Revista de Educación Simón Rodríguez/ www.revistasimonrodriguez.org
Vol. 6 | No. 13| octubre 2026 – enero 2027 | ISSN: 3006-1385
https://doi.org/10.62319/simonrodriguez.v.6i13.263

Artículo de revisión

Inteligencia artificial generativa en educación superior: una revisión integrativa sobre enseñanza, aprendizaje y evaluación

Generative artificial intelligence in higher education: a systematic review on teaching, learning, and assessment

Patricia Elena Ramos La Rosa [1]

Carlos Bernardino Ruiz Huaraz [2]

Teodoro Teófanes Quispealaya Bonelly [3]

Jesus Jacobo Coronado Espinoza [4]

[1]Universidad Nacional José Faustino Sánchez Carrión. Huacho, Perú
Correo: pramos@unjfsc.edu.pe

[2]Universidad Nacional José Faustino Sánchez Carrión. Huacho, Perú
Correo: cruiz@unjfsc.edu.pe

[3]Universidad Nacional José Faustino Sánchez Carrión. Huacho, Perú
Correo: tquispealaya@unjfsc.edu

[4]Universidad Nacional José Faustino Sánchez Carrión. Huacho, Perú
Correo: jcoronado@unjfsc.edu.pe

Cómo citar:
Ramos La Rosa, P. E., Ruiz Huaraz, C. B., Quispealaya Bonelly, T. T., & Coronado Espinoza, J. J. (2026). Inteligencia artificial generativa en la educación superior: una revisión sistemática sobre docencia, aprendizaje y evaluación. *Revista de Educación Simón Rodríguez, 6*(13), 1–20. https://doi.org/10.62319/simonrodriguez.v.6i13.263

Resumen

Contexto: La expansión de la inteligencia artificial generativa (IAG) está transformando la educación superior y plantea nuevos desafíos para la docencia, el aprendizaje y la evaluación. Su incorporación ofrece oportunidades para optimizar procesos educativos, pero también genera riesgos relacionados con la dependencia tecnológica, la descarga cognitiva y la validez de los mecanismos tradicionales de evaluación. **Objetivo:** Analizar la incorporación de la inteligencia artificial generativa en la docencia, el aprendizaje y la evaluación de la educación superior. **Método:** Se realizó una revisión sistemática siguiendo el método PRISMA. Se identificaron, cribaron y seleccionaron 28 estudios publicados entre 2022 y julio de 2026, procedentes de siete bases de datos internacionales y regionales. La evidencia fue analizada considerando las principales aplicaciones, beneficios, riesgos y tendencias metodológicas relacionadas con la incorporación de la IAG en la universidad. **Resultados:** Los hallazgos evidenciaron un campo emergente, caracterizado por el predominio de encuestas transversales y experimentos de corta duración, una fuerte presencia de estudios europeos y asiáticos y una producción iberoamericana limitada. En la docencia se observó una aceptación tecnológica creciente vinculada con la necesidad de fortalecer la alfabetización digital. En el aprendizaje, las ganancias de eficiencia y personalización coexistieron con riesgos de descarga cognitiva y dependencia tecnológica. En la evaluación, las pruebas escritas tradicionales mostraron vulnerabilidades, mientras que los detectores automáticos presentaron baja fiabilidad. **Conclusiones:** La incorporación de la inteligencia artificial generativa en la educación superior requiere un rediseño pedagógico que considere sus posibilidades y limitaciones. Este proceso debe complementarse con políticas institucionales, formación docente y criterios de uso crítico y ético, orientados a preservar la autonomía intelectual, la calidad del aprendizaje y la integridad académica.

Palabras clave: Aprendizaje; Docencia; Educación superior; Evaluación; Inteligencia artificial

Abstract

Context: The rapid expansion of generative artificial intelligence (GenAI) is transforming higher education and posing new challenges for teaching, learning, and assessment. Its integration offers opportunities to optimize

educational processes but also raises concerns regarding technological dependence, cognitive offloading, and the validity of traditional assessment methods. **Objective:** To analyze the integration of generative artificial intelligence into teaching, learning, and assessment in higher education. **Method:** A systematic literature review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A total of 28 studies published between 2022 and July 2026 were identified, screened, and selected from seven international and regional databases. The evidence was analyzed in terms of the main applications, benefits, risks, and methodological trends associated with the integration of GenAI into university education. **Results:** The findings revealed an emerging research field characterized by the predominance of cross-sectional surveys and short-term experiments, a strong concentration of studies conducted in Europe and Asia, and limited scholarly output from Ibero-America. In teaching, increasing acceptance of the technology was observed, alongside a growing need to strengthen digital literacy. In learning, improvements in efficiency and personalization coexisted with risks of cognitive offloading and technological dependence. Regarding assessment, traditional written examinations exhibited vulnerabilities, while automated AI detection tools demonstrated low reliability. **Conclusions:** The integration of generative artificial intelligence into higher education requires pedagogical redesign that accounts for both its potential and its limitations. This process should be supported by institutional policies, teacher professional development, and guidelines for the critical and ethical use of GenAI, aimed at preserving intellectual autonomy, the quality of learning, and academic integrity.

Keywords: Artificial intelligence; Assessment; Higher education; Learning; Teaching

Introducción

La divulgación masiva de los modelos generativos de lenguaje marcó un punto de inflexión en la trayectoria digital de las universidades. ChatGPT sumó un millón de usuarios en cinco días y llevó un debate técnico al centro de la agenda académica global, con implicaciones para la docencia, el aprendizaje y la evaluación. Los mapeos bibliográficos confirmaron un crecimiento acelerado de publicaciones sobre Inteligencia Artificial (IA) en la educación superior, aunque con predominio teórico frente al trabajo empírico (Vélez et al., 2024). A ese desequilibrio se sumó una concentración temática en el diagnóstico y la gestión institucional, que relegó los procesos de enseñanza y de aprendizaje a un plano secundario (Crompton y Burke, 2023). Los análisis tempranos advirtieron, además, efectos sobre el estudiante aún poco documentados (Lo, 2023). Este panorama justifica examinar la base empírica disponible con un diseño transparente.

