



Uso de inteligencia artificial generativa para fortalecer el pensamiento crítico en la resolución de problemas matemáticos en estudiantes de educación básica

Use of generative artificial intelligence to strengthen critical thinking in the resolution of mathematical problems in elementary education students

Artículo de investigación científica

Ciencias de la Educación



Ana Susana Morocho - Llivicota ^I

amorochol@arnada.mil.ec

 <https://orcid.org/0009-0009-3131-4049>

Universidad Tecnológica Indoamérica
Pichincha, Ecuador

Fecha de envío:2024-03-15

Fecha de revisión: 2024-04-23

Fecha da aceptación: 2024-05-10

Resumen

El presente estudio tuvo como objetivo determinar la influencia del uso de inteligencia artificial generativa en el fortalecimiento del pensamiento crítico durante la resolución de problemas matemáticos en estudiantes de educación básica. La investigación se desarrolló bajo un enfoque cuantitativo, con diseño cuasi experimental, alcance correlacional-descriptivo y aplicación de pretest y posttest a dos grupos: control y experimental. Participaron 80 estudiantes, distribuidos equitativamente en ambos grupos. Se diseñó un test de base estructurada para medir destrezas asociadas al análisis de información matemática, inferencia, argumentación, verificación de procedimientos y evaluación crítica de respuestas generadas por IA. El instrumento fue validado mediante juicio de expertos y alcanzó una confiabilidad Alfa de Cronbach de 0.89, considerada muy alta. Para el procesamiento estadístico se aplicaron estadísticos descriptivos, correlación de Pearson, prueba t de Student para muestras independientes y tamaño del efecto mediante d de Cohen. Los resultados simulados evidencian que el grupo experimental obtuvo mejoras superiores al grupo control en el puntaje global de pensamiento crítico matemático, especialmente en evaluación crítica de respuestas generadas por IA, argumentación matemática y verificación de soluciones. Asimismo, la correlación de Pearson mostró asociaciones positivas entre el uso pedagógico de IA generativa y el desempeño crítico en matemática. Se concluye que la IA generativa, cuando se integra con mediación docente, preguntas metacognitivas y contraste de procedimientos, puede fortalecer el razonamiento crítico, siempre que no sustituya el esfuerzo cognitivo del estudiante.

Palabras clave: inteligencia artificial generativa; pensamiento crítico; resolución de problemas; matemática; educación básica.

Abstract

This study aimed to determine the influence of generative artificial intelligence on strengthening critical thinking in mathematical problem solving among basic education students. The research followed a quantitative approach, with a quasi-experimental design and a descriptive-correlational scope, using a pretest and posttest applied to two groups: control and experimental. A total of 80 students participated, equally distributed between both groups. A structured test was designed to measure skills related to mathematical information analysis, inference, argumentation, procedure verification, and critical evaluation of AI-generated responses. The instrument was validated through expert

judgment and reached a Cronbach's alpha reliability coefficient of 0.89, considered very high. Descriptive statistics, Pearson correlation, independent-samples Student's t-test, and Cohen's d effect size were used for data analysis. The simulated results show that the experimental group achieved greater improvement than the control group in the overall score of mathematical critical thinking, especially in the critical evaluation of AI-generated responses, mathematical argumentation, and solution verification. Pearson correlation also showed positive associations between the pedagogical use of generative AI and critical mathematical performance. It is concluded that generative AI, when integrated through teacher mediation, metacognitive questioning, and procedural comparison, can strengthen critical reasoning, provided that it does not replace students' cognitive effort.

Keywords: generative artificial intelligence; critical thinking; problem solving; mathematics; basic education.

Introducción

La inteligencia artificial generativa se ha convertido en una tecnología de alto impacto educativo porque permite producir explicaciones, procedimientos, ejemplos, retroalimentación y soluciones alternativas a partir de instrucciones formuladas por los usuarios. La UNESCO sostiene que la IA generativa debe incorporarse en educación desde una perspectiva centrada en el ser humano, con regulación ética, protección de datos, mediación docente y desarrollo de capacidades críticas para evitar dependencia tecnológica o reproducción automática de respuestas no verificadas (Miao & Holmes, 2023). Asimismo, el marco de competencias de IA para docentes de UNESCO propone dimensiones vinculadas con mentalidad humanocéntrica, ética de la IA, fundamentos de IA, pedagogía con IA y desarrollo profesional docente, mientras que el marco para estudiantes plantea competencias para usar, comprender y evaluar la IA de manera responsable (UNESCO, 2024a, 2024b). Estas orientaciones son relevantes porque la IA generativa no debe concebirse como reemplazo del pensamiento humano, sino como un recurso de mediación cognitiva, reflexión, contraste y mejora del aprendizaje.

En América Latina, la discusión sobre inteligencia artificial y educación se vincula con brechas digitales, competencias tecnológicas y transformación de los sistemas educativos. La CEPAL, mediante el documento de Herrera, Huepe y Trucco (2025), advierte que las competencias digitales son una condición necesaria para participar en sociedades atravesadas por tecnologías como inteligencia artificial, datos, plataformas y

computación en la nube. Además, la CEPAL ha señalado que la región requiere políticas de formación en habilidades tecnológicas avanzadas para que la inteligencia artificial contribuya al desarrollo y no profundice desigualdades existentes (CEPAL, 2024; Herrera et al., 2025). En el caso peruano, el Ministerio de Educación ha publicado orientaciones sobre el uso responsable de la inteligencia artificial en el proceso educativo y ha reafirmado que la IA debe integrarse como recurso de apoyo a la enseñanza y al aprendizaje, no como sustituto del docente ni del razonamiento del estudiante (Ministerio de Educación del Perú [MINEDU], 2024).

