

Inteligencia artificial generativa para autoaprendizaje en educación superior: Diseño y validación de una máquina de ejemplos

Generative artificial intelligence for self-learning in higher education: Design and validation of an example machine



- José Carlos Sánchez-Prieto - *Universidad de Salamanca, USAL (España)*
- Vanessa Izquierdo-Álvarez - *Universidad de Salamanca, USAL (España)*
- María Teresa del Moral-Marcos - *Universidad de Salamanca, USAL (España)*
- Fernando Martínez-Abad - *Universidad de Salamanca, USAL (España)*

RESUMEN

La Inteligencia Artificial (IA) generativa, como tecnología emergente y disruptiva, ha supuesto una revolución en la comunicación hombre-máquina. Esta nueva forma de interactuar con los dispositivos electrónicos abre interesantes posibilidades en el ámbito educativo. El objetivo de este trabajo fue analizar la eficacia de una máquina de ejemplos prácticos interactivos desarrollada con IA generativa para el estudio y repaso de contenidos en enseñanzas universitarias. Bajo un enfoque de investigación evaluativa, se llevó a cabo un proceso de diseño, validación e implementación piloto de cuatro *prompts* desarrollados en la herramienta ChatGPT. Tras el diseño de cada *prompt*, se validó su funcionamiento por parte de tres jueces expertos, que aplicaron un proceso de testeo sistemático. Los *prompts* definitivos fueron pilotados en una muestra de n=192 estudiantes de titulaciones de Ciencias de la Educación, que valoraron la utilidad y su satisfacción general con las máquinas de ejemplos a partir de escalas validadas en estudios previos. Los resultados del testeo mostraron un mejor desempeño de las máquinas de ejemplos con *prompts* más sencillos. Por otra parte, los estudiantes mostraron una satisfacción muy elevada con las máquinas, junto a una elevada percepción sobre su utilidad. Específicamente, mientras que las mujeres mostraron percepciones de utilidad más elevadas que los hombres en alguno de los indicadores medidos, la utilidad percibida fue más elevada en general en los grupos de estudiantes en los que la máquina cometió errores durante el pilotaje. A pesar de las limitaciones de la herramienta, los resultados obtenidos resultan prometedores.

Palabras clave: inteligencia artificial generativa; máquina de ejemplos; ChatGPT; autoaprendizaje; metodología de investigación; educación superior.

ABSTRACT

Generative Artificial Intelligence (AI), as an emerging and disruptive technology, has revolutionised human-machine communication. This new means of interacting with electronic devices has opened up interesting possibilities in the educational field. The objective of this study was to analyse the effectiveness of interactive and practical example-generating machines developed by generative AI for the study and review of content in university education. Using an evaluative research approach, a process of design, validation, and pilot implementation of four prompts developed by the ChatGPT tool was implemented. After designing each prompt, its functionality was validated by three expert judges who applied a systematic testing process. The final prompts were piloted on a sample of 192 students with education sciences degrees, who evaluated the usefulness and their overall satisfaction with the example-generating machines based on scales validated in previous studies. The testing results revealed better performance of the example-generating machines with simpler prompts. Moreover, the students indicated very high satisfaction with the machines along with a high perception of their usefulness. Specifically, while women showed higher perceptions of usefulness than men in a few of the measured indicators, the perceived usefulness was generally higher in the groups of students in which the machine committed errors during the pilot. Despite the limitations of the tool, the results obtained are promising.

Keywords: generative artificial intelligence; example machine; ChatGPT; self-learning; research methodology; higher education.

INTRODUCCIÓN

La Inteligencia Artificial (IA) nace en torno al año 1956 en Estados Unidos, cuando se comienza a hablar de sistemas inteligentes (Hirsch-Kreinsen, 2023; Sánchez, 2024). En los siguientes años se fue avanzando en su uso en distintos ámbitos, como las redes neuronales para procesar grandes cantidades de datos (González y Silveira, 2022; Hirsch-Kreinsen, 2023). No obstante, el gran impacto de la IA generativa se produjo con la aparición en noviembre de 2022 de ChatGPT (Sánchez, 2024), un chatbot creado por OpenAI capaz de generar respuestas coherentes e informativas de tipo humano (Eysenbach, 2023; Lo, 2023).

Para Mintz et al. (2023), más que una revolución en sí misma el cambio se produjo en que cualquier persona podía acceder a esta tecnología. Aunque la IA ya se encontraba presente en la vida cotidiana a través de relojes inteligentes, chatbots, recomendaciones para compras o listas musicales (Kennedy et al., 2023), la irrupción de las IAs generativas ha transformado múltiples aspectos de la vida social, generando nuevas realidades en la cultura, la ciencia, la comunicación o, por supuesto, la educación (Kennedy et al., 2023; Martínez et al., 2019; Pavlik, 2023).

Surden (2019) define la IA como la tecnología que permite “automatizar tareas que normalmente requieren inteligencia humana” (p. 1307). En este sentido, la finalidad de la IA generativa es que los ordenadores sean capaces de aprender a razonar e interpretar como los humanos a través de experiencias previas (Copeland, 2023; Rebelo et al., 2022). Estas tecnologías están pensadas para posibilitar el aprendizaje personalizado, automatizar tareas triviales administrativas, u ofrecer retroalimentación en el momento (Chaudhry y Kazim, 2022; Guan et al., 2020; UNESCO, 2021).

Con la aparición de IAs generativas, la creación, aplicación e investigación sobre estas herramientas se ha extendido a diferentes ámbitos científicos como el sanitario (Casella et al., 2023; Choi et al., 2023), el empresarial (Deike, 2024) o el educativo (Cooper, 2023; Saif et al., 2024).

En el ámbito educativo, la IA generativa presenta un importante potencial para la innovación y mejora. De hecho, la UNESCO (2021) destaca las oportunidades que presenta la IA generativa para afrontar los desafíos de la educación a través de la innovación en los procesos de enseñanza-aprendizaje. Hsu y Ching (2023) añaden que su alcance es casi infinito debido al uso que la IA generativa hace del lenguaje natural de los usuarios. A pesar de que ya se venía haciendo uso de tecnologías de IA en los procesos de enseñanza-aprendizaje, como el aprendizaje adaptativo, los sistemas de tutorías inteligentes, o el big data (Mao et al., 2024), la IA generativa agrega un gran valor cuando se requiere obtener respuesta rápida a tareas específicas como proporcionar texto, imágenes o música (Mao et al., 2024).

A pesar del indudable potencial de estas nuevas herramientas en el sector educativo, las instituciones académicas están reaccionando de distinta forma, alentando su uso en algunos casos, o prohibiéndolo en otros (Ahmad et al.,

2023). El estudio de Lo (2023) destaca que herramientas como ChatGPT tienen el potencial de servir de ayuda a los profesores como tutor virtual de los estudiantes (por ejemplo, para responder preguntas) o como asistente (por ejemplo, para generar material didáctico y ofrecer sugerencias). Sin embargo, su aplicación plantea algunas dificultades como, por ejemplo, generar información incorrecta o falsa en sus interacciones, o eludir los detectores de plagio. También plantea retos o preocupaciones relacionados con aspectos legales y éticos, con la seguridad, la transparencia en el proceso de toma de decisiones de la IA, o incluso problemas en el desarrollo intelectual de los estudiantes (Kalota, 2024). Respecto a este último aspecto, preocupa el uso excesivo o dependencia de la IA que a su vez podría causar problemas en la capacidad de resolución de problemas y en el pensamiento crítico (Alam et al., 2023; Stokel-Walker y Nordeen, 2023).

