
Revista de Estudios y Experiencias en Educación

REXE

journal homepage: <http://revistas.ucsc.cl/index.php/rexe>

Sesgos en la IA y educación superior. Tipologías, impactos y mitigación para la formación universitaria de calidad

Ricardo García-Hormazábal

Universidad Católica de Temuco, Temuco, Chile

Recibido: 07 de enero 2025 - Revisado: 21 de enero 2025 - Aceptado: 24 de febrero 2025

RESUMEN

Este estudio reflexiona sobre la presencia de sesgos en el uso de la inteligencia artificial (IA), sus características particulares y los eventuales efectos que dicho uso generaría como soporte para las actividades de aprendizaje asistido en el desarrollo universitario actual. Para ello, se utilizó un método de análisis documental, recurriendo a revistas y estudios especializados en el desarrollo de la IA y su vinculación con los desafíos formativos actuales en la educación superior. Posterior a ello, se analizan las consecuencias formativas de dichos sesgos; así como las alternativas de mitigación posibles para focalizar estrategias concretas, a fin de que los estudiantes universitarios consideren dichos mecanismos al momento de interactuar en sus procesos de aprendizaje asistido con estos recursos. Los resultados, exponen diversos tipos de sesgos asociados al uso de la IA, referidos principalmente a la manera en que se configura o entrena el modelo de IA, su carácter altamente responsivo y la tendencia a utilizar valores predominantes. Dichos sesgos, que aunque documentados, se encuentran en franco desarrollo; así como las consecuencias formativas, referidas principalmente al desarrollo del pensamiento crítico, la reproducción de estereotipos y favorecer la equidad en educación, en tanto desafíos clave para la educación superior. Finalmente, se discuten los alcances formativos y la necesidad de proyectar un uso de la IA que avance desde un recurso pedagógico a una didáctica complementaria a los procesos formativos en la universidad, donde docentes y estudiantes interactúen con ella en tanto medio más que fin último de la formación académico-profesional.

*Correspondencia: Ricardo García-Hormazábal (R. García-Hormazábal).

 <https://orcid.org/0000-0003-4304-8074> (rgarcia@uct.cl).

Palabras claves: Inteligencia artificial; enseñanza superior; sesgo; calidad de la educación; gestión del conocimiento.

Biases in AI and higher education. Typologies, impacts and mitigation for quality university education

ABSTRACT

This study examines the biases present in the use of artificial intelligence (AI), its specific characteristics, and the potential effects that this use may have on supporting assisted learning activities in contemporary university development. To do this, a documentary analysis method was used, resorting to magazines and studies specialized in the development of AI and its connection with current training challenges in higher education. Next, the study looks at how these biases affect training and explores ways to reduce them, so university students can better understand and use these tools in their learning. The results reveal different kinds of biases linked to using AI, mainly related to how the AI model is set up or trained, its quick responsiveness, and its tendency to reflect common values. These documented biases are still evolving, as are the training consequences, which mainly involve developing critical thinking, reproducing stereotypes, and promoting equity in education—key challenges for higher education. Finally, the training scope and the need to project the use of AI, which advances from a pedagogical resource to a didactic complement to the training processes at the university, are discussed, where teachers and students interact with it as a means rather than the ultimate goal of academic-professional training.

Keywords: Artificial intelligence; higher education; bias; quality of education; knowledge management.

1. Introducción

La educación superior se enfrenta a un entorno dinámico y complejo; caracterizado por rápidos avances tecnológicos y transformaciones sociales que buscan en Latinoamérica, la formación del capital humano (Labraña y Brunner, 2022). Este panorama exige que las instituciones de educación superior adapten sus enfoques pedagógicos, incorporen tecnologías emergentes y promuevan el desarrollo de competencias digitales en docentes y estudiantes, para garantizar la relevancia, calidad y equidad en la formación universitaria (Marín-Megía et al., 2025). La transformación digital en la educación superior es esencial para ampliar el acceso, mejorar la calidad del aprendizaje y fomentar la inclusión (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2023). Además, este proceso es crucial para la equidad educativa, al ofrecer nuevas oportunidades a estudiantes de diversas partes del mundo, especialmente a aquellos de contextos socioeconómicos desfavorecidos (Echeverría y Almendros, 2022). En el mismo sentido, Van Damme, & Zahner (2022) subrayan que las universidades deben convertirse en centros de innovación y adaptabilidad para preparar a los estudiantes para los desafíos de la economía digital; en donde dichos desafíos tensionan la integración de herramientas digitales no solo transformando el proceso educativo, sino que también como un mecanismo que fomenta un aprendizaje más autónomo y colaborativo permanente (Beetham y Sharpe, 2013).

Junto al desafiante escenario anterior y profundizando en la perspectiva formativa de la universidad, la educación superior actual debe responder a la denominada 'triple misión'; es decir, la necesidad de ampliar la formación a una interacción entre la enseñanza, la investigación y la vinculación con la sociedad (Salmi, 2017). De esta forma, los desafíos actuales implican que estas funciones deben actualizarse de manera simultánea para cumplir con las demandas de un mundo globalizado y digitalizado y en donde la integración de tecnologías y la formación en habilidades del siglo XXI son fundamentales para preparar a los futuros profesionales y por cierto, a los docentes de la educación superior (Villareal-Villa et al., 2019).

Paralelo a esto, las instituciones de educación superior deben reconfigurar sus currículos para incluir competencias digitales avanzadas (Romero-Rodríguez et al., 2023), que permitan a los estudiantes adaptarse a un mercado laboral en constante evolución (*Organization for Economic Cooperation and Development* [OECD], 2018). Esta reconfiguración abarca otros escenarios conforme a los cambios que se vienen sucediendo en la sociedad; y en donde la necesidad de transformar la educación superior incluye fomentar la innovación y la investigación mediante el uso de nuevas tecnologías y el desarrollo de competencias digitales en los docentes (Comisión Europea, 2018). Finalmente, destacar que estos cambios y ajustes no son solo institucionales o del sistema universitaria en exclusivo; sino que igualmente, este proceso de transformación requiere una colaboración activa entre los gobiernos, las instituciones educativas y las empresas para asegurar que la educación superior siga siendo relevante en el siglo XXI (Trowler, 2010).

La integración de tecnologías avanzadas, como la IA, está transformando la educación superior, ofreciendo nuevas formas de enseñanza, aprendizaje e interacción; así como la interacción ética que en estos procesos se producen (Gallent et al., 2023). La adopción de tecnologías emergentes es crucial para el desarrollo de sistemas educativos más inclusivos, accesibles y efectivos (OECD, 2023b). La IA y las herramientas digitales están permitiendo personalizar la educación y apoyar tanto a estudiantes como a docentes en sus procesos de aprendizaje y enseñanza (Marín, 2022). Además, la incorporación de estas tecnologías facilita la creación de espacios de aprendizaje colaborativo, favoreciendo el trabajo en equipo y la interacción en tiempo real, lo que fomentaría un enfoque más dinámico y flexible (Miller et al., 2023).

La introducción de tecnologías emergentes también está redefiniendo el rol tradicional del docente. Por ejemplo, las herramientas de IA permiten a los docentes automatizar tareas repetitivas, como la corrección de exámenes, lo que les proporciona más tiempo para enfocarse en la interacción directa con los estudiantes y en actividades de enseñanza más creativas y significativas. En este sentido, el modelo de *flipped classroom* (clase invertida), que promueve la participación activa de los estudiantes antes y después de las clases, se ve potenciado por el uso de herramientas digitales y de IA, que facilitan la preparación de materiales interactivos y el análisis de datos de rendimiento estudiantil (Chou et al., 2022). Además, la IA permitiría la implementación de sistemas de retroalimentación instantánea, lo que facilitaría que los estudiantes reciban comentarios sobre su aprendizaje de manera continua y autónoma, mejorando su desempeño y comprensión (Kerman et al., 2023).