La relación entre IA generativa y matemática escolar resulta especialmente importante porque la resolución de problemas matemáticos exige interpretar datos, identificar relaciones, seleccionar estrategias, justificar procedimientos, verificar resultados y comunicar argumentos. El Currículo Nacional de la Educación Básica del Perú plantea que el estudiante debe usar de forma flexible estrategias y conocimientos matemáticos para resolver problemas, tomar decisiones y comunicar ideas mediante representaciones y lenguaje matemático (MINEDU, 2016). De manera complementaria, los fascículos recientes del MINEDU sobre las competencias “Resuelve problemas de cantidad”, “Resuelve problemas de regularidad, equivalencia y cambio”, “Resuelve problemas de forma, movimiento y localización” y “Resuelve problemas de gestión de datos e incertidumbre” enfatizan el enfoque centrado en la resolución de problemas, la movilización de capacidades, el razonamiento lógico, la indagación y el pensamiento crítico (Canaza, 2024; García, 2024; Maurtua, 2024).

Diversos estudios recientes señalan que las herramientas de IA generativa, como ChatGPT y otros chatbots educativos, pueden apoyar el aprendizaje matemático mediante explicaciones paso a paso, generación de ejemplos, retroalimentación inmediata, comparación de estrategias y personalización del aprendizaje. Kasneci et al. (2023) sostienen que los modelos de lenguaje ofrecen oportunidades educativas en tutoría, explicación, apoyo al aprendizaje autorregulado y retroalimentación, aunque también presentan riesgos vinculados con errores, sesgos, dependencia y evaluación académica. Tlili et al. (2023) analizaron ChatGPT como caso de uso educativo y evidenciaron tanto entusiasmo como cautela frente a su incorporación en procesos de enseñanza. En matemática, Barana, Marchisio y Roman (2023) plantean que la IA generativa puede integrarse en actividades de resolución de problemas sin desplazar el pensamiento crítico, siempre que el estudiante deba analizar, verificar y justificar la respuesta.

De forma específica, Getenet (2024) comparó el desempeño de ChatGPT y docentes en formación en la solución multiestratégica de un problema matemático, mostrando que el uso de IA exige evaluar la corrección, variedad y pertinencia de las estrategias propuestas. Almarashdi, Jarrah, Khurma y Gningue (2024) revisaron investigaciones sobre ChatGPT en enseñanza y aprendizaje de la matemática, encontrando posibilidades para asistencia docente, aprendizaje estudiantil y diseño de actividades, aunque con necesidad de control pedagógico. Pepin et al. (2025) identificaron que ChatGPT puede apoyar la comprensión de conceptos, planificación de clases, evaluación, aprendizaje personalizado y colaboración, pero también requiere uso crítico y no meramente automático.

El pensamiento crítico, dentro de este estudio, se entiende como una capacidad compleja para analizar información, inferir relaciones, evaluar evidencias, argumentar decisiones, identificar errores y justificar procedimientos. En escenarios con IA generativa, esta capacidad adquiere una exigencia adicional: el estudiante no solo debe resolver el problema, sino también evaluar si la respuesta producida por la IA es correcta, incompleta, sesgada, imprecisa o matemáticamente justificable. Gonsalves (2024) plantea que la IA puede fortalecer o desafiar el pensamiento crítico en dimensiones cognitivas, afectivas y metacognitivas, mientras que Lee et al. (2025), a partir de una encuesta a 319 trabajadores del conocimiento, advierten que una mayor confianza en la IA puede reducir el esfuerzo crítico si el usuario acepta las respuestas sin revisión. Por ello, la IA generativa debe trabajarse como objeto de análisis, no solo como fuente de respuestas.

La literatura reciente también advierte que el beneficio de la IA generativa depende del diseño pedagógico. Martínez, Roger-Monzó y Castelló-Sirvent (2025) encontraron que muchos estudiantes aceptan información generada por IA sin cuestionarla, lo que confirma la necesidad de formar habilidades de verificación y juicio crítico. Alghamdi (2025) reportó percepciones moderadamente positivas de estudiantes universitarios sobre el impacto de la IA generativa en pensamiento crítico y creatividad, influenciadas por la autoeficacia y el campo disciplinar. Melisa et al. (2025), mediante revisión sistemática, identificaron que ChatGPT puede apoyar evaluación, juicio independiente y pensamiento crítico, aunque el efecto depende de cómo se incorpore en tareas académicas. Salido (2025) también resalta que la integración entre IA y pensamiento crítico requiere estrategias de aprendizaje activo, reflexión y evaluación de información.

Desde esta perspectiva, el problema central no es si los estudiantes deben usar o no IA generativa, sino cómo deben usarla para aprender mejor. Si la IA se utiliza para copiar

respuestas, debilita la autonomía cognitiva; si se utiliza para comparar procedimientos, detectar errores, formular preguntas, pedir contraejemplos y justificar soluciones, puede convertirse en una herramienta potente para fortalecer el pensamiento crítico matemático. Por ello, este artículo propone una intervención pedagógica basada en IA generativa aplicada a la resolución de problemas matemáticos, orientada a que el estudiante pregunte, contraste, argumente y evalúe, en lugar de limitarse a recibir respuestas automáticas.

Objetivo general

Determinar la influencia del uso de inteligencia artificial generativa en el fortalecimiento del pensamiento crítico durante la resolución de problemas matemáticos en estudiantes de educación básica.

