Ciencia y Educación
(L-ISSN: 2790-8402 E-ISSN: 2707-3378)
Vol. 7 No. 11.1
Edición Especial XI 2026

IMPACTO DE LA INTELIGENCIA ARTIFICIAL EN EL PENSAMIENTO CRÍTICO DURANTE EL APRENDIZAJE DE CIENCIAS SOCIALES EN BACHILLERATO

IMPACT OF ARTIFICIAL INTELLIGENCE ON CRITICAL THINKING DURING SOCIAL SCIENCES LEARNING IN HIGH SCHOOL

Autores: [1]Carmen Laura Benalcázar Arroba y [2]Steven Arturo Torres Burgos.
[1]ORCID ID: https://orcid.org/0009-0008-3742-6765
[2]ORCID ID: https://orcid.org/0000-0001-9299-3254
[1]E-mail de contacto: cbenalcazara@unemi.edu.ec
[2]E-mail de contacto: storresb5@unemi.edu.ec
Afiliación: [1*2*]Universidad Estatal de Milagro, (Ecuador).
Artículo recibido: 26 de Octubre del 2026
Artículo revisado: 28 de Octubre del 2026
Artículo aprobado: 28 de Octubre del 2026
[1]Ingeniera en Mecatrónica, graduada en la Universidad de las Fuerzas Armadas ESPE sede Latacunga, (Ecuador). Con 2 años de experencia en el Area Técnica y 8 años de experiencia laboral en Educación. Magíster en Electrónica y Automatización con mención en Redes Industriales, graduada en la Universidad de las Fuerzas Armadas ESPE sede Latacunga, (Ecuador). Maestrante de la Maestría en Inteligencia Artificial para la Educación, en la Universidad Estatal de Milagro, (Ecuador).
[2]Licenciado en Cultura Física, graduado en la Universidad de Guayaquil, (Ecuador). Magíster en Pedagogía de la Actividad Física, mención en Educación Física Inclusiva, graduado de la Universidad Bolivariana del Ecuador, (Ecuador). PhD en Educación Física por CECEIC, (México).

Resumen

El presente estudio evaluó la efectividad de integrar herramientas de inteligencia artificial (IA) en el área de Ciencias Sociales para desarrollar el pensamiento crítico en estudiantes de Bachillerato de unidades educativas del norte de Guayaquil, Ecuador. Mediante un diseño cuasi-experimental con pre-test y post-test, se implementó una intervención pedagógica de 12 semanas utilizando plataformas como ChatGPT, Claude IA, Gemini, Copilot y DeepSeek en actividades de análisis histórico, argumentación y evaluación de fuentes. La muestra incluyó 120 estudiantes divididos en grupo experimental (n=60) y control (n=60). Los resultados revelaron diferencias estadísticamente significativas ($p<0.001$) en las dimensiones de análisis, evaluación e inferencia del pensamiento crítico, medidas a través del test Watson-Glaser adaptado. El grupo experimental mostró una ganancia media de 9.4 puntos frente a 2.2 puntos del grupo control, con tamaño del efecto muy grande (d de Cohen=2.40). No obstante, también se identificaron desafíos relacionados con la dependencia tecnológica, la verificación de información y la brecha digital. Los hallazgos sugieren que la IA, cuando se integra con estrategias pedagógicas estructuradas y supervisión docente, constituye una herramienta valiosa para fomentar habilidades cognitivas superiores en el área de Ciencias Sociales.

Palabras clave: Inteligencia artificial, Pensamiento crítico, Ciencias Sociales, Bachillerato, ChatGPT.

Abstract

This study evaluated the effectiveness of integrating artificial intelligence (AI) tools in the Social Sciences classroom to develop critical thinking in high school students from educational units in northern Guayaquil, Ecuador. Through a quasi-experimental design with pre-test and post-test, a 12-week pedagogical intervention was implemented using platforms such as ChatGPT, Claude, Gemini, Copilot, and DeepSeek in activities involving historical analysis, argumentation, and source evaluation. The sample included 120 students divided into an experimental group (n=60) and a control group (n=60). Results revealed statistically significant differences ($p<0.001$) in the dimensions of analysis, evaluation, and inference of critical thinking, measured through the adapted Watson-Glaser test. The experimental group showed an average gain of 9.4 points compared to 2.2 points for the control group, with a very large effect size (Cohen's d=2.40). However, challenges related

to technological dependency, information verification, and the digital divide were also identified. Findings suggest that AI, when integrated with structured pedagogical strategies and teacher supervision, constitutes a valuable tool for fostering higher-order cognitive skills in the Social Sciences area.
Keywords: Artificial intelligence, Critical thinking, Social Sciences, High school, ChatGPT.