Y es que, mientras que la IA generativa puede suponer importantes beneficios en la innovación o la contribución a la sociedad digital, existen otras preocupaciones como el posible aumento de las brechas tecno-educativas que pueden aumentar las desigualdades (González y Silveira, 2022), o el desarrollo de prácticas poco éticas en la redacción de textos académicos y en las pruebas de evaluación (Sánchez, 2024). Mintz et al. (2023) señalan otros desafíos importantes como la privacidad de los usuarios, el papel de los docentes y de la pedagogía, el plagio, la explotación comercial de los datos recogidos, los sesgos de respuesta, o el aumento de las desigualdades sociales. De hecho, herramientas basadas en IA generativa como ChatGPT presentan limitaciones importantes en el apoyo de la actividad académica, como la incapacidad para proporcionar referencias bibliográficas precisas, o la tendencia a sugerir fuentes inexistentes (Karakose, 2023).

Independientemente de estas limitaciones, Mao et al. (2024) advierten de la repercusión que producirán las IAs generativas en la educación, señalando que afectarán principalmente a los procesos de enseñanza-aprendizaje, a la interacción docente-discente, y a los procesos de evaluación. En esta línea, Hsu y Ching (2023) señalan diversas potencialidades: para los profesores, señalan que herramientas como ChatGPT pueden apoyar la enseñanza a través de sugerencias, aumentando la calidad de la docencia, mejorando la comunicación con los distintos agentes educativos, o apoyando los procesos de evaluación; para los estudiantes, se abren vías para la creación de entornos de aprendizaje personalizado, espacios para favorecer la creatividad, la lectura y escritura, o la evaluación; respecto a la evaluación, el uso eficaz de la IA generativa puede fomentar la autoevaluación, permitiendo a los estudiantes evaluar críticamente la tecnología, promoviendo la autorreflexión y la autorregulación (Mouta et al., 2023), favoreciendo además la motivación para el aprendizaje de los estudiantes (Ali et al., 2023). Por ello, Kuhail et al. (2023) señalan que el uso de IA generativa promueve un aprendizaje optimizado y eficaz. Incluso el uso de estas herramientas es visto como una forma de involucrar a los estudiantes en su proceso de aprendizaje gracias a la personalización de los contenidos, preparándoles para afrontar los desafíos del siglo XXI (Ruiz-Rojas et al., 2023).

El uso de IAs generativas para la adquisición de conocimientos supone que se pueden generar entornos de aprendizaje personalizados y adaptativos con retroalimentación inmediata para el estudiante basada en su nivel de desempeño (Mao et al., 2024). Algunos estudios empíricos evidencian las posibilidades pedagógicas de la IA generativa, posibilitando la personalización, la práctica o la generación de nuevos contenidos (Sánchez, 2024). En esta línea, Ayuso-del Puerto y Gutiérrez-Esteban (2022) llevaron a cabo experiencias con IA generativa para la formación inicial del profesorado, encontrando mejoras en el autoaprendizaje de los estudiantes, en su motivación, y en la resolución de problemas, consiguiendo experiencias de aprendizaje significativas. Bower et al. (2024) en su estudio con docentes de distintos niveles educativos, evaluaron el impacto de la IA generativa en la enseñanza y la evaluación, encontrando que la mayor parte de docentes identificaba que la IA generativa tenía un impacto importante en ambos aspectos. Así, los resultados de Bower et al. (2024) revelaron la importancia de enseñar previamente a los estudiantes cómo utilizar IA generativa, mostrar cómo es su funcionamiento, o advertir sobre la importancia de contar con habilidades para el pensamiento crítico y la ética. Jauhainen y Guerra (2023) utilizaron ChatGPT para generar y modificar contenido con estudiantes en educación primaria. Los estudiantes se beneficiaron de la personalización de los materiales en función de sus conocimientos y habilidades, aunque se advirtieron problemas en la herramienta, sugiriendo la necesidad de su perfeccionamiento en el futuro para garantizar su utilidad en educación. Algunos estudios con ChatGPT en estudiantes universitarios revelan diferencias de género en cuanto a la actitud, utilidad y facilidad de uso percibidas, siendo en unos casos favorable a los hombres (Stöhr et al., 2024; Yilmaz et al., 2023) y en otros a las mujeres (Raman et al., 2024).

Otro factor que condiciona la integración de las IAs generativas en educación es la desconfianza que despiertan, no solamente por cuestiones éticas (Petricini, 2024) o por su potencial efecto en el desarrollo de la sociedad (Brailas, 2024), sino también por la fiabilidad de sus respuestas y su tendencia a emplear razonamientos abductivos que generan información falsa que la persona puede no reconocer (Illera, 2024). Ese riesgo se incrementa cuando la persona tiene poco conocimiento sobre la materia en cuestión, lo que plantea la cuestión del efecto que puede tener la toma de conciencia sobre la falibilidad de las IAs generativas en el uso de esta tecnología por parte de los estudiantes.

En definitiva, parece necesario que el profesorado de todos los niveles adquiera competencias en el uso de las IAs generativas (Karakose et al., 2023), siendo capaz de integrarlas en sus estrategias didácticas, mejorando los procesos de enseñanza-aprendizaje, y desarrollando las competencias del alumnado en su uso (Bower et al., 2024; Mouta et al., 2023). Así, esta investigación pretende avanzar en el uso de la IA generativa para crear contenidos educativos para los estudiantes, línea abierta en trabajos previos que emplearon esta herramienta para la elaboración de preguntas sobre un tema concreto (Ling y Afzaal, 2024), la realización de textos mentor (Nash

et al., 2023) o como máquina de ejemplos prácticos para el desarrollo de la capacidad de análisis (Mah y Levine, 2023; Trust y Maloy, 2023). Esta investigación se centra en esta última aplicación, con el objetivo de analizar la eficacia de la IA generativa como herramienta generadora de reactivos, o ejemplos prácticos, que faciliten al estudiante el estudio y repaso autónomos. Así, se incluye el diseño de los *prompts*, la validación de contenido de los mismos y una experiencia piloto.

MÉTODO

Este trabajo planteó un diseño de investigación evaluativa que analiza el uso de ChatGPT como máquina de ejemplos para el apoyo al aprendizaje autónomo. Se formularon las siguientes preguntas de investigación:

- PI1. ¿Cómo valoran los estudiantes el potencial de ChatGPT como máquina de ejemplos para el estudio y aprendizaje autónomo?
- PI2. ¿Influye el género en la actitud hacia la máquina de ejemplos de los estudiantes?
- PI3. ¿Los errores en el funcionamiento de la máquina de ejemplos influyen en la valoración de los estudiantes sobre el potencial educativo de la herramienta?

Diseño de la herramienta

Cuatro de los docentes responsables de la asignatura *Metodología de Investigación Educativa* en las cuatro titulaciones de Grado (Primaria, Infantil, Pedagogía y Educación Social) de la Facultad de Educación de la Universidad de Salamanca diseñaron, en equipo, el conjunto de instrucciones para ChatGPT (en adelante, *prompt*).

Este equipo determinó los contenidos que se trabajarían con la máquina de ejemplos, teniendo en cuenta los siguientes criterios: (1) los contenidos tienen que ver con los diseños de investigación en educación; (2), los contenidos tienen una tipología clara y discernible mediante un proceso lógico sencillo; y (3) los contenidos se imparten de una manera similar en las cuatro titulaciones. Estos acuerdos buscaban que el diseño de los *prompts* fuera homogéneo en las cuatro titulaciones, aunque existió alguna salvedad que se comentará más adelante.

Los contenidos de la materia que incluyeron los *prompts* fueron: (1) los tipos de diseños de investigación según su naturaleza experimental o no experimental; (2) los tipos de variables según su función en el diseño de investigación; y (3) los tipos de muestreo según su naturaleza probabilística o no probabilística.