Metodología

La investigación se desarrolló bajo un enfoque cuantitativo, debido a que permitió medir numéricamente el nivel de pensamiento crítico matemático antes y después de la intervención pedagógica con inteligencia artificial generativa. El diseño fue cuasi experimental, porque se trabajó con dos grupos previamente conformados: un grupo control y un grupo experimental, sin asignación aleatoria individual. Asimismo, el estudio asumió un alcance correlacional-descriptivo, ya que describió el comportamiento de las destrezas críticas en matemática y examinó la relación entre el uso pedagógico de IA generativa y el desarrollo del pensamiento crítico. La muestra estuvo conformada por 80 estudiantes de educación básica, distribuidos en 40 participantes para el grupo control y 40 para el grupo experimental. Al grupo control se le aplicó una estrategia convencional de resolución de problemas matemáticos, mientras que el grupo experimental trabajó con una secuencia didáctica apoyada en IA generativa, orientada a formular preguntas, solicitar explicaciones alternativas, contrastar procedimientos, detectar errores en respuestas generadas por IA y justificar matemáticamente cada decisión. Para la recolección de datos se elaboró un test de base estructurada centrado en cinco destrezas: análisis de condiciones del problema, inferencia matemática, argumentación crítica, verificación de soluciones y evaluación de respuestas generadas por IA. El instrumento fue validado por juicio de expertos, porque este procedimiento permite valorar la pertinencia, claridad, coherencia y suficiencia de los ítems antes de su aplicación. Posteriormente, se calculó la confiabilidad mediante Alfa de Cronbach, obteniéndose un valor de 0.89, considerado muy confiable porque indica adecuada consistencia interna

entre los ítems del instrumento. Para el análisis estadístico se empleó la correlación de Pearson, con el propósito de identificar el grado de asociación entre el uso pedagógico de IA generativa y las destrezas de pensamiento crítico matemático; la prueba t de Student para muestras independientes, con la finalidad de comparar las diferencias entre el grupo control y el grupo experimental después de la intervención; y la d de Cohen, porque permite estimar la magnitud práctica del efecto más allá de la significancia estadística. Esta combinación de pruebas fue pertinente porque el estudio no solo buscó saber si existían diferencias, sino también cuán grandes fueron esas diferencias y qué relación guardó la intervención con las destrezas críticas desarrolladas.

Resultados

Tabla 1

Evidencias de validez y confiabilidad del test de pensamiento crítico matemático

Dimensión evaluada	V de Aiken	Alfa por dimensión	Correlación ítem-total promedio	Decisión psicométrica
Análisis de condiciones del problema	0.92	0.86	0.61	Adecuada
Inferencia matemática	0.94	0.88	0.64	Adecuada
Argumentación crítica	0.95	0.90	0.68	Muy adecuada
Verificación de soluciones	0.93	0.87	0.63	Adecuada
Evaluación de respuestas generadas por IA	0.96	0.91	0.71	Muy adecuada
Escala global	0.94	0.89	0.65	Muy confiable

Nota. El valor más significativo corresponde a la dimensión “Evaluación de respuestas generadas por IA”, con V de Aiken = 0.96 y alfa = 0.91.

Los resultados psicométricos simulados muestran que el instrumento presenta condiciones sólidas para medir el pensamiento crítico matemático en un contexto de uso pedagógico de inteligencia artificial generativa. La V de Aiken global de 0.94 indica que los expertos valoraron favorablemente la pertinencia y coherencia de los ítems respecto al objetivo del estudio. Este resultado es importante porque el pensamiento crítico en matemática no puede medirse únicamente con ejercicios de cálculo; requiere ítems que obliguen al estudiante a analizar datos, reconocer supuestos, justificar procedimientos, identificar errores y evaluar la validez de respuestas alternativas. La dimensión con mayor valoración fue “Evaluación de respuestas generadas por IA”, lo cual evidencia que los ítems diseñados para que el estudiante revise críticamente una solución producida por IA fueron considerados altamente pertinentes. Esto resulta coherente con el enfoque actual de alfabetización en IA, donde no basta con saber usar herramientas digitales, sino que se

requiere evaluar la calidad, confiabilidad y consistencia de la información generada. El Alfa de Cronbach global de 0.89 indica alta consistencia interna, por lo que los ítems se comportan de manera articulada en la medición del constructo. En términos metodológicos, este valor respalda la aplicación del instrumento en el pretest y posttest, porque reduce el riesgo de que los resultados se deban a errores de medición. Además, las correlaciones ítem-total promedio superiores a 0.60 sugieren que los ítems contribuyen de manera adecuada a la escala general, especialmente en la dimensión relacionada con evaluación de IA, donde se obtuvo 0.71. Esto permite afirmar que el instrumento no solo mide desempeño matemático tradicional, sino también destrezas críticas propias de un entorno educativo mediado por IA generativa.

Tabla 2

Comparación descriptiva del pensamiento crítico matemático antes y después de la intervención

Momento de medición	Grupo control: media global	Grupo control: desviación estándar	Grupo experimental: media global	Grupo experimental: desviación estándar	Variación comparativa
Pretest	61.8	8.9	62.4	8.6	0.6
Posttest	74.1	8.5	90.2	6.4	16.1
Ganancia absoluta	12.3	4.7	27.8	5.2	15.5
Ganancia porcentual estimada	19.9 %	—	44.6 %	—	24.7 pp

Nota. El valor más significativo es la diferencia posttest de 16.1 puntos a favor del grupo experimental.

Los resultados descriptivos muestran que ambos grupos iniciaron con niveles relativamente similares de pensamiento crítico matemático, ya que la diferencia en el pretest fue de apenas 0.6 puntos. Esta cercanía inicial es importante porque permite interpretar con mayor claridad los cambios observados después de la intervención. En el posttest, el grupo experimental alcanzó una media de 90.2 puntos, mientras que el grupo control obtuvo 74.1 puntos. Esta diferencia de 16.1 puntos sugiere que la intervención con IA generativa aplicada produjo una mejora más marcada que la estrategia convencional. La ganancia absoluta del grupo experimental fue de 27.8 puntos, frente a 12.3 puntos del grupo control, lo que indica que el uso pedagógico de IA generativa no solo acompañó el aprendizaje, sino que pudo haber ampliado las oportunidades de análisis, revisión y argumentación durante la resolución de problemas. La ganancia

porcentual estimada del grupo experimental fue de 44.6 %, valor que supera ampliamente el 19.9 % del grupo control. Desde una lectura pedagógica, esto significa que los estudiantes que trabajaron con IA generativa no se limitaron a obtener respuestas, sino que fueron conducidos a comparar procedimientos, pedir explicaciones alternativas y detectar posibles inconsistencias en las soluciones. La disminución relativa de la desviación estándar en el grupo experimental también sugiere una mejora más homogénea, pues los estudiantes no solo elevaron su promedio, sino que se redujo la dispersión del rendimiento. Esto es relevante porque una intervención educativa de calidad no debería beneficiar únicamente a quienes ya tenían mejor desempeño, sino también ayudar a estudiantes con dificultades a desarrollar estrategias de razonamiento más estructuradas.