Desde la perspectiva docente, la literatura identificó apoyos para la planificación de clases, la elaboración de materiales y la retroalimentación formativa. Los análisis conceptuales vieron en los grandes modelos de lenguaje un medio para aliviar tareas repetitivas y ampliar el tiempo de interacción con el estudiante (Grassini, 2023). También las síntesis sobre ChatGPT describieron beneficios potenciales para la enseñanza personalizada y el diseño de actividades de evaluación formativa, siempre que el profesor conservase el control pedagógico de las decisiones (Baidoo y Owusu, 2023). Más allá de la eficiencia, los programas de investigación

sobre educación con IA propusieron repensar los roles del docente, de la tecnología y del alumno en escenarios híbridos de aprendizaje (Hwang et al., 2020). Tales promesas, sin embargo, exigían comprobación con datos observados, porque el entusiasmo inicial descansó más en opiniones expertas que en evidencia.

En cuanto al aprendizaje, los metaanálisis ofrecieron señales alentadoras, aunque limitadas. Una síntesis cuantitativa de estudios con chatbots educativos reportó efectos positivos moderados sobre los resultados de aprendizaje del estudiante universitario (Wu y Yu, 2024). Ese respaldo se matizó en contextos de desarrollo, donde las ganancias en rendimiento y motivación convivieron con dificultades de accesibilidad y conectividad (Deng y Yu, 2023). En un plano más amplio, las revisiones sistemáticas sobre IA en educación reconocieron oportunidades para la personalización del aprendizaje y la tutoría inteligente, junto con retos éticos, técnicos y de formación del profesorado (Chiu et al., 2023). Este acervo resultaba, no obstante, de utilidad parcial dado que la mayoría de esas síntesis reunió investigaciones previas a los sistemas generativos actuales, de modo que su alcance para explicar los efectos de la Inteligencia Artificial Generativa (IAG) permanecía incierto y exigía actualización con evidencia reciente.

Por otra parte, la evaluación concentró la mayor parte de las alarmas. El primer frente fue la integridad académica, pues se alertó del riesgo de que el estudiante entregara como propios textos generados por sistemas conversacionales (Cotton et al., 2023). Esa inquietud escaló cuando ensayos sobre evaluación universitaria decretaron el agotamiento de las pruebas escritas tradicionales y reclamaron alternativas centradas en el proceso formativo (Rudolph et al., 2023). No faltó una lectura constructiva, pues los análisis sobre evaluación en la era de la IA vieron en las capacidades de estos sistemas una oportunidad para diseñar tareas auténticas y sostenibles (Swiecki et al., 2022). Tales discusiones descansaron, pero, en argumentos y proyecciones, no en observaciones de aulas reales. De ahí que la revisión exigiera reunir estudios que documentaran decisiones institución.

En esta línea y consideradas en conjunto, esas tres vertientes mostraron un campo en formación, con resultados dispersos entre disciplinas, países y tradiciones metodológicas. Las revisiones generales describieron el debate, pero mezclaron estudios teóricos y empíricos sin distinguir bases de datos ni contextos institucionales, lo que dificultaba llevar sus conclusiones a la práctica universitaria cotidiana. A esa limitación se añadió otra geográfica pues la producción iberoamericana permaneció apenas visible en las síntesis internacionales, aunque las decisiones de política académica ya circulaban en la región. Ante esos vacíos, este artículo formuló tres preguntas de investigación: ¿qué características metodológicas presentan las investigaciones empíricas sobre IAG en la educación superior publicadas entre 2022 y 2025?; ¿qué resultados documentan esas investigaciones sobre la docencia, el aprendizaje y la evaluación?; y ¿qué retos y líneas de acción se desprenden de esos resultados para las instituciones universitarias?

Contestar esas preguntas cobra interés para directivos, docentes y estudiantes que deben tomar decisiones sobre currículo, normativa y formación profesional. La síntesis de estudios originales aporta una base documental verificable para orientar políticas de alfabetización

digital, protocolos de integridad académica y rediseño de evaluaciones en universidades iberoamericanas, donde la discusión cuenta aún con producción local limitada. La revisión atiende, en consecuencia, la necesidad académica y práctica, porque la tecnología presiona de inmediato sobre los reglamentos universitarios. Para atenderla, la presente revisión sistemática tuvo como objetivo analizar la incorporación de la IAG en la docencia, el aprendizaje y la evaluación de la educación superior.

Método

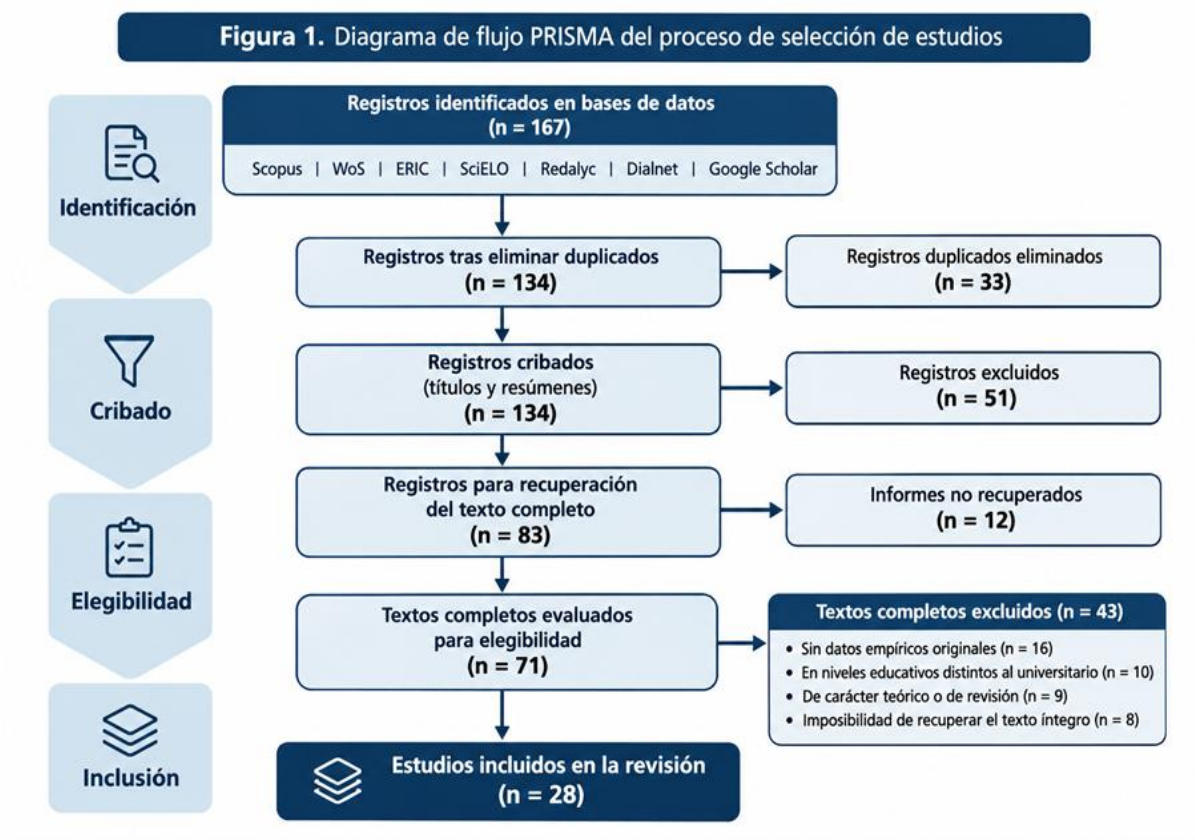
La investigación adoptó el diseño de revisión sistemática, por su capacidad para reunir en un mismo marco analítico estudios de encuesta, experimentos y aproximaciones cualitativas. El protocolo siguió las etapas clásicas de este tipo de revisión: delimitación del problema, búsqueda exhaustiva, evaluación de los datos, análisis y presentación de los resultados. La transparencia del proceso descansó en el diagrama de flujo del modelo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), que permite reconstruir cada decisión de selección. El periodo de consulta comprendió de enero de 2022 a julio de 2026, de manera que la ventana temporal cubriera la aparición y el afianzamiento de los sistemas generativos en las aulas universitarias. La unidad de análisis fue el artículo original de investigación empírica, y la unidad de contexto fue la educación superior formal.