En este contexto, la formación docente juega un papel esencial. Para maximizar el impacto de las tecnologías en el aprendizaje, los docentes deben recibir capacitación continua en el uso efectivo de las herramientas tecnológicas. Los docentes que reciben formación digital adecuada logran una enseñanza más eficaz y son capaces de crear entornos de aprendizaje más inclusivos y personalizados; ya que no solo abordarían el uso técnico de estas herramientas con las capacitaciones necesarias; sino que igualmente se potenciarían las implicaciones pedagógicas, éticas y sociales de su aplicación (Kim et al., 2022; Liu et al., 2022).

La cooperación internacional también es fundamental para superar estos retos. Organizaciones como la UNESCO, la OECD y la Comisión Europea han destacado la necesidad de una colaboración más estrecha entre instituciones educativas, gobiernos y el sector privado para promover un ecosistema digital inclusivo que beneficie a todos los estudiantes, independientemente de su origen a fin de promover un acceso equitativo (Jain & Chhabra, 2022). En un informe reciente de la Comisión Europea (2022), se subraya que las políticas públicas deben apoyarse en la creación de infraestructuras digitales de calidad, la provisión de dispositivos accesibles y la formación digital de todos los actores involucrados en el proceso educativo, para garantizar la igualdad de oportunidades.

Es innegable la integración de tecnologías emergentes en la educación superior y su enorme potencial para transformar los procesos educativos a nivel global. Sin embargo, la relación entre estos recursos y el aprendizaje de los estudiantes es una pieza clave para los análisis más particulares sobre el proceso de construcción de aprendizajes asistido con tecnologías (González et al., 2021; Lo, 2023). Las formas en las que esto ocurre, y, sobre todo, las eventuales limitaciones o consideraciones que es necesario salvaguardar para promover aprendizajes de calidad corresponden al foco central de este trabajo. Específicamente, analizar cómo las características de ChatGPT como apoyo al aprendizaje podría verse potenciado cuando los estudiantes que aprenden, estiman los eventuales sesgos o limitaciones de dichos apoyos, a fin de optimizar esta herramienta como un recurso para el aprendizaje consistente con aspectos éticos, habilidades y por cierto las competencias digitales que se requerirán a futuro en estudiantes y docentes (García-Gutiérrez, 2013; Silva y Miranda, 2020).

2. Antecedentes teóricos

La integración de tecnologías en la educación superior ha traído consigo tanto oportunidades como desafíos, los cuales deben ser abordados de manera estratégica para garantizar su efectividad y equidad. En primer lugar, uno de los principales desafíos es la desigualdad en el acceso a tecnologías. Según el informe de la UNESCO (2020), una gran cantidad de estudiantes en diferentes partes del mundo aún carecen de acceso a infraestructuras tecnológicas adecuadas, como dispositivos electrónicos y conectividad a internet, lo que crea una brecha digital. Esta brecha no solo limita las oportunidades de aprendizaje, sino que también perpetúa las desigualdades sociales, ya que los estudiantes de contextos más desfavorecidos no tienen acceso a los mismos recursos educativos. De esta forma debe garantizarse el acceso equitativo a las tecnologías, independientemente del contexto socioeconómico de los estudiantes (Zhang et al., 2023).

Este desafío de acceso plantea la necesidad urgente de políticas públicas que promuevan la infraestructura digital inclusiva. Frente a esto, la importancia de crear ecosistemas digitales que sean accesibles para todos los estudiantes, particularmente aquellos de zonas rurales o con menos recursos (OECD, 2023). Además, la implementación de tecnologías debe ir acompañada de un fortalecimiento de las habilidades digitales de los docentes y estudiantes, lo que requiere una inversión significativa en formación y capacitación continua. La preparación digital de los docentes es crucial para el éxito de las iniciativas tecnológicas, ya que un uso adecuado de las herramientas tecnológicas puede mejorar significativamente la calidad educativa, mientras que un uso inadecuado puede generar más desigualdades (Hernández y Pérez, 2022).

Además, la incorporación de tecnologías plantea retos pedagógicos, ya que no solo se trata de enseñar a utilizar herramientas digitales, sino también de repensar los enfoques pedagógicos tradicionales. En este sentido, el modelo educativo debe ser flexible y adaptativo

para incorporar metodologías activas como el aprendizaje basado en proyectos (ABP) y la enseñanza híbrida, que promuevan el pensamiento crítico y el aprendizaje autónomo (Sacavino & Candau, 2022). Por tanto, la educación superior debe ser capaz de integrar tecnologías de manera que los estudiantes desarrollen competencias digitales que les permitan no solo interactuar con herramientas tecnológicas, sino también comprender y analizar la información de manera crítica (O'Malley y Cheng, 2023). De esta forma, los estudiantes pueden usar las tecnologías no solo como consumidores de información, sino también como creadores y transformadores del conocimiento.

No obstante, las tecnologías emergentes, como la IA y el aprendizaje automático, también ofrecen enormes oportunidades para mejorar el rendimiento académico, las variables socio-afectivas asociadas y las tendencias a desarrollar pensamiento de orden superior (Deng et al., 2025). La IA, por ejemplo, permite la personalización del aprendizaje a través de algoritmos que analizan el comportamiento y las necesidades de cada estudiante, ofreciendo recursos adaptados a su ritmo y estilo de aprendizaje (Liu et al., 2022). Esto puede ser especialmente beneficioso en entornos con estudiantes con diferentes niveles de habilidades y conocimientos previos. La IA puede apoyar a los docentes en la creación de contenidos educativos más dinámicos y en la evaluación continua del progreso de los estudiantes, lo que permite una intervención temprana cuando se detectan dificultades y en la autorregulación del aprendizaje (Wang et al., 2021).

Sin embargo, aun cuando los beneficios son diversos en el uso de la IA y otras tecnologías avanzadas son evidentes, es importante tener en cuenta los aspectos éticos y de privacidad en un sentido amplio de impacto (Mendoza, 2021). En lo específico, el uso de IA en la educación superior plantea preocupaciones relacionadas con la recolección y el uso de datos personales de los estudiantes. El manejo ético de los datos, la transparencia de los algoritmos utilizados y la protección de la privacidad son aspectos clave que deben ser cuidadosamente gestionados. En este sentido, la colaboración entre instituciones educativas, gobiernos y empresas tecnológicas es esencial para garantizar que las tecnologías se utilicen de manera ética y responsable, evitando posibles abusos y protegiendo los derechos de los estudiantes (Binns, 2023).

Por último, otro desafío significativo es la resistencia al cambio, tanto por parte de los docentes como de las instituciones (Patel et al., 2022). Muchas universidades enfrentan una resistencia estructural y cultural a la integración de nuevas tecnologías en los procesos de enseñanza y aprendizaje (Villegas-José y Delgado-García, 2024). Para superar estos obstáculos, es fundamental que las políticas educativas fomenten una cultura de innovación pedagógica, apoyada por recursos financieros y humanos, que permita a los docentes experimentar con nuevas tecnologías sin temor al fracaso (Sánchez, 2024). Junto a lo anterior, el desafío inminente desde hace algunos años -consiste y consistirá- en que las universidades puedan adoptar enfoques de cambio organizacional que promuevan la colaboración, el aprendizaje continuo y la flexibilidad institucional para facilitar la integración exitosa de las tecnologías (Comisión Europea, 2018).