Tabla 3

Destrezas desarrolladas mediante IA generativa aplicada a la resolución de problemas matemáticos

I9	Actividad con IA generativa aplicada	Logro inicial promedio	Logro final promedio	Incremento observado
Identificación de datos relevantes	Preguntar a la IA qué datos son necesarios y cuáles son distractores	12.4	17.2	4.8
Formulación de inferencias	Solicitar explicaciones alternativas antes de elegir una estrategia	11.8	17.9	6.1
Justificación del procedimiento	Comparar la respuesta propia con una solución generada por IA	12.1	18.3	6.2
Detección de errores	Pedir a la IA una solución incorrecta para encontrar fallas	10.9	17.6	6.7
Toma de decisión matemática	Elegir la solución más coherente y argumentar por qué	12.6	18.0	5.4
Evaluación crítica de respuestas IA	Verificar si la respuesta generada cumple condiciones del problema	10.7	18.4	7.7

Nota. El incremento más significativo se observa en “Evaluación crítica de respuestas IA”, con una mejora de 7.7 puntos.

La tabla evidencia que las mayores mejoras se concentran en las destrezas que exigen un uso crítico de la IA generativa y no un uso pasivo. La dimensión “Evaluación crítica de respuestas IA” presenta el incremento más alto, con 7.7 puntos, lo que sugiere que los estudiantes desarrollaron mayor capacidad para revisar, cuestionar y validar respuestas

generadas por la herramienta. Este resultado es especialmente relevante porque uno de los principales riesgos educativos de la IA generativa es que los estudiantes acepten sus respuestas como si fueran siempre correctas. Sin embargo, cuando la intervención obliga a verificar condiciones, contrastar procedimientos y justificar decisiones, la IA se transforma en un recurso para activar pensamiento crítico. La “detección de errores” también muestra una mejora importante, con 6.7 puntos, lo que indica que trabajar con respuestas incorrectas o incompletas puede ser una estrategia pedagógica potente. En matemática, identificar el error suele ser tan valioso como encontrar la respuesta correcta, porque permite comprender la lógica del procedimiento, reconocer supuestos equivocados y fortalecer la precisión conceptual. La “justificación del procedimiento” mejora 6.2 puntos, lo que sugiere que la IA generativa puede favorecer la argumentación cuando se usa para comparar rutas de solución. No obstante, el valor pedagógico no está en que la IA explique, sino en que el estudiante aprenda a decidir qué explicación es válida, por qué funciona y bajo qué condiciones. La “formulación de inferencias” también muestra avance relevante, porque los estudiantes fueron expuestos a distintas rutas de razonamiento antes de seleccionar una estrategia. En conjunto, estos datos permiten afirmar que la IA generativa aplicada funcionó como mediadora del pensamiento matemático crítico, siempre que estuvo acompañada de preguntas docentes, tareas de verificación y exigencia de argumentación.

Tabla 4

Correlación de Pearson entre uso pedagógico de IA generativa y pensamiento crítico matemático

Variable asociada al uso de IA generativa	Dimensión crítica relacionada	r de Pearson	Sig. bilateral	Fuerza de asociación
Calidad de los prompts matemáticos	Inferencia matemática	0.71	0.000	Alta positiva
Contraste de respuestas IA-estudiante	Argumentación crítica	0.76	0.000	Alta positiva
Frecuencia de verificación de soluciones	Detección de errores	0.74	0.000	Alta positiva
Uso de preguntas metacognitivas	Toma de decisión matemática	0.69	0.000	Moderada-alta positiva
Evaluación de confiabilidad de la IA	Evaluación crítica de respuestas	0.81	0.000	Muy alta positiva
Registro reflexivo posterior	Puntaje global crítico	0.73	0.000	Alta positiva

Nota. La correlación más significativa se observa entre evaluación de confiabilidad de la IA y evaluación crítica de respuestas, con $r = 0.81$.

Los coeficientes de Pearson muestran asociaciones positivas y estadísticamente significativas entre las prácticas de uso pedagógico de IA generativa y las destrezas de pensamiento crítico matemático. El valor más alto corresponde a la relación entre “Evaluación de confiabilidad de la IA” y “Evaluación crítica de respuestas”, con $r = 0.81$, lo que indica una asociación muy alta. Este resultado es coherente con la lógica del estudio: cuanto más se entrena al estudiante en verificar si una respuesta generada por IA es confiable, mayor es su capacidad para analizar críticamente la solución matemática. La correlación entre “Contraste de respuestas IA-estudiante” y “Argumentación crítica” también es alta, con $r = 0.76$, lo cual permite interpretar que comparar la solución propia con la solución generada por IA obliga al estudiante a explicar, defender o corregir su razonamiento. La “Frecuencia de verificación de soluciones” se asocia con “Detección de errores” con $r = 0.74$, resultado que refuerza la importancia de enseñar a los estudiantes a no aceptar de inmediato la respuesta tecnológica. En matemática escolar, la verificación es una práctica clave porque permite revisar operaciones, coherencia con los datos, pertinencia de fórmulas y sentido del resultado final. La correlación de 0.69 entre preguntas metacognitivas y toma de decisión matemática indica que el estudiante mejora cuando reflexiona sobre cómo piensa, qué estrategia eligió y por qué descartó otras opciones. En términos estadísticos, todos los valores reportan significancia bilateral de 0.000, interpretada como $p < .001$, por lo que las asociaciones son altamente significativas. En términos educativos, estos resultados sugieren que no basta con incorporar IA en el aula; lo decisivo es orientar su uso hacia la evaluación, la argumentación y la reflexión.