En cuanto a la búsqueda, combinó las bases de datos de Scopus, Web of Science (WoS), ERIC, SciELO, Redalyc, Dialnet y Google Scholar para localizar textos completos accesibles. La cadena de búsqueda agrupó tres bloques de términos con operadores booleanos: inteligencia artificial generativa, ChatGPT y grandes modelos de lenguaje, en el primer bloque; educación superior, universidad y docentes o estudiantes, en el segundo; y docencia, aprendizaje y evaluación, en el tercero. La búsqueda aplicó las versiones en español y en inglés de los descriptores en cada base, con adaptación de los operadores a la sintaxis propia de cada plataforma. La primera consulta identificó 167 registros, de los cuales 33 resultaron duplicados al contrastar títulos, autores y años en una hoja de cálculo compartida entre revisores. El cribado inicial trabajó entonces con 134 registros únicos.

En vista de lo anterior, los criterios de inclusión exigieron artículos publicados entre 2022 y 2026 en revistas académicas arbitradas, con datos empíricos originales sobre IAG en contextos de educación superior, redacción en español o en inglés y texto completo disponible en acceso abierto. Los criterios priorizaron, además, investigaciones que midieran o documentaran de manera directa prácticas de docencia, procesos de aprendizaje o instrumentos de evaluación. Los criterios de exclusión descartaron textos de opinión, editoriales, revisiones literarias, comunicaciones de congreso sin arbitraje y estudios centrados en niveles educativos distintos del universitario. Conforme a esos parámetros, el cribado de títulos y resúmenes descartó 51 registros y conservó 83 documentos para su recuperación del texto completo, donde 12 informes no pudieron recuperarse. Dos fuentes simultáneas de registro documentaron cada decisión las cuales fueron una matriz de exclusión con motivos tipificados y un acta de acuerdos entre revisores.

Sumado a lo anterior, de la lectura de 71 estudios a textos completos se excluyeron 43 por cuatro motivos: ausencia de datos empíricos originales (16), concentración en niveles educativos ajenos a la universidad (10), carácter teórico o de revisión (9) e imposibilidad de recuperar el texto íntegro (8). El proceso conservó 28 investigaciones que cumplieron la totalidad de los criterios y conformaron el corpus final. La Figura 1 presenta el diagrama de flujo con las cifras de cada etapa, conforme a las directrices del modelo PRISMA. Las tres preguntas de investigación orientaron la organización del corpus pues la descripción de las características metodológicas partió del año, el país, el método y los participantes, y la agrupación de los resultados siguió las dimensiones de docencia, aprendizaje y evaluación. El registro independiente de los hallazgos transversales a las tres dimensiones preservó la trazabilidad del análisis.

Figura 1. *Diagrama de flujo del proceso de identificación, cribado, elegibilidad e incorporación de los estudios.*



La extracción de datos empleó una matriz con seis campos: identificador del estudio, autoría y país, contexto y participantes, dimensión educativa, método e instrumentos, y resultados principales. Dos fases de codificación ordenaron el material, la primera asignó cada estudio a su dimensión dominante y la segunda contrastó las categorías con las preguntas de investigación. El consenso entre revisores resolvió los desacuerdos de clasificación, menores y concentrados en estudios con dos dimensiones simultáneas, durante sesiones de trabajo conjuntas. El análisis descriptivo calculó frecuencias por año, región y tipo de método, y el análisis de contenido elaboró categorías para los resultados de cada dimensión. La matriz completa aparece en la Tabla 1, que sintetiza la totalidad del corpus y permite verificar la

correspondencia entre cada estudio, sus características metodológicas y sus resultados principales.

El rigor del procedimiento descansa en tres decisiones. La combinación de siete bases regionales e internacionales redujo el sesgo de cobertura de las fuentes anglosajonas, aunque la restricción a textos en español e inglés pudo excluir estudios publicados en otras lenguas. La exigencia de acceso abierto facilitó la verificación del texto completo por parte de cualquier lector, si bien dejó fuera investigaciones de calidad publicadas tras muros de pago. La ventana temporal de cinco años delimitó el alcance de las conclusiones a la fase temprana de adopción de los sistemas generativos y a su afianzamiento posterior. El balance resultó adecuado para los propósitos del estudio, dado que reunió la mayor parte de la producción empírica disponible.

Resultados

El corpus reunió 28 investigaciones distribuidas en cinco años: 1 trabajo data del 2022, 10 corresponden al 2023, 10 al 2024, 5 al 2025 y 2 al 2026. Estos datos denotan un crecimiento con rapidez. Dicha expansión coincide con el lanzamiento de ChatGPT y confirma la reacción inmediata de la investigación educativa ante la IAG. Unido a esto, su procedencia geográfica indica que China y Alemania aportan cinco estudios cada una; España y Turquía con dos cada uno; Colombia, Italia, Pakistán, Estados Unidos, Ghana, Suecia, Reino Unido, Australia, Polonia y Malasia registran una contribución individual; y cuatro estudios abarcan varios países o tuvieron alcance internacional. Este reparto conforma un campo con eje asiático y europeo, por lo que la extrapolación de resultados hacia contextos ausentes del corpus exige prudencia.

