

VENTAJAS Y DESVENTAJAS DE LA IA EN LA EDUCACIÓN: EL CASO DE LOS TUTORES INDIVIDUALIZADOS GUIADOS POR LA IA

Ismael Sanz

URJC, Funcas y London School of Economics

Resumen

La irrupción de la inteligencia artificial (IA) generativa en la educación ha abierto nuevas posibilidades para la personalización del aprendizaje a través de tutores individualizados guiados por IA. Este artículo revisa la evidencia empírica más reciente sobre el impacto de estas herramientas en contextos escolares reales, haciendo hincapié en los efectos sobre el alumnado con bajo rendimiento y en entornos con recursos limitados. A partir del análisis de ensayos aleatorizados y estudios de caso internacionales, se examinan tanto las ventajas potenciales—como la capacidad para ofrecer apoyo personalizado a gran escala y mejorar la equidad educativa— como los riesgos asociados a su implementación: desinformación, sesgos, pereza cognitiva o dependencia tecnológica. El artículo ilustra estos retos y oportunidades mediante el análisis de experiencias recientes, como el uso de Khanmigo o el despliegue de tutores conversacionales en contextos de bajos ingresos, y concluye con recomendaciones para integrar estas tecnologías bajo criterios pedagógicos y de equidad.

Palabras clave: inteligencia artificial en educación, tutoría personalizada, IA generativa, equidad educativa, aprendizaje adaptativo, riesgos de la IA, innovación educativa.

Abstract

The emergence of generative artificial intelligence (AI) in education has opened up new possibilities for personalized learning through AI-guided individualized tutoring. This article reviews the latest empirical evidence on the impact of these tools in real-life school settings, with a particular focus on the effects on low-performing students and those in resource-limited environments. Based on the analysis of randomized controlled trials and international case studies, the article examines both the potential benefits—such as the ability to offer large-scale personalized support and improve educational equity—and the risks associated with their implementation: misinformation, bias, cognitive laziness, and technological dependence. The article illustrates these challenges and opportunities by analyzing recent experiences, such as the use of Khanmigo or the deployment of conversational tutors in low-income settings, and concludes with recommendations for integrating these technologies under pedagogical and equity criteria.

Keywords: artificial Intelligence in education, personalized tutoring, generative AI, educational equity, adaptive learning, AI risks, educational innovation.

JEL classification: I21, I28, O33.

I. INTRODUCCIÓN

La inteligencia artificial (IA) tiene el potencial de transformar profundamente los procesos educativos, especialmente mediante el uso de tutores personalizados que adaptan el contenido al progreso de cada estudiante (Sevilla *et al.*, 2024). Para comprender el papel de estas nuevas

herramientas, es necesario situarlas en el contexto más amplio de la incorporación de la tecnología en la educación. Durante las últimas dos décadas, se han desarrollado múltiples iniciativas con resultados mixtos. Como señalan Escueta *et al.* (2020), si bien ciertos programas—especialmente aquellos basados en aprendizaje asistido por ordenador (CAL, por sus siglas en inglés)— han mostrado

efectos positivos en contextos concretos, muchas otras intervenciones han fracasado por falta de coherencia pedagógica o implementación deficiente.

Además, recientes estudios han puesto de relieve que el uso intensivo y desregulado de la tecnología puede tener efectos adversos sobre el rendimiento. Gorjón y Osés (2022), utilizando datos de PISA 2018, (Programme for International Student Assessment, en inglés) muestran que el alumnado que hace un uso muy intensivo de las TIC (tecnologías de la Información y la comunicación) en el aula —más de 1 o 2 veces por semana— obtiene resultados significativamente inferiores en matemáticas, lectura y ciencias incluso después de controlar por sus características socioeconómicas. La asociación entre el empleo de herramientas digitales y el rendimiento académico tiene una forma de “U” invertida, con un impacto positivo al principio con poco uso, pero negativo cuando se supera el óptimo. Esta relación refuerza la idea de que no es la cantidad de tecnología lo que importa, sino su uso pedagógicamente guiado. En este mismo número de *Papeles de Economía Española*, Marcenaro y López (2025) usan los datos de PISA de 2015, 2018 y 2022 para concluir también que el uso de recursos digitales en educación secundaria puede tener impactos negativos sobre el rendimiento académico cuando no está debidamente regulado. Según estos autores, el impacto de las TIC depende menos de su disponibilidad que de su integración pedagógica efectiva y de la existencia de un entorno regulador que guíe su utilización.

La irrupción de la IA generativa plantea, en este marco, un salto cualitativo. Por primera vez, la tecnología no solo distribuye contenido, sino que genera interacciones en lenguaje natural, ofreciendo explicaciones, retroalimentación y acompañamiento adaptativo en tiempo real. Cuevas-Ruiz et al. (2025) subrayan que estos modelos de lenguaje representan una evolución en la interfaz entre alumnos, docentes y contenidos. Su capacidad para generar tutorías conversacionales y personalizadas puede marcar un punto de inflexión, especialmente si se integran en un marco pedagógico explícito.

Esta transformación ocurre, además, en un momento de presión estructural en los sistemas

educativos. Los docentes señalan una sobrecarga creciente de tareas administrativas y dificultades para atender adecuadamente a la diversidad del alumnado (OCDE, 2023). En este contexto, la IA puede convertirse en una herramienta de apoyo, no como sustituto, sino como catalizador para liberar tiempo docente y ampliar su capacidad de instrucción.

Este artículo analiza el potencial de la IA como guía de tutorías individualizadas con una estructura en seis apartados. En la segunda sección, se contextualizan las aplicaciones y promesas de la IA generativa en el ámbito educativo, revisando la evidencia reciente sobre su potencial transformador. En tercer lugar, se analizan los resultados de estudios empíricos que evalúan el impacto de tutores individualizados guiados por IA, prestando especial atención a su efecto sobre el aprendizaje de alumnos con bajo rendimiento, y sintetizando las ventajas que estas herramientas pueden ofrecer cuando se integran adecuadamente en el entorno escolar. En la cuarta sección, se abordan los riesgos asociados al uso de IA en educación, incluyendo la desinformación, los sesgos, la delegación excesiva del esfuerzo cognitivo o la dependencia tecnológica. En el quinto apartado, se presenta el caso de Khanmigo como ejemplo de integración pedagógica de un tutor conversacional basado en IA, con implicaciones prácticas para su adopción en sistemas escolares. La sección sexta presenta las principales conclusiones del artículo.

II. APLICACIONES, PROMESAS Y DESAFÍOS DE LA IA GENERATIVA EN LAS AULAS

La incorporación de tecnología en los sistemas escolares no es un fenómeno reciente. Desde hace décadas, Gobiernos e instituciones han invertido en dispositivos, plataformas y programas digitales con la expectativa de mejorar el aprendizaje. Escueta et al. (2020) señalan que el simple acceso a tecnología no es suficiente para mejorar el aprendizaje. Muchas intervenciones fracasan cuando se enfocan únicamente en el componente técnico, sin un diseño pedagógico sólido ni un rol claro para el profesorado.

Además, la evidencia acumulada sugiere que un uso excesivo o poco estructurado de las TIC puede tener efectos contraproducentes. El trabajo de Gorjón y Osés (2022) muestra que los estudiantes que hacen un uso muy intensivo de la tecnología en el aula —en tareas repetitivas, apps o navegación para actividades escolares— obtienen resultados notablemente inferiores en matemáticas, lectura y ciencias incluso tras tener en cuenta las características socioeconómicas. Este resultado pone de manifiesto que la tecnología educativa, lejos de ser una panacea, puede convertirse en un factor de distracción o dependencia si no se integra cuidadosamente en un marco pedagógico.

En este contexto, la aparición de la IA generativa —modelos como GPT-4 capaces de generar texto, mantener conversaciones y resolver problemas complejos— representa un cambio de paradigma. Estas herramientas permiten personalizar la interacción con los estudiantes, adaptando el contenido, el ritmo y el tipo de apoyo en función de sus respuestas. Sevilla *et al.* (2024) señalan que este tipo de tutoría conversacional automatizada puede facilitar la práctica docente, el desarrollo de habilidades metacognitivas y la activación de conocimientos previos, siempre que el diseño de la interfaz y los principios pedagógicos sean coherentes.

Lo novedoso no es solo la tecnología, sino su potencial para simular aspectos clave de la tutoría humana a gran escala. Esta capacidad resulta especialmente relevante en un contexto donde, según la OCDE (2023), más del 40 por 100 del profesorado declara tener dificultades para atender adecuadamente a estudiantes con distintos niveles de rendimiento en la misma aula. El potencial de la IA generativa para ofrecer un acompañamiento personalizado podría contribuir a aliviar esta presión, especialmente en entornos con ratios de alumnos por clase elevadas o escasez de recursos humanos especializados, como ocurre en el caso de docentes de matemáticas, física o informática.

La OCDE realizó en 2023 una de las primeras aproximaciones estructuradas a los cambios que puede implicar la IA generativa para la enseñanza, el aprendizaje y el rol docente, tanto dentro como fuera del aula. El Informe de este organismo inter-

nacional titulado *Generative AI in the classroom: From hype to reality* analiza el potencial transformador de los modelos de lenguaje generativos en los sistemas escolares. La OCDE (2023) plantea que, aunque la educación ha sido tradicionalmente uno de los sectores más rezagados en la adopción tecnológica, la aparición de modelos de uso libre como ChatGPT —de acceso universal y capacidad creciente— podría representar un punto de inflexión, más allá de anteriores promesas incumplidas (*blockchain*, realidad virtual, etc.). La hipótesis central es que, en combinación con otras tecnologías, la IA generativa puede contribuir a reconfigurar la experiencia educativa hacia prácticas más flexibles, individualizadas y colaborativas, si bien su éxito dependerá de cómo se integre en los marcos pedagógicos existentes.

El informe estructura sus aportaciones en torno a cinco dimensiones clave de la práctica docente: 1) retroalimentación, 2) interacción, 3) compromiso cognitivo, 4) contenido disciplinar, y 5) desarrollo socioemocional. Para cada una de ellas, se analiza el potencial de la IA generativa según el modelo SAMR (*substitution, augmentation, modification, redefinition*), identificando si su uso se limita a sustituir tareas, las mejora funcionalmente, permite rediseñarlas o introduce nuevas prácticas antes inimaginables.

Por ejemplo, en el ámbito de la retroalimentación, la IA generativa podría ayudar a los docentes a proporcionar comentarios más frecuentes, individualizados y específicos, gracias a su capacidad para analizar el trabajo del alumnado, detectar errores, ofrecer sugerencias y generar explicaciones adaptadas. Esta capacidad se podría traducir en una retroalimentación más oportuna, con efectos positivos tanto en el rendimiento como en la motivación.