3. Formación Universitaria: Capacitación en Herramientas de IA como ChatGPT

La integración de herramientas basadas en IA, como ChatGPT, en la formación universitaria actual es fundamental (Chiu et al., 2023; García-Peñalvo et al., 2023; Wang et al., 2024) así como sus implicancias éticas (Flores-Vivar y García-Peñalvo, 2022; Nguyen, et al., 2023) para preparar a los estudiantes para el futuro digital (Romero-Rodríguez et al., 2023). La capacitación en el uso de estas tecnologías no solo mejora la eficiencia en la realización de tareas académicas, sino que también fomenta el desarrollo de habilidades digitales esenciales para la vida profesional (Jackman et al., 2021; Livingstone et al., 2023). De acuerdo con el informe de la OECD (2023b), la educación superior debe abordar la creciente demanda de habilidades tecnológicas a través de la integración de IA en sus currículos, proporcionando a los estudiantes los conocimientos necesarios para usar estas herramientas de manera ética y efectiva. En este sentido, se debe enfatizar la importancia de promover una comprensión crítica sobre las ventajas y limitaciones de la IA, ya que su uso indebido podría llevar a una confianza excesiva en los resultados generados sin un análisis reflexivo (Walter, 2024). Es crucial que los programas académicos incluyan cursos que no solo enseñen a los estudiantes a usar estas herramientas, sino que también los eduquen sobre los aspectos éticos, sociales y técnicos de su utilización (Pérez et al., 2023).

Las tecnologías como ChatGPT tienen el potencial de transformar el aprendizaje en la educación superior, ofreciendo ventajas significativas tanto para estudiantes como para docentes (Lo, 2023). Dentro de las ventajas más destacadas podemos mencionar su capacidad para personalizar la enseñanza y enfocarse en un apoyo preciso para cada estudiante (García, 2023). Un segundo aspecto derivado de la ventaja anterior es la capacidad para adaptar el contenido educativo a las necesidades y ritmos de aprendizaje individuales, lo que permite a los estudiantes avanzar a su propio paso y recibir respuestas a preguntas en tiempo real, mejorando su experiencia educativa (Liu et al., 2023). En tercera instancia, ChatGPT cuenta con la ventaja de servir como una herramienta para la redacción, la investigación y la creación de contenido académico, mejorando la eficiencia de los estudiantes al proporcionarles resúmenes rápidos y sugerencias sobre cómo abordar temas complejos (Chiappe et al., 2025). Respecto de esto último, el uso de IA como ChatGPT en tareas de investigación permitiría a los estudiantes acceder a información más rápidamente, lo que facilitaría un proceso de aprendizaje más dinámico y profundo (Jenkins et al., 2022). De esta forma, la posibilidad de automatización de tareas repetitivas en ChatGPT permitiría que los estudiantes puedan centrarse en actividades cognitivamente más desafiantes, como el análisis crítico y la síntesis de información, dejando las tareas más repetitivas a la IA (Chang et al., 2023). Como es posible de asociar, estas ventajas corresponden a las tareas predominantes que realizan los estudiantes durante -al menos- sus primeros años de formación universitaria.

4. Riesgos de Dependencia: Consecuencias para el Pensamiento Crítico y la Autonomía

A pesar de las diversas ventajas que ofrece ChatGPT, uno de los principales riesgos asociados; eventualmente consideradas desventajas para los procesos formativos (Kuleto et al., 2021) consiste en la dependencia excesiva de la tecnología, lo que puede tener efectos adversos en el desarrollo de habilidades críticas e independientes; sobre todo, a la eventual disminución de la capacidad de cuestionar la información y desarrollar habilidades de investigación autónoma (Smith y Johnson, 2023). Este fenómeno, conocido como 'pensamiento de confirmación', ocurre cuando los estudiantes aceptan sin cuestionar las respuestas proporcionadas por la IA, sin realizar un análisis profundo ni considerar diferentes perspectivas. La dependencia de la IA podría, por tanto, llevar a una pérdida de la capacidad para realizar investigaciones originales o para desarrollar habilidades de pensamiento crítico esenciales en

el entorno académico y profesional (Zhai et al., 2024). Además, tal dependencia puede generar un desajuste en la preparación de los estudiantes para enfrentar problemas complejos que requieren un enfoque más independiente y creativo. Junto a ello, un segundo riesgo se refiere a la relación posible entre desinformación y uso de IA; aspecto concomitante al riesgo anteriormente desarrollado. Aprender con la IA implica lograr niveles de certezas fundadas (uso del pensamiento crítico) y en donde es necesario lograr niveles de uso avanzados con estas tecnologías que permitan la detección temprana de la desinformación en lugar de buscar soluciones completamente automatizadas y eventualmente erróneas (Montoro-Montarroso et al., 2023).

Conforme a lo anterior, UNESCO (2023) ya mencionó que la educación debe equilibrar el uso de tecnologías avanzadas con el desarrollo de habilidades cognitivas que promuevan la autonomía y el pensamiento crítico, para evitar que la tecnología sustituya el juicio humano en los procesos de aprendizaje o induzca errores por desconocimiento. Estos aspectos, están en franco desarrollo. Sin embargo, ¿de qué manera podría la IA afectar la capacidad reflexiva del estudiante?, o más bien, ¿Qué tipo de sustitución o criterios implícitos podría generar la IA sin ser detectada en el trabajo de los estudiantes? frente a estas preguntas, es que cobra relevancia el análisis de los sesgos en ChatGPT desde una perspectiva psicológico-educativa.

5. ChatGPT y Sesgos

El sesgo en la inteligencia artificial (IA) se refiere a la tendencia de los sistemas de IA a producir resultados que reflejan prejuicios o parcialidades presentes en los datos con los que fueron entrenados, lo que puede conducir a decisiones injustas o discriminatorias. Este fenómeno puede manifestarse en diversas áreas, como el reconocimiento facial, donde se ha observado que algunos sistemas presentan mayores tasas de error al identificar a personas de piel más oscura, debido a un entrenamiento basado predominantemente en imágenes de personas de piel clara. Además, estudios han demostrado que las personas pueden heredar los sesgos de la IA, adoptando decisiones influenciadas por los prejuicios presentes en los algoritmos. Por lo tanto, es crucial abordar y mitigar estos sesgos en el desarrollo y despliegue de tecnologías de IA para garantizar resultados más justos y equitativos (Ramírez, 2023).

De esta forma, los sesgos en los modelos de IA como ChatGPT son errores implícitos y sistemáticos que emergen de los datos de entrenamiento y los algoritmos de procesamiento que se definen en dicho modelo, afectando sus parámetros, y, por ende, la representación de los datos que este modelo genera. Esto, debido a que estos sistemas utilizan técnicas de aprendizaje automático, en particular el aprendizaje profundo, para identificar e imitar patrones, estilos y estructuras que se encuentran en los datos de entrada del modelo y que eventualmente son transferidos por los mecanismos con los que dichos sistema se entrenan (Baidoo-Anuand y Owusu, 2023). De esta forma, ocurriría una transferencia de sesgo desde su diseño o bien, en el proceso de entrenamiento y del método de refinamiento o ajuste del modelo establecido, en donde los creadores humanos podrían transferir dichos patrones (Von Krogh et al., 2023).

Estos sesgos en la génesis de los modelos pueden estar relacionados con la selección de los datos, el diseño algorítmico y la interpretación de las entradas de los usuarios. Por tanto, los sesgos en los modelos de IA pueden reflejar estereotipos y prejuicios presentes en los datos históricos con los que los modelos son entrenados, lo que a menudo resulta en respuestas que perpetúan desigualdades y parcialidades (Binns, 2022). En específico, el entrenamiento de ChatGPT se basa en grandes corpus de texto extraídos de diversas fuentes de internet, las cuales pueden incluir sesgos de género, raza, clase social, y otros factores que no siempre están equilibrados o son representativos de todas las culturas y perspectivas (Günther et al.,

2023). Estos sesgos pueden influir en la calidad y la imparcialidad de las respuestas generadas, ya que el modelo puede reflejar, sin intención, las preconcepciones y los sesgos presentes en los textos de entrenamiento, resultando en respuestas potencialmente sesgadas que afectan la fiabilidad y equidad del sistema (Vincent, 2023).