A este aspecto se une el análisis del contexto y los participantes, en este sentido se evidencia un predominio del estudiante en 17 estudios que centraron la muestra en universitarios, cuatro en docentes (Colombia, tres países latinoamericanos, Suecia y un equipo internacional de evaluadores), tres combinaron a ambos actores, tres analizaron materiales (evaluaciones, textos y reseñas en línea) y uno reunió participantes de niveles educativos diversos. Los tamaños oscilaron entre 24 docentes suecos y 6300 universitarios alemanes, extremos que corresponden a una entrevista cualitativa y a una encuesta nacional. Sobresalen subpoblaciones acotadas como futuros docentes de lenguas, estudiantes de programación, cursos de escritura en lengua extranjera, usuarios de generación Z y sujetos con alfabetización desigual en IA. Este panorama denota un desbalance dado que la voz estudiantil concentra la mayor evidencia y la perspectiva de los profesores, actor que decide la integración curricular de la IAG, ocupa un lugar minoritario en el corpus.

Respecto a los métodos empleados en los estudios analizados, se pudo constatar que se combinaron enfoques cuantitativos, cualitativos y mixtos. Las encuestas con modelos de aceptación tecnológica dominaron el grupo cuantitativo, como ilustra el estudio de adopción estudiantil basado en la teoría unificada de aceptación y uso de la tecnología (Strzelecki, 2023). Los diseños cualitativos aportaron profundidad interpretativa mediante entrevistas con docentes sobre sus prácticas de evaluación (Farazouli et al., 2024), y los diseños mixtos

integraron cuestionarios con grupos focales en la formación inicial de los profesores (Gibert et al., 2024). Los experimentos aleatorizados y cuasiexperimentales permitieron aislar efectos sobre variables cognitivas y motivacionales en aulas reales. El conjunto se inclinó, en suma, por las encuestas transversales, los experimentos controlados y las aproximaciones cualitativas y mixtas. Ninguna investigación reportó seguimientos superiores a un año académico, lo que acotó la capacidad del corpus para describir trayectorias formativas extendidas.

En lo que concierne a las dimensiones abordadas en los estudios del corpus, en la docencia, los resultados giraron en torno a la aceptación tecnológica, la formación profesional y las nuevas competencias. La encuesta a 1175 docentes de Ecuador, República Dominicana y Venezuela reportó aceptación generalizada de la IA y alta intención de uso, con influencia de las creencias pedagógicas (Cabero et al., 2025). Ese patrón coincidió con la distancia generacional entre estudiantes de la generación Z y docentes previos, con mayor escepticismo (Chan y Lee, 2023). Los estudios formativos documentaron que la alfabetización en IA y el dominio del diseño de instrucciones mejoraron la calidad de las interacciones académicas (Knoth et al., 2024) y sostuvieron la incorporación curricular con formación docente en cada sesión (Walter, 2024). Las percepciones docentes, no obstante, oscilaron entre el entusiasmo por la eficiencia y la preocupación por el aprendizaje de los estudiantes.

Sobre la dimensión del aprendizaje, coincidieron ganancias cognitivas y señales de dependencia. Los experimentos reportaron mejoras en el pensamiento computacional (Yilmaz y Yilmaz, 2023), en habilidades críticas y creativas (Essel et al., 2024) y en el pensamiento crítico complejo (Suriano et al., 2025). Las encuestas documentaron efectos adversos del uso intensivo como la procrastinación, pérdida de memoria y menor rendimiento (Abbas et al., 2024), y correlaciones negativas entre uso frecuente y pensamiento crítico, mediadas por la descarga cognitiva (Gerlich, 2025). Los diseños con medición de procesos precisaron que los sistemas redujeron la carga mental y la autorregulación (Stadler et al., 2024; Fan et al., 2025), beneficiaron a participantes con menor desempeño inicial (Darvishi et al., 2024) y reorientaron la escritura académica (Yan, 2023). Las percepciones agregaron valoraciones positivas de personalización (Chan y Hu, 2023) y cautela frente a sesgos y dependencia (Tlili et al., 2023).

En la dimensión evaluación, los resultados fueron contundentes respecto de la vulnerabilidad de las pruebas tradicionales. La evaluación comparativa de asignaturas de ingeniería australianas mostró que ChatGPT alcanzó calificaciones aprobatorias en gran parte de las pruebas, con desempeño variable según el formato de la tarea (Nikolic et al., 2023). La prueba sistemática de 14 detectores de textos automáticos desaconsejó su uso en decisiones disciplinarias, por la baja precisión ante traducciones y estrategias de enmascaramiento (Weber et al., 2023). La combinación de juicio docente y software detector identificó solo una parte del contenido generado (Perkins et al., 2024). A su vez, la desafección académica predijo la disposición estudiantil a delegar trabajos escritos en la IA (Playfoot et al., 2024). Los resultados, en conjunto, trasladaron la discusión desde la detección punitiva hacia el rediseño evaluativo y la formación en integridad.

En el marco de lo expuesto, la lectura transversal del corpus precisó tres patrones. El primero, situó la producción empírica en fase temprana, con predominio de encuestas transversales y experimentos de sesión única o de corta duración. El segundo mostró efectos bifrontes pues los apoyos inmediatos y la personalización convivieron con riesgos de dependencia, descarga cognitiva y simplificación del razonamiento. El tercero, ubicó la evaluación como dimensión crítica, porque la mayoría de los instrumentos tradicionales resultó susceptible de respuestas automáticas sin mecanismos de verificación fiables. La Tabla 1 sintetiza los 28 estudios con su identificador, país, contexto, dimensión, método y resultados.