En el caso del compromiso cognitivo, la OCDE (2023) destaca que la IA puede facilitar la generación de representaciones múltiples de un mismo contenido (gráficos, esquemas, analogías, etc.), así como contextualizar el aprendizaje en ejemplos significativos y cercanos a los intereses del alumnado. Además, puede contribuir a desarrollar habilidades metacognitivas, al invitar a los estudiantes

a reflexionar sobre su comprensión, identificar lagunas de conocimiento y planificar su propio proceso de aprendizaje. Otro de los ámbitos donde hay un elevado potencial es en la personalización del aprendizaje. Mediante sistemas de análisis de datos y aprendizaje adaptativo, la IA generativa puede facilitar trayectorias individualizadas de progresión, ajustadas a la velocidad y estilo de aprendizaje de cada alumno, permitiendo que los estudiantes avancen según su dominio real del contenido, en lugar de seguir un ritmo uniforme para toda la clase. Ciertos grupos tradicionalmente rezagados por el enfoque *one-size-fits-all* podrían beneficiarse especialmente de estas tecnologías. Entre ellos se encuentran estudiantes con necesidades educativas especiales, alumnos inmigrantes con barreras lingüísticas, residentes en zonas rurales con recursos limitados, o aquellos con bajo nivel socioeconómico. Para estos colectivos, un tutor personalizado basado en IA podría suponer un refuerzo adaptado a sus circunstancias, accesible y constante, que complementa el trabajo docente.

Fuera del aula, la IA generativa puede ser un motor de aprendizaje autónomo. Puede facilitar el desarrollo de tareas cognitivamente exigentes (resúmenes, esquemas, simulaciones), ofrecer soporte lingüístico en redacción o lectura, fomentar la indagación sobre temas de interés y actuar como un asistente que estimula la curiosidad del estudiante (OCDE, 2023). Una capacidad de responder de forma inmediata, flexible y a demanda que puede ayudar al alumnado en el proceso educativo.

En este sentido, la OCDE (2023) concluye con una batería de recomendaciones para sistemas educativos, entre las que destacan: 1) formar a los docentes en el uso pedagógico, ético y eficaz de la IA; 2) diseñar entornos de evaluación que minimicen el plagio y promuevan el pensamiento crítico; 3) desarrollar estándares de calidad y protección de datos; y 4) fomentar la cocreación de contenidos con apoyo de la IA, pero bajo supervisión humana. La incorporación de la IA generativa en los sistemas escolares no puede entenderse como una simple cuestión técnica. Es una cuestión de *teachology*, en la que el diseño pedagógico, la profesionalidad docente y la arquitectura institucional serán determinantes. El potencial de estas herramientas para

democratizar el aprendizaje, reducir desigualdades y mejorar resultados es considerable, pero su realización dependerá del juicio humano, la formación docente y la gobernanza educativa.

En relación a las posibilidades de individualización de enseñanza que ofrece la tecnología, vamos a comenzar en esta sección revisando la evidencia empírica que analiza las tutorías en pequeños grupos incluso sin uso de la IA, para continuar después en la siguiente sección centrándonos en las que están guiadas por la IA. Una experiencia especialmente relevante en el contexto de tutorías remotas en pequeños grupos es la documentada por Fesler et al. (2023), quienes evaluaron el programa Air Tutors en varios centros escolares del estado de Texas (Estados Unidos). Esta intervención se centró en estudiantes de 4.º a 6.º grado con bajo rendimiento en matemáticas, la mayoría de ellos pertenecientes a colectivos históricamente desfavorecidos: un 67 por 100 del alumnado era hispano y un 21 por 100 afroamericano.

El programa consistía en sesiones semanales de tutoría *online* —alrededor de tres horas semanales durante un período de entre 15 y 21 semanas— impartidas por tutores altamente cualificados a través de videollamadas, en grupos reducidos de tres a cuatro estudiantes. Las sesiones incluían el uso de una pizarra interactiva y materiales didácticos digitales, diseñados para fomentar la participación activa y el compromiso del alumnado con la tarea, con énfasis en la construcción de una relación significativa entre tutor y alumno.

El impacto se evaluó mediante un diseño experimental aleatorizado estratificado, con estimación de efectos a través de modelos de regresión y análisis bayesiano complementario. Los resultados muestran mejoras estadísticamente significativas en el rendimiento académico en matemáticas, medido mediante el test estandarizado *MAP*, (*measures of academic progress*, en inglés) con un efecto promedio de 0,13 desviaciones estándar para el conjunto de escuelas participantes. Este efecto se duplicó en la escuela media donde las sesiones se realizaron durante el horario lectivo, alcanzando 0,26 d.e., lo que sugiere que la integración curricular del programa refuerza su eficacia.

Además del rendimiento académico, también se observaron mejoras en indicadores psicosociales, como el aumento de la confianza en matemáticas y el sentimiento de pertenencia escolar, recogidas mediante encuestas. Los estudiantes indicaron que experimentaron una mejora en la percepción de la relación con su tutor, lo cual se interpreta como un elemento clave en la eficacia del programa. El análisis cualitativo de los datos sugiere que esta mejora en la confianza y en la percepción de apoyo contribuye a una mayor implicación en la tarea y persistencia frente a las dificultades.

Uno de los resultados más interesantes del estudio es que los efectos positivos se produjeron sin necesidad de introducir una nueva tecnología disruptiva, sino a través del uso intensivo, estructurado y pedagógicamente guiado de recursos ya disponibles —como las plataformas de videoconferencia, materiales didácticos digitales y pizarras virtuales—, en combinación con tutores cuidadosamente seleccionados y formados para adaptarse a las necesidades del alumnado.

Esta aproximación refuerza la idea de que el valor añadido de las tutorías personalizadas no reside únicamente en la sofisticación tecnológica de las herramientas utilizadas, sino en la calidad de la interacción educativa, la regularidad del acompañamiento, la pertinencia de los contenidos, y su integración con el currículo escolar. Aunque el programa Air Tutors no se basa en IA generativa, su diseño comparte varios principios clave con las intervenciones que incorporan IA: personalización del apoyo, retroalimentación inmediata, foco en el alumnado rezagado, y un esfuerzo explícito por reforzar la motivación intrínseca del estudiante.

En este sentido, el estudio de Fesler *et al.* (2023) constituye un punto de referencia útil para pensar en cómo integrar herramientas basadas en IA en modelos híbridos que mantengan el componente humano como elemento esencial del aprendizaje personalizado, y cuyo diseño ponga el énfasis en el acompañamiento pedagógico constante y la adaptación a las características específicas de cada grupo de estudiantes.

Otro estudio experimental sobre instrucción individualizada basada en tecnología es el reali-

zado por De Barros y Ganimian (2024), que se centran en estudiantes de 6.º a 8.º grado de escuelas públicas del estado de Rajastán, en la India. La intervención, llevada a cabo en 15 escuelas modelo, consistió en el uso de la plataforma educativa Mindspark durante nueve meses, con sesiones de 45 minutos diarias en horario lectivo, centradas en matemáticas. Los estudiantes fueron asignados aleatoriamente a dos condiciones: el grupo de control accedía a contenidos correspondientes al currículo de su grado escolar, mientras que el grupo de tratamiento recibía contenidos personalizados según su nivel de competencia real, diagnosticado previamente mediante una prueba estandarizada.

Los resultados muestran que la instrucción adaptada no produjo efectos significativos sobre el rendimiento medio del alumnado (efecto estimado de 0,05 d. e., pero no significativo). Sin embargo, se observaron efectos positivos y significativos para el 25 por 100 de estudiantes con menor desempeño inicial: en este subgrupo, el tratamiento aumentó el rendimiento en matemáticas en 0,22 d.e., especialmente en los dominios de números y resolución de problemas, que coincidían con los contenidos más frecuentemente abordados por la plataforma. Estos efectos son relevantes dado el bajo nivel de uso promedio de la herramienta: la mediana fue de 329 minutos acumulados en todo el curso escolar, lo que indica que incluso un uso relativamente limitado de la plataforma puede generar beneficios significativos si se adapta al nivel real del estudiante.

El estudio destaca que el principal mecanismo de impacto fue el ajuste del nivel de instrucción, más que otros componentes como el tipo de ejercicios o el diseño de la interfaz. Además, la intervención fue implementada por personal del propio sistema escolar, sin necesidad de recursos adicionales externos. Los autores concluyen que el aprendizaje personalizado a través de tecnología puede contribuir a mejorar el rendimiento de los estudiantes más rezagados, sin afectar negativamente al resto, siempre que se integre de forma realista y eficiente en la jornada escolar. Esta evidencia es especialmente valiosa en contextos con alta heterogeneidad de niveles y recursos limitados, donde adaptar la enseñanza al punto de partida del alumno resulta clave para evitar el rezago acumulativo.

CUADRO N.º 1

EVALUACIONES EXPERIMENTALES RECIENTES SOBRE TUTORÍAS INDIVIDUALIZADAS Y PERSONALIZACIÓN DEL APRENDIZAJE MEDIANTE TECNOLOGÍA EN CONTEXTOS ESCOLARES

AUTOR	PUBLICACIÓN	PAÍS	MEDIDA DE OUTPUT	DIRECCIÓN DEL EFECTO	TAMAÑO DEL EFECTO
Fesler, Gu y Chojnacki.	<i>Mathematica & Air Tutors</i> (marzo 2023).	Estados Unidos (Texas).	Resultados en matemáticas (MAP), confianza matemática, sentido de pertenencia y calidad de la relación tutor-alumno.	Positiva. Mejora en rendimiento en MAP, confianza matemática y vínculo con el tutor.	0,13 d. e. (media general); 0,26 d.e. en la escuela media con sesiones durante el horario lectivo.
De Barros y. Ganimian.	<i>Journal of Research on Educational Effectiveness</i> , 2024, Vol. 17, Issue 2.	India (Rajastán).	Test estandarizado de matemáticas adaptado al currículo indio.	Nulo en media; positivo y significativo para estudiantes en el cuartil inferior del rendimiento inicial.	0,05 d. e. en media; 0,22 d. e. para el cuartil inferior.
Robinson, Pollard, Novicoff, White y Loeb.	<i>Educational Evaluation and Policy Analysis</i> (2024).	Estados Unidos.	Competencias de alfabetización temprana (DIBELS, K-2).	Positiva, especialmente en alumnos de bajo rendimiento inicial.	+0,05 d.e. (general); +0,08 d. e. (sin alumnos con necesidades especiales); +0,15 d.e. (grupo 1:1 de bajo rendimiento inicial).
Bhatt, Guryan, Khan, LaForest-Tucker y Mishra.	<i>NBER Working Paper</i> N.º 32510 (2024).	Estados Unidos (Chicago y Nueva York).	Test estandarizado de matemáticas, GPA, suspensos.	Mejora significativa en rendimiento matemático; efectos replicables y persistentes tras la intervención.	+0,23 d.e. (TOT en matemáticas); +0,24 puntos en GPA; -22% en suspensos de matemáticas.
Araya, Arias Ortiz, Botta y Cristia.	<i>Economics of Education Review</i> , 2025.	Chile.	4.º de Primaria (matemáticas).	Plataforma de matemáticas integrada en horario escolar, con apoyo externo.	+0,27 d. e. (SIMCE); +0,13 a +0,18 d. e. (pruebas externas).

El estudio de Robinson *et al.* (2024) proporciona una de las primeras evaluaciones rigurosas del impacto de la tutoría virtual individualizada en alumnos de educación infantil y primaria (grados K-2). El estudio se desarrolló en una red de escuelas concertadas en Texas (Estados Unidos). Mediante un experimento aleatorizado, se asignó a los alumnos a sesiones de tutoría 1:1, 2:1 o a un grupo

de control. La intervención, aunque no basada en IA generativa, comparte también elementos clave con los tutores automatizados: personalización, interacción frecuente y atención adaptada al nivel del estudiante.