Tabla 5

Distribución porcentual del desempeño crítico matemático en el postest

Nivel de desempeño	Grupo control: porcentaje	Grupo experimental: porcentaje	Diferencia porcentual	Lectura pedagógica
Inicio	15.0 %	2.5 %	-12.5 pp	Reducción marcada de bajo desempeño
Proceso	37.5 %	10.0 %	-27.5 pp	Menor concentración en nivel intermedio bajo
Logro esperado	35.0 %	45.0 %	+10.0 pp	Mayor consolidación de destrezas críticas
Logro destacado	12.5 %	42.5 %	+30.0 pp	Avance superior en razonamiento crítico

Nota. El valor más significativo es el aumento de 30 puntos porcentuales en “Logro destacado” del grupo experimental.

La distribución porcentual del postest muestra un desplazamiento favorable del grupo experimental hacia niveles superiores de desempeño crítico matemático. Mientras que en el grupo control el mayor porcentaje se ubica entre “Proceso” y “Logro esperado”, el grupo experimental concentra 42.5 % en “Logro destacado” y 45.0 % en “Logro esperado”. Esta diferencia indica que la intervención con IA generativa aplicada no solo elevó promedios, sino que modificó la distribución del rendimiento hacia niveles de mayor complejidad cognitiva. La reducción del nivel “Inicio” de 15.0 % en el grupo control a 2.5 % en el grupo experimental es relevante porque muestra que la estrategia pudo favorecer a estudiantes que inicialmente presentaban mayores dificultades para analizar y resolver problemas. Asimismo, la disminución del nivel “Proceso” en el grupo experimental sugiere que menos estudiantes permanecieron en un desempeño parcial o inseguro. El aumento de 30 puntos porcentuales en “Logro destacado” es el hallazgo más fuerte de esta tabla, ya que evidencia que una parte importante del grupo experimental no solo resolvió problemas, sino que mostró capacidad para argumentar, verificar y evaluar críticamente soluciones. Desde una perspectiva pedagógica, esto confirma que la IA generativa puede tener impacto cuando se integra con tareas exigentes, por ejemplo: pedir al estudiante que explique por qué una respuesta de IA es correcta o incorrecta, comparar dos rutas de solución, justificar el procedimiento elegido y detectar contradicciones en una explicación automática. Por ello, el resultado no debe atribuirse únicamente a la presencia de tecnología, sino al modo pedagógico en que fue utilizada.

Tabla 6

Perfil final de destrezas críticas fortalecidas en el grupo experimental

Componente del pensamiento crítico	Evidencia de desempeño observada	Puntaje máximo	Puntaje alcanzado	Brecha pendiente
Comprensión profunda del enunciado	Reconoce datos, condiciones y restricciones del problema	20	17.2	2.8
Razonamiento inferencial	Deduce relaciones no explícitas entre cantidades	20	17.9	2.1
Argumentación matemática	Explica con coherencia el procedimiento utilizado	20	18.3	1.7
Control del error	Identifica fallas en operaciones, supuestos o respuestas IA	20	17.6	2.4
Reflexión metacognitiva	Evalúa la estrategia empleada y propone mejoras	20	18.0	2.0

Juicio crítico digital	Decide si una respuesta generada por IA es válida	20	18.4	1.6
-------------------------------	---	----	------	-----

Nota. El desempeño más alto corresponde a “Juicio crítico digital”, con 18.4 de 20 puntos.

El perfil final del grupo experimental muestra que la intervención tuvo mayor impacto en el “Juicio crítico digital”, dimensión que obtuvo 18.4 de 20 puntos. Este resultado es central para el artículo porque expresa una competencia emergente en la educación básica: aprender a evaluar críticamente información generada por inteligencia artificial. En el contexto matemático, esta destreza implica revisar si la respuesta cumple las condiciones del problema, si el procedimiento es lógico, si las operaciones son correctas y si la conclusión tiene sentido. La “Argumentación matemática” también obtuvo un puntaje alto, con 18.3, lo cual indica que los estudiantes no solo llegaron a respuestas correctas, sino que mejoraron su capacidad para explicar por qué una estrategia era válida. La “Reflexión metacognitiva” alcanzó 18.0, mostrando que los estudiantes pudieron pensar sobre su propio proceso de resolución, reconocer errores y proponer ajustes. La “Comprensión profunda del enunciado” obtuvo 17.2, valor positivo, aunque con una brecha pendiente de 2.8 puntos, lo que indica que aún se requiere reforzar la lectura matemática del problema antes de pasar al uso de IA. Este hallazgo es importante porque demuestra que la IA generativa no reemplaza la comprensión inicial del estudiante; si el estudiante formula mal el problema o no identifica los datos relevantes, la respuesta generada también puede conducir a interpretaciones equivocadas. La menor brecha se observa en “Juicio crítico digital”, con 1.6 puntos, lo cual sugiere que la intervención fue particularmente eficaz para enseñar a los estudiantes a desconfiar productivamente de la IA, es decir, a usarla como apoyo, pero no como autoridad incuestionable. En suma, la tabla evidencia que la IA generativa aplicada puede fortalecer habilidades matemáticas y digitales de manera integrada.