Tabla 1. *Estudios sobre IAG en la educación superior considerados en la revisión (2022-2026).*

ID	Autor (año) / País	Contexto y participantes	Método e instrumentos	Dimensión	Principales resultados
1	Boude y Meza (2026) / Colombia	115 docentes de universidades colombianas.	Encuesta a docentes universitarios con análisis descriptivo.	Docencia	El profesor reconoce utilidad para planificar y producir materiales. Reclama formación continua y lineamientos institucionales de uso ético.
2	Cabero et al. (2026) / España	254 estudiantes universitarios españoles.	Encuesta ex post facto con ecuaciones estructurales.	Transversal	Las dimensiones del modelo explicaron el comportamiento de uso. Los usuarios habituales valoran la utilidad y el apoyo a tareas académicas.
3	Cabero et al. (2025) / Ecuador, Rep. Dominicana y Venezuela	1175 docentes universitarios de tres países.	Encuesta ex post facto con modelo UTAUT2.	Docencia	Aceptación generalizada de la IA y alta intención de uso. Las creencias pedagógicas constructivistas favorecieron la adopción tecnológica.
4	Fan et al. (2025) / China	117 universitarios en tarea académica con cuatro condiciones de apoyo.	Experimento aleatorizado con analíticas de aprendizaje.	Aprendizaje	ChatGPT mejoró el desempeño de tarea, aunque redujo la autorregulación y provocó pereza metacognitiva.
5	Gerlich (2025) / Alemania	666 participantes de niveles educativos diversos.	Diseño mixto: encuesta, ANOVA y entrevistas.	Aprendizaje	El uso frecuente de herramientas de IA se asoció con menor pensamiento crítico, mediado por la descarga cognitiva.
6	Kaya y Adıgüzel (2025) / Turquía	378 participantes: 346 estudiantes y 32 docentes.	Encuesta con modelo UTAUT2.	Docencia	Los docentes mostraron mayor aceptación que los estudiantes. Las expectativas de desempeño y esfuerzo predijeron la intención de uso.
7	Suriano et al. (2025) / Italia	213 universitarios en medición de pensamiento crítico	Estudio cuantitativo con análisis de mediación.	Aprendizaje	Las actitudes y la confianza hacia ChatGPT favorecieron el compromiso y el desempeño en pensamiento crítico complejo.
8	Abbas et al. (2024) / Pakistán	165 (estudio 1) y 494 (estudio 2) estudiantes universitarios.	Encuesta con modelado de ecuaciones estructurales.	Aprendizaje	El uso excesivo de ChatGPT se asoció con procrastinación, pérdida de memoria y menor rendimiento académico.

ID	Autor (año) / País	Contexto y participantes	Método e instrumentos	Dimensión	Principales resultados
9	Darvishi et al. (2024) / Estados Unidos	1625 universitarios en sistema de retroalimentación entre pares.	Experimento aleatorizado en curso real.	Aprendizaje	La asistencia de IA benefició a estudiantes con menor desempeño, retirado el apoyo, los comportamientos aprendidos decayeron.
10	Essel et al. (2024) / Ghana	125 estudiantes en curso con aula invertida.	Cuasiexperimento con pretest y postest y grupos experimental y control.	Aprendizaje	ChatGPT fortaleció el pensamiento crítico, creativo y reflexivo, con señales iniciales de dependencia del sistema.
11	Farazouli et al. (2024) / Suecia	24 docentes universitarios de ciencias sociales y humanidades.	Entrevistas cualitativas con análisis temático.	Evaluación	Los docentes modificaron pruebas domiciliarias y percibieron los chatbots, a la vez, como amenaza y recurso para la evaluación.
12	Gibert et al. (2024) / España	Estudiantes de formación inicial de profesorado de Español como lengua extranjera.	Estudio mixto con cuestionario y grupos focales.	Docencia	Percepción positiva de ChatGPT para el aprendizaje de lenguas, con demanda de formación en IA dentro del currículo docente.
13	Knoth et al. (2024) / Alemania	Estudiantes con distintos niveles de alfabetización en IA.	Entrevistas y experimentos de diseño de instrucciones.	Docencia	La alfabetización en IA determinó la calidad de las estrategias de prompting y de los resultados obtenidos.
14	Perkins et al. (2024) / Internacional	15 docentes evaluadores y 22 trabajos experimentales.	Experimento con juicio docente y detector Turnitin.	Evaluación	El detector señaló parte del contenido generado. El juicio docente reportó solo la mitad de los casos de uso indebido.
15	Playfoot et al. (2024) / Reino Unido	160 estudiantes universitarios de grado.	Encuesta transversal con escalas de desafección.	Aprendizaje	La desafección hacia el título predijo la disposición a usar IA para trabajos escritos, factor de riesgo de conducta indebida.
16	Stadler et al. (2024) / Alemania	91 universitarios asignados a ChatGPT o a un buscador.	Experimento aleatorizado con medición de carga cognitiva.	Aprendizaje	Los grandes modelos de lenguaje redujeron el esfuerzo mental, aunque disminuyeron la calidad y la diversidad del razonamiento.

ID	Autor (año) / País	Contexto y participantes	Método e instrumentos	Dimensión	Principales resultados
17	Walter (2024) / Alemania	Docentes y estudiantes de programas universitarios.	Estudio de caso exploratorio con intervención formativa.	Docencia	La instrucción en alfabetización digital, prompting y pensamiento crítico apoyó la integración pedagógica de la IA.
18	Chan y Hu (2023) / China	399 estudiantes universitarios.	Encuesta mixta con preguntas abiertas y cerradas.	Docencia y aprendizaje	Beneficios percibidos en personalización y eficiencia; preocupaciones por precisión, privacidad, ética y dependencia.
19	Chan y Lee (2023) / China	399 estudiantes de generación Z y 184 docentes.	Encuesta con preguntas abiertas y cerradas.	Docencia	Los estudiantes mostraron mayor interés en la IA generativa que sus docentes, ambas partes pidieron formación y políticas claras.
20	Chan y Zhou (2023) / China	405 estudiantes universitarios.	Encuesta con análisis factorial confirmatorio.	Aprendizaje	El valor percibido se asoció con la intención de uso. El costo percibido mostró relación negativa con la intención.
21	Nikolic et al. (2023) / Australia	Evaluaciones de 10 asignaturas de 7 universidades.	Benchmarking de respuestas de ChatGPT calificadas por docentes.	Evaluación	ChatGPT alcanzó calificaciones aprobatorias en gran parte de las evaluaciones, con desempeño variable según el formato de tarea.
22	Strzelecki (2023) / Polonia	539 estudiantes universitarios.	Encuesta con modelo UTAUT2 y ecuaciones estructurales.	Aprendizaje	El hábito y las expectativas de desempeño predijeron la intención y el uso de ChatGPT entre estudiantes.
23	Tlili et al. (2023) / Internacional	Reseñas y comentarios en línea de usuarios tempranos	Estudio de caso instrumental cualitativo.	Aprendizaje	El discurso público mostró entusiasmo educativo, con cautela frente a sesgos, dependencia y consecuencias para la evaluación.
24	von Garrel y Mayer (2023) / Alemania	6300 estudiantes de universidades alemanas.	Encuesta nacional transversal.	Aprendizaje	Dos tercios usaban herramientas de IA, sobre todo para búsquedas y comprensión. El uso en evaluaciones fue menor.