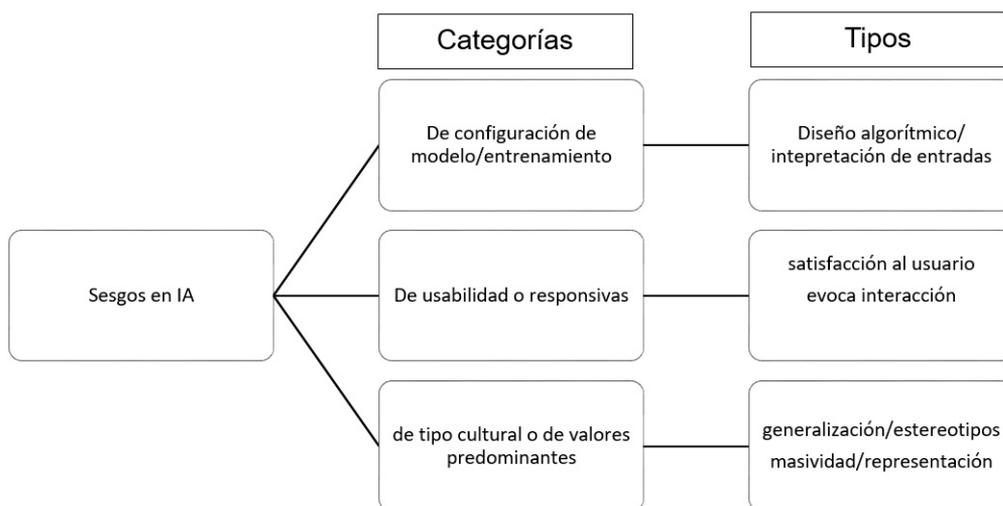
Otro tipo de sesgo, más bien de tipo cultural, social o respecto de la procedencia de los estudiantes, se ve influenciado, por ejemplo, por aquellos que pertenecen a sectores minoritarios o marginados. Esto, debido a que, en dichos casos, es posible apreciar que tienden a ser acusados mayormente de violar normas de integridad académica al usar IA en comparación a estudiantes no pertenecientes a estas características (Farrelly & Baker, 2023). De ahí, la importancia de generar análisis de los sesgos sobre las características de los estudiantes desde una perspectiva ética de la IA, que priorice la alfabetización y empoderamiento de los estudiantes en situación de desventaja, bajo un enfoque -por qué no- etnográfico sobre su uso (Michel-Villarreal et al., 2023).

Finalmente, un tercer tipo de sesgo documentado obedece a criterios más bien de usabilidad. Esto se refiere al uso directamente ejercido por un usuario sobre la IA ya que dependiendo de las necesidades que se esperan resolver y las eventuales solicitudes académicas, en este contexto es que la usabilidad obedece más bien a la posibilidad de automatizar tareas de aprendizaje, aumentar productividad y eficiencia; así como contribuir al aprendizaje adaptativo (Gligorea et al., 2023). Sin embargo, hay limitaciones o sesgos en este aspecto referidos a inconsistencias fácticas, eventuales dificultades para el desarrollo de la comprensión profunda y problemas evidentes de seguridad; todos aspectos importantes de considerar en el contexto educativo (Elbanna & Armstrong, 2024).

Conforme a lo anterior y desde una perspectiva más explicativa conforme a los alcances de dichos sesgos para contexto educativo, es que a continuación, se expresa esquemáticamente tres tipos de sesgos asociados a ChatGPT producto del análisis documental realizado, a fin de analizar sus alcances en este escenario.

Figura 1

Categorías y tipos de sesgos en la IA.



Fuente: Elaboración propia.

Los sesgos asociados a la categoría configuración/entrenamiento de las IA pueden originarse en varias etapas del desarrollo y operación de los modelos. Una de las principales fuentes de sesgo es la disponibilidad de datos, que puede ser limitante o no representativa de la diversidad de contextos y perspectivas. De acuerdo con [Mehrabi et al. \(2022\)](#), cuando los conjuntos de datos usados para entrenar IA como ChatGPT no son lo suficientemente diversos o contienen sesgos implícitos, esto puede afectar la capacidad del modelo para generalizar de manera justa y precisa. Además, la confirmación implícita es un fenómeno en el que los modelos tienden a generar respuestas que refuerzan patrones preexistentes o suposiciones previas, lo que podría perpetuar creencias erróneas y reforzar la parcialidad en el aprendizaje automático ([Zhou et al., 2023](#)). Por ejemplo, si los datos utilizados para entrenar a ChatGPT incluyen predominantemente opiniones tecnocéntricas, esto puede hacer que el modelo favorezca soluciones basadas en tecnología sin considerar enfoques más holísticos o interdisciplinarios ([O'Neil, 2022](#)). Estos sesgos configuracionales pueden ser sutiles, pero tienen un impacto significativo en las respuestas y en cómo los usuarios perciben la objetividad de la IA ([Binns, 2022](#)).

Además de los sesgos presentes en los datos y la configuración, los sesgos en la categoría usabilidad o responsivas también son una preocupación central. Los sesgos de respuesta pueden tomar varias formas, entre ellas, el sesgo de optimismo, donde el modelo tiende a generar respuestas excesivamente positivas o esperanzadoras, incluso cuando no son del todo precisas o realistas ([Zhang et al., 2022](#)). Otro sesgo común es el sesgo de autoridad, en el que ChatGPT asume que sus respuestas son definitivas o autoritativas, lo que podría llevar a los usuarios a confiar ciegamente en la información proporcionada sin cuestionarla ([Hao, 2022](#)). Además, el sesgo de simplificación se refiere a la tendencia del modelo a ofrecer respuestas simplificadas que omiten matices y detalles importantes, lo que puede resultar en una visión parcial o incompleta de los temas tratados ([Vincent, 2023](#)). Por último, el sesgo de selección de información puede influir en las respuestas de ChatGPT, ya que el modelo podría priorizar ciertas fuentes o perspectivas sobre otras, dependiendo de la estructura de los datos con los que ha sido entrenado ([Günther et al., 2023](#)). Estos sesgos de respuesta pueden distorsionar la calidad y la integridad de la información proporcionada, afectando la capacidad del modelo para ofrecer una asistencia precisa y equitativa.

Finalmente, la categoría referida al sesgo cultural en la inteligencia artificial (IA) se refiere a la tendencia de los sistemas de IA a reflejar y perpetuar los prejuicios y estereotipos presentes en los datos culturales dominantes con los que fueron entrenados; debido a que no pueden reconocer las diversidades culturales en las respuestas, los sistemas de IA pueden contribuir a la perpetuación de estereotipos y a la falta de inclusión ([Liu et al., 2023b](#)). Además, este sesgo puede dificultar el acceso a una educación verdaderamente global, ya que los usuarios de contextos no occidentales pueden enfrentar dificultades para encontrar información que sea relevante y adecuada a su entorno cultural y social ([Hao, 2022](#)). Este sesgo cultural/valórico en los modelos de IA es otro aspecto crítico a considerar, especialmente dado que los sistemas como ChatGPT operan en un contexto globalizado. Según el informe de la [UNESCO \(2023\)](#), los modelos de IA como ChatGPT a menudo están entrenados en textos que reflejan perspectivas y valores dominantes, típicamente de sociedades occidentales, lo que puede llevar a respuestas que no consideren adecuadamente las particularidades culturales de diferentes regiones o contextos. Este sesgo cultural puede ser especialmente problemático en aplicaciones educativas y de información, donde los estudiantes o usuarios de diversas culturas pueden sentirse marginados o mal representados. Al no reconocer las diversidades culturales en las respuestas, los sistemas de IA pueden contribuir a la perpetuación de estereotipos y a la falta de inclusión ([Liu et al., 2023](#)). Además, este sesgo puede dificultar el acceso a una educación verdaderamente global, ya que los usuarios de contextos

no occidentales pueden enfrentar dificultades para encontrar información que sea relevante y adecuada a su entorno cultural y social (Hao, 2022). Por lo tanto, es crucial desarrollar modelos de IA que sean conscientes de estas diferencias culturales y que se ajusten para ser más inclusivos y representativos de la diversidad global (Zhou et al., 2023).