Los resultados muestran un impacto positivo y significativo: una mejora de 0,05 d. e. para el

• CUADRO N.º 1 (continuación)
EVALUACIONES EXPERIMENTALES RECIENTES SOBRE TUTORÍAS INDIVIDUALIZADAS Y PERSONALIZACIÓN DEL APRENDIZAJE MEDIANTE TECNOLOGÍA EN CONTEXTOS ESCOLARES

INTERVENCIÓN	COMENTARIOS	MÉTODO	OBJETIVO	TAMAÑO DE MUESTRA
Tutoría <i>online</i> en grupos pequeños (3-4 alumnos) con tutores altamente cualificados mediante videollamada y pizarra interactiva.	3 horas semanales durante 15-21 semanas; mayores efectos en sesiones integradas en horario escolar; mejoras en confianza y pertenencia (autoinforme).	RCT estratificado; regresión multivariable y análisis bayesiano complementario.	Estudiantes de 4.º a 6.º grado con bajo rendimiento inicial; 67 % hispanos, 21 % afroamericanos.	276 estudiantes con datos completos (184 tratamiento, 92 control); 153 participaron en al menos una sesión.
Instrucción individualizada en plataforma digital (Mindspark), durante el horario escolar.	Mediana de uso de 329 minutos en el curso; efecto concentrado en los contenidos más abordados por la plataforma.	RCT con asignación aleatoria en 15 escuelas.	Identificar efectos de la instrucción personalizada sobre estudiantes con distintos niveles de rendimiento.	1.438 estudiantes de 6.º a 8.º grado en escuelas públicas (tratamiento y control asignados por lotería).
Tutoría virtual 1:1 y 2:1 integrada durante el día escolar.	Efectos moderados comparados con tutoría presencial, pero escalable, efectiva y coste-efectiva, especialmente en formato 1:1.	Experimento aleatorizado con grupo control.	Evaluar el impacto de la tutoría virtual individualizada en alfabetización temprana.	2.085 estudiantes (grados K-2).
Tutoría híbrida 4-a-1: parejas con tutor presencial alternando con práctica en CAL durante 50 min. diarios.	El modelo reduce costes (-30%) y uso de personal (-50%) sin pérdida de eficacia; efectos persistentes el año siguiente	RCT con asignación aleatoria en 7 institutos públicos	Evaluar si el uso estratégico de CAL permite escalar modelos de tutoría manteniendo eficacia.	~4.000 estudiantes de 9.º grado en dos cohortes (2018-19 y 2019-20).
Plataforma integrada en clase; sesiones 2-3 veces/semana; apoyo de coordinador externo.	Mejora significativa en matemáticas en horario lectivo regular.	RCT (24 escuelas públicas).	Evaluar impacto de integrar plataformas digitales con apoyo externo durante horario escolar.	2.310 estudiantes en 4.º de Primaria de escuelas urbanas públicas en Santiago de Chile.

Nota: d. e. = desviaciones estándar; GPA = Grade Point Average.

conjunto de estudiantes. El efecto más fuerte (0,15 d. e.) se observó entre los alumnos con menor rendimiento inicial en el grupo 1:1, lo que refuerza el argumento de que el formato más intensivo y personalizado amplifica los beneficios de la tutoría remota. Aunque la magnitud del efecto es más moderada que en intervenciones presenciales intensivas, el estudio destaca que la tutoría vir-

tual bien estructurada puede ser una herramienta eficaz, escalable y coste-efectiva para mejorar la alfabetización temprana, especialmente cuando se adapta a las necesidades del alumnado más vulnerable.

Una contribución relevante al debate sobre el uso de tecnología en modelos de tutoría escalables es la

de Bhatt *et al.* (2024), quienes evalúan el impacto de incorporar aprendizaje asistido por ordenador (CAL) en programas de tutoría de alta dosis en escuelas públicas de secundaria en Estados Unidos. El estudio se basa en un experimento aleatorizado llevado a cabo en tres institutos públicos de Chicago y cuatro de Nueva York durante los cursos escolares 2018-2019 y 2019-2020, con estudiantes de 9.º grado.

La intervención se estructuró en un modelo de tutoría 4-a-1: cuatro estudiantes compartían una mesa, alternando cada día entre trabajar en parejas con un tutor presencial y realizar actividades de matemáticas en una plataforma de CAL durante una sesión diaria de 50 minutos. Este modelo, denominado *Saga Technology*, buscaba reducir los costes y requerimientos de personal del modelo tradicional de tutoría diaria 2-a-1. El coste por estudiante se redujo en un 30 por 100, y el número de tutores necesarios en un 50 por 100, manteniendo una implementación durante el horario lectivo habitual.

Los resultados del experimento muestran un impacto significativo en los resultados académicos: el efecto tratamiento sobre los tratados (TOT, TOT, por sus siglas en inglés) fue de 0,23 d. e. en matemáticas, una magnitud comparable a la obtenida en el modelo de tutoría diaria 2-a-1 evaluado por Guryan *et al.* (2023). También se observaron mejoras en matemáticas (+0,24 puntos) y una reducción del 22 por 100 en la tasa de suspensos en esta asignatura. Además, los efectos se replicaron parcialmente en el segundo año del estudio (2019-2020), y se encontraron efectos positivos persistentes sobre el rendimiento en matemáticas en el año posterior a la intervención.

A diferencia de otros estudios centrados en tutores virtuales con IA conversacional, este trabajo explora un enfoque híbrido en el que la tecnología no sustituye al tutor, sino que libera parte de su tiempo, lo que permite aumentar la escala de implementación sin sacrificar la eficacia. Los autores subrayan que, incluso sin interacción personalizada con un sistema inteligente, el uso estratégico del CAL en combinación con la tutoría humana puede generar mejoras sustanciales y sostenidas en contextos escolares reales.

El estudio de Araya *et al.* (2025) también ofrece evidencia rigurosa sobre el impacto del uso integrado de plataformas digitales en el horario escolar en contextos de países en desarrollo. La intervención, implementada en 24 escuelas públicas urbanas de Santiago de Chile, consistió en sustituir parcialmente las clases tradicionales de matemáticas por sesiones estructuradas en el laboratorio de informática utilizando la plataforma Conectaldeas. Esta herramienta permitía la realización de ejercicios interactivos, ofrecía *feedback* inmediato a los estudiantes, y generaba información en tiempo real sobre el progreso del grupo para los docentes.

Un elemento distintivo del programa fue la incorporación de un coordinador externo en cada escuela, encargado de facilitar el uso de la plataforma y apoyar al profesorado en su integración curricular. Las sesiones se desarrollaban en horario lectivo regular (entre dos y tres veces por semana), lo que representa un cambio importante respecto a muchas intervenciones tecnológicas, habitualmente implementadas como complemento fuera del horario escolar. Además, la intervención no requería formación intensiva del profesorado ni modificaciones profundas del currículo.

Los resultados son contundentes: los estudiantes del grupo de tratamiento obtuvieron una mejora promedio de 0,27 d. e. en la prueba nacional estandarizada de matemáticas (SIMCE), una mejora sustancial si se considera que partían de una desventaja de 0,6 d. e. respecto a la media nacional. También se registraron mejoras consistentes (entre 0,13 y 0,18 d. e.) en pruebas estandarizadas contratadas por los investigadores. Estos efectos fueron superiores a los encontrados en intervenciones similares realizadas en otras regiones de América Latina. No obstante, el impacto fue limitado a la competencia matemática: no se observaron efectos sobre el rendimiento en lectura, ni sobre variables motivacionales como el interés intrínseco por la asignatura. Además, se halló un efecto negativo en la preferencia por el trabajo colaborativo, posiblemente derivado del énfasis competitivo de la plataforma (que premiaba el desempeño individual).

Este estudio refuerza la idea de que la tecnología puede mejorar significativamente el apren-

dizaje cuando se integra en el núcleo del horario escolar y se acompaña de un esquema de implementación claro y apoyado institucionalmente. Aunque el programa no utilizó IA ni personalización algorítmica, la experiencia en Chile aporta lecciones relevantes sobre las condiciones organizativas necesarias para que una herramienta tecnológica tenga un impacto sostenido sobre el aprendizaje.

El cuadro n.º 1 sintetiza los estudios analizados en esta sección sobre las tutorías individualizadas.

III. LAS VENTAJAS DEL USO DE LAS TUTORÍAS PERSONALIZADAS GUIADAS POR IA

La IA tiene la capacidad para ofrecer tutorías individualizadas a gran escala. Este objetivo parecía inalcanzable en el pasado por su coste, la tecnología actual permite imaginar escenarios donde cada estudiante pueda interactuar con un asistente que le explique conceptos, le ayude a resolver dudas o le dé apoyo conforme a su ritmo de progreso. El impacto potencial sería especialmente significativo para los alumnos con más dificultades, que en un modelo tradicional tienden a quedar rezagados. La tecnología permite responder a la diversidad real de niveles y estilos de aprendizaje que existen dentro de una misma aula.

Los beneficios de la IA no se materializarán por sí solos. Para que esta oportunidad tenga un efecto real en términos de mejora del aprendizaje y de reducción de desigualdades, es necesario un esfuerzo institucional sostenido. Entre las prioridades que identifica están la evaluación rigurosa de los programas piloto, el desarrollo de herramientas que respondan a criterios pedagógicos sólidos, la formación del profesorado y el acompañamiento para su integración en el aula. Como desarrolla en su artículo en este mismo número de *Papeles de Economía Española* Ildelfonso Méndez, no basta con introducir tecnología; es imprescindible diseñar ecosistemas de aprendizaje donde esta tecnología tenga sentido y esté al servicio de una estrategia educativa ambiciosa. Roldán (2025), también señala que es importante que las herramientas que proporciona la IA no se dejen únicamente en manos de plataformas comerciales. Las Administraciones

públicas y las instituciones educativas tienen que guiar la implementación de la IA en las aulas para evitar que las desigualdades de oportunidades entre estudiantes favorecidos y desfavorecidos se amplíen. Hay que aprovechar las posibilidades y el potencial que ofrece la IA en la educación para cerrar brechas y no para abrir otras nuevas.

El estudio experimental llevado a cabo por Bastani *et al.* (2024) en Turquía constituye una contribución sólida y relevante a la literatura sobre los efectos educativos de los tutores basados en IA generativa. A diferencia de muchos estudios centrados en entornos universitarios o simulados, esta intervención se diseñó e implementó en colaboración con el Ministerio de Educación de Turquía en condiciones reales de aula, con 3.200 estudiantes de secundaria en un contexto de recursos limitados. Su objetivo era analizar no solo si los tutores IA mejoran el rendimiento académico, sino cómo el diseño de la interfaz —es decir, la manera en que el estudiante interactúa con el modelo— condiciona los efectos de la herramienta sobre el aprendizaje.

La intervención consistió en comparar tres grupos: 1) un grupo de control sin acceso a IA; 2) un grupo con acceso a un tutor IA basado en GPT-4 en su versión estándar (GPT-Base); y 3) un grupo con acceso a una versión modificada de GPT-4 con principios pedagógicos integrados (GPT-Tutor). Ambos modelos se utilizaron para asistir a los estudiantes en la resolución de problemas matemáticos. Mientras que el GPT-Base respondía de forma directa a cualquier pregunta —incluyendo la resolución completa del ejercicio si se le solicitaba—, el GPT-Tutor fue configurado para no ofrecer respuestas completas, sino pistas parciales, contraejemplos, reformulaciones y preguntas orientadoras, siguiendo un enfoque de *scaffolding* (andamiaje) inspirado en la pedagogía de Vygotsky y en prácticas de tutoría humana. El andamiaje educativo es una estrategia en la que el profesor proporciona apoyo temporal al alumno para que pueda realizar una tarea que aún no domina por completo. A medida que el estudiante gana autonomía, ese apoyo se reduce progresivamente. Su objetivo es fomentar el aprendizaje activo y el desarrollo de habilidades complejas, guiando al alumno paso a paso según su nivel.