ID	Autor (año) / País	Contexto y participantes	Método e instrumentos	Dimensión	Principales resultados
25	Weber et al. (2023) / Internacional	14 detectores de textos evaluados con muestras estudiantiles.	Prueba comparativa de precisión con textos humanos y generados.	Evaluación	Ningún detector identificó textos generados de manera fiable. La traducción y el enmascaramiento redujeron la precisión.
26	Yan (2023) / China	Estudiantes de un practicante de escritura en lengua extranjera.	Estudio de caso cualitativo con registros y entrevistas.	Aprendizaje y evaluación	ChatGPT aportó ideas y vocabulario, con dudas sobre precisión, originalidad y efectos sobre el esfuerzo propio.
27	Yilmaz y Yilmaz (2023) / Turquía	45 estudiantes de programación.	Experimento con pretest y postest y grupo control	Aprendizaje	El grupo que usó ChatGPT mejoró el pensamiento computacional, la autoeficacia programadora y la motivación.
28	Mohd et al. (2022) / Malasia	302 estudiantes universitarios malasios.	Encuesta con ecuaciones estructurales y redes neuronales.	Transversal	El rendimiento esperado y la influencia social predijeron la intención de uso de chatbots educativos.

Discusión

La primera pregunta indagó por las características del campo, y las respuestas proyectan un área joven, apoyada en encuestas y concentrada en pocas regiones. Ese perfil explica la prudencia institucional que los estudios describen con sin síntesis estables, las universidades alternaron entre prohibiciones y silencios normativos. Los resultados sobre percepciones ambiguas y prácticas diversas de Sullivan et al. (2023) concuerdan con el argumento de que las instituciones necesitaban políticas de alcance institucional, por encima de sanciones puntuales. La coincidencia se extiende a las orientaciones internacionales abordadas por Miao y Holmes (2023), que pidieron regular las herramientas, proteger los datos personales y fijar límites de edad, con una visión humanista del cambio educativo. En consecuencia, la fase temprana del campo no invalida la acción institucional, sino que la vuelve necesaria, porque las decisiones tomadas durante la incertidumbre fijaron precedentes duraderos para la gobernanza académica.

Asimismo, otro grupo de resultados concernió a la integridad académica. La revisión documentó pruebas tradicionales vulnerables ante respuestas automáticas y detectores poco fiables, de modo que el control disciplinario pierde capacidad frente al fenómeno. Esta situación guarda sintonía con la tesis del posplagio de Eaton (2023), que plantea una gradación de usos legítimos y contrarios a la honestidad académica y exige reformular las categorías mismas del plagio. Los resultados también refuerzan la advertencia sobre los sesgos de los detectores que realizan Liang et al. (2023), que penalizan de manera desproporcionada los textos de autores no nativos del inglés, con riesgos de injusticia procedimental en universidades multilingües. De ahí que los datos respalden decisiones de rediseño antes que tecnología de vigilancia, dado que la sanción sin validación técnica agrava la inequidad y no protege el aprendizaje que las evaluaciones pretenden certificar.

Por otro lado, las ganancias cognitivas a corto plazo convivieron con costos de autorregulación, y esta tensión constituye el aporte más relevante de la revisión. Los experimentos del corpus mostraron mejoras inmediatas de desempeño junto con descensos del esfuerzo mental y del monitoreo metacognitivo, un patrón que Zhai et al. (2024) como revisores de sistemas de diálogo vincularon con dependencia excesiva y atrofia gradual de habilidades sin andamiaje. El patrón coincide con las notas cautelarias sobre la desvalorización del esfuerzo intelectual y la delegación del pensamiento en máquinas que describen Selwyn (2022) y que tienen efectos de largo alcance sobre la formación profesional. La coincidencia denota lecturas longitudinales de los beneficios reportados, porque la satisfacción inmediata no garantiza aprendizaje transferible. La agenda formativa debería, por consiguiente, enseñar a delegar tareas de manera deliberada y a reasumir el control cognitivo en momentos pedagógicos planificados.

De igual manera, otro conjunto de resultados versó sobre competencias docentes. Los estudios del corpus asociaron la calidad de las interacciones académicas con la alfabetización en IA y con el dominio del diseño de instrucciones, lo que sitúa la formación profesional en el centro de cualquier estrategia institucional. Ese resultado se alinea con las capacidades que Markauskaite et al. (2022) definieron para aprender en un mundo con IA, entre ellas el juicio

evaluativo y la autorregulación asistida. Refuerza, a su vez, el diagnóstico realizado por Holmes y Tuomi (2022) de que los sistemas educativos llegaron a la era generativa con preparación docente desigual y con competencias tecnológicas concentradas en grupos reducidos. Para las universidades iberoamericanas, la comparación propone programas de desarrollo profesional con casos de aula y criterios de calidad, y no catálogos de herramientas.

En otro aspecto, la fragilidad de las pruebas escritas ante respuestas automáticas invita a leer la crisis de la evaluación como crisis de sentido y no solo de vigilancia. Los discursos del campo de Bearman et al. (2023) dividieron a los actores entre optimistas y pesimistas, y ese marco dicotómico impidió apreciar los usos pedagógicos intermedios documentados por el corpus. La comparación refleja que la discusión sobre fines debe preceder a la discusión sobre instrumentos, tal como propuso la agenda de investigación para la educación terciaria realizada por Lodge et al. (2023), que pidió estudiar el cambio de evaluación dentro de sistemas de enseñanza completos. Los datos de la revisión apuntan en esa dirección pues las tareas que exigían juicio personal, contexto local y proceso visible resistieron mejor la delegación automática. El rediseño, por tanto, aparece como tarea pedagógica antes que tecnológica.

Sumado a lo anterior, los experimentos del corpus aportaron indicios para un diseño instruccional deliberado de la IAG. Los efectos positivos aparecieron cuando las herramientas se incorporaron con andamiajes, metas claras y mediciones de proceso, y se diluyeron en los usos espontáneos sin estructura pedagógica. Ese patrón guarda relación con la advertencia histórica sobre la escritura automatizada realizada por Sharples (2022), que anticipó una delegación creciente de la autoría y exigió nuevas formas de garantía de la autenticidad académica. También conecta con la propuesta de Denny et al. (2024) de incorporar los modelos generativos en la enseñanza de la computación mediante tareas que simulen escenarios profesionales reales. La comparación permite formular una hipótesis formativa que indica que la IAG amplifica la calidad del diseño pedagógico previo, de manera que enriquece los escenarios bien construidos y agrava los escenarios débiles.