En el proceso cooperativo de diseño de los *prompts* se partió de las directrices establecidas por Korzynski et al. (2023) y Reynolds y McDonell (2021). Así, la estructura de cada *prompt* se dividió en tres secciones:

- **Contexto:** párrafo de introducción para informar a ChatGPT del objetivo que iban a tener las siguientes secciones de texto.
- **Base de conocimiento:** texto introducido entre triples tildes que ofrece una explicación del contenido de la asignatura para el que se quieren obtener ejemplos. De esta manera los estudiantes no tienen que aportar a ChatGPT sus propias definiciones de los contenidos ni la herramienta recurrirá a su propia base de datos, sino que los ejemplos se generarán a partir de una serie de definiciones elaboradas y validadas por el profesorado de las asignaturas.
- **Instrucciones:** texto explicativo de las funciones que debe realizar ChatGPT a partir de la base de conocimiento. Inicialmente, la máquina genera un ejemplo de una investigación que incluye el contenido trabajado. A partir del ejemplo, el alumnado debe intentar adivinar cuál es su diseño de investigación, cuáles son las variables, o cuál es el muestreo presentado. En caso de respuesta correcta por parte del estudiante, ChatGPT aporta *feedback* positivo, y en caso de error devuelve una pista adicional dando la oportunidad de intentarlo de nuevo. Una vez adivinado el ejemplo, el estudiante puede optar por pedir más ejemplos o finalizar la conversación.

Para facilitar el empleo gratuito de la máquina, se empleó la versión ChatGPT 3.5 para el diseño y entrenamiento de los *prompts*. Inicialmente se intentaron unificar los tres contenidos trabajados en un solo *prompt*, de manera que la máquina ofreciera ejemplos de mayor complejidad. No obstante, se observó que ChatGPT tenía dificultades con las operaciones lógicas y para identificar las respuestas correctas en sus ejemplos, por lo que se optó por separar los contenidos en *prompts* independientes más sencillos para simplificar las bases de conocimiento y reducir la complejidad lógica de la tarea solicitada. Finalmente, se diseñaron cuatro *prompts* para trabajar los contenidos de la materia ligados a la competencia para la identificación y clasificación de los elementos básicos de la investigación (Tabla 1).

Tabla 1
Contenidos de los prompts

Nombre del <i>prompt</i>	Contenido
Variables en diseños experimentales	• Variable dependiente e independiente
Variables en diseños no experimentales	• Variable criterio y explicativa

Nombre del <i>prompt</i>	Contenido
Diseños de investigación	<ul style="list-style-type: none"> • Diseños experimentales: <ul style="list-style-type: none"> • Pretest-postest con grupo de control • Sólo postest con grupo de control • Un solo grupo • Diseños pre-experimentales: <ul style="list-style-type: none"> • Transversal • Longitudinal
Muestreo	<ul style="list-style-type: none"> • Probabilístico <ul style="list-style-type: none"> • Aleatorio simple • Estratificado • Por conglomerados • No probabilístico <ul style="list-style-type: none"> • Por disponibilidad • Por bola de nieve • Por criterios • Por cuotas

Fuente: elaboración propia.

Validación de contenido

Una vez diseñados los cuatro *prompts*, y tras pruebas internas por parte del equipo de diseño, fueron sometidos a un proceso de validación de contenido por jueces expertos.

Para ello, se contactó con tres investigadores del área de Métodos de Investigación en Educación, que probaron los cuatro *prompts* y realizaron un informe identificando errores en los ejemplos y en las respuestas posteriores guiados por las siguientes instrucciones:

https://drive.google.com/file/d/1DGuSxkFwT5To-j7pGcYOguoTKC6Gu3GY/view?usp=drive_link

En total los jueces realizaron 10 pruebas para cada uno de los *prompts*, cinco de ellas dando la respuesta correcta al ejemplo recibido y cinco de ellas dando una respuesta incorrecta. Así, en total, cada juez experto probó 40 ejemplos de la máquina, para un total de 120 ejemplos de prueba entre los tres jueces. Cada uno de los intentos debía llevarse a cabo abriendo una conversación nueva en ChatGPT e introduciendo nuevamente las tres secciones del *prompt*. Los resultados de las pruebas fueron registrados en una hoja de cálculo con los siguientes apartados:

- Nombre: nombre del *prompt*.
- Número: número de prueba.
- Modalidad: intento de respuesta correcta o incorrecta.
- Ejemplo: texto del ejemplo aportado por ChatGPT.

- Evaluación del ejemplo: adecuación del ejemplo aportado por ChatGPT (correcto/incorrecto).
- Respuesta: texto de la respuesta dada por el juez a ChatGPT.
- *Feedback*: texto del *feedback* de ChatGPT como reacción a la respuesta del juez.
- Evaluación del *feedback*: adecuación del *feedback* aportado por el programa.
- Observaciones: espacio destinado a comentarios cualitativos.

Tras recibir las valoraciones de los tres jueces se realizaron correcciones para obtener las versiones definitivas de los *prompts*, y se puso a disposición de los estudiantes, en el campus virtual, un documento con el código de los cuatro *prompts* para su utilización

(<https://drive.google.com/file/d/1sXgoiR7Qzkmp6NiuYLJCS428EoqFInYq/view?usp=sharing>).

Prueba piloto

Para la validación empírica de los *prompts* se diseñó una prueba piloto con los estudiantes de cada titulación, realizada en marzo de 2024, en la que se explicó el funcionamiento de las máquinas durante una sesión de la asignatura. Estas sesiones, de aproximadamente 30 minutos, incluyeron: (1) breve introducción aclarando aspectos básicos sobre el funcionamiento de ChatGPT y objetivo de las máquinas de ejemplos diseñadas; (2) presentación y prueba de cada *prompt*, explicando su estructura (contexto, base de conocimiento e instrucciones) y guiando a los estudiantes mientras los probaban en ChatGPT en sus propios equipos; y (3) práctica libre para que los estudiantes probaran la implementación de los *prompts* en sus ordenadores autónomamente. En todos los casos el profesorado hizo una demostración del funcionamiento, generando al menos dos ejemplos de prueba, que los estudiantes replicaron en paralelo en sus ordenadores, resolviendo cualquier incidencia surgida durante el proceso.

Esta prueba piloto tuvo una doble función: (1) comprobar el funcionamiento de las máquinas de ejemplos en un número elevado de ordenadores y la capacidad de los estudiantes para detectar posibles errores en caso de producirse; y (2) obtener información sobre la valoración de los estudiantes.

Para recoger esta información se aplicó un instrumento compuesto por dos secciones. La primera sección para recoger los datos de identificación de los participantes: Género (hombre, mujer, otro), Grado (Educación Infantil, Educación Primaria, Educación Social o pedagogía) y edad. La segunda sección incluyó dos escalas validadas en estudios previos para evaluar innovaciones (Martínez-Abad y Hernández-Ramos, 2017; Olmos et al., 2014): utilidad del uso de la IA y satisfacción general con la herramienta. Se adaptó el contenido de las escalas para realizar una referencia explícita a la IA. Así, se aplicaron 12 ítems (Tabla 2) con escala de

respuesta tipo Likert de 0 a 10 puntos (0=totalmente en desacuerdo - 10=totalmente de acuerdo).