Durante la fase de práctica con tutor, los dos grupos que usaban IA mostraron mejoras significativas frente al grupo de control: +0,137 puntos con GPT-Base y +0,361 con GPT-Tutor. Sin embargo, en el examen posterior sin asistencia, solo el grupo GPT-Tutor logró mantener un rendimiento equivalente al del grupo de control, mientras que el grupo GPT-Base obtuvo peores resultados (-0,054 puntos), lo que representa una caída del 17 por 100. Es decir, aquellos que resolvieron los ejercicios anteriores con ayuda de GPT-Base aprendieron menos que quienes trabajaron sin ningún tipo de asistencia.

Este resultado tiene implicaciones relevantes: demuestra que el uso de tutores IA no garantiza por sí mismo un mejor aprendizaje, y que, de hecho, puede ser contraproducente si no se guía adecuadamente. El GPT-Base, al ofrecer respuestas completas, llevó a los estudiantes a adoptar una actitud pasiva, delegando el esfuerzo cognitivo en la máquina. Este fenómeno ha sido descrito como una forma de delegación excesiva del esfuerzo cognitivo, un efecto que comentaremos con más detalle en la siguiente sección.

El efecto negativo del GPT-Base fue especialmente acusado entre los estudiantes con menor rendimiento inicial, mientras que el GPT-Tutor mostró un efecto positivo justamente en este mismo grupo. Así, el diseño pedagógico no solo mejoró los resultados globales, sino que redujo las desigualdades, actuando como mecanismo compensador. Este resultado es especialmente relevante para políticas educativas centradas en la equidad, ya que pone de manifiesto que el diseño cuidadoso de la tecnología puede contribuir a cerrar brechas, mientras que un diseño acrítico puede agravarlas.

Más allá de los resultados cuantitativos, el estudio también analizó las interacciones entre los estudiantes y el modelo de IA. Se observó que quienes utilizaban GPT-Base formulaban preguntas superficiales y orientadas a obtener directamente la respuesta (por ejemplo, "¿cuál es la solución?"), mientras que los usuarios de GPT-Tutor desarrollaban interacciones más ricas, solicitando aclaraciones, interpretando las pistas y reformulando estrategias. En términos cualitativos, se generó un

entorno de aprendizaje más profundo, en el que el estudiante asumía un rol activo y reflexivo en lugar de actuar como simple receptor de información.

Otro elemento destacado del estudio es su enfoque de implementación. La intervención fue sencilla desde el punto de vista técnico: no se necesitó *software* propietario ni infraestructuras costosas, y se emplearon equipos informáticos ya disponibles en los centros. Sin embargo, el impacto educativo dependió enteramente del diseño pedagógico del sistema. El valor de la IA en educación no reside en la sofisticación del modelo *per se*, sino en la intencionalidad educativa que guía su uso.

En definitiva, Bastani *et al.* (2024) ofrecen evidencia sólida de que el diseño instruccional de los tutores IA es un determinante clave de su efectividad. El uso de modelos como GPT-4 puede tener efectos positivos o negativos en función de cómo se estructura la interacción con el estudiante. Cuando se limita a dar respuestas, puede inhibir el aprendizaje autónomo; cuando se configura para apoyar el razonamiento y fomentar la autorregulación, puede amplificar el aprendizaje y mejorar la equidad. No basta con introducir IA en el aula, es necesario acompañarla de criterios pedagógicos explícitos y de una supervisión docente activa.

Otra intervención reciente sobre tutores individualizados guiados por IA en contextos de bajos ingresos es la evaluada por Henkel *et al.* (2024), quienes analizaron el impacto del tutor conversacional Rori en once escuelas de Ghana pertenecientes a la red Rising Academies. El estudio se llevó a cabo con cerca de 500 estudiantes de 3^o a 8^o curso de primaria, que completaron tanto una prueba inicial como una prueba final. Las escuelas fueron asignadas aleatoriamente al grupo de tratamiento (5 escuelas, 236 estudiantes) y al grupo de control (6 escuelas, 241 estudiantes).

Rori es un tutor de matemáticas basado en IA, accesible mediante WhatsApp y diseñado para funcionar en teléfonos móviles básicos conectados a redes de baja capacidad. El contenido está estructurado en más de 500 microlecciones coherentes con el *Global Proficiency Framework for Mathematics*. Cada microlección incluye una explicación breve,

ejercicios prácticos y un sistema de andamiaje: si el estudiante falla, recibe primero una pista y luego la solución explicada. El sistema permite interacciones mediante lenguaje natural, lo que favorece una experiencia similar a una tutoría personalizada.

Los estudiantes del grupo de tratamiento utilizaron Rori durante dos sesiones semanales de 30 minutos en el horario de estudio autónomo (*study hall*), con supervisión técnica por parte del profesorado. La intervención no requirió formación adicional, personal extra ni cambios curriculares. Los resultados muestran una mejora estadísticamente significativa en el rendimiento matemático del grupo de tratamiento. La diferencia implica un efecto de 0,36 d. e., un impacto elevado en el contexto de la economía de la educación. Los autores afirman que este tamaño del efecto es aproximadamente equivalente a un año adicional de aprendizaje.

El coste estimado de la intervención fue de unos 5 dólares por estudiante, lo que refuerza su viabilidad como una solución coste-efectiva y escalable en sistemas educativos con recursos limitados. Aunque el estudio abarca solo el primer año de implementación, sus resultados iniciales respaldan el potencial de tutores IA conversacionales como Rori para mejorar de manera significativa los resultados de aprendizaje en países de renta baja.

En una línea complementaria, aunque con un enfoque distinto, se sitúa el estudio de Thomas *et al.* (2024), que analiza el impacto de un modelo híbrido de tutoría, en el que se combina la personalización algorítmica de un *software* de IA con la interacción directa con tutores humanos. Frente a soluciones como la de Henkel (2024) —diseñadas para funcionar en condiciones de infraestructura mínima—, la propuesta de Thomas *et al.* se centra en entornos urbanos de Estados Unidos, con estudiantes de secundaria en situación de vulnerabilidad (principalmente afroamericanos y latinos) en centros escolares de Pensilvania y California. La intervención parte de la hipótesis de que la IA puede detectar en tiempo real patrones de dificultad, y que esta información puede aprovecharse para guiar la acción de tutores humanos, que complementan la instrucción automatizada con apoyo emocional, motivacional y pedagógico.

El programa fue evaluado mediante tres casos cuasiexperimentales independientes, en los que se comparó el grupo tratado (que recibió tutoría híbrida) con estudiantes que solo utilizaron *software* de matemáticas adaptativo. Los resultados muestran mejoras significativas tanto en rendimiento como en la implicación con la herramienta, con efectos especialmente pronunciados entre el alumnado con peor desempeño inicial. En concreto, los tutores humanos, informados sobre el aprendizaje del alumnado por los datos del sistema, lograban intervenir de forma más eficaz y personalizada que si operaran de manera ciega o genérica. Esta sinergia entre la IA y la inteligencia humana es una de las claves del éxito del modelo.

El estudio destaca también que este enfoque híbrido puede ser escalable. El coste marginal anual por alumno —en torno a 700 dólares— es sensiblemente inferior al de una tutoría exclusivamente humana, lo que lo convierte en una alternativa viable para distritos escolares con recursos limitados, pero con acceso a infraestructura básica. Además, Thomas *et al.* (2024) subrayan la importancia del diseño del panel de control para el tutor y de mantener una ratio baja de alumnos por tutor, para garantizar que el apoyo sea verdaderamente personalizado.

En conjunto, el estudio de Thomas *et al.* (2024) muestra que la combinación estratégica de IA y apoyo humano puede no solo mejorar el aprendizaje, sino también hacerlo de forma más coste-efectiva y sostenible. Sus conclusiones coinciden con las de Henkel *et al.* (2024) en que la clave del éxito no está únicamente en el poder de procesamiento del algoritmo, sino en cómo se integra esa capacidad en una arquitectura pedagógica robusta y centrada en el estudiante.

La IA puede ayudar a aliviar algunas de las cargas administrativas que afectan a los docentes. El tiempo dedicado a preparar clases, corregir tareas o buscar materiales podría reducirse significativamente si se aprovechan herramientas que automaticen parte de esas funciones. Esto, a su vez, permitiría a los profesores centrarse más en lo esencial: acompañar, motivar, observar y atender de forma más cercana a cada alumno.

Otro estudio que analiza los tutores individualizados guiados por IA es el de Oreopoulos *et al.* (2024). El programa que evalúan Oreopoulos *et al.* se dirige a estudiantes de primaria y secundaria de EE.UU., y tiene como objetivo central fomentar una mayor profundización y dominio de los contenidos de matemáticas. Para ello, se introduce una plataforma tecnológica que permite personalizar el aprendizaje y ajustar la secuencia de ejercicios y explicaciones según el nivel de cada estudiante. Lo más interesante del diseño del programa es que no se limita al uso de tecnología, sino que se combina con un componente humano clave: la figura del entrenador o *coach*, que se reúne semanalmente con los docentes participantes. Estos entrenadores proporcionan orientación proactiva sobre cómo utilizar las herramientas de aprendizaje asistido por ordenador (CAL) de forma eficaz en el aula. Su papel no es meramente técnico, sino pedagógico: ayudan a los profesores a interpretar los datos generados por la plataforma, a identificar patrones en el progreso de los alumnos, a planificar intervenciones específicas con estudiantes que presentan dificultades, y a establecer metas realistas y personalizadas para cada grupo. De este modo, la tecnología no sustituye al profesor, sino que se convierte en una herramienta que amplifica su capacidad de enseñar, de manera más eficiente y dirigida.

Oreopoulos *et al.* (2024) implementan dos experimentos aleatorios controlados (RCT, por sus siglas en inglés) para evaluar el impacto de esta intervención, mostrando evidencia de relaciones causales sólidas. Los resultados indican que hay mejoras en el rendimiento en matemáticas que oscilan entre el 0,12 y 0,22 de una desviación estándar, dependiendo del tiempo de práctica semanal con el programa CAL. Esta heterogeneidad en los resultados es especialmente relevante, y está relacionada con la intensidad del uso de la herramienta. Los estudiantes que practicaron al menos 35 minutos semanales con la plataforma mostraron mejoras significativamente mayores. Es el umbral mínimo de uso necesario para obtener beneficios relevantes.

La fidelidad en la implementación del programa —es decir, la medida en que los docentes integran

de manera consistente y estructurada el uso del CAL en sus clases— es un factor crítico para su éxito. Aquellas aulas en las que se logró mantener un nivel elevado de práctica con la herramienta registraron efectos positivos comparables a los de programas de tutoría intensiva uno a uno. Este resultado es interesante si se considera que el programa evaluado por Oreopoulos *et al.* (2024) se apoya mayoritariamente en recursos existentes y no requiere de una inversión extraordinaria en personal adicional.