Sumário

Este estudo avaliou a eficácia da integração de ferramentas de inteligência artificial (IA) ao currículo de Ciências Sociais para promover o pensamento crítico entre alunos do ensino médio em instituições de ensino na zona norte de Guayaquil, Equador. Utilizando um delineamento quase-experimental de pré-teste e pós-teste, implementou-se uma intervenção pedagógica de 12 semanas, empregando plataformas como ChatGPT, Claude AI, Gemini, Copilot e DeepSeek em atividades que envolviam análise histórica, argumentação e avaliação de fontes. A amostra foi composta por 120 alunos, divididos em um grupo experimental (n=60) e um grupo de controle (n=60). Os resultados revelaram diferenças estatisticamente significativas (p<0,001) nas dimensões de pensamento crítico — análise, avaliação e inferência —, medidas por meio de uma versão adaptada do teste Watson-Glaser. O grupo experimental apresentou um ganho médio de 9,4 pontos, em comparação com 2,2 pontos no grupo de controle, o que representa um tamanho de efeito muito grande (d de Cohen = 2,40). No entanto, também foram identificados desafios relacionados à dependência tecnológica, à verificação de informações e à exclusão digital. As conclusões sugerem que a IA, quando integrada a estratégias pedagógicas estruturadas e à supervisão docente, constitui uma ferramenta valiosa para o desenvolvimento de habilidades cognitivas de ordem superior no ensino de Ciências Sociais.
Palavras-chave: Inteligência artificial, Pensamento crítico, Ciências sociais, Ensino médio, ChatGPT.

Introducción

La irrupción de la inteligencia artificial generativa en el ámbito educativo durante los últimos años ha generado transformaciones profundas en las metodologías de enseñanza, particularmente en la educación secundaria (García et al., 2024). Herramientas como ChatGPT han alcanzado cifras superiores a 769 millones de usuarios activos mensuales, consolidándose como la plataforma de IA más utilizada globalmente con una cuota de mercado del 40.5% en 2025.

Paralelamente, otras alternativas como Claude, desarrollado por Anthropic, Gemini de Google, Microsoft Copilot y DeepSeek han fragmentado el panorama tecnológico, ofreciendo capacidades especializadas en análisis textual, razonamiento complejo y generación de contenido educativo. Asimismo, en el contexto ecuatoriano, el sistema educativo de Bachillerato General Unificado atiende a estudiantes de 15 a 17 años distribuidos en tres cursos formativos que buscan desarrollar competencias interdisciplinarias vinculadas con el pensamiento crítico, la ciudadanía activa y la preparación para la educación superior (Ministerio de Educación del Ecuador, 2022).

No obstante, el desarrollo del pensamiento crítico constituye una de las competencias más desafiantes en la formación académica contemporánea. Diversos estudios han documentado niveles medio-bajos en habilidades de análisis, evaluación e inferencia entre estudiantes latinoamericanos (Puche, 2024), situación que se replica en niveles educativos previos. Por otro lado, el pensamiento crítico, entendido como la capacidad de identificar, analizar y evaluar información relevante para llegar a conclusiones lógicas y fundamentadas, representa una habilidad transversal reconocida

por el Foro Económico Mundial como competencia clave para el siglo XXI. Particularmente en el área de Ciencias Sociales, esta capacidad resulta esencial para desarrollar el pensamiento histórico, que implica contextualizar fuentes, discernir sesgos autorales y construir narrativas complejas sobre procesos sociales.

De igual forma, la integración de tecnologías educativas digitales en aulas de Ciencias Sociales ha transitado por diversas etapas. Inicialmente, plataformas como Kahoot y recursos multimedia básicos permitieron dinamizar procesos evaluativos y mejorar la participación estudiantil (García et al., 2024). Con la masificación de internet, surgieron herramientas colaborativas basadas en la web que facilitaron el aprendizaje activo y el trabajo en equipo mediado por TIC (Mota et al., 2020). Sin embargo, la aparición de modelos de lenguaje de gran escala (LLMs) a partir de 2022 marcó un punto de inflexión cualitativo.

En consecuencia, ChatGPT ha impulsado la evolución con mejoras desde el año anterior, demostró capacidades sin precedentes para generar respuestas coherentes, explicar conceptos complejos y asistir en tareas académicas diversas. Investigaciones recientes han explorado el impacto de la IA en la educación desde múltiples perspectivas. Delgado et al. (2024) analizaron las percepciones de docentes de educación primaria, secundaria y superior sobre los beneficios y limitaciones de la IA, identificando tanto oportunidades pedagógicas como preocupaciones éticas. Parreño et al. (2024) documentaron cómo la IA puede constituirse en herramienta para el desarrollo del pensamiento crítico en instituciones educativas, aunque advierten sobre el riesgo de dependencia que limite la reflexión autónoma. Específicamente en bachillerato, García et al. (2024) reportaron efectos positivos en capacidades cognitivas mediante enfoque mixto, mientras que Martínez et al. (2023) realizaron una revisión sistemática sobre el impacto de la IA en métodos de evaluación en educación primaria y secundaria. En el contexto ecuatoriano, la evaluación Ser Estudiante 2023-2024 evidenció resultados mixtos en bachillerato: mientras Biología mostró incrementos de +23 puntos, Ciencias Sociales registró avances modestos de +2 puntos, revelando la necesidad de estrategias innovadoras que fortalezcan competencias analíticas en áreas sociales (INEVAL, 2024).

Adicionalmente, el Ministerio de Educación implementó en 2024-2025 una nueva malla curricular que incorpora transversalmente materias de Cívica, Ética, Integridad, Educación Financiera, Socioemocional y Vial, creando un marco propicio para integrar herramientas tecnológicas que fomenten el pensamiento crítico desde perspectivas históricas y sociales. A pesar del creciente interés en la IA educativa, persisten vacíos significativos en la investigación empírica sobre su efectividad en contextos específicos.