No menos importante, la concentración geográfica del corpus constituye una limitación visible y, a la vez, un resultado con implicaciones regionales. Once de los 28 estudios procedieron de Europa y 9 de Asia y Oceanía, y la producción iberoamericana se redujo a dos investigaciones: una encuesta multinacional a docentes y un estudio mixto sobre formación inicial. La distribución replica el mapa global del uso de ChatGPT en la educación superior realizado por Ansari et al. (2024), cuya cartografía documentó mayor densidad de estudios en Norteamérica, Asia y Europa, con testimonios escasos de América Latina y África. La comparación refuerza la necesidad de guías contextualizadas, como la introducción regional al uso de ChatGPT elaborada para los actores de la educación superior de América Latina y el Caribe (Sabzalieva y Valentini, 2023). La escasez local no implica ausencia de uso, sino de datos con los cuales gobernarlo.

De este modo, las tres preguntas quedaron respondidas, y la agenda exige precisión metodológica. El corpus adoleció de estudios longitudinales, de réplicas entre países y de investigaciones centradas en docentes, un déficit que las metasíntesis del campo atribuyeron

a la novedad del fenómeno y a la presión editorial por publicaciones rápidas (Bond et al., 2024). A esa carencia metodológica se suma la necesidad de principios éticos operativos, que conviertan declaraciones genéricas en criterios verificables para el diseño de tareas, la evaluación y la gestión de datos estudiantiles (Nguyen et al., 2023). La presente revisión presenta, además, limitaciones propias como la restricción a textos en español e inglés, la exigencia de acceso abierto y la ventana temporal de cinco años acotaron el universo documental, por lo que los resultados deben leerse como fotografía de la fase temprana y no como cuadro definitivo.

Conclusiones

Esta revisión describe la incorporación de la IAG en la docencia, el aprendizaje y la evaluación de la educación superior. Los estudios analizados muestran un campo joven, con predominio de encuestas transversales y experimentos de corta duración, fuerte presencia europea y asiática y producción iberoamericana reducida. En la docencia, la aceptación tecnológica avanza junto con demandas de formación en alfabetización digital y diseño de instrucciones. En el aprendizaje, las ganancias de eficiencia y personalización conviven con riesgos de descarga cognitiva y de debilitamiento de la autorregulación. En la evaluación, las pruebas escritas tradicionales resultan vulnerables y los detectores automáticos no ofrecen garantías suficientes, de modo que el rediseño pedagógico se perfila como la respuesta más sólida disponible.

En sintonía con lo anterior, las implicaciones prácticas orientan tres acciones simultáneas para las universidades. Las instituciones deben aprobar políticas que definan usos legítimos, protejan los datos personales y establezcan criterios de integridad aplicables a cada facultad. Los programas de desarrollo profesional deben incorporar la alfabetización en IA como competencia docente transversal, con casos de aula, criterios de calidad de las instrucciones y espacios de reflexión sobre la delegación cognitiva. Las unidades académicas deben rediseñar la evaluación hacia tareas auténticas, procesos visibles y acompañamiento continuo, y aprovechar los sistemas generativos para la retroalimentación formativa bajo supervisión docente. Esa triple vía evita el rechazo acrítico y la adopción entusiasta sin criterios. Futuras investigaciones deberían desarrollar diseños longitudinales, réplicas entre países iberoamericanos y estudios centrados en docentes, con indicadores verificables de uso, aprendizaje e integridad académica.

Acerca de

Contribución de los autores: Los autores participaron en la conceptualización del estudio, el diseño metodológico, la búsqueda y selección documental, la extracción y análisis de la información, la redacción del manuscrito y la revisión crítica de su contenido.

Financiamiento: Los autores declaran que no recibió financiamiento externo para esta investigación.

Conflicto de interés: Los autores declaran no tener conflicto de intereses.

Certificación ética: Por tratarse de una revisión sistemática de estudios publicados, se debe garantizar la integridad académica, la transparencia en la selección y análisis de la evidencia, así como el reconocimiento y la citación adecuada de las fuentes consultadas.

Historia del artículo: Artículo recibido: 16 de septiembre 2026 | Arbitrado: 30 de septiembre 2026 | Aceptado: 15 de octubre 2026 | Publicado: 17 de noviembre 2026

Referencias

Abbas, M., Jam, F. A., y Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education, 21*, 10. https://doi.org/10.1186/s41239-024-00444-7

Ansari, A. N., Ahmad, S., y Bhutta, S. M. (2024). Mapping the global evidence around the use of ChatGPT in higher education: A systematic scoping review. *Education and Information Technologies, 29*, 11281-11321. https://doi.org/10.1007/s10639-023-12223-4

Baidoo, D., y 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 AI, 7*(1), 52-62. https://doi.org/10.61969/jai.1337500

Bearman, M., Ryan, J., y Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education, 86*(2), 369-385. https://doi.org/10.1007/s10734-022-00937-2

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., y Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, 4. https://doi.org/10.1186/s41239-023-00436-z

Boude, O. R., y Meza, J. (2026). Percepciones docentes sobre la integración de la inteligencia artificial generativa en la educación superior. *Formación universitaria, 19*(3), 213-222. https://doi.org/10.4067/s0718-50062026000300213

Cabero, J., Barroso, J., Guillén, F. D., y Palacios, A. (2025). Creencias pedagógicas docentes y su aceptación de la inteligencia artificial en la educación superior: un estudio comparativo entre países. *Aula Abierta, 54*(3), 257-268. https://doi.org/10.17811/RIFIE.21273

Cabero, J., Palacios, A., Llorente, C., y Barroso, J. (2026). Aceptación de ChatGPT en Educación Superior: actitudes y percepciones del modelo UTAUT2. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 24*(1). https://doi.org/10.15366/reice2026.24.1.001

Chan, C. K. Y., y Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, 43. https://doi.org/10.1186/s41239-023-00411-8

Chan, C. K. Y., y Lee, K. K. W. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and Millennial generation teachers? Smart Learning Environments, 10, 60. https://doi.org/10.1186/s40561-023-00269-3

Chan, C. K. Y., y Zhou, W. (2023). An expectancy value theory (EVT) based instrument for measuring student perceptions of generative AI. *Smart Learning Environments, 10*, 64. https://doi.org/10.1186/s40561-023-00284-4

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., y Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Cotton, D. R. E., Cotton, P. A., y Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Crompton, H., y Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20*, 22. https://doi.org/10.1186/s41239-023-00392-8

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., y Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education, 210*, 104967. https://doi.org/10.1016/j.compedu.2023.104967

Deng, X., y Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability, 15*(4), 2940. https://doi.org/10.3390/su15042940