6. Consecuencias de los sesgos en ChatGPT

Conforme a los sesgos detallados anteriormente, es innegable pensar en las eventuales consecuencias que esto podría generar en los procesos formativos o de aprendizaje en los estudiantes. A continuación, se describen tres grandes consecuencias derivadas de estos sesgos. En primera instancia, los sesgos presentes en los modelos de IA, como ChatGPT, pueden tener un impacto significativo en la forma en que los estudiantes perciben y comprenden ciertos temas. Según Noble (2022), los sesgos implícitos en las respuestas generadas por la IA pueden distorsionar la información proporcionada, lo que puede influir negativamente en el desarrollo del pensamiento crítico de los estudiantes. Si los estudiantes dependen demasiado de la información proporcionada por ChatGPT sin cuestionarla, podrían internalizar ideas erróneas o inexactas, lo que afectaría su capacidad para pensar de manera independiente y reflexiva. Este fenómeno es especialmente crítico en el ámbito académico, donde se fomenta la capacidad de análisis y reflexión profunda (Zhang et al., 2023). Además, los sesgos en la información proporcionada por la IA pueden reforzar creencias preconcebidas, limitando la capacidad de los estudiantes para desarrollar una comprensión matizada y crítica de los temas tratados (Günther et al., 2023). Como resultado, los estudiantes pueden quedar atrapados en una visión distorsionada de los temas, lo que obstaculiza su aprendizaje y su capacidad para abordar problemas complejos de manera eficaz.

En segunda instancia, un riesgo significativo asociado a los sesgos en ChatGPT es la reproducción y amplificación de estereotipos de género, raza, o cultura. Los modelos de IA, al ser entrenados con grandes cantidades de datos de texto, pueden replicar patrones de prejuicio que están presentes en la sociedad. Según O'Neil (2022), los modelos de IA como ChatGPT pueden aprender y reforzar estereotipos, reproduciendo en sus respuestas sesgos históricos sobre género, etnia y otras características sociales. Por ejemplo, si el modelo está entrenado en textos que contienen un sesgo de género, puede generar respuestas que refuerzan los roles tradicionales de género, como la representación de mujeres en ciertos campos profesionales de manera estereotipada. De acuerdo con Mehrabi et al. (2022), este tipo de sesgo no solo afecta la calidad de la información, sino que también contribuye a la perpetuación de desigualdades y estigmas en la sociedad. En el contexto educativo, esto puede afectar la forma en que los estudiantes se relacionan con temas diversos, limitando su capacidad para explorar y comprender diversas perspectivas culturales y sociales. La amplificación de estos estereotipos podría influir en la manera en que los estudiantes se posicionan en torno a cuestiones de identidad y diversidad, lo que es particularmente problemático en la educación superior, donde se busca promover un pensamiento inclusivo y plural (Liu et al., 2023). La posibilidad de que la IA pueda reivindicar los derechos de los grupos pertenecientes a grupos socioculturales minoritarios se transforma en un desafío importante para el desarrollo de la IA; sobre todo frente a las evidencias asociadas a la mejora educativa en el rendimiento de dichos grupos, la motivación asociada al aprendizaje y su efecto en la sociabilidad activa de dichos grupos (Salas-Pilco et al., 2022). Desde este punto de vista, son los mismos algoritmos computacionales los que debieran disminuir el racismo sistémico en su operar, a fin de contribuir a la gobernanza digital conforme a los requerimientos de la política pública actual (Fountain, 2022).

Como tercer elemento, la influencia de los sesgos en ChatGPT también tiene consecuencias significativas en la equidad educativa. Los estudiantes provenientes de diferentes contextos culturales, lingüísticos o socioeconómicos pueden verse desfavorecidos por las respuestas sesgadas generadas por la IA. Según Liu et al. (2023), los modelos de IA pueden tender a ofrecer respuestas que reflejan una visión homogénea, basada en datos predominantemente occidentales o de contextos más privilegiados. Esto puede dar lugar a una falta de representación de diversas perspectivas, exacerbando las desigualdades preexistentes entre estudiantes de diferentes orígenes (Zhou et al., 2023). Además, el sesgo en la IA puede contribuir a que los estudiantes de entornos menos favorecidos no encuentren respuestas que se alineen con sus contextos o que ignoren problemas específicos de sus realidades, como las barreras lingüísticas o la falta de acceso a recursos educativos (Binns, 2022). Esto puede afectar negativamente su experiencia educativa y crear una brecha más amplia entre los estudiantes que tienen acceso a información personalizada y aquellos que dependen de sistemas de IA que no consideran sus necesidades y circunstancias particulares. En términos de equidad educativa, es fundamental que las herramientas como ChatGPT se diseñen para ser inclusivas y sensibles a la diversidad de experiencias y contextos, a fin de garantizar que todos los estudiantes tengan las mismas oportunidades de aprendizaje y desarrollo (Vincent, 2023).

7. Mitigación de sesgos en ChatGPT. Técnicas para la formación universitaria de calidad

Luego de presentar el escenario educativo actual y la veloz influencia de la IA en los procesos de aprendizaje, es que es necesario proponer algunas alternativas que permitan contrabalancear los sesgos presentes en ChatGPT o en otros recursos de IA. Para ello, se describen cinco amplias estrategias, cuyo sentido principal son desarrollar en los estudiantes (y en los docentes) competencias específicas de uso con IA, que permitan equilibrar estos sesgos a fin de por una parte, comprender los eventuales sesgos, sus tipos y características; así como utilizar herramientas que metodológicamente reduzcan estos efectos en los procesos de interacción con la IA al momento de avanzar en el aprendizaje personalizado, tanto de estudiantes como docentes (Marín, 2022).

Como primera estrategia para mitigar los sesgos en ChatGPT es utilizar el entrenamiento con datos diversos y balanceados que representen una amplia gama de perspectivas culturales, sociales, lingüísticas y económicas. El sesgo en los modelos de IA surge principalmente cuando los datos de entrenamiento son homogéneos o no reflejan adecuadamente las realidades diversas de los usuarios (Binns, 2022; Mehrabi et al., 2022). Para reducir estos sesgos, es necesario incorporar datos representativos de distintos grupos, incluidas diversas etnias, géneros, orientaciones sexuales, y contextos socioeconómicos. Según el informe de la UNESCO (2023), un enfoque que promueva la inclusión y diversidad de datos no solo mejora la precisión y relevancia de las respuestas, sino que también contribuye a la equidad educativa, asegurando que los modelos no excluyan o discriminen a ciertos grupos. Implementar un enfoque más inclusivo y representativo en la recopilación de datos de entrenamiento permitirá generar respuestas que reflejen mejor las realidades y experiencias globales, ayudando a evitar sesgos y promoviendo una interacción más justa y respetuosa con los usuarios (Liu et al., 2023).

La segunda estrategia, se refiere al uso de filtros de moderación para mitigar los sesgos en ChatGPT. Estos filtros permiten controlar y corregir respuestas que puedan ser inapropiadas o sesgadas, funcionando como una segunda capa de verificación. Los filtros pueden detectar respuestas que reproduzcan estereotipos, contenidos ofensivos o cualquier tipo de sesgo implícito. Según Binns (2022), los filtros de moderación deben ser diseñados de manera que no solo eliminen el lenguaje ofensivo, sino que también identifiquen y neutralicen respuestas que puedan estar influenciadas por sesgos culturales, de género o raciales. Además, los filtros

deben ser constantemente actualizados y ajustados para adaptarse a nuevas formas de sesgo que puedan surgir a medida que los modelos continúan siendo entrenados con datos adicionales (Zhou et al., 2023). Una moderación efectiva también debe garantizar que el modelo mantenga un equilibrio entre evitar sesgos y permitir una diversidad de opiniones, asegurando que las respuestas proporcionadas sean tanto precisas como respetuosas (Binns, 2022).