Tabla 2
Ítems del cuestionario

Dimensión	Ítem	Texto
Utilidad		<i>El uso de la IA me ayudará...</i>
	U01	en la memorización y reproducción de los contenidos sobre metodología de la investigación
	U02	en la comprensión de los conceptos e ideas básicas sobre metodología
	U03	en la generalización de los conocimientos teóricos a situaciones reales
	U04	en la resolución de problemas prácticos
		<i>El uso de la IA...</i>
	U05	favorecerá mi aprendizaje en la asignatura
	U06	me ayudará en la mejora de mis resultados académicos en la asignatura
	U07	se adaptará a mi ritmo de aprendizaje
Satisfacción general	U08	se adaptará a mis necesidades concretas durante el proceso de enseñanza-aprendizaje
	U09	ayudará a que aumente mi interés por los contenidos de la asignatura
		<i>A nivel general...</i>
	S01	si volviera a cursar la asignatura me gustaría contar con este recurso
	S02	a pesar de las limitaciones considero que el recurso es satisfactorio
	S03	este tipo de herramientas son útiles para fomentar el aprendizaje autónomo

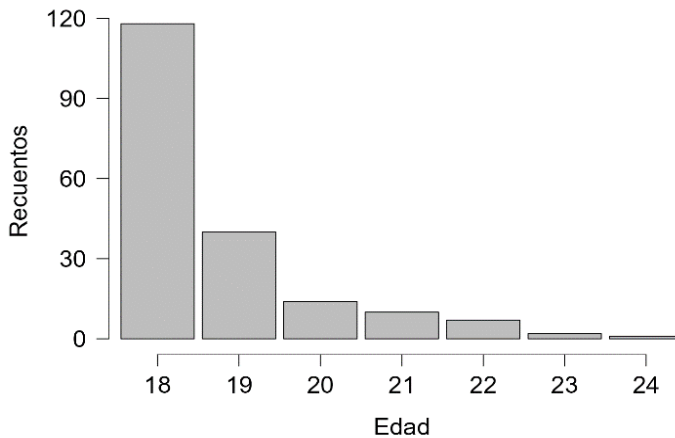
Participantes

El cuestionario se desarrolló en la herramienta Google Forms¹, y fue distribuido electrónicamente a los estudiantes, aplicando un muestreo no probabilístico por disponibilidad o conveniencia.

En total participaron 192 estudiantes: 32 del Grado en Educación Infantil (16.6 %), 98 del Grado en Educación Primaria (51 %), 34 del Grado en Educación Social (17.7 %) y 28 del Grado en Pedagogía (14.5 %). En cuanto al género, se obtuvo un 18.9 % de hombres y un 81 % de mujeres, y el 1 % de estudiantes no definió su género.

La edad media de la muestra se situó en 18.74 años, con una desviación típica de 1.2 (Figura 1).

Figura 1
Distribución de la variable edad



Análisis de datos

Inicialmente, se analizó la fiabilidad de la herramienta a partir de los resultados de la validación de contenido obteniendo los porcentajes de ejemplos adecuados y de interacciones correctas por parte de ChatGPT. Posteriormente, se estudió la calidad de los *prompts* a partir de la valoración de los estudiantes. Además del análisis descriptivo de los ítems de ambas escalas, se aplicaron contrastes de hipótesis para dos grupos independientes para comprobar las diferencias de valoración en función del género de los estudiantes y de si en la prueba piloto ChatGPT cometió errores o no. Dada la falta de normalidad (test de Shapiro-Wilk), se aplicó el test de la U de Mann-Whitney.

Se empleó en los análisis de datos el software libre JASP 18.0.3 (Halter, 2020), partiendo de un nivel alfa del 5 %.

RESULTADOS

Validación de contenido

La Tabla 3 recoge el porcentaje de ejemplos adecuados y de respuestas correctas obtenidas por los jueces expertos. Se consideraron errores los casos en los que los ejemplos no se entendieran claramente o no permitieran la correcta identificación del elemento. De acuerdo con estos criterios, la mayoría de *prompts* generaron ejemplos adecuados, observándose errores únicamente en los casos de identificación de variables. Concretamente, el error que cometía la máquina era hacer explícito, junto con el ejemplo propuesto, el tipo de las variables del ejemplo, impidiendo que los alumnos lo adivinaran. Esta cuestión se solucionaba tras indicarle a ChatGPT que formulara un ejemplo nuevo sin incluir los tipos de variables en el texto.

Tabla 3

Porcentaje de respuestas correctas durante la validación de contenido

Nombre del <i>prompt</i>	Ejemplo generado	<i>Feedback</i> tras la respuesta
VARIABLES en diseños experimentales	83,3 %	100 %
VARIABLES en diseños no experimentales	96.7 %	100 %
DISEÑOS de investigación	100 %	76.7 %
Muestreo	100 %	63.4 %

En cuanto al *feedback* ofrecido a los estudiantes, se consideraron errores aquellos casos en los que ChatGPT indicara como erróneas respuestas correctas de los estudiantes y viceversa. También se consideraron errores aquellas ocasiones en las que ChatGPT identificara erróneamente el tipo de ejemplo dado o no aportara información suficiente para que el estudiante pudiera responder correctamente. En este caso, se observó un mayor porcentaje de errores, siendo los fallos más frecuentes cuanto más extensas fueran las bases del conocimiento e instrucciones incluidas en el *prompt*. Dada la simplicidad de las instrucciones y bases del conocimiento en el *prompt* de variables, la respuesta ofrecida al estudiante tras el intento de identificación de los ejemplos resultó correcta en la totalidad de los casos. Sin embargo, en el caso de los tipos de diseño, en los que el programa debía identificar correctamente cinco tipos de diseño distintos agrupados en experimentales (tres tipos) y no experimentales (dos tipos), el porcentaje de respuestas correctas se redujo al 76.7 %. En estos casos, los errores se producían al generar enunciados demasiado concisos en los que no se ofrecía información suficiente para discernir entre los distintos diseños posibles. También se dieron algunos casos en los que el programa denominaba tipos de

diseños experimentales como si fueran no experimentales. En todos los casos, la respuesta proporcionada por la máquina de ejemplos permitía la identificación del error fácilmente. Los problemas más serios aparecieron en la máquina de ejemplos para los muestreos. En este caso, se ofreció un *feedback* correcto en el 63,4 % de los casos, observándose errores constantes al distinguir entre muestreo probabilístico aleatorio simple y muestreo no probabilístico por disponibilidad o conveniencia. Las valoraciones completas de los jueces pueden consultarse en: https://docs.google.com/spreadsheets/d/1MZCY8ixi1ZSjA7laAkr3mGZONXDJeFAe/edit?usp=drive_link&oid=111124780597519622319&rtopof=true&sd=true

Para solucionar estos errores, dado que existían siete tipos diferentes de muestreos agrupados en probabilísticos (tres tipos) y no probabilísticos (cuatro tipos), se optó por dividir este *prompt* en dos independientes (uno para muestreos probabilísticos y otro para muestreos no probabilísticos) para disminuir la dificultad de la tarea. Estas nuevas máquinas fueron sometidas a una segunda ronda de validación por parte del equipo de diseño, que aplicó 90 pruebas (45 intentos con respuesta correcta y 45 intentos con respuesta incorrecta), no encontrando errores.

En el siguiente enlace se ofrece un ejemplo del funcionamiento de los *prompts* y de su interacción con las respuestas correctas e incorrectas de los estudiantes para los tres grandes bloques de contenidos:

<https://docs.google.com/document/d/1rJxZwRl6OArSbjdmC7zxEQxfC6zdJWJokl8gTWERv2s/edit?usp=sharing>

Prueba piloto

Una vez recogidos los datos se realizó el análisis descriptivo de las puntuaciones obtenidas para la dimensión Utilidad (Omega de McDonald .897) y Satisfacción General (Omega de McDonald .845). Los resultados muestran una valoración positiva de la herramienta por parte de los estudiantes (Tabla 4), con medias por encima del 7 en todos los indicadores salvo en el ítem U09, referido al impacto de la herramienta en el aumento del interés hacia los contenidos. Este constituye, junto con U01 (utilidad para la memorización y reproducción de los contenidos), los dos ítems con valoraciones más bajas.

En contraste los indicadores U03 y U04, referidos a la generalización de contenidos a situaciones reales y la resolución de problemas prácticos, obtuvieron las puntuaciones más altas en la dimensión de utilidad.

Resultan también llamativas las puntuaciones obtenidas en la dimensión satisfacción general, que presentan en todos los ítems puntuaciones por encima de 8, en coherencia con las medias en ambos factores.