El estudio identifica también una serie de factores que explican por qué algunos docentes logran una implementación más efectiva del programa que otros. Entre ellos se encuentran: la alta participación inicial del profesorado, una estrategia clara de implementación para fomentar la práctica regular del alumnado, y la disposición de los docentes a observar de cerca el progreso de sus estudiantes y a intervenir cuando detectan dificultades. Además, juegan un papel importante el apoyo institucional al programa, la dedicación exclusiva o prioritaria al mismo, la convicción por parte del profesorado de que el programa es eficaz, y la implicación activa en su desarrollo. Algunos docentes logran que sus estudiantes practiquen más tiempo con los programas CAL que otros, poniendo de relieve la importancia de factores como el convencimiento, la implicación y el acompañamiento profesional. En definitiva, Oreopoulos *et al.* (2024) proporcionan evidencia sólida sobre el impacto positivo que puede tener una combinación de aprendizaje asistido por ordenador (CAL) con una estrategia de apoyo sostenido al profesorado. La IA puede actuar como tutor personalizado eficaz cuando se integra en un marco pedagógico claro, con apoyo institucional, compromiso docente y acompañamiento continuo. Su artículo muestra la efectividad de un programa que utiliza principalmente recursos ya existentes para facilitar un aprendizaje más personalizado, y que, por tanto, convierten a esta intervención en una opción escalable y sostenible.

Otra aportación clave que se centra en los docentes y tutores en la literatura sobre tutorías individualizadas asistidas por IA es la de Wang *et al.* (2024), quienes evalúan *Tutor CoPilot*, un sistema que ofrece apoyo experto en tiempo real a tutores humanos durante sesiones de tutoría en matemá-

ticas. A diferencia de otros programas centrados directamente en el estudiante, este enfoque combina inteligencia artificial generativa con la mediación activa del tutor, buscando amplificar sus capacidades pedagógicas.

El estudio se llevó a cabo mediante un ensayo aleatorizado con 900 tutores y 1.800 estudiantes de K-12 procedentes de comunidades históricamente desfavorecidas en Estados Unidos. Los tutores asignados al grupo de tratamiento tuvieron acceso a Tutor CoPilot, una interfaz que proporciona sugerencias de intervención pedagógica, como preguntas orientadoras, estrategias de andamiaje y recordatorios para evitar dar respuestas directamente. El análisis muestra que los estudiantes cuyos tutores utilizaron esta herramienta de Tutor CoPilot tuvieron una probabilidad 4 puntos porcentuales mayor de dominar los contenidos matemáticos, con beneficios especialmente destacados entre los estudiantes atendidos por tutores con menos experiencia previa (de hasta 9 puntos porcentuales).

El sistema también resultó notablemente eficiente en términos de coste: el uso medio anual por tutor fue de aproximadamente 20 dólares. Además del impacto cuantitativo, los autores analizaron más de 550.000 mensajes intercambiados entre tutores y estudiantes, observando que el acceso a Tutor CoPilot se asoció con un aumento en el uso de estrategias pedagógicas alineadas con buenas prácticas docentes, como formular preguntas que fomentan el razonamiento del estudiante. Las entrevistas cualitativas revelan que los tutores valoraron el apoyo en tiempo real, aunque señalaron como desafío la adecuación del nivel de las sugerencias generadas. En conjunto, este estudio refuerza la idea de que los sistemas de IA generativa pueden actuar como “copilotos pedagógicos”, ayudando a escalar la calidad de la enseñanza sin reemplazar el rol humano.

Entre los estudios empíricos más recientes sobre el impacto de la IA generativa en entornos reales se encuentra el trabajo de De Simone *et al.* (2025) en el Banco Mundial. Estos autores implementan un ensayo aleatorizado con un modelo de lenguaje generativo en un entorno educativo de un país en desarrollo. La intervención analizada por De Simone *et al.* (2025) se desarrolló en Benín City (Nigeria)

y consistió en un programa de tutoría después del horario escolar de seis semanas de duración, dirigido a estudiantes de primer curso de secundaria en nueve centros públicos.

Durante seis semanas, los alumnos participaron en sesiones de tutoría después del horario escolar, en las que interactuaban con Microsoft Copilot para reforzar sus competencias en lengua inglesa, habilidades digitales y conocimientos sobre inteligencia artificial. Los estudiantes participaron en doce sesiones de 90 minutos en laboratorios de informática, guiadas por profesores. El 52 por 100 del alumnado elegible mostró interés en participar, y la asignación al tratamiento se realizó de manera aleatoria, lo que permite estimar efectos causales mediante un diseño experimental robusto.

Los resultados muestran un impacto significativo: el grupo tratado obtuvo mejoras de 0,31 d. e. en los resultados finales de competencias en lengua inglesa, habilidades digitales y conocimientos sobre inteligencia artificial, con un efecto específico de 0,23 d. e. en competencias de lengua inglesa, el objetivo principal del programa. Este progreso en el aprendizaje equivale a entre 1,5 y 2 años de aprendizaje convencional (De Simone *et al.*, 2025). Los efectos positivos se observaron en todo el espectro de rendimiento inicial, aunque fueron más pronunciados entre los estudiantes con mejor desempeño previo. También se observaron mayores beneficios entre las alumnas, lo que permitió compensar un déficit inicial de rendimiento en comparación con los chicos.

El estudio también destaca la relación entre la dosis de la intervención y la respuesta: por cada día adicional de asistencia al programa se estima una mejora de 0,031 d. e. en los resultados. Prolongando la intervención a un curso completo con niveles similares de asistencia (72 por 100), se proyecta una mejora total de 1,55 d. e. En escenarios más optimistas, con asistencia plena, los efectos podrían superar las 2 d.e. En el diseño del programa también se tuvo en cuenta la supervisión docente. Aunque la IA proporcionaba el apoyo principal, los profesores recibieron formación específica y actuaban como guías durante las sesiones, asegurando un uso adecuado de la herramienta y evitando una dependencia excesiva. Esta

CUADRO N.º 2

EVALUACIONES COMPARADAS DE TUTORES GUIADOS EN INTELIGENCIA ARTIFICIAL: DISEÑOS, CONTEXTOS Y EFECTOS EDUCATIVOS

AUTOR	PUBLICACIÓN	PAÍS	MEDIDA DE OUTPUT	DIRECCIÓN DEL EFECTO	TAMAÑO DEL EFECTO
Bastani, Bastani, Sungu, Ge, Kabakcı y Mariman.	<i>Wharton School Research Paper</i> (2024).	Turquía.	Puntuación normalizada en ejercicios y examen.	GPT-Tutor mejora significativamente la práctica y evita efectos negativos en el examen; GPT-Base mejora la práctica, pero perjudica el rendimiento posterior.	GPT-Tutor: +0,361 (práctica), -0,004 (examen); GPT-Base: +0,137 (práctica), -0,054 (examen).
Henkel, Horne-Robinson, Kozhakhmetova y Lee.	<i>Effective and Scalable Math Support</i> (2024).	Ghana.	Rendimiento en matemáticas (test estandarizado)	Positivo y significativo. Mejora superior en grupo con Rori.	0,36 d.e. en crecimiento de puntuación ($p < 0.001$)
Thomas, Marsh, Allbright, Johnson y Jennings.	<i>AIED 2024</i> (Springer CCIS, vol. 2150).	EE. UU. (Pensilvania y California).	Rendimiento en matemáticas y nivel de <i>engagement</i> .	Positivo; mayor impacto en estudiantes con bajo rendimiento inicial.	No se reporta efecto en d. e., pero mejora estadísticamente significativa frente a grupo control.
De Simone, Ogundeyi, Adesanya, Bello, Adeniran, Bello, Daramola y Jhingran.	<i>World Bank Policy Research Working Paper</i> N.º. 10747 (febrero 2025).	Nigeria (Benin City).	Prueba estandarizada en lengua inglesa, competencias digitales e IA.	Positivo; especialmente en inglés y entre alumnas y estudiantes con buen desempeño inicial.	+0,31 d. e. global; +0,23 d. e. en inglés; +0,031 d. e. por sesión adicional.
Oreopoulos, Gibbs, Jensen y Price.	<i>EdWorking Paper</i> (2024).	EE.UU. (Texas, Tennessee).	Desempeño en matemáticas (STAAR).	Positivo; mejora significativa del rendimiento, especialmente con mayor práctica.	0,12-0,22 d. e. (según tiempo de práctica semanal).
Wang, Ribeiro, Robinson, Loeb y Demzsky.	Stanford University (2024).	Estados Unidos.	Dominio de contenidos matemáticos (<i>exit tickets</i>), calidad pedagógica en tutorías.	Positiva, especialmente para tutores menos experimentados.	Aumento de 4 puntos porcentuales en dominio general ($p < 0,01$); hasta 9 puntos porcentuales en estudiantes con tutores menos efectivos.

CUADRO N.º 2 (continuación)

EVALUACIONES COMPARADAS DE TUTORES GUIADOS EN INTELIGENCIA ARTIFICIAL: DISEÑOS, CONTEXTOS Y EFECTOS EDUCATIVOS

INTERVENCIÓN	COMENTARIOS	MÉTODO	OBJETIVO	TAMAÑO DE MUESTRA
Tutor virtual con GPT-4, en versión estándar (GPT-Base) o pedagógica (GPT-Tutor), en sesiones de matemáticas con <i>laptops</i> en clase.	GPT-Tutor usa pistas, errores comunes y andamiaje. GPT-Base ofrece respuestas completas. Efecto negativo del GPT-Base más acusado en alumnos con menor rendimiento inicial.	RCT con asignación aleatoria a nivel de aula.	Evaluar cómo el diseño de la interacción con tutores IA (GPT-Base vs GPT-Tutor) afecta el rendimiento y el aprendizaje autónomo.	839 estudiantes.
Tutor conversacional por WhatsApp. Dos sesiones semanales de 30 min con Rori durante <i>study hall</i> .	Funciona en móviles básicos. Interacción por lenguaje natural. Coste de 5 dólares por estudiante. No requiere cambios curriculares ni formación extra.	RCT con asignación por escuela.	Evaluar el impacto de un tutor conversacional IA accesible y escalable en aprendizaje matemático en primaria.	477 estudiantes con datos completos (236 tratamiento, 241 control).
Tutoría híbrida: <i>software</i> adaptativo + tutores humanos informados por datos del sistema.	El uso de <i>dashboards</i> y ratio baja tutor/alumnos (1:4) permite personalización efectiva. Coste ≈700 dólares/año/alumno.	Tres estudios cuasiexperimentales con grupo control.	Evaluar si la combinación de IA y apoyo humano mejora el rendimiento matemático y la implicación.	≈ 500 estudiantes en tres centros escolares urbanos.
Tutoría con Copilot (ChatGPT-4) tras el horario escolar: 12 sesiones de 90 min en 6 semanas, con supervisión docente.	Sin <i>software</i> propietario ni bancos de preguntas. Coste de 48 dólares por alumno. Alta rentabilidad educativa y viabilidad en entornos con limitaciones.	RCT con asignación aleatoria individual en 9 escuelas públicas.	Evaluar impacto de tutores IA generativos en contextos escolares reales de renta baja.	857 estudiantes elegibles; 447 participantes; análisis con 414 estudiantes con datos completos.
<i>Uso de Khan Academy en aula con apoyo semanal de "khoaaches" a docentes.</i>	Efectos mayores con más de 35 minutos de práctica semanal; la implementación de calidad es clave.	Dos experimentos aleatorios controlados (RCTs).	Evaluar el efecto del aprendizaje asistido por ordenador (CAL) con apoyo pedagógico semanal a profesores sobre el rendimiento matemático.	224 docentes, 10.979 estudiantes.
Tutor CoPilot, apoyo pedagógico en tiempo real mediante IA generativa (GPT-4) a tutores humanos durante tutorías virtuales.	Uso de estrategias pedagógicas efectivas, coste muy bajo (\$20 anual por tutor). Mayor impacto en tutores con menor experiencia inicial.	RCT prerregistrado con análisis cualitativo y NLP sobre 550.000 mensajes.	Evaluar si la IA puede amplificar la capacidad pedagógica del tutor humano en tiempo real, mejorando resultados de estudiantes de comunidades vulnerables.	900 tutores y 1800 estudiantes (K-12).

integración entre tecnología y supervisión pedagógica es clave para el éxito del modelo, como también se señala en otras experiencias previas con aprendizaje adaptativo (Muralidharan *et al.*, 2019).

