079.2024.2326956>

- Crawford, K. (2023). *Atlas de inteligencia artificial: Poder, política y costos planetarios*. Fondo de Cultura Económica Argentina.
- Cybenko, A.K. y Cybenko, G. (2018). AI and fake news. *IEEE Intelligent Systems*, 33(5), 1-5. <https://www.doi.org/10.1109/MIS.2018.2877280>
- Dahlin, E. (2019). Are Robots Stealing our Jobs? *Socius*, 5, 1–14. <https://doi.org/10.1177/2378023119846249>
- Dhar, P. (2020). The carbon impact of artificial intelligence. *Nature Machine Intelligence*, 2, 423–425. <https://doi.org/10.1038/s42256-020-0219-9>
- Darwin, R. D., Mukminatien, N., Suryati, N., Laksmi, E.D., y Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, 11(1), 2290342. <https://doi.org/10.1080/2331186X.2023.2290342>
- Dawson, P. (2024). Remote Proctoring: Understanding the Debate. *Second Handbook of Academic Integrity* (pp. 1511-1526), Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54144-5_150
- Del Bonifro, F., Gabbrielli, M., Lisanti, G., y Zingaro, S. P. (2020). Student dropout prediction. Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, Proceedings, Part I 21 (pp. 129-140). Springer International Publishing. https://doi.org/10.1007/978-3-030-52237-7_11
- Desbrières, L. (2023). ChatGPT: la résistance s'organise. *Lire*, (525), 24-24. <https://www.proquest.com/openview/6529479031f7aea046040f6488e2d1af/1?pq-origsite=gscholar&cbl=5001983>
- Dilmegani, C. (2024). Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2024. AIMultiple, <https://research.aimultiple.com/ai-bias/>
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200. <https://www.paulekman.com/wp-content/uploads/2013/07/An-Argument-For-Basic-Emotions.pdf>
- Elliot, D. y Soifer, E. (2022). AI Technologies, privacy, and security. *Frontiers in Artificial Intelligence*, 5, 60. <https://doi.org/10.3389/frai.2022.826737>
- Ersoy, E. (2024). EU artificial intelligence regulation. *Computational Intelligence and Blockchain in Complex Systems* (pp. 113-119), Morgan Kaufmann. <https://doi.org/10.1016/B978-0-443-13268-1.00001-7>
- Facione, P. (2007). Pensamiento Crítico: ¿Qué es y por qué es importante? *Insight assessment*, 22, 23-56. <http://www.eduteka.org/PensamientoCriticoFacione.php>
- Floridi, L. (2023). *The Ethics of Artificial Intelligence: principles, challenges, and opportunities*. Oxford University Press. <https://books.google.es/books?id=EKLLEAAAQBAJ>
- Frenkel, S. y Thompson, S. A. (2023). 'Not for Machines to Harvest': Data Revolts Break Out Against AI. *International New York Times*, NA-NA. <https://link.gale.com/apps/doc/A757592811/AONE?u=anon~21cbfe25&sid=googleScholar&xid=84be1335>