Tabla 4
Estadísticos descriptivos

	Media	DT	n
U01	7.189	1.865	190
U02	7.812	1.656	191
U03	7.863	1.653	190
U04	7.979	1.721	190
U05	7.641	1.713	192
U06	7.482	1.759	191
U07	7.670	1.696	191
U08	7.674	1.736	190
U09	6.900	1.996	190
S01	8.266	1.718	192
S02	8.272	1.341	191
S03	8.288	1.496	191
U	7.595	1.275	181
S	8.277	1.329	190

En cuanto a las diferencias por género (Tabla 5), se encontraron diferencias significativas en tres ítems (U03, U05 y U09), con tamaños del efecto pequeños. En todos los casos las mujeres obtuvieron puntuaciones medias por encima de los hombres. Estos ítems hacen referencia a la utilidad a la hora de generalizar los contenidos y aplicarlos a situaciones reales, favorecer el aprendizaje y aumentar el interés respectivamente. En cuanto a las puntuaciones promedio de ambos factores no se encuentran diferencias significativas.

Tabla 5
Estadísticos descriptivos de los ítems y factores con diferencias significativas en función del género

	Grupo	Media	DT	n	W	p	d
U03	Hombre	7.286	1.655	35	1928.500	.008	-0.280
	Mujer	7.980	1.632	153			
U05	Hombre	7.083	1.519	36	2023.500	.010	-0.270
	Mujer	7.740	1.726	154			
U09	Hombre	6.222	2.179	36	2120.500	.029	-0.230
	Mujer	7.046	1.924	153			

Para finalizar, se comprobó la influencia de la existencia de errores durante la prueba piloto. Para ello los docentes documentaron las incidencias surgidas durante la demostración del funcionamiento de los *prompts*. En este sentido, aparecieron algunos errores en tres de los grupos (Grado en Maestro en Educación Infantil, Grado en Pedagogía y grupo de tarde del Grado en Maestro en Educación Primaria) y no se registró ningún error en los dos grupos restantes (Grado en Educación Social y grupo de mañana del Grado en Maestro en Educación Primaria). En todos los casos los errores eran cuestiones menores referidas a erratas en el *feedback* de ChatGPT similares a las vistas en el proceso de validación, que fueron aprovechados para explicar cómo identificarlas y corregirlas.

La Tabla 6 muestra los tres ítems en los que se encontraron diferencias significativas entre la valoración de los estudiantes de los grupos con y sin errores. Estos ítems hacían referencia a la utilidad a la hora de memorizar y reproducir el contenido (U01), favorecer el aprendizaje (U05) y aumentar el interés (U09). A nivel general se han encontrado diferencias significativas en la dimensión utilidad. En todos estos casos, los estudiantes en cuyas sesiones se registraron errores valoraron más positivamente las máquinas de ejemplos.

Tabla 6

Estadísticos descriptivos de los ítems con diferencias significativas en función de los errores

	Grupo	Media	DT	n	W	p	d
U01	Sin errores	6.782	1.913	87	3385.500	0.003	-0.244
	Con errores	7.534	1.759	103			
U05	Sin errores	7.295	1.716	88	3463.500	0.003	-0.243
	Con errores	7.933	1.662	104			
U09	Sin errores	6.341	1.982	88	2996.500	< .001	-0.332
	Con errores	7.382	1.888	102			
U	Sin errores	7.348	1.201	83	2983.500	0.002	-0.266
	Con errores	7.804	1.304	92			

Para reducir la heterogeneidad de la muestra se decidió replicar el contraste de hipótesis anterior solamente con los alumnos del Grado de Maestro en Educación Primaria, dado que los alumnos se asignaron al grupo de mañana o de tarde de manera aleatoria y la aparición de errores se produjo solamente en la explicación en el grupo de tarde.

En este caso (Tabla 7), se encontraron diferencias significativas en cinco ítems: los tres anteriores referidos a la utilidad para la memorización (U01), el aprendizaje (U05) y el interés (U09), y dos ítems adicionales relativos a la capacidad de la IA

para adaptarse a su ritmo de aprendizaje (U07) y la satisfacción con el recurso a pesar de las limitaciones (S02). En cuanto a las puntuaciones escalares, al igual que con la muestra completa, se encontraron diferencias significativas en la dimensión utilidad. En todos los casos los estudiantes con errores registrados en sus sesiones valoraron más positivamente las máquinas de ejemplos.

Tabla 7

Estadísticos descriptivos de los ítems con diferencias significativas en función de los errores

	Grupo	Media	DT	n	W	p	d
U01	Sin errores	6.352	1.925	54	626.000	< .001	-0.473
	Con errores	7.864	1.850	44			
U05	Sin errores	7.333	1.671	54	837.000	0.011	-0.295
	Con errores	8.091	1.750	44			
U07	Sin errores	7.130	2.102	54	905.500	0.041	-0.238
	Con errores	7.977	1.772	44			
U09	Sin errores	6.056	2.032	54	550.500	< .001	-0.526
	Con errores	7.767	1.837	43			
S02	Sin errores	8.130	1.229	54	908.500	0.040	-0.235
	Con errores	8.614	1.280	44			
U	Sin errores	7.240	1.061	51	646.000	0.001	-0.397
	Con errores	7.963	1.375	42			

CONCLUSIONES

Siguiendo la estela de trabajos previos (Ali et al., 2023; Ayuso-del Puerto y Gutiérrez-Esteban, 2022; Bower et al., 2024; Mao et al., 2024; Mouta et al., 2023; Ruiz-Rojas et al., 2023; Sánchez, 2024), esta investigación ha buscado aumentar las evidencias en torno a la eficacia del uso de la IA generativa en los procesos de enseñanza-aprendizaje, en concreto en el contexto universitario. Para ello, se creó una máquina de ejemplos que permitió a los estudiantes de la materia Métodos de Investigación en Educación estudiar y repasar el contenido de manera autónoma, a partir de la generación automática de ejemplos prácticos orientados.

Atendiendo a la primera pregunta de investigación, la valoración por parte de los estudiantes de la utilidad de la herramienta se sitúa en puntuaciones elevadas, alcanzando la satisfacción global niveles muy altos. El valor de estos resultados debería ser abordado con cautela, teniendo en cuenta que estudios previos muestran la alta valoración que habitualmente realizan los estudiantes de la integración de

recursos tecnológicos en los procesos de enseñanza-aprendizaje, principalmente cuando estos son emergentes o disruptivos (e.g., Ayala et al., 2023; Cabero-Almenara et al., 2018; Cabero-Almenara y Fernández-Robles, 2018; Ruiz-Campo et al., 2023; Zhang et al., 2023), como es el caso de la IA generativa. A pesar de este sesgo de valoración, las altas valoraciones obtenidas pueden estar reflejando que los estudiantes son conscientes del potencial de la herramienta creada para autorregular su aprendizaje (Mouta et al., 2023) adaptándose a sus necesidades y ritmos de aprendizaje individuales (Jauhiainen y Guerra, 2023; Ruiz-Rojas et al., 2023), de manera que la transmisión de conocimientos sea más eficaz (Kuhail et al., 2023).

En relación con la segunda cuestión de investigación, se observó un nivel superior de utilidad percibida y de satisfacción con la máquina de ejemplos en las mujeres. Estos resultados contrastan con estudios previos, que muestran habitualmente que los hombres son más entusiastas y proactivos con el empleo de nuevas tecnologías (e.g., Aranda et al., 2019; Ruiz-Campo et al., 2023; Stöhr et al., 2024; Yilmaz et al., 2023; Zhang et al., 2023). Específicamente, sí es consistente con Raman et al. (2024), donde las mujeres valoran más la facilidad de uso y los beneficios directos de ChatGPT.