Desde el punto de vista de la eficiencia, el programa evaluado por De Simone *et al.* (2025) se sitúa entre las intervenciones educativas más coste-efectivas documentadas en países de renta baja y media. Con un coste estimado de 48 dólares por alumno, se lograron 3,2 años equivalentes de escolarización por cada 100 dólares invertidos. Además, la intervención se implementó con un modelo gratuito y sin necesidad de bancos de preguntas predefinidos, lo que facilita su escalabilidad en contextos con recursos limitados. De Simone *et al.* (2025) subrayan también que la intervención tuvo lugar en condiciones adversas, incluyendo interrupciones de internet y cortes eléctricos, reforzando la robustez de los resultados obtenidos y el potencial del modelo para contextos con limitaciones estructurales.

El estudio de De Simone *et al.* (2025) se enmarca dentro de una literatura emergente sobre el uso de modelos de lenguaje generativo en educación. La experiencia en Nigeria proporciona evidencia relevante sobre el potencial de estas herramientas en secundaria, con alumnado más vulnerable y sistemas educativos más frágiles. En conjunto, los resultados del estudio respaldan la idea de que los tutores virtuales basados en IA, cuando están bien diseñados y acompañados por un entorno de implementación adecuado, pueden generar mejoras sustanciales en los aprendizajes.

El cuadro n.º 2 sintetiza los artículos revisados en la sección tercera, la evidencia empírica sobre tutorías individualizadas guiadas específicamente por IA.

IV. RIESGOS DEL USO DE IA GENERATIVA: VERACIDAD EN LA INFORMACIÓN, SESGOS Y PEREZA COGNITIVA

El apartado anterior muestra experiencias del potencial de los tutores individuales guiados por IA como instrumento de mejora en el aprendizaje de los estudiantes, especialmente del alumnado

rezagado. Pero también es necesario atender a los riesgos asociados al uso de la IA en los centros educativos. El informe de la OCDE (2023) alerta sobre otros riesgos sistémicos: sesgos algorítmicos y falta de transparencia en los modelos, erosión de habilidades socioemocionales, dependencia tecnológica, pérdida de control sobre los datos personales y desinformación. Las respuestas generadas por la IA generativa aunque bien redactadas, pueden ser erróneas, incompletas o sesgadas, lo que exige una alfabetización crítica del alumnado y una guía constante por parte del profesorado.

A este respecto, un enfoque fundamental es mitigar los efectos negativos de la desinformación en entornos digitales. En un estudio reciente, Berger *et al.* (2025) comparan los efectos de dos intervenciones breves: por un lado, el *fact-checking*, que proporciona correcciones explícitas a determinadas noticias falsas; y por otro, una intervención de alfabetización mediática, consistente en diez consejos prácticos para identificar contenidos engañosos en redes sociales. El experimento aleatorizado se llevó a cabo en Alemania con más de 3.000 participantes y midió efectos inmediatos y a corto plazo (dos semanas después) sobre la credibilidad percibida de contenidos, conocimiento factual y actitudes.

Los resultados muestran que ambas intervenciones redujeron la credibilidad de *fakes* sobre vacunas COVID cuando estos eran corregidos. Sin embargo, solo la alfabetización mediática redujo la credibilidad de desinformación no corregida (como la relativa a suplementos nutricionales), tanto inmediatamente como dos semanas después. Asimismo, ambas intervenciones mejoraron el conocimiento factual de los participantes de forma inmediata, pero solo la alfabetización mediática lo hizo de forma sostenida.

En cuanto a las actitudes, la intervención de alfabetización mediática incrementó significativamente la disposición de los participantes a vacunarse o recibir dosis de refuerzo frente al COVID-19. Ninguna de las dos intervenciones afectó negativamente la percepción de publicaciones verídicas: los participantes no se volvieron más escépticos en general, sino más selectivos.

Uno de los mecanismos identificados es que la alfabetización mediática aumenta la atención activa y la evaluación crítica. En una extensión del experimento, se observó que quienes recibieron esta intervención dedicaron más tiempo a leer publicaciones y buscaron más información adicional. En cambio, los participantes expuestos al *fact-checking* actuaron como receptores pasivos de correcciones sin desarrollar habilidades transferibles. Por ejemplo, los efectos de la alfabetización mediática sobre la credibilidad de *fakes news* se mantuvieron dos semanas después, mientras que los del *fact-checking* se desvanecieron.

Las implicaciones educativas del artículo de Berger *et al.* (2025) son evidentes. En contextos donde los estudiantes consumen información a través de redes sociales, fortalecer habilidades críticas para evaluar información resulta más efectivo y escalable que confiar únicamente en correcciones externas. La alfabetización mediática se presenta así como una estrategia pedagógica de bajo coste, fácil implementación y efectos sostenidos, especialmente útil para poblaciones con menor formación informativa. El trabajo de Berger *et al.* (2024) ofrece, por tanto, una advertencia oportuna: en la medida en que los sistemas educativos incorporen tutores IA, también deberán desarrollar marcos éticos y pedagógicos que prevengan los posibles efectos adversos sobre la calidad de la información que los estudiantes consumen y reproducen. Los artículos de Rodríguez (2025), Molina y Cobo (2025) y Lastra y Flynn (2025) en este mismo número de *Papeles de Economía Española* inciden en esta conclusión en el sentido de que la IA no es, en sí misma, una garantía de mejora educativa; su impacto dependerá del ecosistema institucional y formativo en el que se integre.

Una de las propuestas más integrales y contextualizadas para la incorporación de inteligencia artificial generativa en sistemas educativos en desarrollo es la formulada por Levy Yeyati *et al.* (2025). En su trabajo, los autores plantean un marco para integrar herramientas basadas en modelos generativos en aulas de América Latina, bajo criterios de complementariedad, gradualidad y equidad. Frente a enfoques entusiastas adoptados en otros contextos, el estudio subraya que cualquier inte-

gración de IA en educación debe tener en cuenta las condiciones estructurales de los sistemas: desigualdades de acceso, brechas de género, limitaciones formativas del profesorado y falta de conectividad en muchas escuelas.

Levy Yeyati *et al.* (2025) analizan evidencia cualitativa y cuantitativa, con datos provenientes del programa de pensamiento computacional y del *Gender Dashboard* de Ceibal (Uruguay), documentando patrones de uso y desigualdades en la adopción. Por ejemplo, los varones presentan una mayor participación en actividades de robótica, y las docentes mujeres —mayoría en la fuerza laboral educativa— muestran menores tasas de uso de IA, en parte debido a barreras culturales, percepciones propias y menor acceso a formación. En respuesta, los autores recomiendan intervenciones específicas como programas de formación dirigidos a mujeres, estrategias de adopción sensibles al género y modelos híbridos escalables que combinen la supervisión docente con la asistencia de chatbots generativos. El estudio concluye que, si se aplica con cuidado pedagógico e institucional, la IA generativa puede contribuir a reducir las desigualdades, fortalecer la preparación docente y aumentar el compromiso del alumnado.

En términos de implementación, Levy Yeyati *et al.* (2025) abogan por una integración progresiva. El marco propuesto parte del principio de que la IA debe complementar, no sustituir, al profesorado. Los autores sugieren comenzar con usos dirigidos al profesorado, como la planificación de clases o la generación de materiales, y solo introducir aplicaciones dirigidas a estudiantes una vez que se haya proporcionado la formación adecuada. Este enfoque gradual preserva el rol central del docente y permite una apropiación crítica de las herramientas por parte de la comunidad educativa. El enfoque de estos autores está claramente alineado con una concepción de la IA como palanca para la equidad. Subrayan que su valor no reside en tratar a todos por igual, sino en permitir una respuesta más personalizada a quienes suelen quedar invisibilizados: los alumnos con brechas de aprendizaje, aquellos con menor participación verbal o los que requieren más tiempo para procesar la información.

CUADRO N.º 3

DESINFORMACIÓN, SESGOS Y RIESGOS COGNITIVOS DEL USO DE IA EN EDUCACIÓN: EVIDENCIA EXPERIMENTAL Y TEÓRICA

AUTOR	PUBLICACIÓN	PAÍS	MEDIDA DE OUTPUT	DIRECCIÓN DEL EFECTO	TAMAÑO DEL EFECTO
Berger, Kerkhof, Mindl y Münster.	<i>Journal of Public Economics</i> (2025).	Alemania.	Credibilidad percibida de noticias, conocimiento factual, actitudes hacia vacunas.	<i>Media literacy</i> reduce credibilidad de <i>fakes</i> corregidos y no corregidos; <i>fact-checking</i> solo en <i>fakes</i> corregidos. Solo <i>media literacy</i> tiene efectos duraderos.	Efecto sostenido de <i>media literacy</i> en conocimiento y actitudes; <i>fact-checking</i> limitado. ↑ disposición a vacunarse solo con <i>media literacy</i> .
Fan et al.	<i>British Journal of Educational Technology</i> (2024).	China y Australia.	Calidad del texto final, motivación intrínseca, uso de estrategias autorregulatorias.	ChatGPT mejora la calidad del texto, pero reduce el uso de estrategias metacognitivas. No mejora la motivación intrínseca.	Mejora significativa en calidad de escritura con ChatGPT; reducción en planificación, monitoreo y autoevaluación.
Oakley, Johnston, Chen, Jung y Sejnowski.	<i>The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI.</i> (Springer Nature, en prensa 2025).	EE. UU., Nueva Zelanda, Taiwán, Australia.	Formación de esquemas, consolidación de memoria, transferencia de aprendizaje, codificación profunda, razonamiento.	La dependencia de IA educativa y ayudas externas debilita la memoria declarativa y procedural, dificulta la formación de esquemas internos y reduce la capacidad de transferencia y automatización del conocimiento.	No aplica (ensayo teórico, sin datos empíricos).

En definitiva, la propuesta de Levy Yeyati et al. (2025) refuerza la tesis central de este artículo: la IA puede contribuir a mejorar los aprendizajes y reducir desigualdades solo si se despliega bajo un enfoque pedagógico robusto, adaptado al contexto y centrado en el docente. La clave no está en automatizar la enseñanza, sino en crear las condiciones institucionales y formativas que permitan a los docentes aprovechar el potencial de la IA para enseñar mejor.