- Gallent-Torres, C. y Comas-Forgas, R. (2024). La llama de Prometeo: IA e integridad académica. *Cuadernos de pedagogía*, 549, 16-27, <https://dialnet.unirioja.es/servlet/articulo?codigo=9255110>
- Galli, M. G. y Kanobel, M. C. (2023). ChatGPT en Educación Superior: explorando sus potencialidades y sus limitaciones. *Revista Educación Superior y Sociedad (ESS)*, 35(2), 174-195. <https://doi.org/10.54674/ess.v35i2.815>
- Gendron, Y., Andrew, J., y Cooper, C. (2022). The perils of artificial intelligence in academic publishing. *Critical Perspectives on Accounting*, 87. <https://doi.org/10.1016/j.cpa.2021.102411>
- Gervet, T., Koedinger, K., Schneider, J. y Mitchell, T. (2020). When is deep learning the best approach to knowledge tracing? *Journal of Educational Data Mining*, 12(3), 31-54. <https://doi.org/10.5281/zenodo.4143614>
- Glikson, E. y Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660, <https://doi.org/10.5465/annals.2018.0057>
- Gómez, W. O. A. (2023). La inteligencia artificial y su incidencia en la educación: Transformando el aprendizaje para el siglo XXI. *Revista internacional de pedagogía e innovación educativa*, 3(2), 217-229. <https://doi.org/10.51660/ripie.v3i2.133>
- González-González, C. S. (2023). El impacto de la Inteligencia Artificial en la educación: transformación de la forma de enseñar y de aprender. *Revista Currículum*, 36, 51-60. <https://doi.org/10.25145/j.qurricul.2023.36.03>
- Grynbaum, M. M. y Mac, R. (2023). The Times Sues OpenAI and Microsoft. *The New York Times*, B1-B1. <https://tinyurl.com/y6te2vyf>
- Guenduez, A. A. y Mettler, T. (2023). Strategically constructed narratives on artificial intelligence: What stories are told in governmental artificial intelligence policies? *Government Information Quarterly*, 40(1), 101719. <https://doi.org/10.1016/j.giq.2022.101719>
- Haejoong, K. I. M. y Sangmin, L. E. E. (2021). Do less teaching, do more coaching: toward critical thinking for ethical applications of artificial intelligence. *Journal of Learning and Teaching in Digital Age*, 6(2), 97-100. <https://dergipark.org.tr/en/download/article-file/1311877>
- Herrera de las Heras, R. (2022). Aspectos legales de la inteligencia artificial: personalidad jurídica de los robots, protección de datos y responsabilidad civil. *Aspectos legales de la inteligencia artificial*, 1-124. <http://digital.casalini.it/9788411221627>
- Holmes, W. y Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542-570. <https://doi.org/10.1111/ejed.12533>
- Hopcan, S., Türkmen, G., y Polat, E. (2023). Exploring the artificial intelligence anxiety and machine learning attitudes of teacher candidates. *Education and Information Technologies*, 29, 7281-7301. <https://doi.org/10.1007/s10639-023-12086-9>
- Hu, Q. y Rangwala, H. (2020). Towards Fair Educational Data Mining: A Case Study on Detecting At-Risk Students. *International Educational Data Mining Society*. <https://files.eric.ed.gov/fulltext/ED608050.pdf>
- Huang, M. H. y Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49, 30-50. <https://doi.org/10.1007/s11747-020-00749-9>

Huang, X., Yusoff, Z. M., Nor, M. Z. B. M. y Labanieh, M. F. (2024). The Legal Challenges and Regulatory Responses to Artificial Intelligence (AI) in China. 12th UUM International Legal Conference 2023 (UUMILC 2023) (pp. 335-347). Atlantis Highlights in Social Sciences, Education and Humanities. <https://www.atlantis-press.com/proceedings/uumilc-23/125997171>

Huawei, S. y Aryadoust, V. (2023). A systematic review of automated writing evaluation systems. *Education and Information Technologies*, 28(1), 771-795. <https://doi.org/10.1007/s10639-022-11200-7>

Hutchinson, B. y Mitchell, M. (2019). 50 years of test (un)fairness: Lessons for machine learning. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 49–58. <https://doi.org/10.1145/3287560.3287600>

Islam, M. (2023). Are AI content detectors reliable in 2023? *Meta Minds* <https://www.linkedin.com/pulse/ai-content-detectors-reliable-2023-mirajul-islam>

Kai, S., Andres, J. M. L., Paquette, L., Baker, R. S., Molnar, K., Watkins, H., y Moore, M. (2017). Predicting Student Retention from Behavior in an Online Orientation Course. *International Educational Data Mining Society* <https://files.eric.ed.gov/fulltext/ED596601.pdf>

Khalil, M. y Erkan, E. (2023). Will ChatGPT Get You Caught? Rethinking of Plagiarism Detection. *International Conference on Human-Computer Interaction* (pp. 475-487). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34411-4_32

Kizilcec, R. F. y Lee, H. (2022). Algorithmic fairness in education. *The ethics of artificial intelligence in education* (pp. 174-202). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780429329067-10/algorithmic-fairness-education-ren%C3%A9-kizilcec-hansol-lee>

Kong, S. C., Cheung, W. M. Y. y Zhang, G. (2021). Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. *Computers and Education: Artificial Intelligence*, 2, 100026. <https://doi.org/10.1016/j.caeai.2021.100026>