Se abre una incógnita en torno a estos resultados, que podrían deberse simplemente a las peculiaridades de los estudiantes del ámbito de las Ciencias de la Educación. Otra posibilidad es que, a diferencia de otras tecnologías emergentes, la propia naturaleza de la IA generativa y la manera en la que se interactúa con la misma provoque que las mujeres sientan un mayor apego y satisfacción de uso. Dado lo novedoso de estos efectos, se abre una interesante línea de investigación que debe ser ampliada en estudios futuros.

En cuanto a la tercera cuestión de investigación, a diferencia de otras tecnologías disruptivas como el metaverso (Ruiz-Campo et al., 2023) o la realidad aumentada (Cabero-Almenara et al., 2018; Cabero-Almenara y Fernández-Robles, 2018), la creatividad (o margen de error) programada en la IA generativa lleva a que cometa errores en sus interacciones. Nuestros resultados evidencian, de manera sólida, que los estudiantes que participaron en pruebas piloto en las que la máquina generó ejemplos con algún error valoraron mejor la utilidad de la herramienta y se mostraron más satisfechos con la misma. Dado que este tipo de comportamientos erráticos no se encuentran en otras tecnologías, no existen antecedentes que orienten o justifiquen los resultados aquí obtenidos, más allá del efecto pedagógico que tiene el propio error del estudiante y su consciencia sobre el mismo (Krause-Wichmann et al., 2023; Williams et al., 2016). Las evidencias aquí obtenidas apuntan a que el error de la máquina ayuda al estudiante a ser más consciente de estos errores, como si de un error propio se tratara, consiguiendo un efecto de aprendizaje basado en el error.

Los resultados obtenidos ponen en evidencia el potencial de las IAs generativas como herramientas de apoyo al autoaprendizaje, presentando máquinas de ejemplos desarrolladas empleando ChatGPT y validadas de manera empírica que permiten al

alumnado trabajar contenidos de manera más flexible y adaptada a sus necesidades y ritmos de aprendizaje.

La relevancia de estos resultados debe interpretarse en el marco del fuerte debate sobre el reto que supone para la sociedad la inclusión de estas tecnologías y su impacto en el desarrollo cultural (Destéfano et al., 2024). Desde un punto de vista educativo, podemos considerar que las IAs generativas son una herramienta tecnológica puesta al servicio de una tarea concreta cuyo uso, especialmente desde una perspectiva constructivista, tiene un impacto en el desarrollo de la cognición humana, lo que reclama poner el foco no solamente en los resultados conseguidos a corto plazo sino en los posibles efectos a medio o largo plazo (Illera, 2024). En este sentido algunos autores han manifestado su preocupación sobre los efectos de estas tecnologías sobre el desarrollo del pensamiento crítico y la expresión escrita (Law, 2024; Thiga, 2024), lo que pone de manifiesto la difícil tarea de desarrollar una pedagogía artificial (Díaz y Nussbaum, 2024) capaz de afrontar los retos que trae consigo este nuevo panorama educativo.

Dentro de este debate, el desarrollo de actividades que integren máquinas de ejemplos y otras modalidades de autoaprendizaje reactivo, en las que el alumno tiene que analizar de manera crítica los mensajes producidos por la inteligencia artificial y determinar no solamente cuál es la respuesta correcta, sino la veracidad de la información generada, resultan útiles a la hora de paliar los potenciales efectos negativos sobre el pensamiento crítico (Illera, 2024). Esta utilidad puede evidenciarse en esta investigación, por la mejor valoración de las máquinas de ejemplos realizada por los estudiantes que han visto errores y han sido capaces de comprobar su capacidad para detectarlos y solventarlos.

A pesar de evidenciar la utilidad de estos recursos, durante la investigación se han encontrado cuestiones que pueden dificultar la integración de estas herramientas en los procesos de enseñanza-aprendizaje. En primer lugar, se constata la falta de formación de docentes y estudiantes en el uso correcto de estas herramientas (Bower et al., 2024; Lo, 2023), no sólo en el manejo básico del programa y la generación de *prompts*, sino también en cuanto a los riesgos asociados al uso de IAs generativas por su falta de fiabilidad en las respuestas, problemas de privacidad y usos poco éticos (Mintz et al. 2023; Sánchez, 2024). Esta última cuestión quedó claramente reflejada durante la prueba piloto, donde se comprobó el desconocimiento que tienen los estudiantes de estos riesgos, estableciéndose debates en torno a cuestiones como la incapacidad para generar citas bibliográficas reales (Karakose, 2023), la naturaleza probabilística del algoritmo (Kortemeyer, 2023) o las alucinaciones (Amézquita Zamora, 2023).

En segundo lugar, se encontraron problemas relacionados con la capacidad de la IA generativa seleccionada en cuestiones que entrañaban procesos lógicos y la interacción con el usuario, dificultando el desarrollo de los *prompts*. Aunque las IAs generativas tienen un gran potencial para transformar las actividades humanas y alterar los procesos de enseñanza-aprendizaje, están en un proceso temprano de

desarrollo y es necesario que sigan evolucionando y mejorando sus interacciones naturales (Jauhiainen y Guerra, 2023).

Así mismo, este estudio también ha contado con algunas limitaciones relevantes. En cuanto al diseño de investigación, cabe destacar que la muestra de estudiantes es perteneciente a una sola universidad y rama del conocimiento, lo que dificulta la generalización de los resultados obtenidos. Por otro lado, se ha aplicado una investigación evaluativa en la que existe poco control sobre las variables, lo que dificulta el establecimiento de relaciones causales. Por último, debido a las limitaciones de la duración de la prueba piloto y la naturaleza evaluativa de la investigación, solamente se han podido analizar las percepciones de los estudiantes, quedando pendiente analizar el posible impacto del uso de las máquinas de ejemplos sobre el rendimiento de los estudiantes.

En cuanto a las limitaciones relacionadas con la herramienta, destacamos los problemas de interacción con los usuarios, la dificultad para ejecutar operaciones complejas y la existencia de errores puntuales en la generación de ejemplos. También cabe mencionar la adaptación de ChatGPT a las conversaciones y estilo comunicativo de cada sujeto, que provoca que los *prompts* funcionen diferente para cada usuario aumentando la probabilidad de aparición de errores. En este sentido hay que considerar que los *prompts* han sido diseñados para la versión 3.5 de ChatGPT, de uso gratuito y ampliamente conocida por los estudiantes, aunque la dependencia de esta versión constituye una limitación a considerar de cara al futuro.

Los resultados y limitaciones encontrados abren la puerta a futuras investigaciones. En cuanto al diseño, se propone ampliar la muestra a otras poblaciones y aplicar diseños con un mayor nivel de experimentalidad y control, incorporando una medida de rendimiento. En cuanto a la herramienta, resulta de interés continuar refinando las máquinas de ejemplos, empleando IAs generativas más avanzadas para conseguir ejemplos más completos, integrados y con menos errores, así como el diseño de un *prompt* que funcione en distintas IAs generativas para reducir así la dependencia de versiones específicas de una herramienta concreta. Otra línea futura se asocia a mejorar la experiencia del usuario a través de un software específico que se comunique con la IA a través de una API y que permita su integración en el campus virtual institucional. Por último, dados los resultados obtenidos, sería interesante profundizar en el efecto de los errores sobre la adopción de IAs en contextos educativos, y profundizar en el estudio de la satisfacción diferencial por género.