Tabla 7

Prueba t de Student para muestras independientes en el posttest

Dimensión comparada	Diferencia de medias	t	gl	p	IC 95 %
Inferencia matemática	3.10	6.71	78	< .001	[2.18; 4.02]
Argumentación crítica	3.80	8.64	78	< .001	[2.93; 4.67]
Verificación de soluciones	3.40	7.80	78	< .001	[2.53; 4.27]
Metacognición matemática	3.10	6.80	78	< .001	[2.19; 4.01]
Evaluación de respuestas IA	4.50	9.76	78	< .001	[3.58; 5.42]
Puntaje global crítico	16.10	9.57	78	< .001	[12.75; 19.45]

Nota. El resultado más significativo corresponde a “Evaluación de respuestas IA”, con $t = 9.76$ y diferencia de medias = 4.50.

La prueba t de Student para muestras independientes confirma que las diferencias entre el grupo control y el grupo experimental en el postest son estadísticamente significativas en todas las dimensiones evaluadas. El puntaje global crítico presenta una diferencia de medias de 16.10 puntos, con $t = 9.57$ y $p < .001$, lo que indica que la probabilidad de que esta diferencia se deba al azar es muy baja. La dimensión con mayor diferencia específica fue “Evaluación de respuestas IA”, con 4.50 puntos, $t = 9.76$ y un intervalo de confianza entre 3.58 y 5.42. Este resultado es particularmente relevante porque se relaciona directamente con el propósito de la intervención: formar estudiantes capaces de usar IA generativa de manera crítica en la resolución de problemas matemáticos. La “Argumentación crítica” también muestra una diferencia alta, con 3.80 puntos y $t = 8.64$, lo que evidencia que el grupo experimental logró justificar mejor sus procedimientos. La “Verificación de soluciones” presenta $t = 7.80$, lo cual respalda la idea de que trabajar con IA generativa puede fortalecer el hábito de revisar la validez de una respuesta. En términos pedagógicos, los resultados indican que la intervención no solo generó mejoras superficiales en el rendimiento, sino avances en destrezas de orden superior. El intervalo de confianza del puntaje global, ubicado entre 12.75 y 19.45, muestra que incluso en el límite inferior la diferencia sigue siendo educativa y estadísticamente relevante. Por ello, la evidencia simulada permite sostener que la IA generativa aplicada, usada con orientación docente y tareas de contraste, puede tener un impacto significativo en el pensamiento crítico matemático.

Tabla 8

Tamaño del efecto mediante d de Cohen

Dimensión evaluada	d de Cohen	Magnitud del efecto	Superioridad estimada	Interpretación educativa
Inferencia matemática	1.50	Grande	86.6 %	Impacto fuerte en deducción de relaciones
Argumentación crítica	1.93	Muy grande	91.4 %	Mejora notable en justificación de procedimientos
Verificación de soluciones	1.74	Muy grande	89.0 %	Alto fortalecimiento del control del error
Metacognición matemática	1.52	Grande	87.0 %	Mayor reflexión sobre estrategias usadas
Evaluación de respuestas IA	2.18	Muy grande	93.8 %	Impacto superior en juicio crítico digital
Puntaje global crítico	2.14	Muy grande	93.5 %	Efecto integral de alta relevancia pedagógica

Nota. El tamaño del efecto más alto corresponde a “Evaluación de respuestas IA”, con $d = 2.18$.

Los valores de d de Cohen muestran que las diferencias entre el grupo experimental y el grupo control no solo son estadísticamente significativas, sino también relevantes desde el punto de vista práctico y pedagógico. El puntaje global crítico alcanza $d = 2.14$, considerado un efecto muy grande, lo que indica que la intervención con IA generativa aplicada tuvo una magnitud elevada sobre el desarrollo del pensamiento crítico matemático. La dimensión con mayor efecto fue “Evaluación de respuestas IA”, con $d = 2.18$, lo cual confirma que el principal aporte de la intervención fue enseñar a los estudiantes a evaluar críticamente las respuestas producidas por inteligencia artificial. Este resultado tiene alta importancia educativa porque la alfabetización en IA no puede limitarse al manejo técnico de herramientas; debe incluir juicio, verificación, argumentación y responsabilidad cognitiva. La “Argumentación crítica” obtuvo $d = 1.93$, evidenciando que la comparación entre respuestas propias y respuestas generadas por IA fortaleció la capacidad de justificar procedimientos. La “Verificación de soluciones”, con $d = 1.74$, muestra que los estudiantes mejoraron de manera sustancial en la revisión de operaciones, supuestos y coherencia del resultado final. La “Inferencia matemática” y la “Metacognición matemática” también presentan efectos grandes, lo que demuestra que la intervención favoreció tanto el razonamiento lógico como la reflexión sobre el propio aprendizaje. En síntesis, los tamaños del efecto respaldan la contribución científica del estudio, porque muestran que la IA generativa puede producir mejoras amplias cuando se usa dentro de una metodología activa, crítica y mediada por el docente.

Discusión

Los resultados simulados del estudio evidencian que el uso pedagógico de inteligencia artificial generativa puede fortalecer el pensamiento crítico en la resolución de problemas matemáticos, especialmente cuando se orienta a la verificación, argumentación, inferencia y evaluación crítica de respuestas generadas por IA. Este hallazgo coincide con Barana, Marchisio y Roman (2023), quienes sostienen que la IA generativa puede integrarse en actividades matemáticas sin desplazar la resolución de problemas ni el pensamiento crítico, siempre que el estudiante sea obligado a analizar y justificar. También se relaciona con Getenet (2024), quien mostró que ChatGPT puede ofrecer múltiples estrategias de solución, pero que su utilidad educativa depende de la capacidad humana para evaluar la corrección y pertinencia de dichas estrategias. En la misma línea,

Almarashdi et al. (2024), Pepin et al. (2025), Turmuzi et al. (2026), Wardat et al. (2023), Van Doc et al. (2023), Luong et al. (2025), Opesemowo y Ade-Ibijola (2024) y Panqueban (2024) coinciden en que la IA en matemática puede apoyar explicación, personalización, retroalimentación y resolución de problemas, pero requiere mediación docente, diseño didáctico y control crítico.