Primero, la mayoría de los estudios se concentran en educación superior, dejando escasa evidencia sobre bachillerato, etapa crítica para consolidar habilidades de pensamiento de orden superior. Segundo, existe limitada investigación sobre la integración de IA en el área de Ciencias Sociales, tradicionalmente dominada por enfoques memorísticos que priorizan contenidos factuales sobre análisis crítico (Oviedo, 2023). Tercero, aunque plataformas como ChatGPT dominan el mercado, otras alternativas emergentes ofrecen características diferenciadas que requieren evaluación comparativa en contextos educativos reales. El

problema específico radica en determinar si la integración sistemática de herramientas de IA en el aula de Ciencias Sociales contribuye efectivamente al desarrollo del pensamiento crítico en estudiantes de bachillerato del norte de Guayaquil. Esta zona urbana presenta características socioeconómicas heterogéneas, con instituciones educativas fiscales, fiscomisionales y particulares que atienden a poblaciones estudiantiles con diversos niveles de acceso tecnológico.

La pregunta central de investigación se formula así: ¿En qué medida el uso guiado de plataformas de IA generativa mejora las habilidades de pensamiento crítico, medidas a través de dimensiones de análisis, evaluación, inferencia, deducción y argumentación, en estudiantes de tercer año de Bachillerato durante un período de intervención de 12 semanas? La relevancia teórica del estudio se fundamenta en aportar evidencia empírica sobre la relación entre IA educativa y desarrollo cognitivo en un contexto latinoamericano subrepresentado en la literatura internacional.

Mientras que la mayoría de las investigaciones provienen de países desarrollados con infraestructuras tecnológicas consolidadas, el presente estudio documenta la viabilidad y efectividad de estas herramientas en un sistema educativo en desarrollo, caracterizado por recursos limitados y heterogeneidad institucional (Naranjo et al., 2023). Desde la perspectiva práctica, los resultados pueden orientar políticas educativas y prácticas docentes. El Ministerio de Educación del Ecuador ha manifestado interés en incorporar tecnologías emergentes en la malla curricular; sin embargo, carece de lineamientos específicos sobre el uso pedagógico de IA. Este estudio ofrece un modelo replicable de integración que incluye actividades estructuradas, rúbricas de evaluación y estrategias para mitigar riesgos asociados con desinformación y dependencia tecnológica, problemáticas identificadas en la Guía UNESCO para el uso de IA generativa en educación (Miao y Holmes, 2024). La justificación metodológica reside en el diseño cuasi-experimental riguroso, que permite establecer relaciones causales minimizando sesgos inherentes a contextos educativos naturales donde la aleatorización completa es impracticable (Ramos, 2021). Adicionalmente, el uso del test Watson-Glaser, instrumento con más de 80 años de validación psicométrica y confiabilidad entre 0.73-0.83 medida con Alfa de Cronbach, garantiza la objetividad en la medición de las dimensiones del pensamiento crítico.

Posteriormente, la justificación social se vincula con la formación de ciudadanos críticos capaces de enfrentar desafíos del siglo XXI. En la era de la desinformación, desarrollar habilidades para evaluar críticamente fuentes, identificar sesgos y construir argumentos fundamentados constituye una necesidad democrática urgente (López y Sánchez, 2025). El área de Ciencias Sociales requiere revitalización metodológica que conecte contenidos históricos con competencias digitales contemporáneas.

Asimismo, para contribuir a este trabajo y fortalecer la presente investigación se describen el objetivo general: Evaluar la efectividad de la integración de herramientas de inteligencia artificial en el desarrollo del pensamiento crítico en estudiantes de tercer año de Bachillerato del área de Ciencias Sociales en unidades educativas del norte de Guayaquil. De igual forma, los objetivos específicos: 1. Medir los niveles iniciales y finales de pensamiento crítico (dimensiones de análisis, evaluación, inferencia, deducción y argumentación) en

grupos experimental y control mediante la aplicación del test Watson-Glaser adaptado al contexto ecuatoriano. 2. Diseñar e implementar una intervención pedagógica estructurada de 12 semanas que integre herramientas de IA en actividades de análisis de fuentes históricas, evaluación de argumentos y construcción de narrativas críticas en el grupo experimental. 3. Comparar estadísticamente los resultados del post-test entre grupo experimental y control para determinar diferencias significativas atribuibles a la intervención con IA. 4. Identificar fortalezas, limitaciones y desafíos asociados con el uso de IA en el aula desde las perspectivas de estudiantes y docentes participantes.

Materiales y Métodos

Se implementó un diseño cuasi-experimental con grupos no equivalentes y mediciones pre-test y post-test. Este diseño resulta apropiado para contextos educativos naturales donde la asignación aleatoria de participantes es impracticable debido a la organización institucional en secciones preestablecidas (Cook y Campbell, 1979); (Ramos, 2021). El esquema metodológico se representa así: Grupo Experimental O_1 X O_2; Grupo Control O_1 - O_2, donde O_1 representa el pre-test, X la intervención con IA, y O_2 el post-test.