Kong, S. C., Cheung, W. M. Y. y Zhang, G. (2022). Evaluating an AI literacy programme for empowering and developing concepts, literacy and ethical awareness in senior secondary students. *Education and Information Technologies*, 28, 4703-4724. <https://doi.org/10.1007/s10639-022-11408-7>

Lacoste, A., Luccioni, A., Schmidt, V. Y DANDRES, T. (2019). Quantifying the Carbon Emissions of Machine Learning. arXiv preprint arXiv: 1910.09700v2. <https://doi.org/10.48550/arXiv.1910.09700>

Lee, I. y Perret, B. (2022). Preparing high school teachers to integrate artificial intelligence methods into STEM classrooms. *Proceedings of the AAAI Conference on AI*, 36(11), 12783–12791. <https://doi.org/10.1609/aaai.v36i11.21557>

Lee, I., Ali, S., Zhang, H., Dipaola, D., y Breazeal, C. (2021). Developing middle school students' artificial intelligence literacy. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 191197. <https://doi.org/10.1145/3408877.3432513>

Lee, J. y Park, J. (2023). AI as “Another I”: Journey map of working with artificial intelligence from AI-phobia to AI-preparedness. *Organizational Dynamics*, 52(3), 100994. <https://doi.org/10.1016/j.orgdyn.2023.100994>

- Li, J. y Huang, J-S. (2020). Dimensions of artificial Intelligence anxiety based on the integrated fear acquisition theory. *Technology in Society*, 63, 101410. <https://www.doi.org/10.1016/j.techsoc.2020.101410>
- Li, L., Sha, L., Li, Y., Rakovic, M., Rong, J., Jokimovic, S., ... y Chen, G. (2023). Moral machines or tyranny of the majority? A systematic review on predictive bias in education. LAK23: 13th international learning analytics and knowledge conference (pp. 499-508). <https://doi.org/10.1145/3576050.3576119>
- Liu, M., Ren, Y., Nyagoga, L. M., Stonier, F., Wu, Z., y Yu, L. (2023). Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools. *Future in Educational Research*, 1(1), 72-101. <https://doi.org/10.1002/fer3.10>
- López Meneses, E., Díaz Díaz, N., Vázquez Cano, E., y Lagares Rodríguez, J. A. (2022). Inteligencia artificial y formación docente en ecologías de aprendizaje. Dykinson. <http://digital.casalini.it/9788411228114>
- Mahmut, D. y Şen, N. D. Ş. (2023). Reflection of artificial intelligence applications on students' critical and analytical abilities. Scientific Collection «InterConf», 184, 459-467. <https://archive.interconf.center/index.php/conference-proceeding/article/view/5124>
- Martínez, S.J. (2022). Deshonestidad académica en la era digital. *Interacción digital UVEG*, 4(8), 1-16. <https://d18xmwa1oz3vx.cloudfront.net/revista/images/8/04Conocimientos-y-saberes-Deshonestidad-academica-en-la-era-digital.pdf>
- Melo, E., Silva, I., Costa, D. G., Viegas, C. M., y Barrios, T. M. (2022). On the use of explainable artificial intelligence to evaluate school dropout. *Education Sciences*, 12(12), 845. <https://doi.org/10.3390/educsci12120845>
- Menzies, R.G. y Clarke, J.C. (1995). The etiology of phobias: a non-associative account. *Clinical Psychological Review*, 15(1), 23-48. [https://doi.org/10.1016/0272-7358\(94\)00039-5](https://doi.org/10.1016/0272-7358(94)00039-5)
- Miao, F., Holmes, W., Huang, R., y Zhang, H. (2021). *Inteligencia artificial y educación: Guía para las personas a cargo de formular políticas*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000379376>
- Mokander, J. (2023). Auditing of AI: Legal, ethical and technical approaches. *Digital Society*, 2(3), 49. <https://doi.org/10.1007/s44206-023-00074-y>
- Mucci, F. y Marazziti, D. (2023). Artificial Intelligence in Neuropsychiatry: A Potential Beacon in an Ocean of Uncertainty? *Clinical Neuropsychiatry*, 20(6), 467. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10852407/>
- Mujezinovic, D. (2023). AI content detectors don't work, and that's a big problem, *Make Use Of*, <https://www.makeuseof.com/ai-content-detectors-dont-work/>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., y Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Ouyang, F., Zheng, L. y Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, 27(6), 7893-7925. <https://doi.org/10.1007/s10639-022-10925-9>