Otra de las posibles desventajas es el del *offloading* cognitivo: que los estudiantes deleguen en

exceso la comprensión, la memoria o el razonamiento, reduciendo su implicación activa y su capacidad crítica. Este fenómeno, ya observado con tecnologías previas, como los GPS, podría agravarse si no se establece una cultura de uso reflexivo y metacognitivo de las herramientas IA. En línea con las advertencias sobre el riesgo de que la IA generativa debilite el juicio humano, otro estudio reciente, el de Fan et al. (2024), proporciona evidencia adicional sobre cómo el uso de herramientas como ChatGPT puede afectar los procesos autorregulatorios del aprendizaje. Mediante un experimento aleatorizado en entorno de

CUADRO N.º 3 (continuación)

DESINFORMACIÓN, SESGOS Y RIESGOS COGNITIVOS DEL USO DE IA EN EDUCACIÓN: EVIDENCIA EXPERIMENTAL Y TEÓRICA

INTERVENCIÓN	COMENTARIOS	MÉTODO	OBJETIVO	TAMAÑO DE MUESTRA
1) <i>Fact-checking</i> sobre publicaciones falsas seleccionadas; 2) intervención de alfabetización mediática con 10 consejos prácticos.	<i>Media literacy</i> mejora atención crítica, tiempo de lectura y búsqueda activa. No aumenta escepticismo hacia información verídica. Intervención de bajo coste, escalable y efectiva incluso dos semanas después.	RCT (experimento aleatorizado.)	Comparar efectividad de <i>fact-checking</i> y alfabetización mediática en reducir credibilidad de noticias falsas.	~3.000 adultos.
Apoyo a la redacción con: 1) ChatGPT, 2) tutor humano, 3) herramienta analítica, 4) sin ayuda.	ChatGPT facilita la tarea, pero genera "pereza metacognitiva". Requiere estrategias docentes para fomentar el uso reflexivo y no dependiente.	RCT (experimento aleatorizado en laboratorio).	Evaluar efectos del uso de IA, tutores humanos y herramientas digitales sobre motivación, autorregulación y rendimiento en tareas de escritura.	117 estudiantes universitarios.
No hay intervención experimental; se analiza el uso intensivo de IA educativa y metodologías constructivistas como base del <i>cognitive offloading</i> .	Vinculan la reversión del efecto Flynn con el abandono de la instrucción directa; defienden el valor de la memoria interna frente al <i>outsourcing</i> cognitivo; advierten sobre la superficialidad de aprendizajes mediados por IA sin supervisión.	Ensayo teórico basado en literatura neurocientífica.	Analizar cómo el uso intensivo de IA y la dependencia de ayudas externas afectan la consolidación de conocimiento y aprendizaje profundo.	No aplica.

laboratorio, se compararon cuatro tipos de apoyos al aprendizaje durante una tarea de redacción: (i) un chatbot basado en IA generativa (ChatGPT), (ii) un tutor humano experto, (iii) herramientas de analítica de escritura, y (iv) un grupo sin apoyo adicional. El objetivo era analizar diferencias en motivación intrínseca, procesos de autorregulación y rendimiento en la tarea. Los 117 estudiantes universitarios participantes fueron asignados aleatoriamente a cada uno de los cuatro grupos, y se recogieron datos de motivación, comportamiento autorregulatorio y desempeño académico. El grupo que trabajó con ChatGPT mostró

una mejora significativa en la calidad del texto final, lo que evidencia que la herramienta puede tener efectos positivos inmediatos sobre el desempeño. Sin embargo, esta mejora no se tradujo en un progreso en el conocimiento duradero ni en una mayor capacidad de transferencia. Tampoco se observaron diferencias significativas en la motivación intrínseca entre grupos, lo que sugiere que el uso de IA no eleva necesariamente el compromiso interno con la tarea.

Los patrones de autorregulación del aprendizaje fueron distintos según el tipo de apoyo recibido.

En particular, los estudiantes que usaron ChatGPT mostraron una menor frecuencia de estrategias metacognitivas como la planificación, la monitorización y la autoevaluación. Esta tendencia fue interpretada por los autores como una forma de “pereza metacognitiva”: al recibir respuestas bien estructuradas de manera inmediata, los estudiantes tendían a delegar el esfuerzo cognitivo en la herramienta, reduciendo su implicación activa en el proceso de aprendizaje.

Desde un punto de vista teórico, este fenómeno conecta con la noción de *cognitive offloading* (Risko y Gilbert, 2016), según la cual los individuos externalizan tareas mentales para reducir carga cognitiva. Si bien esta estrategia puede ser útil en contextos de sobrecarga, también puede debilitar las capacidades internas de razonamiento cuando se convierte en un hábito. En el caso de ChatGPT, su facilidad de uso y aparente autoridad pueden reducir el umbral de dificultad percibido por los estudiantes, limitando su disposición a revisar, cuestionar o reelaborar la información recibida. El uso de IA generativa también puede reducir el umbral de dificultad percibido por los estudiantes, limitando así el desencadenamiento de procesos analíticos más exigentes —los llamados procesos del “Sistema 2” en la psicología cognitiva (Alter et al., 2007).

Estos resultados refuerzan la importancia de diseñar estrategias pedagógicas que incorporen la IA como apoyo, pero no como sustituto del esfuerzo metacognitivo del alumno. Fan et al. (2024) recomiendan, por ejemplo, que los docentes definan claramente qué tareas deben realizarse con ayuda de IA y cuáles requieren un enfoque más autónomo. También proponen establecer andamiajes explícitos que fomenten la reflexión crítica sobre las respuestas generadas por el modelo, promoviendo así una cultura de uso activo y consciente de la IA en el aula.

En definitiva, este estudio subraya que el impacto de la IA sobre el aprendizaje no puede medirse únicamente en función de resultados a corto plazo. Para que un tutor personalizado basado en IA contribuya realmente a un aprendizaje sostenible, debe integrarse dentro de una ecología educativa que refuerce la motivación intrínseca, el pensamiento crítico y la autorregulación. De lo

contrario, corremos el riesgo de generar una ilusión de competencia, en la que el alumno mejora en tareas específicas, pero pierde autonomía y capacidad de transferencia, elementos clave para el aprendizaje a lo largo de la vida.

Oakley et al. (2025) introducen otra dimensión relevante y poco explorada en el debate sobre la IA generativa en educación: el impacto que tiene en los procesos de memoria y en la consolidación del aprendizaje a largo plazo. A partir de una revisión interdisciplinaria basada en la neurociencia cognitiva, argumentan que el uso excesivo de ayudas externas, como los tutores de IA, puede debilitar los sistemas de memoria declarativa y procedimental, que son fundamentales para el desarrollo de esquemas internos, la intuición experta y el pensamiento flexible.

Oakley et al. (2025) no presentan nueva evidencia empírica, sino que sintetiza resultados recientes de la literatura sobre aprendizaje y memoria. En particular, se sostiene que la dependencia de herramientas como ChatGPT puede fomentar el *cognitive offloading* que ya hemos comentado—la delegación de tareas mentales en dispositivos externos— y obstaculizar la formación de esquemas robustos. Esta práctica compromete la codificación profunda necesaria para el aprendizaje duradero, al limitar el uso del sistema declarativo y dificultar la transición hacia el sistema procedimental, en el que el conocimiento se automatiza y se vuelve intuitivo.

Oakley et al. (2025) vinculan esta preocupación con el retroceso del efecto Flynn —la disminución de los coeficientes intelectuales en países desarrollados desde los años setenta—, sugiriendo que la subutilización de la memoria interna y la externalización excesiva del conocimiento podrían estar debilitando las estructuras cognitivas necesarias para el razonamiento complejo y la transferencia. A nivel teórico, los autores conectan esta problemática con el papel del esfuerzo metacognitivo y con la activación del sistema 2 del pensamiento analítico, que suele verse inhibido cuando la IA proporciona soluciones completas y sin fricción cognitiva. En consecuencia, advierten que un uso pasivo de la IA en contextos educativos podría comprometer el

desarrollo de habilidades profundas y generar una ilusión de competencia sin comprensión real.

En línea con Fan *et al.* (2024), que mostraban una reducción en la autorregulación metacognitiva en estudiantes que usaban ChatGPT sin orientación, Oakley *et al.* (2025) subrayan que el verdadero valor educativo de la IA no radica en ofrecer respuestas, sino en su potencial para promover el esfuerzo mental, la recuperación activa y la formación de conexiones significativas. Por ello, Oakley *et al.* (2025) proponen que la integración de tutores de IA en el aprendizaje debe ir acompañada de un diseño instruccional explícito que estimule la participación activa del alumno, refuerce la memoria interna y evite la sobredependencia de recursos externos.

En cuadro el n.º 3 muestra las principales conclusiones de los artículos de esta sección cuarta sobre los riesgos de los usos de la IA en la educación, entre ellos la desinformación, los sesgos y la delegación del esfuerzo cognitivo.

V. KHANMIGO COMO EJEMPLO DE INTEGRACIÓN DE IA GENERATIVA EN TUTORÍAS PERSONALIZADAS

La Academia Khan es una plataforma ampliamente reconocida por su papel en la difusión de contenidos educativos gratuitos, y cuyo impulsor Salman Amin Khan fue galardonado con el Premio Princesa de Asturias por su Cooperación Internacional en 2019. La Academia Khan ha sido uno de los primeros actores globales en integrar la IA generativa en una solución concreta de tutoría personalizada. Khanmigo, su tutor conversacional, que representa un paso significativo en la evolución del aprendizaje asistido por ordenador (CAL), combinando la entrega de contenido estructurado con interacciones en lenguaje natural que simulan una conversación con un tutor humano.

El enfoque de Khanmigo parte de una constatación común a muchos sistemas educativos: las diferencias individuales en nivel de desempeño y ritmo de aprendizaje generan una tensión constante en el aula, donde algunos estudiantes avanzan

sin haber dominado los contenidos previos, mientras que otros podrían progresar más rápidamente si tuvieran un entorno más flexible. Esta situación no solo dificulta la atención individualizada en aulas numerosas, sino que perpetúa desigualdades en el progreso académico.

Khanmigo, como destaca el artículo de Oreopoulos *et al.* (2024) comentado en la sección tercera, se integra como solución parcial a este problema. En combinación con la plataforma de Khan Academy —que incluye vídeos explicativos, ejercicios interactivos y retroalimentación inmediata—, el tutor IA ofrece acompañamiento en tiempo real, detecta errores comunes, formula preguntas de seguimiento y guía al estudiante hacia la comprensión conceptual. El objetivo no es únicamente proporcionar respuestas correctas, sino fomentar procesos de razonamiento, reflexión y autocomprobación, en línea con prácticas pedagógicas de calidad.

A diferencia de otras herramientas que simplemente presentan contenido, Khanmigo busca crear una experiencia de aprendizaje con más diálogo, con interacciones que imitan la tutoría uno a uno. No obstante, como se ha señalado en otros estudios (Bastani *et al.*, 2024; Wang *et al.*, 2024), la efectividad de este tipo de sistemas depende críticamente del diseño pedagógico y de la supervisión docente. Khanmigo no sustituye al profesor: su función es ampliar su capacidad de seguimiento, adaptarse al nivel del alumno y proporcionar apoyo inmediato cuando no es posible una tutoría humana directa.

Lanzado en marzo de 2023 y desarrollado con tecnología GPT-4, Khanmigo se ha concebido como una interfaz pedagógica que combina rigor de contenido y acompañamiento conversacional. Además de su uso con estudiantes, Khanmigo ha sido diseñado como una herramienta de gran valor para los docentes. Proporciona apoyo para planificar clases, redactar correos electrónicos multilingües a las familias, elaborar rúbricas, y generar ejercicios personalizados en función del tema trabajado en clase. Su objetivo no es automatizar la enseñanza, sino reducir la carga administrativa y permitir que el profesorado pueda dedicar más tiempo a lo esencial: observar, acompañar y enseñar. Esta

función se alinea con los planteamientos de Levy Yeyati (2025) y del Informe de la OCDE (2023), que abogan por una IA que actúe como amplificadora del trabajo docente, no como su sustituta.