El enfoque fue cuantitativo con alcance descriptivo-explicativo, buscando establecer relaciones causales entre la variable independiente (uso de herramientas de IA) y la dependiente (nivel de pensamiento crítico). La población objetivo comprendió estudiantes de tercer año de Bachillerato General Unificado de cuatro unidades educativas del norte de Guayaquil: dos instituciones fiscales, una fiscomisional y una particular. Esta selección buscó representar la diversidad socioeconómica del sector. Mediante muestreo no probabilístico intencional, se seleccionaron secciones completas que cumplieran los criterios de inclusión: estudiantes regulares de 16 a 18 años, matriculados en la asignatura de Historia y Ciencias Sociales, con acceso a dispositivos tecnológicos (smartphone, tableta o computadora) y conectividad a internet. La muestra final quedó conformada por 120 estudiantes: 60 en el grupo experimental (distribuidos en dos secciones de 30 estudiantes cada una) y 60 en el grupo control (igualmente distribuidos en dos secciones). La distribución por género fue equilibrada (52% mujeres, 48% hombres) y la edad promedio fue 16.8 años (DE=0.6).

Variable independiente: Integración de herramientas de IA en actividades de enseñanza-aprendizaje de Ciencias Sociales. Operacionalmente, se definió como el uso sistemático y guiado de cinco plataformas (ChatGPT, Claude, Gemini, Copilot, DeepSeek) en tareas estructuradas durante 12 semanas consecutivas, con frecuencia de dos sesiones semanales de 80 minutos cada una.

Variable dependiente: Nivel de pensamiento crítico, medido a través del test Watson-Glaser en sus cinco dimensiones: (1) Inferencia, (2) Reconocimiento de supuestos, (3) Deducción, (4) Interpretación, y (5) Evaluación de argumentos (Watson y Glaser, 2002); Pérez et al., 2025). Test Watson-Glaser Critical Thinking Appraisal (Forma E): Instrumento estandarizado con alta confiabilidad (α=0.73-0.83) que evalúa pensamiento crítico mediante 40 reactivos de opción múltiple. El tiempo de aplicación fue de 30 minutos en modalidad digital. Se utilizó una versión adaptada lingüísticamente al español ecuatoriano, conservando la estructura original y los criterios de puntuación (1 punto por respuesta correcta), con rango total de 0-40 puntos. Cuestionario de

percepciones estudiantiles: Instrumento de 15 ítems con escala Likert (1=Totalmente en desacuerdo, 5=Totalmente de acuerdo) diseñado para evaluar facilidad de uso de las plataformas de IA, utilidad percibida, confianza en la información, y autopercepción de mejora en habilidades críticas. Aplicado al grupo experimental en la semana 12. Rúbrica de evaluación docente: Matriz de valoración con cuatro criterios (análisis de fuentes, calidad argumentativa, uso crítico de IA, originalidad del pensamiento) y tres niveles de desempeño (inicial, en desarrollo, logrado).

El estudio se desarrolló en cinco fases entre septiembre y diciembre de 2025. Fase 1: Preparación (Semanas 1-2): Obtención de permisos institucionales, consentimientos informados y capacitación de cuatro docentes participantes. Fase 2: Pre-test (Semana 3): Aplicación simultánea del test Watson-Glaser a ambos grupos. Fase 3: Intervención (Semanas 4-15): El grupo experimental participó en 24 sesiones estructuradas que integraron herramientas de IA en el currículo regular.

Las actividades incluyeron análisis comparativo de interpretaciones históricas generadas por diferentes plataformas, evaluación crítica de argumentos, verificación de información mediante contraste entre respuestas de IA y fuentes académicas, construcción colaborativa de ensayos argumentativos, y análisis de sesgos en narrativas históricas. Cada sesión seguía esta estructura: (a) introducción del tema histórico-social (10 minutos), (b) planteamiento de pregunta problema (5 minutos), (c) interacción guiada con plataformas de IA (40 minutos), (d) discusión plenaria (20 minutos), (e) síntesis (5 minutos). El grupo control continuó con metodología tradicional. Fase 4 - Post-test (Semana 16): Aplicación del test Watson-Glaser y cuestionario de percepciones. Fase 5 - Análisis (Semanas 17-18): Procesamiento estadístico mediante SPSS versión 27. Los datos cuantitativos se analizaron mediante estadística descriptiva e inferencial. Inicialmente se calcularon medidas de tendencia central (media, mediana) y dispersión (desviación estándar) para caracterizar las puntuaciones. La normalidad se verificó mediante la prueba de Shapiro-Wilk.

Para evaluar la equivalencia inicial se aplicó prueba t de Student para muestras independientes. El análisis principal consistió en pruebas t para muestras relacionadas (intragrupo) e independientes (intergrupos). El tamaño del efecto se calculó mediante d de Cohen. El nivel de significancia fue $\alpha=0.05$. Los datos del cuestionario se analizaron mediante estadística descriptiva. El estudio cumplió con principios éticos fundamentales de investigación educativa. Se obtuvo aprobación del Comité de Ética y autorización de las unidades participantes. Los representantes legales firmaron consentimientos informados. Se garantizó anonimato mediante codificación numérica.

Resultados y Discusión

Los resultados no afectaron calificaciones académicas. Al finalizar, el grupo control recibió capacitación en uso de IA educativa. Se orientó sobre uso ético de IA, privacidad de educativa, dos docentes con experiencia superior a 10 años en Ciencias Sociales de bachillerato, y un psicólogo educativo con experticia en evaluación de pensamiento crítico revisaron la pertinencia, claridad y suficiencia de los ítems. Se empleó el coeficiente V de Aiken, obteniéndose valores superiores a 0.80 en todos los criterios evaluados, lo cual indica validez de contenido aceptable. Las observaciones de los expertos permitieron ajustar la redacción de tres ítems del

cuestionario, mejorando su comprensibilidad para estudiantes de 16-18 años. Antes de presentar los resultados principales, resulta pertinente caracterizar demográficamente la muestra para contextualizar los hallazgos. La Tabla 1 presenta la distribución de participantes según variables sociodemográficas relevantes. Como se observa, ambos grupos mantuvieron composiciones similares en términos de género, edad y tipo de institución educativa, lo cual refuerza la validez del diseño cuasi-experimental al minimizar fuentes potenciales de confusión.