Una tercera estrategia para mitigar los sesgos en ChatGPT es a través de la adaptación al contexto específico del usuario. Este enfoque implica diseñar el modelo para que pueda personalizar sus respuestas de acuerdo con las necesidades y el contexto cultural del usuario. Según el estudio de Liu et al. (2023), permitir que ChatGPT adapte sus respuestas a diferentes contextos no solo mejora la relevancia y la aplicabilidad de la información, sino que también ayuda a reducir el riesgo de sesgos al ofrecer respuestas más contextualizadas y sensibles a las características particulares del usuario. Esto podría incluir ajustar el tono, la complejidad del lenguaje o la perspectiva cultural, asegurando que los usuarios de diversos orígenes reciban respuestas que sean comprensibles y apropiadas para su contexto (Mehrabi et al., 2022). La adaptación al contexto también fomenta una experiencia más inclusiva, donde los usuarios no se sienten alienados por respuestas que no se alinean con sus realidades o identidades.

Como cuarta estrategia de mitigación de los sesgos, emerge el diseño para la neutralidad y claridad. Un diseño centrado en la neutralidad garantiza que las respuestas generadas por el modelo sean equilibradas y no favorezcan un punto de vista particular ni simplifiquen excesivamente los temas complejos (Zhang et al., 2022). Es crucial que ChatGPT ofrezca respuestas claras y comprensibles, evitando respuestas autoritarias que puedan parecer definitivas o dogmáticas. Según el informe de la UNESCO (2023), la capacidad de generar respuestas claras y bien fundamentadas es esencial para aumentar la confianza de los usuarios en la herramienta, permitiéndoles evaluar y utilizar la información de manera crítica. Además, un diseño neutral fomenta la transparencia en las respuestas generadas, lo que permite que los usuarios comprendan cómo se formulan las respuestas y qué fuentes o datos están influyendo en ellas (O'Neil, 2022). Esta claridad y transparencia no solo son importantes para mantener la confianza en la herramienta, sino también para asegurar que los usuarios tengan una experiencia más enriquecedora y objetiva.

Finalmente, la quinta estrategia para mitigar los eventuales sesgos y sobre todo para garantizar que los estudiantes hagan un uso responsable de ChatGPT y otras herramientas de IA, es fundamental proporcionarles una educación sólida sobre el uso crítico de estas tecnologías. Según Binns (2022) y Liu et al. (2023), los estudiantes deben ser instruidos no solo en cómo utilizar las herramientas de IA, sino también en cómo cuestionar y evaluar la información proporcionada por estos modelos. Esta educación crítica es esencial para evitar la dependencia excesiva de las respuestas de la IA y fomentar el desarrollo de habilidades de pensamiento crítico y análisis independiente. La enseñanza sobre la ética de la IA, los sesgos y la fiabilidad de las respuestas puede ayudar a los estudiantes a tomar decisiones informadas sobre cuándo y cómo utilizar herramientas como ChatGPT (Zhou et al., 2023). Formar a los estudiantes en el uso crítico de la IA no solo mejora su competencia tecnológica, sino que también promueve una comprensión más profunda de los problemas sociales y éticos que implica el uso de estas tecnologías, lo que es esencial para su desarrollo académico y profesional (Günther et al., 2023).

8. Conclusiones

La interacción entre la IA y la educación superior se perfila como un eje transformador de los procesos formativos. En un futuro cercano, herramientas como ChatGPT seguirán evolucionando hacia sistemas más personalizados y adaptativos, facilitando el aprendizaje autónomo y complementando el trabajo docente en la medida en que sus atributos se vinculen con la autonomía, motivación y otros factores psicoeducativos más fehacientemente. No obstante, este avance requiere que las instituciones de educación superior sean capaces de integrar la IA no solo como un recurso tecnológico, sino como un componente fundamental en la planificación curricular, asegurando que responda a las necesidades de los estudiantes y fomente competencias clave como el pensamiento crítico y la creatividad. La IA se convertirá en un agente mediador que amplifique las posibilidades de aprendizaje, aunque también demandará una constante actualización por parte de los docentes y gestores educativos para maximizar su impacto.

ChatGPT, como ejemplo de IA generativa, presenta diversos tipos de sesgos, como aquellos derivados de los datos de entrenamiento (sesgos culturales, de género o lingüísticos) y los inherentes a la forma en que la herramienta prioriza y genera información. Estos sesgos pueden afectar negativamente el aprendizaje al perpetuar visiones parciales, inexactas o incluso erróneas sobre determinados temas, lo que podría limitar la capacidad de los estudiantes para desarrollar un entendimiento global y crítico. En contextos educativos, la exposición repetida a respuestas sesgadas puede reforzar estereotipos o desinformación, lo que subraya la necesidad de fomentar una interacción supervisada y reflexiva con la IA para mitigar estos efectos.

Por otra parte, la mitigación de los sesgos en herramientas de IA como ChatGPT requiere un enfoque dual: técnico y educativo. Desde un punto de vista técnico, es fundamental mejorar los algoritmos de entrenamiento y diversificar las bases de datos utilizadas, mientras que en el ámbito educativo se debe trabajar en el desarrollo de competencias digitales tanto en estudiantes como en docentes. Estas competencias incluyen la capacidad de identificar posibles sesgos, validar información a través de fuentes confiables y utilizar la IA como una herramienta complementaria y no como una fuente única de conocimiento. En este sentido, las instituciones de educación superior juegan un rol esencial al promover alfabetización digital avanzada y enfoques éticos en el uso de la tecnología.

El estudio de los sesgos en la IA y su impacto en la educación superior abre múltiples líneas de investigación futuras. Entre ellas, destaca la necesidad de analizar cómo diferentes contextos socioculturales condicionan la percepción y el uso de la IA en entornos educativos, así como la efectividad de estrategias pedagógicas para integrar herramientas como ChatGPT de manera ética y responsable. Asimismo, será crucial investigar la evolución de las tecnologías de IA generativa y su capacidad para adaptarse a necesidades formativas específicas, lo que podría transformar paradigmas educativos tradicionales. Finalmente, los estudios futuros deberían explorar modelos de gobernanza y regulación que aseguren un uso equitativo y accesible de estas herramientas, con miras a consolidar una educación superior de calidad para todos, bajo un compromiso de la sociedad en su conjunto.