NOTAS

1. https://docs.google.com/forms/d/e/1FAIpQLSfEptXcAzDmo36siQMvkbrUj_y4ThdLfkHxvGeoWG4ryvV4WA/viewform

REFERENCIAS

- Ahmad, N., Murugesan, S. y Kshetri, N. (2023). Generative Artificial Intelligence and the Education Sector. *Computer*, 56(6), 72-76. <https://doi.org/10.1109/MC.2023.3263576>
- Alam, F., Lim, M. A. y Zulkipli, I. N. (2023). Integrating AI in Medical Education: Embracing Ethical Usage and Critical Understanding. *Frontiers in Medicine*, 10, Article 1279707. <https://doi.org/10.3389/fmed.2023.1279707>
- Ali, J., Shamsan, M. A. A., Hezam, T. y Mohammed, A. A. Q. (2023). Impact of ChatGPT on learning motivation: Teachers and students' voices. *Journal of English Studies in Arabia Felix*, 2(1), 41-49. <https://doi.org/10.56540/jesaf.v2i1.51>
- Amézquita Zamora, J. A. (2023). Uso responsable de ChatGPT en el aula: Cómo convertirlo en un aliado en los procesos educativos. *Company Games & Business Simulation Academic Journal*, 3(2), 69-86. <http://uajournals.com/ojs/index.php/businesssimulationjournal/article/view/1511>
- Aranda, L., Rubio, L., Di Giusto, C. y Dumitrache, C. (2019). Evaluation in the use of TIC in students of the University of Malaga: gender differences. *INNOEDUCA. International Journal of Technology And Educational Innovation*, 5(1), 63-71. <https://doi.org/10.24310/innoeduca.2019.v5i1.5175>
- Ayala, E., López, R. E. y Menéndez, V. H. (2023). Implementación holística de tecnologías digitales emergentes en educación superior. *Edutec. Revista Electrónica de Tecnología Educativa*, (83), 153-172. <https://doi.org/10.21556/edutec.2023.83.2707>
- Ayuso-del Puerto, D. y Gutiérrez-Esteban, P. (2022). La Inteligencia Artificial como Recurso Educativo durante la Formación Inicial del Profesorado. *RIED-Revista Iberoamericana de Educación a Distancia*, 25(2), 347-362. <https://doi.org/10.5944/ried.25.2.32332>
- Bower, M., Torrington, J., Lai, J. W. M., Petocz, P. y Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-12405-0>
- Brañas, A. (2024). Postdigital Duoethnography: An Inquiry into Human-Artificial Intelligence Synergies. *Postdigital Science and Education*, 6, 486-515. <https://doi.org/10.1007/s42438-024-00455-7>
- Cabero-Almenara, J. y Fernández-Robles, B. (2018). Las tecnologías digitales emergentes entran en la Universidad: RA y RV. *RIED-Revista Iberoamericana de Educación a Distancia*, 21(2), 119-138. <https://doi.org/10.5944/ried.21.2.20094>
- Cabero-Almenara, J., Vázquez-Cano, E. y López-Meneses, E. (2018). Uso de la Realidad Aumentada como Recurso Didáctico en la Enseñanza Universitaria. *Formación Universitaria*, 11(1), 25-34. <https://doi.org/10.4067/S0718-50062018000100025>
- Cascella, M., Montomoli, J., Bellini, V. y Bignami, E. (2023). Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems*, 47(1), Article 33. <https://doi.org/10.1007/s10916-023-01925-4>
- Chaudhry, M. A. y Kazim, E. (2022). Artificial Intelligence in Education (AIED): a high-level academic and industry note 2021. *AI and Ethics*, 2(1), 157-165. <https://doi.org/10.1007/s43681-021-00074-z>
- Choi, E. P. H., Lee, J. J., Ho, M. H., Kwok, J. Y. Y. y Lok, K. Y. W. (2023).

- Chatting or cheating? The impacts of ChatGPT and other artificial intelligence language models on nurse education. *Nurse Education Today*, 125, Article 105796. <https://doi.org/10.1016/j.nedt.2023.105796>
- Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, 32(3), 444-452. <https://doi.org/10.1007/s10956-023-10039-y>
- Copeland, B. (2023). Artificial intelligence (AI). Definition, Examples, Types, Applications, Companies, & Facts. *Encyclopedia Britannica*. <https://www.britannica.com/technology/artificial-intelligence>
- Deike, M. (2024). Evaluating the performance of ChatGPT and Perplexity AI in Business Reference. *Journal of Business & Finance Librarianship*, 29(2), 125-154. <https://doi.org/10.1080/08963568.2024.2317534>
- Destéfano, M., Trifonova, A. y Barajas, M. (2024). Teaching AI to the next generation: A humanistic approach. *Digital Education Review*, 45, 115-123. <https://doi.org/10.1344/der.2024.45.115-123>
- Díaz, B. y Nussbaum, M. (2024). Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence. *Computers & Education*, 217, Article 105071. <https://doi.org/10.1016/j.compedu.2024.105071>
- Eysenbach, G. (2023). The role of ChatGPT, generative language models, and artificial intelligence in medical education: A conversation with ChatGPT and a call for papers. *JMIR Medical Education*, 9(1), Article e46885. <https://doi.org/10.2196/46885>
- González, R. A. y Silveira, M. H. (2022). Educación e Inteligencia Artificial: Nodos temáticos de inmersión. *EduTec: Revista Electrónica de Tecnología Educativa*, 82, 59-77. <https://doi.org/10.21556/edutec.2022.82.2633>
- Guan, C., Mou, J. y Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, 4(4), 134-147. <https://doi.org/10.1016/j.ijis.2020.09.001>
- Halter, C. P. (2020). *Quantitative Analysis: With JASP open-source software*. Independently published.
- Hirsch-Kreinsen, H. (2023). Artificial intelligence: A “promising technology”. *AI & Society*. <https://doi.org/10.1007/s00146-023-01629-w>
- Hsu, Y. C. y Ching, Y. H. (2023). Generative Artificial Intelligence in Education, Part One: the Dynamic Frontier. *TechTrends*, 67, 603-607. <https://doi.org/10.1007/s11528-023-00863-9>
- Illera, J. (2024). AI in the discourse of the relationships between technology and education. *Digital Education Review*, 45, 1-7. <https://doi.org/10.1344/der.2024.45.1-7>
- Jauhainen, J. S. y Guerra, A. G. (2023). Generative AI and ChatGPT in School Children's Education: Evidence from a School Lesson. *Sustainability*, 15(18), Article 14025. <https://doi.org/10.3390/su151814025>
- Kalota, F. (2024). A Primer on Generative Artificial Intelligence. *Education Sciences*, 14(2), Article 172. <https://doi.org/10.3390/educsci14020172>
- Karakose, T. (2023). The utility of ChatGPT in educational research—potential opportunities and pitfalls. *Educational Process: International Journal*, 12, 7-13. <https://doi.org/10.22521/edupij.2023.122.1>
- Karakose, T., Demirkol, M., Yirci, R., Polat, H., Ozdemir, T. Y. y Tülübaşı, T. A. (2023). A conversation with ChatGPT about digital leadership and technology integration: Comparative analysis based on human–AI collaboration. *Administrative*