Tabla 1. *Características demográficas de la muestra por grupo.*

VARIABLE	CATEGORÍA	EXPERIMENTAL N (%)	CONTROL N (%)
Género	Femenino	31 (51.7%)	31 (51.7%)
	Masculino	29 (48.3%)	29 (48.3%)
Edad	16 años	18 (30.0%)	17 (28.3%)
	17 años	35 (58.3%)	36 (60.0%)
	18 años	7 (11.7%)	7 (11.7%)
Tipo de institución	Fiscal	30 (50.0%)	30 (50.0%)
	Fiscomisional	15 (25.0%)	15 (25.0%)
	Particular	15 (25.0%)	15 (25.0%)

Fuente: elaboración propia

La distribución homogénea entre grupos

Fuente: Elaboración propia. Previo a la intervención, resultaba fundamental verificar estadísticamente que ambos grupos partieran de niveles equivalentes de pensamiento crítico. La Tabla 2 presenta los resultados del pre-test aplicado en la semana 3, desagregados por dimensión del test Watson-Glaser. El análisis mediante prueba t de Student para muestras independientes confirmó la ausencia de diferencias significativas en todas las mediciones basales. Los datos de la Tabla 2 confirman la equivalencia estadística inicial entre grupos. Ninguna de las cinco dimensiones del pensamiento crítico ni la puntuación total mostraron diferencias significativas (todos los valores $p > 0.05$), con estadísticos t cercanos a cero que indican medias prácticamente idénticas. Este hallazgo resulta fundamental para la validez interna del diseño cuasi-experimental, dado que establece que cualquier divergencia observada en el post-test puede atribuirse razonablemente a la intervención pedagógica con IA y no a diferencias preexistentes entre grupos. Adicionalmente, los valores de desviación estándar similares entre grupos sugieren homogeneidad en la variabilidad individual, lo cual refuerza la comparabilidad de las muestras. Tras la intervención de 12 semanas, ambos grupos experimentaron cambios en sus niveles de pensamiento crítico, aunque con magnitudes marcadamente diferentes. La Tabla 3 presenta el análisis comparativo detallado de las puntuaciones pre-test y post-test para cada

evidencia que el proceso de selección logró conformar muestras comparables. Específicamente, ambos grupos presentaron composición de género prácticamente idéntica (51.7% mujeres en ambos), distribución etaria similar (mayoría de 17 años con 58.3% experimental vs 60.0% control), y representación equivalente de tipos institucionales (50% fiscal, 25% fiscomisional, 25% particular en ambos grupos). Esta similitud demográfica reduce la probabilidad de que diferencias observadas en pensamiento crítico se deban a variables extrañas relacionadas con características de los participantes.

Tabla 2. *Puntuaciones del pre-test por dimensión y comparación entre grupos.*

Dimensión	Experimental M (DE)	Control M (DE)	t	P
Inferencia	4.6 (1.2)	4.5 (1.3)	0.41	0.682
Reconocimiento de supuestos	4.1 (1.4)	4.0 (1.5)	0.37	0.712
Deducción	4.3 (1.3)	4.2 (1.4)	0.39	0.694
Interpretación	4.7 (1.2)	4.6 (1.3)	0.40	0.688
Evaluación de argumentos	4.7 (1.1)	4.8 (1.2)	-0.45	0.654
Puntuación Total	22.4 (4.2)	22.1 (4.5)	0.38	0.705

Fuente: Elaboración propia.

grupo, incluyendo las ganancias absolutas y los tamaños del efecto correspondientes.

Tabla 3. *Comparación pre-test y post-test por grupo con tamaños del efecto.*

Dimensión	Grupo	Pre-test M (DE)	Post-test M (DE)	Ganancia	t	p	d Cohen
Puntuación Total	Experimental	22.4 (4.2)	31.8 (3.6)	+9.4	18.62	<0.001	2.40
Puntuación Total	Control	22.1 (4.5)	24.3 (4.2)	+2.2	4.35	<0.001	0.56
Inferencia	Experimental	4.6 (1.2)	6.3 (1.1)	+1.7	10.45	<0.001	1.48
Inferencia	Control	4.5 (1.3)	4.9 (1.3)	+0.4	2.31	0.024	0.31
Interpretación	Experimental	4.7 (1.2)	7.1 (0.9)	+2.4	14.23	<0.001	2.24
Interpretación	Control	4.6 (1.3)	5.2 (1.2)	+0.6	3.48	0.001	0.48
Evaluación	Experimental	4.7 (1.1)	6.8 (1.0)	+2.1	12.87	<0.001	2.01
Evaluación	Control	4.8 (1.2)	5.2 (1.1)	+0.4	2.52	0.014	0.35

Fuente: Elaboración propia.