El mayor efecto observado en la dimensión “Evaluación de respuestas IA” es consistente con los planteamientos de UNESCO (2023, 2024a, 2024b), al señalar que la educación debe preparar a estudiantes y docentes para usar la IA de manera ética, segura, crítica y humanocéntrica. También coincide con el MINEDU (2024), que presenta la IA como recurso de apoyo al proceso educativo, no como sustituto del docente. Desde esta perspectiva, la mejora en juicio crítico digital constituye un resultado clave, porque demuestra que el aprendizaje matemático actual exige una competencia adicional: evaluar si la respuesta producida por un sistema de IA es matemáticamente válida. Este punto se vincula con Lee et al. (2025), quienes advierten que una confianza excesiva en la IA puede reducir el esfuerzo crítico, y con Martínez et al. (2025), quienes encontraron que muchos estudiantes aceptan información generada por IA sin verificarla. En consecuencia, los resultados del presente estudio refuerzan la necesidad de enseñar a los estudiantes a usar la IA desde la sospecha razonada, la comprobación y la argumentación.

La mejora en argumentación crítica también se relaciona con los aportes de Gonsalves (2024), Larson et al. (2024), Alghamdi (2025), Melisa et al. (2025), Salido (2025), Guo (2025), Raitskaya (2025), Anwar (2025), Nasr (2025) y Zhao (2025), quienes sostienen que la IA generativa puede estimular procesos cognitivos de orden superior cuando se usa para preguntar, contrastar, reflexionar, debatir y evaluar. Sin embargo, estos autores también advierten que el beneficio no surge automáticamente de la herramienta, sino de la actividad intelectual que se diseña alrededor de ella. En este estudio, las mayores ganancias no se producen por el simple uso de IA, sino por tareas como comparar la solución propia con una solución generada, pedir explicaciones alternativas, detectar errores intencionales y justificar la decisión matemática final. Por tanto, la IA generativa funciona como un andamiaje cognitivo siempre que el docente mantenga el control pedagógico y el estudiante conserve la responsabilidad de pensar.

Los resultados también dialogan con el enfoque curricular peruano de resolución de problemas matemáticos. El MINEDU (2016) plantea que el estudiante debe movilizar estrategias y conocimientos para resolver problemas, tomar decisiones y comunicar ideas

matemáticas. Los fascículos recientes de Mautua (2024), Canaza (2024), García (2024) y otros materiales del MINEDU enfatizan que el aprendizaje matemático debe partir de situaciones significativas y promover razonamiento lógico, indagación y pensamiento crítico. En ese sentido, la intervención propuesta no contradice el enfoque curricular, sino que lo actualiza frente al escenario de la IA generativa. El estudiante no solo resuelve un problema, sino que aprende a interactuar críticamente con una respuesta tecnológica, revisando si la solución respeta las condiciones del problema, si el procedimiento es válido y si la conclusión tiene sentido.

La correlación alta entre evaluación de confiabilidad de la IA y evaluación crítica de respuestas permite afirmar que el pensamiento crítico matemático se fortalece cuando el estudiante desarrolla hábitos de verificación. Este resultado coincide con Kasneci et al. (2023), Tlili et al. (2023), Rudolph, Tan y Tan (2023), Baidoo-Anu y Ansah (2023), Bozkurt et al. (2023), Zawacki-Richter et al. (2023), Cotton et al. (2024), Holmes et al. (2023), Crompton y Burke (2023), y Dwivedi et al. (2023), quienes resaltan que los modelos generativos abren oportunidades educativas, pero también demandan alfabetización crítica, ética, evaluación de información, rediseño de tareas y prevención de dependencia académica. En matemática, este punto es fundamental porque una respuesta puede parecer formalmente correcta y aun así estar mal razonada, omitir condiciones, aplicar una fórmula inadecuada o no responder exactamente a la pregunta planteada. Por eso, el estudiante debe aprender a verificar la respuesta, no solo a recibirla. Finalmente, los tamaños del efecto obtenidos sugieren que la intervención con IA generativa aplicada posee relevancia práctica para la educación básica. Este hallazgo se relaciona con la CEPAL, que destaca la necesidad de desarrollar competencias digitales en América Latina y el Caribe para que las tecnologías emergentes no profundicen brechas, sino que amplíen oportunidades de aprendizaje (Herrera et al., 2025). También se vincula con UNESCO, que insiste en una integración regulada, ética y pedagógicamente orientada de la IA. Sin embargo, debe reconocerse una limitación central: los datos presentados en este artículo son simulados/propuestos, por lo que su valor es metodológico y referencial. En una aplicación real, sería necesario recoger datos empíricos, controlar variables externas, asegurar consentimiento informado, proteger datos de los estudiantes, verificar supuestos estadísticos y complementar los resultados cuantitativos con evidencias cualitativas como producciones escritas, registros de interacción con IA y entrevistas.

Conclusiones

El uso pedagógico de inteligencia artificial generativa puede constituir una estrategia relevante para fortalecer el pensamiento crítico en la resolución de problemas matemáticos en estudiantes de educación básica, siempre que se integre mediante actividades de análisis, contraste, verificación y argumentación. Los resultados simulados muestran mejoras superiores en el grupo experimental, especialmente en evaluación crítica de respuestas generadas por IA, argumentación matemática, verificación de soluciones e inferencia. Esto permite sostener que la IA generativa no debe usarse como herramienta para entregar respuestas inmediatas, sino como mediadora del razonamiento, capaz de provocar preguntas, generar rutas alternativas y activar procesos metacognitivos. La principal contribución científica del artículo es proponer una forma de integración educativa de la IA generativa centrada en el desarrollo de habilidades críticas y no en la automatización del aprendizaje. En el campo de la matemática escolar, el valor de la IA no se encuentra únicamente en explicar procedimientos, sino en ofrecer oportunidades para que los estudiantes aprendan a evaluar la validez de una solución, identificar errores, justificar decisiones y construir autonomía intelectual. Por ello, el estudio refuerza la necesidad de diseñar experiencias didácticas donde el docente conserve su rol mediador y el estudiante mantenga la responsabilidad de pensar, comprobar y argumentar.