Denny, P., Prather, J., Becker, B. A., Finnie, J., Hellas, A., Leinonen, J., Luxton, A., Reeves, B. N., Santos, E. A., y Sarsa, S. (2024). Computing education in the era of generative AI. *Communications of the ACM, 67*(2), 56-67. https://doi.org/10.1145/3624720

Eaton, S. E. (2023). Postplagiarism: Transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. *International Journal for Educational Integrity, 19*, 23. https://doi.org/10.1007/s40979-023-00144-1

Essel, H. B., Vlachopoulos, D., Essuman, A. B., y Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence, 6*, 100198. https://doi.org/10.1016/j.caeai.2023.100198

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., y Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489-530. https://doi.org/10.1111/bjet.13544

Farazouli, A., Cerratto, T., Bolander, K., y McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education, 49*(3), 363-375. https://doi.org/10.1080/02602938.2023.2241676

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006

Gibert, I., Bellot, A. R., y Iglesia, S. (2024). Integración de ChatGPT en la formación inicial de profesorado de ELE: percepciones sobre el uso de la IA en el contexto académico. *Doblele: Revista de Lengua y Literatura, 10*(10), 82-100. https://doi.org/10.5565/rev/doblele.148

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences, 13*(7), 692. https://doi.org/10.3390/educsci13070692

Holmes, W., y Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education, 57*, 542-570. https://doi.org/10.1111/ejed.12533

Hwang, G. J., Xie, H., Wah, B. W., y Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/https://doi.org/10.1016/j.caeai.2020.100001

Kaya, M. H., y Adıgüzel, T. (2025). Exploring the acceptance of ChatGPT in higher education: A comprehensive quantitative study of university students and faculty. *Frontiers in Education, 10*, 1652292. https://doi.org/10.3389/feduc.2025.1652292

Knoth, N., Tolzin, A., Janson, A., y Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 6*, 100225. https://doi.org/10.1016/j.caeai.2024.100225

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., y Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Lodge, J. M., Thompson, K., y Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology, 39*(1), 1-8. https://doi.org/10.14742/ajet.8695

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Croft, D., y Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? Computers and Education: Artificial Intelligence, 3, 100056. https://doi.org/10.1016/j.caeai.2022.100056

Miao, F., y Holmes, W. (2023). Guía para el uso de IA generativa en educación e investigación. UNESCO. https://doi.org/10.54675/EWZM9535

Mohd, N. I., A. Iahad, N., Yusof, A. F., y A. Al-Sharafi, M. (2022). AI-Based Chatbots Adoption Model for Higher-Education Institutions: A Hybrid PLS-SEM-Neural Network Modelling Approach. *Sustainability*, *14*(19), 12726. https://doi.org/10.3390/su141912726

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., y Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies, 28*(9), 4221-4241. https://doi.org/10.1007/s10639-022-11316-w

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., y Sandison, C. (2023). ChatGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education*, *48*(4), 559-614. https://doi.org/10.1080/03043797.2023.2213169

Perkins, M., Roe, J., Postma, D., McGaughran, J., y Hickerson, D. (2024). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. *Journal of Academic Ethics, 22*(1), 89-113. https://doi.org/10.1007/s10805-023-09492-6

Playfoot, D., Quigley, M., y Thomas, A. G. (2024). Hey ChatGPT, give me a title for a paper about degree apathy and student use of AI for assignment writing. *The Internet and Higher Education, 62*, 100950. https://doi.org/10.1016/j.iheduc.2024.100950

Rudolph, J., Tan, S., y Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching, 6(1), 342-363. https://edu.usn.no/getfile.php/13759799-1678453059/edu.usn.no/ChatGPT%20-%20Bullshit%20spewer%20or%20the%20end%20of%20traditional%20assessments%20in%20higher%20education.pdf

Sabzalieva, E., y Valentini, A. (2023). ChatGPT e inteligencia artificial en la educación superior: Guía de inicio rápido. UNESCO IESALC. https://unesdoc.unesco.org/ark:/48223/pf0000385146_spa

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education, 57*(4), 620-631. https://doi.org/10.1111/ejed.12532

Sharples, M. (2022). Automated essay writing: An AIED opinion. *International Journal of Artificial Intelligence in Education, 32*(4), 1119-1126. https://doi.org/10.1007/s40593-022-00300-7

Stadler, M., Bannert, M., y Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, 108386. https://doi.org/10.1016/j.chb.2024.108386

Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments, 32*(9), 5142-5155. https://doi.org/10.1080/10494820.2023.2209881

Sullivan, M., Kelly, A., y McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching, 6*(1), 31-40. https://www.researchgate.net/publication/369378950_ChatGPT_in_higher_education_Considerations_for_academic_integrity_and_student_learning

Suriano, R., Plebe, A., Acciai, A., y Fabio, R. A. (2025). Student interaction with ChatGPT can promote complex critical thinking skills. *Learning and Instruction, 95*, 102011. https://doi.org/10.1016/j.learninstruc.2024.102011

Swiecki, Z., Khosravi, H., Chen, G., Martinez, R., Lodge, J. M., Milligan, S., Selwyn, N., y Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 3*, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., y Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*, 15. https://doi.org/10.1186/s40561-023-00237-x

Vélez, R., Muñoz, D., Leal, P., y Ruiz, A. (2024). Uso de inteligencia artificial en educación superior y sus implicancias éticas: Mapeo sistemático de literatura. *Hachetetepé. Revista Científica de Educación y Comunicación, (28), 1105*. https://doi.org/10.25267/Hachetetepe.2024.i28.1105

von Garrel, J., y Mayer, J. (2023). Artificial intelligence in studies-use of ChatGPT and AI-based tools among students in Germany. *Humanities and Social Sciences Communications, 10*, 799. https://doi.org/10.1057/s41599-023-02304-7

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. *International Journal of Educational Technology in Higher Education, 21*, 15. https://doi.org/10.1186/s41239-024-00448-3

Weber, D., Anohina, A., Bjelobaba, S., Foltýnek, T., Guerrero, J., Popoola, O., Šigut, P., y Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, 26. https://doi.org/10.1007/s40979-023-00146-z

Wu, R., y Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*, 10–33. https://doi.org/10.1111/bjet.13334

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*(11), 13943-13967. https://doi.org/10.1007/s10639-023-11742-4

Yilmaz, R., y Yilmaz, F. G. K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence, 4*, 100147. https://doi.org/10.1016/j.caeai.2023.100147

Zhai, C., Wibowo, S., y Li, L. D. (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

Este artículo está bajo la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0).