La Tabla 3 evidencia hallazgos contundentes. El grupo experimental mostró una ganancia total de 9.4 puntos (equivalente al 42% de mejora respecto al nivel basal), con un tamaño del efecto de 2.40 que se clasifica como muy grande según los criterios convencionales de Cohen ($d > 0.8$). Este resultado implica que la intervención generó cambios sustanciales y prácticos, no meramente estadísticamente significativos. En contraste, el grupo control experimentó una ganancia modesta de 2.2 puntos (10% de mejora), con tamaño del efecto mediano ($d=0.56$), atribuible principalmente al efecto de maduración natural, exposición continua al currículo regular y posible efecto test-retest. El análisis dimensional revela que las dimensiones de Interpretación y Evaluación de argumentos registraron las mayores ganancias en el grupo experimental (+2.4 y +2.1 puntos respectivamente), con tamaños del efecto superiores a 2.0 que indican transformaciones profundas en estas competencias. Estas dimensiones implican habilidades complejas de análisis contextual, discernimiento de evidencias válidas y juicio crítico sobre la solidez de razonamientos, capacidades directamente estimuladas por las actividades de contraste entre respuestas de diferentes plataformas de IA y verificación sistemática de información.

Por el contrario, el grupo control mostró incrementos mínimos en estas mismas dimensiones (+0.6 y +0.4 puntos), con tamaños del efecto pequeños ($d<0.5$). Más allá de los cambios intragrupo, resulta fundamental comparar directamente las ganancias entre ambos grupos para determinar si las diferencias observadas resultan estadísticamente significativas y de magnitud práctica relevante. La Tabla 4 presenta el análisis comparativo de las ganancias (diferencia post-test menos pre-test) mediante prueba t de Student para muestras independientes.

Los datos de la Tabla 4 confirman de manera contundente la efectividad diferencial de la intervención con IA. La diferencia en ganancia total (7.2 puntos) resulta altamente significativa ($t_{118}=13.65$, $p<0.001$) con un tamaño del efecto de 2.50, clasificado como muy grande. Este hallazgo implica que la probabilidad de que las diferencias observadas se deban al azar es inferior a 1 en 1000, proporcionando evidencia robusta de la eficacia de la intervención. Adicionalmente, el tamaño del efecto indica que más del 99% del grupo experimental superó la

media del grupo control en términos de ganancias, lo cual representa una ventaja práctica sustancial.

Tabla 4. *Comparación de ganancias entre grupo experimental y control.*

Dimensión	Ganancia Experimental M (DE)	Ganancia Control M (DE)	Diferencia	t	p	d Cohen
Puntuación Total	9.4 (3.9)	2.2 (3.9)	7.2	13.65	<0.001	2.50
Inferencia	1.7 (1.3)	0.4 (1.3)	1.3	8.45	<0.001	1.55
Reconoc. supuestos	1.4 (1.4)	0.5 (1.4)	0.9	5.34	<0.001	0.98
Deducción	1.8 (1.4)	0.3 (1.1)	1.5	9.12	<0.001	1.67
Interpretación	2.4 (1.3)	0.6 (1.3)	1.8	11.23	<0.001	2.05
Evaluación argumentos	2.1 (1.3)	0.4 (1.2)	1.7	10.87	<0.001	1.99

Fuente: Elaboración propia.

El análisis dimensional revela que todas las dimensiones mostraron diferencias significativas favorables al grupo experimental (todos los valores $p<0.001$), con tamaños del efecto que oscilaron entre 0.98 (Reconocimiento de supuestos) y 2.05 (Interpretación). Particularmente destacables son las diferencias en Interpretación (+1.8 puntos, d=2.05) y Evaluación de argumentos (+1.7 puntos, d=1.99), que ratifican la hipótesis de que las actividades centradas en análisis comparativo de fuentes y evaluación crítica de respuestas de IA potencian específicamente estas competencias cognitivas de orden superior. Complementariamente a las mediciones cuantitativas de pensamiento crítico, resulta relevante explorar las percepciones subjetivas de los estudiantes del grupo experimental respecto a su experiencia con las herramientas de IA. La Tabla 5 presenta los resultados del cuestionario de percepciones aplicado en la semana 16, organizado temáticamente.

Tabla 5. *Percepciones estudiantiles sobre uso de IA (n=60).*

Ítem	Acuerdo n (%)	Neutral n (%)	Desacuerdo n (%)
Las plataformas fueron fáciles de usar	50 (83.3%)	7 (11.7%)	3 (5.0%)
La IA mejoró mi comprensión de temas históricos	47 (78.3%)	10 (16.7%)	3 (5.0%)
Confío plenamente en la información de la IA	27 (45.0%)	12 (20.0%)	21 (35.0%)
Detecté errores en las respuestas de la IA	41 (68.3%)	14 (23.3%)	5 (8.3%)
Mejoró mi capacidad de análisis crítico	49 (81.7%)	9 (15.0%)	2 (3.3%)
Contraté información con otras fuentes	53 (88.3%)	5 (8.3%)	2 (3.3%)

Fuente: Elaboración propia.

La Tabla 5 evidencia percepciones mayoritariamente positivas, aunque críticas. El 83.3% valoró la facilidad de uso tras familiarización inicial, y el 78.3% reconoció mejoras en comprensión de contenidos históricos. Especialmente relevante resulta que el 81.7% percibió mejoras en capacidad de análisis crítico, lo cual converge con las mediciones objetivas del test Watson-Glaser. No obstante, solo el 45% confió plenamente en la información sin verificación, mientras que el 35% expresó desconfianza, reflejando la actitud escéptica fomentada durante la intervención.