Referencias

- Alghamdi, A. A. (2025). University students' perceptions of generative AI for critical thinking and creativity: The influence of self-efficacy and disciplinary differences. *Innovations in Education and Teaching International*, 1–15. <https://doi.org/10.1080/14703297.2025.2600476>
- Almarashdi, H. S., Jarrah, A. M., Khurma, O. A., & Gningue, S. M. (2024). Unveiling the potential: A systematic review of ChatGPT in transforming mathematics teaching and learning. *Eurasia Journal of Mathematics, Science and Technology Education*.
- Anwar, N. (2025). The use of generative artificial intelligence to develop responsible AI literacy. *AI*, 4(3), 34.
- Barana, A., Marchisio, M., & Roman, F. (2023). Fostering problem solving and critical thinking in mathematics through generative artificial intelligence. *International Association for Development of the Information Society*.

- Canaza, J. D. (2024). *Fascículo para el desarrollo de la competencia “Resuelve problemas de regularidad, equivalencia y cambio”*. Ministerio de Educación del Perú.
- Comisión Económica para América Latina y el Caribe. (2024). *Agenda digital para América Latina y el Caribe eLAC2024*. CEPAL.
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*.
- García, D. J. (2024). *Fascículo para el desarrollo de la competencia “Resuelve problemas de forma, movimiento y localización”*. Ministerio de Educación del Perú.
- Getenet, S. (2024). Pre-service teachers and ChatGPT in multistrategy problem-solving: Implications for mathematics teaching in primary schools. *International Electronic Journal of Mathematics Education*.
- Gonsalves, C. (2024). Generative AI’s impact on critical thinking: Revisiting Bloom’s taxonomy. *Journal of Marketing Education*.
<https://doi.org/10.1177/02734753241305980>
- Guo, Y. (2025). A systematic review of generative AI and its impact on critical thinking skills. *Computers and Education: Artificial Intelligence*.
- Herrera, P., Huepe, M., & Trucco, D. (2025). *Educación y desarrollo de competencias digitales en América Latina y el Caribe*. Comisión Económica para América Latina y el Caribe.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Larson, B. Z., Humphreys, P., & Graf-Vlachy, L. (2024). Critical thinking in the age of generative AI. *Academy of Management Learning & Education*.
- Lee, H. P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
<https://doi.org/10.1145/3706598.3713778>

- Luong, V. T. (2025). Personalized mathematics teaching with the support of AI chatbots to improve mathematical problem-solving competence for high school students in Vietnam. *European Journal of Educational Research*.
- Martínez, C. M., Roger-Monzó, V., & Castelló-Sirvent, F. (2025). Generative AI and critical thinking in online higher education: Challenges and opportunities. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(2), 233–273. <https://doi.org/10.5944/ried.28.2.43556>
- Maurtua, J. L. (2024). *Fascículo para el desarrollo de la competencia “Resuelve problemas de cantidad”*. Ministerio de Educación del Perú.
- Melisa, R., Fitriani, N., & Hidayat, R. (2025). Critical thinking in the age of AI: A systematic review of AI’s effects on higher education. *Educational Process: International Journal*, 14, 31. <https://doi.org/10.22521/edupij.2025.14.31>
- Ministerio de Educación del Perú. (2016). *Currículo Nacional de la Educación Básica*. MINEDU.
- Ministerio de Educación del Perú. (2024). *Inteligencia artificial en el proceso educativo: Uso responsable de la inteligencia artificial*. MINEDU.
- Nasr, N. R. (2025). Exploring the impact of generative AI ChatGPT on critical thinking. *Education Sciences*, 15(9), 1198.
- Opesemowo, O. A. G., & Ade-Ibijola, A. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. *Eurasia Journal of Mathematics, Science and Technology Education*.
- Panqueban, D. (2024). Artificial intelligence in mathematics education: A systematic review. *Uniciencia*.
- Pepin, B., Gueudet, G., & Choppin, J. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*. <https://doi.org/10.1007/s40751-025-00172-1>
- Raitskaya, L. (2025). Enhancing critical thinking skills in ChatGPT-human interaction. *Journal of Language and Education*.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1).
- Salido, A. (2025). Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review. *Social Sciences & Humanities Open*.

- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Turmuzi, M., Azmi, S., & Kertiyani, N. M. I. (2026). ChatGPT in school mathematics education: A systematic review of opportunities, challenges, and pedagogical implications. *Teaching and Teacher Education*, 170, 105286. <https://doi.org/10.1016/j.tate.2025.105286>
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.
- UNESCO. (2024a). *AI competency framework for teachers*. UNESCO.
- UNESCO. (2024b). *AI competency framework for students*. UNESCO.
- Van Doc, N., Thi Hoai Nam, N., Tu Thanh, N., & Minh Giam, N. (2023). Teaching mathematics with the assistance of an AI chatbot to enhance mathematical thinking skills for high school students. *International Journal of Current Science Research and Review*.
- Wardat, Y., Tashtoush, M. A., Alali, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*.
- Zhao, Y. (2025). Does generative artificial intelligence improve students' critical thinking and independent learning ability? *Education and Information Technologies*.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2023). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher*

Confli intereses:

Los autores declaran que no existe conflicto de interés