- Parinandi, S., Crosson, J., Peterson, K. y Nadarevic, S. (2024). Investigating the politics and content of US State artificial intelligence legislation. *Business and Politics*, 26(2), 240-262. <https://doi.org/10.1017/bap.2023.40>
- Pelánek, R. (2022). Adaptive, Intelligent, and Personalized: Navigating the Terminological Maze Behind Educational Technology. *International Journal of Artificial Intelligence in Education*, 32(1), 151–173. <https://doi.org/10.1007/s40593-021-00251-5>
- Poulton, R., Waldie, K.E., Menzies, R.G., Craske, M. G., y Silva, P.A. (2001). Failure to overcome 'innate' fear: a developmental test of the non-associative model of fear acquisition. *Behaviour Research and Therapy*, 39(1), 29–43. [https://doi.org/10.1016/S0005-7967\(99\)00156-4](https://doi.org/10.1016/S0005-7967(99)00156-4)
- Rachman, S. (1977). The conditioning theory of fear-acquisition: a critical examination, *Behaviour Research and Therapy*, 15(5), 375–387. [https://doi.org/10.1016/0005-7967\(77\)90041-9](https://doi.org/10.1016/0005-7967(77)90041-9)
- Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human behavior and emerging technologies*, 1(1), 33-36. <https://doi.org/10.1002/hbe2.117>
- Rusandi, M. A., Saripah, A. I., Khairun, D. Y., y Mutmainnah, M. (2023). No worries with ChatGPT: building bridges between artificial intelligence and education with critical thinking soft skills. *Journal of Public Health*, 45(3), e602-e603. <https://doi.org/10.1093/pubmed/fdad049>
- Ryan-Mosley, T. (2024). Vuelta al mundo por las regulaciones de la IA en 2024. *MIT technology Review*. <https://www.technologyreview.es/s/16069/vuelta-al-mundo-por-las-regulaciones-de-la-ia-en-2024>
- Sabzalieva, E. y Valentini, A. (2023). ChatGPT and artificial intelligence in higher education: quick start guide. UNESCO. International Institute for Higher Education in Latin America and the Caribbean (IESALC). <https://tinyurl.com/2s3d2sk6>
- Sánchez, I. V. M. D. O., Bravo, M. G. E., Reyes, A. T. C., Marín, H. J. V., y Chacha, A. G. O. (2023). *EduTrends: Navegando en la Era Digital de la Educación*. Editorial Investigativa Latinoamericana (SciELa).
- Self, J. (2016). The Birth of IJAIED. *International Journal of Artificial Intelligence in Education*, 26(1), 4–12. <https://doi.org/10.1007/s40593-015-0040-5>
- Shams, R. A., Zowghi, D., y Bano, M. (2023). AI and the quest for diversity and inclusion: a systematic literature review. *AI and Ethics*, 1-28. <https://doi.org/10.1007/s43681-023-00362-w>
- Shen, Y., Heacock, L., Elía, J., Hentel, K.D., Reig, B., Shih, G., y Moy, L. (2023). ChatGPT and other large language models are double edged swords. *Radiology*, 307(2), e230163. <https://www.doi.org/10.1148/radiol.230163>
- Simbeck, K. (2023). They shall be fair, transparent, and robust: auditing learning analytics systems. *AI and Ethics*, 4, 555-571. <https://doi.org/10.1007/s43681-023-00292-7>
- Slater, G. B. (2024). Dread and the automation of education: From algorithmic anxiety to a new sensibility. *Review of Education, Pedagogy, and Cultural Studies*, 46(1), 170-182. <https://doi.org/10.1080/10714413.2023.2299521>
- Smith, H. (2020). Algorithmic bias: should students pay the price? *AI & society*, 35(4), 1077-1078. <https://doi.org/10.1007/s00146-020-01054-3>

Soller, A., Martínez, A., Jermann, P., y Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education*, 15(4), 261-290. <https://content.iospress.com/articles/international-journal-of-artificial-intelligence-in-education/jai15-4-02>