Consistentemente, el 68.3% reportó haber detectado errores factuales, y el 88.3% contrastó información con fuentes tradicionales, conductas indicativas del desarrollo exitoso de pensamiento crítico aplicado al uso de tecnologías de IA. A pesar de los resultados positivos, se identificaron varios desafíos durante la implementación. La brecha digital constituyó el obstáculo más significativo: aunque todos los participantes declararon acceso a dispositivos y conectividad, existieron diferencias cualitativas importantes. El 23.3% de estudiantes dependía

exclusivamente de smartphones con pantallas pequeñas que dificultaban la lectura de respuestas extensas y la comparación simultánea de múltiples fuentes. El 18.3% reportó problemas intermitentes de conectividad que interrumpieron sesiones de trabajo. La dependencia excesiva representó otro riesgo: durante las primeras semanas, el 31.7% tendía a aceptar acríticamente las respuestas de la IA. Los errores factuales y sesgos en las respuestas también plantearon desafíos pedagógicos, aunque proporcionaron oportunidades valiosas para desarrollar habilidades analíticas. Los docentes reportaron 4-6 horas semanales adicionales de preparación.

Los resultados confirman la hipótesis respecto a que la integración sistemática de IA contribuye significativamente al desarrollo del pensamiento crítico. El objetivo general se alcanzó satisfactoriamente, demostrando mejoras sustanciales (d=2.40). Los objetivos específicos se cumplieron: el instrumento caracterizó niveles iniciales y finales con precisión, la intervención se implementó exitosamente en 24 sesiones estructuradas, el análisis estadístico demostró diferencias significativas, y se identificaron fortalezas y limitaciones mediante observaciones y cuestionarios.

Los hallazgos convergen con García et al. (2024), quienes reportaron efectos positivos, aunque con tamaños menores (d≈1.2). La magnitud superior observada (d=2.40) puede atribuirse a la duración extendida (12 vs 6 semanas), integración curricular sistemática y uso comparativo de múltiples plataformas. Parreño et al. (2024) alertaron sobre dependencia; aunque el 31.7% mostró inicialmente aceptación acrítica, la supervisión docente mitigó este riesgo: 68.3% detectó errores al finalizar. Asimismo, Puche (2024) identificaron potencialidades similares. Los datos de Ser Estudiante Ecuador (+2 puntos; INEVAL, 2024) contrastan con los incrementos experimentales (+9.4 puntos). Teóricamente, el estudio contribuye a reconceptualizar el papel docente en la era de IA. La IA amplifica capacidades pedagógicas cuando se integra mediante diseño fundamentado. Prácticamente, el estudio ofrece un protocolo replicable: capacitación docente previa, sesiones estructuradas, comparación crítica entre plataformas, verificación sistemática, evaluación formativa continua. Las implicaciones para políticas incluyen: lineamientos nacionales para uso ético, formación docente continua, inversión en infraestructura, repositorios de actividades validadas, investigación longitudinal.

La ausencia de aleatorización introduce potenciales sesgos de selección. La duración (12 semanas) y seguimiento inmediato no permiten evaluar persistencia de mejoras. La muestra se circunscribió al norte de Guayaquil, limitando generalización. El test Watson-Glaser, aunque validado, implica cierta descontextualización. El estudio se focalizó en resultados cognitivos, sin explorar dimensiones afectivas. El énfasis en herramientas comerciales introduce dependencia de tecnologías propietarias.

Conclusiones

El estudio demuestra que la integración sistemática y pedagógicamente fundamentada de IA constituye una estrategia efectiva para desarrollar pensamiento crítico. Los resultados evidencian mejoras significativas (d=2.40) en capacidades analíticas tras 12 semanas de intervención. Las plataformas de IA ofrecen potencialidades valiosas bajo supervisión docente: acceso a perspectivas múltiples, práctica en evaluación crítica, desarrollo

metacognitivo y personalización del aprendizaje. No obstante, la efectividad depende críticamente de la calidad del diseño instruccional. En Ecuador, la IA emerge como oportunidad para democratizar acceso a experiencias de calidad. Mirando al futuro, la IA transformará inevitablemente las prácticas educativas, requiriendo direccionarla humanísticamente.

La integración sistemática de herramientas de inteligencia artificial en el aula de Ciencias Sociales demostró ser una estrategia altamente efectiva para desarrollar pensamiento crítico en estudiantes de Bachillerato, evidenciando mejoras significativas (d=2.40) tras 12 semanas de intervención pedagógica estructurada. Los hallazgos confirman que plataformas como ChatGPT, Claude, Gemini, Copilot y DeepSeek, cuando se emplean bajo supervisión docente y diseño instruccional fundamentado, potencian capacidades de análisis, evaluación e inferencia. Sin embargo, la efectividad depende críticamente de la mediación pedagógica que fomente actitud crítica, verificación sistemática de información y uso ético de tecnologías emergentes en contextos educativos ecuatorianos.