Referencias

- Baidoo-Anu, D. & Owusu, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of IA*, 7(1), 52-62, <http://dx.doi.org/10.2139/ssrn.4337484>.
- Beetham, H., & Sharpe, R. (2013). *Rethinking learning in the digital age: A framework for the future of education*. Routledge.
- Binns, R. (2022). The challenges of bias in AI systems: A systematic review. *Journal of Artificial Intelligence Research*, 45(4), 285-310. <https://doi.org/10.1007/s10115-022-01521-w>.
- Binns, R. (2023). Ethical considerations in the use of artificial intelligence in education. *International Journal of Educational Technology*, 46(2), 131-144. <https://doi.org/10.1080/17439884.2023.2047789>.
- Comisión Europea (2018). *Educación y formación 2020: La dimensión europea en la educación superior*. <https://ec.europa.eu/education-policies>.
- Chang, A., Kim, H., & Lee, Y. (2023). Artificial Intelligence and learning: A comprehensive review. *Journal of Educational Technology*, 35(2), 157-175. <https://doi.org/10.1016/j.jedtech.2023.01.005>.
- Chiappe, A., Sanmiguel, C., & Sáez Delgado, F. M. (2025). IA generativa versus profesores: reflexiones desde una revisión de la literatura. *Pixel-Bit. Revista De Medios Y Educación*, 72, 119–137. <https://doi.org/10.12795/pixelbit.107046>.
- Chiu, T., Xia, Q., Zhou, X., Chai, CH., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, <https://doi.org/10.1016/j.caeai.2022.100118>.
- Chou, P., Chen, W., & Wang, C. (2022). Exploring the use of flipped classrooms for enhancing student engagement in the digital era. *Computers & Education*, 178, 104393. <https://doi.org/10.1016/j.compedu.2021.104393>.
- Deng, R., Jiang M., Yu X., Lu Y. & Liu S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 277, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>.
- Echeverría, J. & Almendros, L. S. (2022). Transformación digital y educación inclusiva. *Revista Institucional | UPB*, 61(161), 137–153. <https://revistas.upb.edu.co/index.php/revista-institucional/article/view/8105>.
- Elbanna, S. & Armstrong, L. (2024). "Exploring the integration of ChatGPT in education: adapting for the future". *Management & Sustainability: An Arab Review*, 3(1), 16-29. <https://doi.org/10.1108/MSAR-03-2023-0016>.
- Farrelly, T. & Baker, N. (2023). Generative Artificial Intelligence: Implications and Considerations for Higher Education Practice. *Education Science*, 13, 1109. <https://doi.org/10.3390/educsci13111109>.
- Flores-Vivar, J. y García-Peñalvo, F. (2022). *Reflexiones sobre la ética, potencialidades y retos de la Inteligencia Artificial en el marco de la Educación de Calidad (ODS4)*, 74, v.XXXI, 37-47, <https://doi.org/10.3916/C74-2023-03>.
- Gallent Torres, C., Zapata González, A., & Ortego Hernando, J. L. (2023). El impacto de la inteligencia artificial generativa en educación superior: una mirada desde la ética y la integridad académica. *RELIEVE - Revista Electrónica De Investigación Y Evaluación Educativa*, 29(2). <https://doi.org/10.30827/relieve.v29i2.29134>.

- García, O. (2023). Uso y percepción de ChatGPT en la educación superior. *Revista De Investigación En Tecnologías De La Información*, 11(23), 98–107. <https://doi.org/10.36825/RITI.11.23.009>.
- García-Gutiérrez, J. (2013). Aproximación ética a la competencia digital. Los niveles de uso y sentido en ámbitos educativos virtuales. *Education in the Knowledge Society (EKS)*, 14(3), 121–145. <https://doi.org/10.14201/eks.11354>.
- García Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 9–39. <https://doi.org/10.5944/ried.27.1.37716>.
- González, Y., Manzano, O. y Torres, M. (2021). Tecnologías disruptivas en educación virtual. *Revista REDIPE*, 10(7), 185-200, <https://dialnet.unirioja.es/servlet/articulo?codigo=8116507>.
- Gligorea, I., Cioca, M., Oancea, R., Gorski, A. -T., Gorski, H., & Tudorache, P. (2023). Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review. *Education Sciences*, 13(12), 1216. <https://doi.org/10.3390/educsci13121216>.
- Günther, M., Reiter, A., & Müller, L. (2023). Bias and fairness in machine learning: Exploring the implications of AI decision-making. *Journal of Computer Science and Technology*, 27(1), 12-28. <https://doi.org/10.1109/JCST.2023.00445>.
- Hao, K. (2022). AI systems and their impact on decision-making: A deeper look at algorithmic biases. *AI & Society*, 37(2), 553-569. <https://doi.org/10.1007/s00146-022-01342-6>.
- Hernández, A., & Pérez, J. (2022). Digital literacy in higher education: The role of teachers in the digital transformation process. *Journal of Education and Technology*, 29(3), 198-213. <https://doi.org/10.1007/s12672-022-00377-6>.
- Jackman, J.A., Gentile, D.A., Cho, N.J. et al. (2021). Addressing the digital skills gap for future education. *Nat Hum Behav* 5, 542–545, <https://doi.org/10.1038/s41562-021-01074-z>.
- Jain, P., & Chhabra, P. (2022). Digital education and the digital divide: Ensuring equal opportunities for students. *Journal of Educational Technology & Society*, 25(4), 65-80. <https://www.jstor.org/stable/48605088>.
- Jenkins, H., Li, Y., & Zhang, T. (2022). AI in higher education: Enhancing learning through artificial intelligence tools. *International Journal of Educational Research*, 101, 54-67. <https://doi.org/10.1016/j.ijer.2022.09.002>.
- Kerman, N., Banihashem, S., Karami, M., Er, E., van Ginkel, S., Noroozi, O. (2023). Online peer feedback in higher education: A synthesis of the literature. *Education and Information Technologies*, 29:763–813 <https://doi.org/10.1007/s10639-023-12273-8>.
- Kim, M., Kim, Y., & Lim, C. (2022). A study on teacher training for digital education: Implications for the future of teaching. *Educational Technology Research & Development*, 70(2), 443-463. <https://doi.org/10.1007/s11423-022-10060-3>.
- Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M. D., Păun, D., & Mihoreanu, L. (2021). Exploring Opportunities and Challenges of Artificial Intelligence and Machine Learning in Higher Education Institutions. *Sustainability*, 13(18), 10424. <https://doi.org/10.3390/su131810424>.
- Labraña, J., y Brunner, J. J. (2022). Transformación de la educación superior latinoamericana y su impacto en la idea de la universidad: Del acceso de élite a la masificación y universalización del acceso. *Perfiles Educativos*, 44(176). <https://doi.org/10.22201/ii-sue.24486167e.2022.176.60539>.

- Liu, M., Wivagg, J., & Noddings, N. (2022). Digital tools in higher education: How teachers are responding to challenges and opportunities. *International Journal of Educational Technology*, 49(1), 100-115. <https://doi.org/10.1007/s10758-021-09576-4>.
- Liu, X., Zhao, M., & Wu, C. (2023). Personalized learning through AI: Opportunities and challenges in higher education. *Educational Technology Research and Development*, 71(4), 123-140. <https://doi.org/10.1007/s11423-023-10001-w>.
- Liu, X., Zhao, M., & Wu, C. (2023b). Understanding cultural bias in AI: Implications for global applications. *International Journal of AI & Education*, 24(2), 144-160. <https://doi.org/10.1007/s40593-023-00308-7>.
- Livingstone, S., Mascheroni, G., & Stoilova, M. (2023). The outcomes of gaining digital skills for young people's lives and wellbeing: A systematic evidence review. *New Media & Society*, 25(5), 1176-1202. <https://doi.org/10.1177/14614448211043189>.
- Lo, C.K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Science*, 13, 410. <https://doi.org/10.3390/educsci13040410>.
- Marín, M. (2022). Digital transformation in higher education: Opportunities and challenges. *Technology, Pedagogy, and Education*, 31(1), 97-113. <https://doi.org/10.1080/1475939X.2022.2072492>.
- Marín-Megía, f.; López-Meneses, E.; Luque-de-la-Rosa, A.; Sánchez-Amate, J. J. (2025). Evolución de las competencias digitales docentes en la educación superior: un análisis de la producción científica. *Campus Virtuales*, 14(2), 35-56. <https://doi.org/10.54988/cv.2025.2.1605>.
- Mehrabi, N., Morstatter, F., Saxena, N., & Lerman, K. (2022). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 55(3), 1-35. <https://doi.org/10.1145/3497322>.
- Mendoza, O. (2021). El derecho de protección de datos personales en los sistemas de inteligencia artificial. *Revista IUS*, 15(48), 179-207, <https://doi.org/10.35487/rius.v15i48.2021.743>.
- Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D., Thierry-Aguilera, R., & Gerardou, F. (2023). Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT. *Education Sciences*, 13(9), 856. <https://doi.org/10.3390/educsci13090856>.
- Miller, T., Phillips, L., & Chen, S. (2023). Collaborative learning and technology integration in higher education. *Learning, Media and Technology*, 48(1), 22-38. <https://doi.org/10.1080/17439884.2022.2076234>.
- Montoro-Montarroso, A., Cantón-Correa, J., Rosso, P., Chulvi, B., Panizo-Lledot, A., Huer-tas-Tato, J., Calvo-Figueras, B., Rementería, M., Gómez-Romero, J. (2023). Fighting disinformation with artificial intelligence: fundamentals, advances and challenges. *Profesional De La información*, 32(3). <https://doi.org/10.3145/epi.2023.may.22>.
- Nguyen, A., Ngo, H.N., Hong, Y. et al. (2023). Ethical principles for artificial intelligence in education. *Educ Inf Technol*, 28, 4221-4241, <https://doi.org/10.1007/s10639-022-11316-w>.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO]. (2020). *Global education monitoring report 2020: Inclusion and education: All means all*. <https://doi.org/10.54676/JJNK6989>.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO]. (2023). *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*. <https://unesdoc.unesco.org/ark:/48223/pf0000366994>.