Southworth, J., Migliaccio, K., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., y Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, 4, 100127. <https://doi.org/10.1016/j.caeai.2023.100127>

Steinbauer, G., Kandlhofer, M., Chklovski, T., Heintz, F., y Koenig, S. (2021). A differentiated discussion about AI education K-12. *KI-Künstliche Intelligenz*, 35, 131–137. <https://doi.org/10.1007/s13218-021-00724-8>

Su, J., Ng, D. T. K., y Chu, S. K. W. (2023). Artificial Intelligence (AI) literacy in early childhood education: the challenges opportunities. *Computer and Education: Artificial Intelligence*, 4, 100124. <https://doi.org/10.1016/j.caeai.2023.100124>

Sureda, J., Reynes, J. Comas, R. (2016). Anti-academic fraud regulations in Spanish universities. *Revista de la educación superior*, 45(178), 31-44. <http://dx.doi.org/10.1016/j.resu.2016.03.002>

Susnjak, T. (2022). ChatGPT: The end of online exam integrity? arXiv Preprint arXiv: 2212.09292. <https://arxiv.org/abs/2212.09292>

Takaki, P., Dutra, M. L., De Araújo, G., y Júnior, E. M. D. (2022). A Proposed Framework for Evaluating the Academic-failure Prediction in Distance Learning. *Mobile Networks and Applications*, 27(5), 1958–1966. <https://doi.org/10.1007/s11036-022-01965-z>

Tapia, T., Orenday, T.M., y Gómez, F. M. (2023). Percepciones sobre corrupción y deshonestidad académica entre estudiantes universitarios. *Psicología, educación y sociedad. Revista de investigación y difusión*, 2(3), 1-21. <https://doi.org/10.5281/zenodo.8000549>

Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, LIX (236): 433–460. <https://doi.org/10.1093/mind/LIX.236.433>

Uto, M. y Okano, M. (2021). Learning automated essay scoring models using item-response-theory-based scores to decrease effects of rater biases. *IEEE Transactions on Learning Technologies*, 14(6), 763-776. <https://doi.org/10.1109/TLT.2022.3145352>

Van Giffen, B., Herhausen, D. y Fahse, T. (2022). Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *Journal of Business Research*, 144, 93-106. <https://doi.org/10.1016/j.jbusres.2022.01.076>

Varsha, P. S. (2023). How can we manage biases in artificial intelligence systems—A systematic literature review. *International Journal of Information Management Data Insights*, 3(1), 100165. <https://doi.org/10.1016/j.ijime.2023.100165>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gómez, A. N., ... y Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. 31st Conference on Neural Information Processing Systems (NIPS 2017). https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vera, F. (2023). Integración de la Inteligencia Artificial en la Educación superior: Desafíos y oportunidades. *Revista Electrónica Transformar*, 4(1), 17-34. <https://www.revistatransformar.cl/index.php/transformar/article/view/84>

Walter, Y. (2024). Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences. *Discover Artificial Intelligence*, 4(1), 14. <https://doi.org/10.1007/s44163-024-00109-4>

Wang, C. Y., Song, Y., Wu, C. Y. y Yang, P. T. (2020). The Moderating Effect of Artificial Intelligence Phobia on the Relationship between Trust and Product Promotion Effectiveness: An Exploratory Study. Proceedings of the 2020 11th International Conference on E-Education, E-Business, E-Management, and E-Learning (pp. 356-359). <https://doi.org/10.1145/3377571.3377594>

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3), 129-140. <https://doi.org/10.1080/17470216008416717>

Xu, W. y Ouyang, F. (2022). The application of AI technologies in STEM education: a systematic review from 2011 to 2021. *International Journal of STEM Education*, 9(1), 59. <https://doi.org/10.1186/s40594-022-00377-5>

Yang, W. (2022). Artificial intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers and Education: Artificial Intelligence*, 3, 100061. <https://doi.org/10.1016/j.caeai.2022.100061>

Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, 16(3), e12962. <https://doi.org/10.1111/soc4.12962>

Recibido: 12/04/2024

Versión corregida: 04/08/2024

Aceptado: 05/08/2024

Publicado online: 12/08/2024