Khanmigo también presenta un diseño cuidadoso en cuanto a seguridad y gobernanza. Cuenta con filtros de contenido, registro de conversaciones, alertas para palabras sensibles y supervisión por parte de educadores, todo ello con el fin de asegurar un uso responsable por parte del alumnado. Además, el sistema no proporciona directamente las respuestas a los problemas planteados, sino que promueve estrategias de razonamiento mediante preguntas orientadoras y pistas parciales, lo que previene el riesgo de “pereza cognitiva” o *offloading* detectado en estudios como el de Fan *et al.* (2024).

Este enfoque pedagógico dialoga también con el diseño instruccional de otras intervenciones exitosas, como Tutor CoPilot (Wang *et al.*, 2024), donde el modelo de IA actúa como copiloto del tutor humano y le sugiere buenas prácticas durante la interacción. En el caso de Khanmigo, este rol de copiloto se extiende tanto al estudiante como al profesor, ofreciendo apoyo en ambas direcciones. El sistema puede ser usado también fuera del aula, como refuerzo individual o como herramienta de autoaprendizaje, replicando las condiciones de accesibilidad y escalabilidad que se han valorado positivamente en programas como Rori (Henkel *et al.*, 2024) o TutorAI (Bhatt *et al.*, 2024).

En definitiva, Khanmigo ejemplifica una de las propuestas más completas y coherentes de integración de IA generativa en el entorno escolar. Su valor radica no solo en la sofisticación del modelo subyacente, sino en la arquitectura educativa que lo acompaña: un ecosistema de contenidos de calidad, formación docente, accesibilidad progresiva, supervisión institucional y principios pedagógicos explícitos. Como otras experiencias revisadas en este artículo, demuestra que el impacto positivo de la IA en educación depende menos de la tecnología en sí que del marco en que se inserta. La clave está en utilizar estos sistemas para ampliar las oportunidades de aprendizaje, reforzar la equidad y complementar el trabajo docente, y no en delegar en ellos tareas educativas sin orientación. Khanmigo es, en

este sentido, una de las experiencias más prometedoras de cómo la inteligencia artificial puede integrarse con sentido en las prácticas educativas cotidianas.

VI. CONCLUSIONES

El análisis de los tutores individualizados guiados por IA en contextos educativos reales muestra que no estamos ante una tecnología neutra ni inevitablemente positiva. El impacto de estos sistemas no depende tanto de su sofisticación técnica como del diseño pedagógico e institucional que los acompaña. Los estudios revisados muestran que pueden mejorar significativamente el aprendizaje, especialmente entre el alumnado más rezagado, pero también que mal diseñados o aplicados sin orientación pueden debilitar procesos clave como la autorregulación, la memoria a largo plazo o el pensamiento crítico.

Más allá de sus efectos inmediatos sobre el rendimiento, la IA educativa plantea preguntas de fondo sobre qué significa aprender en la era digital. Si el conocimiento está disponible de forma instantánea, ¿qué lugar ocupa el esfuerzo cognitivo? Si una herramienta responde con precisión a nuestras dudas, ¿qué ocurre con la necesidad de recordar, de practicar o de equivocarse? Este escenario exige avanzar hacia un nuevo pacto pedagógico que incorpore la IA no como sustituto, sino como catalizador del aprendizaje significativo. Ello implica formar a los docentes no solo en el uso técnico de las herramientas, sino en su integración crítica y ética en el aula. También requiere sistemas de supervisión y gobernanza que aseguren equidad, transparencia y protección de los datos del alumnado. Y, sobre todo, invita a redefinir las finalidades de la educación en un entorno donde el conocimiento es abundante, pero la atención, la motivación y la capacidad de pensar de forma autónoma son recursos cada vez más escasos.

En este marco, los tutores IA no deben concebirse como una solución cerrada, sino como una tecnología de propósito general cuya eficacia dependerá de las decisiones políticas, pedagógicas y organizativas que se tomen en torno a su implementación.

Los próximos años serán clave para evitar que esta innovación, en lugar de ampliar oportunidades, refuerce las desigualdades preexistentes o genere nuevas formas de dependencia intelectual. La tarea no es solo adaptar la escuela a la IA, sino asegurarse de que la IA se adapte a los fines de la educación.

BIBLIOGRAFÍA

- Alter, A. L., Oppenheimer, D. M., Epley, N. y Eyre, R. N. (2007).** Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136(4), 569–576. <https://doi.org/10.1037/0096-3445.136.4.569>
- Araya, R., Arias Ortiz, E., Botton, N. y Cristia, J. (2025).** Integrating learning platforms within regular school time: Experimental evidence from Chilean primary schools. *Economics of Education Review*, 106, 102647. <https://doi.org/10.1016/j.econedurev.2025.102647>
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö. y Mariman, R. (2024).** Generative AI can harm learning (*The Wharton School Research Paper*). Disponible en SSRN: <https://ssrn.com/abstract=4895486> o <http://dx.doi.org/10.2139/ssrn.4895486>
- Berger, L. M., Kerkhof, A., Mindl, F. y Münster, J. (2025).** Debunking “fake news” on social media: Immediate and short-term effects of fact-checking and media literacy interventions. *Journal of Public Economics*, 245, 105345. <https://doi.org/10.1016/j.jpubeco.2025.105345>
- Bhatt, R., Kulkarni, M., Muralidharan, K., Singh, A., Sood, A. y Yadav, S. (2024).** Can Technology Facilitate Scale? Evidence from a Randomized Evaluation of High Dosage Tutoring (*NBER Working Paper N° 32510*). National Bureau of Economic Research. <https://www.nber.org/papers/w32510>
- Cuevas-Ruiz, P., Rello, L., Sanz, I. y Sevilla, A. (2025).** Bridging literacy gaps: The impact of AI-driven personalised learning on reading skills and educational equity (*EdWorkingPaper: 25-1209*). Annenberg Institute at Brown University. <https://doi.org/10.26300/68mk-mk89>
- De Barros, A. y Ganimian, A. J. (2024).** Which Students Benefit from Computer-Based Individualized Instruction? Experimental Evidence from Public Schools in India. *Journal of Research on Educational Effectiveness*, 17(2), 318–343. <https://doi.org/10.1080/19345747.2023.2191604>
- Escueta, M., Nickow, A. J., Oreopoulos, P. y Quan, V. (2020).** Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, 58(4), 897–996. <https://doi.org/10.1257/jel.20191507>
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X. y Gašević, D. (2024).** 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>
- Fesler, L., Gu, A. y Chojnacki, G. (2023).** Air tutors’ online tutoring: Math knowledge impacts and participant math perceptions. Middle Years Math Grantee Report Series. *Mathematica*. <https://eric.ed.gov/?id=ED628638>
- Lastra, C. & Flynn, D. J., (2025).** Desinformación entre adolescentes en España: Desafíos y oportunidades de mejora en la escuela y online. *Papeles de Economía Española*.
- Gorjón, L., y Osés, A. (2022).** The negative impact of information and communication technologies overuse on student performance: Evidence from OECD countries. *Journal of Educational Computing Research*, 61(4), 933–961. <https://doi.org/10.1177/07356331221133408>
- Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K. A., ... y Pollack, H. A. (2023).** Not too late: Improving academic outcomes among adolescents. *American Economic Review*, 113(3), 738–765. <https://doi.org/10.1257/aer.20210434>
- Henkel, O., Horne-Robinson, H., Kozhakhmetova, N. y Lee, A. (2024).** Effective and scalable math support: Experimental evidence on the impact of an AI-math tutor in Ghana. En E. André et al. (eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky* (AIED 2024) pp. 373–381. Springer.
- Levy Yeyati, E., Robano, V., Pereiro, E., Porto, C. y Koleszar, V. (2025).** Generative AI in education: A framework for leveraging digital tools in Latin American classrooms. (*School of Government Working Paper N° 202050327*). Universidad Torcuato Di Tella.

- Marcenaro, Ó. D. y López, L. A. (2025).** Sobre el uso de recursos digitales y el rendimiento educativo: De las políticas educativas a la práctica. *Papeles de Economía Española*, (2025 en este mismo número).
- Méndez, I. (2025).** Las tecnologías de la información y la comunicación en España: presente y futuro. *Papeles de Economía Española*, (2025 en este mismo número).
- Molina, E. y Cobo, C. (2025).** *Repensar la política docente en la era de la IA: Evidencia para la acción*. Papeles de Economía Española (2025 en este mismo número).
- Muralidharan, K., Singh, A. y Ganimian, A. J. (2019).** Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4), 1426–1460.
- Oakley, B., Johnston, M., Chen, K.-Z., Jung, E. y Sejnowski, T. (2025).** The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI. En *The Future of Artificial Intelligence: Economics, Society, Risks and Global Policy*. Springer Nature, en prensa.
- OECD (2023).** *Generative AI in the classroom: From hype to reality?* EDU/EDPC(2023)11. Organisation for Economic Co-operation and Development. [https://one.oecd.org/document/EDU/EDPC\(2023\)11/en/pdf](https://one.oecd.org/document/EDU/EDPC(2023)11/en/pdf)
- Oreopoulos, P., Gibbs, C. R., Jensen, M. y Price, J. P. (2024).** Teaching Teachers to Use Computer Assisted Learning Effectively: Experimental and Quasi-Experimental Evidence. *EdWorkingPaper* N°. 24-1036. Annenberg Institute at Brown University. <https://doi.org/10.26300/znjb-1v83>
- Risko, E. F. y Gilbert, S. J. (2016).** Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Robinson, C. D., Pollard, C., Novicoff, S., White, S. y Loeb, S. (2024).** The effects of virtual tutoring on young readers: Results from a randomized controlled trial. *Educational Evaluation and Policy Analysis*. <https://doi.org/10.3102/01623737241288845>
- Rodríguez, F. (2025).** La inteligencia artificial como herramienta transformadora en la educación: Aplicación, impacto potencial y políticas. Papeles de Economía Española, (2025 en este mismo número).
- Roldán, T. (2025, 7 de febrero).** La IA revolucionará la educación (para bien). *El País*, p. 13.
- Sevilla, A., Cuevas, P., Rello, L., & Sanz, I. (2024).** El impacto del aprendizaje asistido por tecnología en la educación no universitaria. *Panorama Social*, (39), 73–93
- De Simone, L., Ogundeyi, A., Adesanya, A., Bello, M., Adeniran, A., Bello, A., Daramola, A. y Jhingran, D. (2025).** Generative AI for Education: Experimental Evidence from Nigeria. *Policy Research Working Paper* N°. 10747. World Bank. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099237502082424179/idu0d44e3ad04f055b0b550ae9401780c899f6f7>
- Thomas, D., Marsh, J. A., Allbright, T., Johnson, R. y Jennings, J. (2024).** *Scaling Human Intelligence with Artificial Intelligence: A Hybrid Tutoring Model Supporting Students' Math Learning and Engagement*. AIED 2024: Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky. Communications in Computer and Information Science, 2150, 276–287. Springer.
- Wang, R. E., Ribeiro, A. T., Robinson, C. D., Loeb, S. y Demszky, D. (2024).** Tutor CoPilot: A human-AI approach for scaling real-time expertise. Stanford University. <https://osf.io/8d6ha>