- Sciences, 13(7), Article 157. <https://doi.org/10.3390/admsci13070157>
- Kennedy, B., Tyson, A. y Saks, E. (2023). *Public awareness of artificial intelligence in everyday activities: Limited enthusiasm in U.S. over AI's growing influence in daily life*. Pew Research Center. <https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/>
- Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course? *Physical Review Physics Education Research*, 19(1), Article 010132. <https://doi.org/10.1103/PhysRevPhysEducRes.19.010132>
- Korzynski, P., Mazurek, G., Krzypkowska, P. y Kurasniski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review*, 11(3), 25-37. <https://doi.org/10.15678/EBER.2023.110302>
- Krause-Wichmann, T., Greisel, M., Wekerle, C., Kollar, I. y Stark, R. (2023). Promoting future teachers' evidence-informed reasoning scripts: Effects of different forms of instruction after problem-solving. *Frontiers in Education*, 8, Article 1001523. <https://doi.org/10.3389/educ.2023.1001523>
- Kuhail, M. A., Alturki, N., Alramlawi, S. y Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technology*, 28, 973-1018. <https://doi.org/10.1007/s10639-022-11177-3>
- Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6, Article 100174. <https://doi.org/10.1016/j.caeo.2024.100174>
- Ling, J. y Afzaal, M. (2024). Automatic question-answer pairs generation using pre-trained large language models in higher education. *Computers and Education: Artificial Intelligence*, 6, Article 100252. <https://doi.org/10.1016/j.caeai.2024.100252>
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), Article 410. <https://doi.org/10.3390/educsci13040410>
- Mah, C. y Levine, S. (2023, February 19). How to use ChatGPT as an example machine. *Cult of Pedagogy*. <https://www.cultofpedagogy.com/chatgpt-example-machine/>
- Mao, J., Chen, B. Y. y Liu, J. C. (2024). Generative Artificial Intelligence in Education and Its Implications for Assessment. *Techtrends*, 8(1), 58-66. <https://doi.org/10.1007/s11528-023-00911-4>
- Martinez, D., Malyska, N., Streilein, B., Caceres, R., Campbell, W., Dagli, C., Gadepally, V., Greenfield, K., Hall, R., King, A., Lippmann, R., Miller, B., Reynolds, D., Richardson, F., Sahin, C., Tran, A., Trepagnier, P. y Zipkin, J. (2019). *Artificial intelligence: Short history, present developments, and future outlook*. Massachusetts Institute of Technology.
- Martínez-Abad, F. y Hernández-Ramos, J. P. (2017). Flipped Classroom con píldoras audiovisuales en prácticas de análisis de datos para la docencia universitaria: Percepción de los estudiantes sobre su eficacia. En S. Pérez Aldeguer, G. Castellano Pérez y A. Pina Calafi (Eds.), *Propuestas de Innovación Educativa en la Sociedad de la Información* (pp. 92-105). AdayaPress. <https://doi.org/10.58909/ad17267178>
- Mintz, J., Holmes, W., Liu, L. y Perez-Ortiz, M. (2023). Artificial Intelligence and K-12 Education: Possibilities, Pedagogies and Risks. *Computers in the Schools*, 40(4),

- 325-333. <https://doi.org/10.1080/07380569.2023.2279870>
- Mouta, A., Pinto-Llorente, A. M. y Torrecilla-Sánchez, E. M. (2023). Uncovering Blind Spots in Education Ethics: Insights from a Systematic Literature Review on Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, 2023. <https://doi.org/10.1007/s40593-023-00384-9>
- Nash, B. L., Hicks, T., Garcia, M., Fassbender, W., Alvermann, D., Boutelier, S., McBride, C., McGrail, E., Moran, C., O'Byrne, I., Piotrowski, A., Rice, M. y Young, C. (2023). Artificial Intelligence in English Education: Challenges and Opportunities for Teachers and Teacher Educators. *English Education*, 55(3), 201-206. <https://doi.org/10.58680/ee202332555>
- Olmos, S., Martínez-Abad, F., Torrecilla, E. M. y Mena, J. J. (2014). Análisis psicométrico de una escala de percepción sobre la utilidad de Moodle en la universidad. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 20(2), Artículo 1. <https://doi.org/10.7203/relieve.20.2.4221>
- Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*, 78(1), 84-93. <https://doi.org/10.1177/10776958221149577>
- Petricini, T. (2024). What would Aristotle do? Navigating generative artificial intelligence in higher education. *Postdigital Science and Education*, 6, 439-445. <https://doi.org/10.1007/s42438-024-00463-7>
- Raman, R., Mandal, S., Das, P., Kaur, T., Sanjanasri, J. P. y Nedungadi, P. (2024). Exploring University Students' Adoption of ChatGPT Using the Diffusion of Innovation Theory and Sentiment Analysis with Gender Dimension. *Human Behavior and Emerging Technologies*, 2024, Article 085910. <https://doi.org/10.1155/2024/3085910>
- Rebello, A. D. P., Inês, G. D. O. y Damion, D. V. (2022). The Impact of Artificial Intelligence on the Creativity of Videos. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 18(1), 1-27. <https://doi.org/10.1145/3462634>
- Reynolds, L. y McDonell, K. (2021). Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. En *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 314). Association for Computing Machinery (ACM). <https://doi.org/10.1145/3411763.3451760>
- Ruiz-Campo, S., Matias-Batalla, D., Boronat-Clavijo, B. y Acevedo-Duque, Á. (2023). Los metaversos como herramienta docente en la formación de profesores de educación superior. *RELATEC. Revista Latinoamericana de Tecnología Educativa*, 22(1), 135-153. <https://doi.org/10.17398/1695-288X.22.1.135>
- Ruiz-Rojas, L. I., Acosta-Vargas, P., De-Moreta-Llovet, J. y Gonzalez-Rodriguez, M. (2023). Empowering Education with Generative Artificial Intelligence Tools: Approach with an Instructional Design Matrix. *Sustainability*, 15(15), Article 11524. <https://doi.org/10.3390/su15151524>
- Saif, N., Khan, S. U., Shaheen, I., ALotaibi, F. A., Alnfai, M. M. y Arif, M. (2024). ChatGPT; validating Technology Acceptance Model (TAM) in education sector via ubiquitous learning mechanism. *Computers in Human Behavior*, 154, Article 108097. <https://doi.org/10.1016/j.chb.2023.108097>
- Sánchez, M. M. (2024). La inteligencia artificial como recurso docente: usos y posibilidades para el profesorado. *Educar*, 60(1), 33-47. <https://doi.org/10.5565/rev/educar.1810>

- Stöhr, C., Ou, A. W. y Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. *Computers and Education. Artificial Intelligence*, 7, Article 100259. <https://doi.org/10.1016/j.caeai.2024.100259>
- Stokel-Walker, C. y Nordeen, V. R. (2023). The promise and peril of generative AI. *Nature*, 614, 214-216. <https://doi.org/10.1038/d41586-023-00340-6>
- Surden, H. (2019). Artificial intelligence and law: An overview. *Georgia State University Law Review*, 35(4), 1305-1337. <https://readingroom.law.gsu.edu/gsulr/vol35/iss4/8>
- Thiga, M. M. (2024). Generative AI and the development of critical thinking skills. *Iconic Research and Engineering Journals*, 7(9), 83-90. <https://bit.ly/3WEyQRt>
- Trust, T. y Maloy, R. W. (2023). AI in the history classroom: Strategies for rethinking teaching and learning. *Teaching History*, 57(3), 10-15. <https://search.informit.org/doi/10.3316/informit.322842139237015>
- UNESCO. (2021). *Artificial Intelligence and Education. Guidance for Policy-makers*. United Nations Educational, Scientific and Cultural Organization. <https://doi.org/10.54675/PCSP7350>
- Williams, C. K., Tremblay, L. y Carnahan, H. (2016). It pays to go off-track: Practicing with error-augmenting haptic feedback facilitates learning of a curve-tracing task. *Frontiers in Psychology*, 7, Article 2010. <https://doi.org/10.3389/fpsyg.2016.02010>
- Yilmaz, H., Maxutov, S., Baitekova, A. y Balta, N. (2023). Student attitudes towards chat GPT: a technology acceptance model survey. *International Educational Review*, 1(1), 57-83. <https://doi.org/10.58693/ier.114>
- Zhang, C., Schießl, J., Plöbl, L., Hofmann, F. y Gläser-Zikuda, M. (2023). Acceptance of artificial intelligence among pre-service teachers: a multigroup analysis. *International Journal of Educational Technology in Higher Education*, 20, Article 49. <https://doi.org/10.1186/s41239-023-00420-7>

Fecha de recepción del artículo: 1 de junio de 2024

Fecha de aceptación del artículo: 1 de agosto de 2024

Fecha de aprobación para maquetación: 5 de septiembre de 2024

Fecha de publicación en OnlineFirst: 23 de septiembre de 2024

Fecha de publicación: 1 de enero de 2025