- Organization for Economic Co-operation and Development [OECD] (2018). *The future of education and skills 2030: OECD report*. OECD Publishing. https://www.oecd.org/con- tent/dam/oecd/en/publications/reports/2018/06/the-future-of-education-and-skill- s_5424dd26/54ac7020-en.pdf.
- Organization for Economic Co-operation and Development [OECD] (2023). *Skills in Latin America: Insights from the Survey of Adult Skills (PIAAC)*. OECD Skills Studies, OECD Publishing, Paris, <https://doi.org/10.1787/5ab893f0-en>.
- Organization for Economic Co-operation and Development [OECD] (2023b). *Country Di- gital Education Ecosystems and Governance: A Companion to Digital Education Outlook 2023*. OECD Publishing, Paris. <https://doi.org/10.1787/906134d4-en>.
- O'Malley, P., & Cheng, R. (2023). Active learning and digital tools: Transforming higher edu- cation for the 21st century. *Learning, Media and Technology*, 48(1), 22-38. <https://doi.org /10.1080/17439884.2022.2076234>.
- O'Neil, C. (2022). *Weapons of math destruction: How big data increases inequality and threa- tens democracy*. Crown Publishing Group.
- Patel, V., Sharma, R., & Chhabra, P. (2022). Overcoming resistance to technology adoption in higher education: A framework for educational institutions. *International Journal of Educational Innovation*, 44(1), 12-27. <https://doi.org/10.1016/j.joe.2022.03.002>.
- Pérez, J., Hernández, L., & Garcia, M. (2023). Integrating AI tools in university curricula: Challenges and solutions. *Journal of Higher Education Innovation*, 9(1), 42-57. <https://doi. org/10.1016/j.jhe.2023.03.006>.
- Ramírez, R. (2023). Sesgos y discriminaciones sociales de los algoritmos en Inteligencia Ar- tificial: una revisión documental. *Entretextos*, 15(39), 1-17. <https://doi.org/10.59057/ibe- roleon.20075316.202339664>.
- Romero-Rodríguez, J., Ramírez-Montoya, M.S., Buenestado-Fernández, M. & Lara-Lara, F. (2023). Use of ChatGPT at University as a Tool for Complex Thinking: Students' Perceived Usefulness. *Journal of new approaches in educational research*, 12(2), 323-339, <https:// doi.org/10.7821/naer.2023.7.1458>.
- Sacavino, S. & Candau, M. (2022). Enseñanza Híbrida: desafíos y potencialidades. *Es- tudios pedagógicos (Valdivia)*, 48(2), 257-266. <https://dx.doi.org/10.4067/S0718- 07052022000200257>.
- Salas-Pilco, S. Z., Xiao, K., & Oshima, J. (2022). Artificial Intelligence and New Technologies in Inclusive Education for Minority Students: A Systematic Review. *Sustainability*, 14(20), 13572. <https://doi.org/10.3390/su142013572>.
- Salmi, J. (2017). The challenge of university transformation: A global perspective. *Higher Edu- cation Policy*, 30(1), 21-38. <https://doi.org/10.1057/s41307-017-0042-4>.
- Sánchez, M. P. (2024). La inteligencia artificial como recurso docente: usos y posibilidades para el profesorado. *Educar 2024*, 60(1) 33-47, <https://doi.org/10.5565/rev/educar.1810>.
- Silva, J., & Miranda, P. (2020). Presencia de la competencia digital docente en los progra- mas de formación inicial en universidades públicas chilenas. *Revista De Estudios Y Expe- riencias En Educación*, 19(41), 149-165. <https://revistas.ucsc.cl/index.php/rexe/article/ view/819>.
- Smith, A., & Johnson, R. (2023). The impact of AI on critical thinking in students: A longitu- dinal study. *Technology, Pedagogy, and Education*, 32(3), 89-104. <https://doi.org/10.1080/ 1475939X.2023.1870227>.

- Trowler, P. (2010). *Student engagement and higher education: Theoretical perspectives and practical approaches*. Routledge.
- Van Damme, D. & D. Zahner (eds.) (2022). *Does Higher Education Teach Students to Think Critically?*. OECD Publishing, Paris. <https://doi.org/10.1787/cc9fa6aa-en>.
- Villarreal-Villa, S., García-Guliany, J., Hernández-Palma, H., y Steffens-Sanabria, E. (2019). Competencias Docentes y Transformaciones en la Educación en la Era Digital. *Formación universitaria*, 12(6), 3-14. <https://dx.doi.org/10.4067/S0718-50062019000600003>.
- Villegas-José, V. y Delgado-García, M. (2024). Inteligencia artificial: revolución educativa innovadora en la Educación Superior. *PIXEL-BIT*, 71, 159-177, <https://doi.org/10.12795/pixelbit.107760>.
- Vincent, J. (2023). AI and the limits of objectivity: Exploring algorithmic bias in language models. *Nature Machine Intelligence*, 5(1), 34-47. <https://doi.org/10.1038/s41586-023-00553-1>.
- Von Krogh, G., Roberson, Q. & Gruber, M. (2023). Recognizing and Utilizing Novel Research Opportunities with Artificial Intelligence. *Academy of Management Journal*, 66(2), <https://doi.org/10.5465/amj.2023.4002>.
- Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *Int J Educ Technol High Educ* 21, 15. <https://doi.org/10.1186/s41239-024-00448-3>
- Wang, Q., Peng, Y. and Wang, H. (2021). A Curation Activity-Based Self-Regulated Learning Promotion Approach as Scaffolding to Improving Learners' Performance in STEM Courses. *Journal of Educational Computing Research*, 60(4). <https://doi.org/10.1177/073563312111056532>.
- Wang, S., Wang, F., Zhu, Z., Wang, J., Du, Z. (2024). *Artificial intelligence in education: A systematic literature review*. *Expert Systems with Applications*, 252, Part A. <https://doi.org/10.1016/j.eswa.2024.124167>.
- Zhai, Ch., Wibowo, S. & Li, L. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(28) <https://doi.org/10.1186/s40561-024-00316-7>.
- Zhang, Y., Chen, T., & Wang, S. (2022). Optimism bias and AI: The role of algorithms in shaping perceptions. *Journal of Artificial Intelligence Ethics*, 4(3), 203-217. <https://doi.org/10.1007/s43681-022-00048-4>.
- Zhou, L., Li, H., & Zhang, W. (2023). Technocentric bias in AI models: A critical analysis of algorithmic fairness. *Journal of AI and Society*, 38(5), 677-693. <https://doi.org/10.1007/s00146-023-01555-0>.
- Zhang, L., Zhao, M., & Wong, R. (2023). Addressing the digital divide in education: A review of policies and practices in developing countries. *Educational Technology Research & Development*, 71(1), 77-90. <https://doi.org/10.1007/s11423-022-10060-3>.



Este trabajo está sujeto a una licencia de Reconocimiento 4.0 Internacional Creative Commons (CC BY 4.0).