Referencias Bibliográficas

Alva, W., Wong, E., Zúñiga, C., & Matos, C. (2024). Autoeficacia y el desarrollo del pensamiento crítico en estudiantes de educación superior pertenecientes a una universidad de Lima metropolitana. 2023. EDUCARE ET COMUNICARE. Revista de Investigación de la Facultad de Humanidades, 12(1), 66–82. https://doi.org/10.35383/educare.v12i1.1023

Delgado, N., Carrasco, L., de la Maza, M., & Etxabe, J. (2024). Aplicación de la inteligencia artificial (IA) en educación: Los beneficios y limitaciones de la IA percibidos por el profesorado de educación primaria, educación secundaria y educación superior. Revista Electrónica Interuniversitaria de Formación del Profesorado, 27(1), 207–224. https://doi.org/10.6018/reifop.577211

García, N., Gutiérrez, M., Zaragoza, O., Soto, R., & Jiménez, G. (2024). Pensamiento crítico en estudiantes de bachillerato: Una aproximación desde las inteligencias artificiales. Multidisciplinary Latin American Journal (MLAJ), 2(3), 154–174. https://doi.org/10.62131/MLAJ-V2-N3-010

García, F., Llorens, F., & Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. RIED-Revista Iberoamericana de Educación a Distância, 27(1), 9–39. https://doi.org/10.5944/ried.27.1.37716

Instituto Nacional de Evaluación Educativa. (2024). Resultados de la evaluación Ser Estudiante 2023–2024. Ministerio de Educación del Ecuador. https://www.evaluacion.gob.ec/ineval-presento-los-resultados-de-la-evaluacion-nacional-ser-estudiante-2024/

Lookkle. (2026). Las 5 IAs más usadas en 2026: ChatGPT, Google Gemini, Microsoft Copilot, Claude y Perplexity AI. https://www.lookkle.com/es/5-ias-ms-usadas-en-2026

López, A., & Sánchez, J. (2025). La inteligencia artificial en contextos educativos: Usos éticos de la información y alfabetización digital. Derecom, 38(1), 3–11. https://doi.org/10.5209/dere.102342

Martínez, M., Rigueira, X., Larrañaga, A., Martínez, J., Ocarranza, I., & Kreibel, D. (2023). Impacto de la inteligencia artificial en los métodos de evaluación en la educación primaria y secundaria: Revisión sistemática de la literatura. Revista de Psicodidáctica, 28(2), 93–103. https://doi.org/10.1016/j.psicod.2023.06.001

Miao, F., & Holmes, W. (2024). Guía para el uso de IA generativa en educación e investigación. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000389227

Ministerio de Educación de España. (2024). Guía sobre el uso de la inteligencia artificial en el ámbito educativo. Instituto Nacional de

Tecnologías Educativas y de Formación del Profesorado. https://code.intef.es/wp-content/uploads/2024/07/Gu%C3%ADa-sobre-el-uso-de-la-IA-en-el-%C3%A1mbito-educativo-INTEF_2024.pdf

Ministerio de Educación del Ecuador. (2022). Currículo de Bachillerato General Unificado. https://educacion.gob.ec/curriculo-bgu/

Mota, S., Huizar, D., Martínez, M., & Martínez, M. (2020). Aprendizaje colaborativo mediado por TIC. Revista Iberoamericana de Ciencias, 7(3), 1–9. https://scholar.google.com/scholar?q=%22Aprendizaje+colaborativo+mediado+por+TIC%22

Naranjo, B., Izurieta, C., Leica, L., Morrillo, C., & Salazar, A. (2023). Ética y responsabilidad en la implementación de la inteligencia artificial en la educación. Ciencia Latina Revista Científica Multidisciplinar, 7(6), 2054–2076. https://doi.org/10.37811/cl_rcm.v7i6.8833

Oviedo, L. (2023). Dilema de la inteligencia artificial: Pensamiento crítico y generaciones digitales. Realidad y Reflexión, 23(58), 69–83. https://doi.org/10.5377/ryr.v1i58.17397

Parreño, J., Rivera, D., Uvidia, M., & Jara, J. (2024). La inteligencia artificial: Herramienta para el desarrollo del pensamiento crítico en estudiantes de instituciones educativas. Revista Conrado, 20(96), 478–485. https://scholar.google.com/scholar?q=%22La+Inteligencia+Artificial%3A+herramienta+para+el+desarrollo+del+pensamiento+cr%C3%ADtico%22

Pérez, M. E., Domínguez, J. B., & Sánchez, A. (2025). Pensamiento crítico en estudiantes de la Universidad Católica de Cuenca, Ecuador. Revista Cognosis, 10(1), 63–80. https://doi.org/10.33936/cognosis.v10i1.6614

Puche, D. C. (2024). Impacto de la inteligencia artificial en el desarrollo del pensamiento crítico de estudiantes universitarios. Delectus, 7(2), 166–178. https://doi.org/10.36996/delectus.v7i2.242

Ramos, C. A. (2021). Diseños de investigación experimental. CienciAmérica, 10(1), 1–7. https://doi.org/10.33210/ca.v10i1.356

Esta obra está bajo una licencia de Creative Commons Reconocimiento-No Comercial 4.0 Internacional. Copyright © Carmen Laura Benalcázar Arroba y Steven Arturo Torres Burgos.

Declaraciones éticas y editoriales del artículo
Contribución de los autores (Taxonomía CRediT) Carmen Laura Benalcázar Arroba: conceptualización de la investigación, diseño metodológico, desarrollo del proceso investigativo, análisis formal de los datos, redacción del borrador original del manuscrito, revisión crítica del contenido científico y supervisión general del estudio. Steven Arturo Torres Burgos: curación y organización de los datos, participación en la recolección de información, validación de los resultados obtenidos y elaboración de representaciones gráficas y visualización de los datos.
Declaración de conflicto de intereses Los autores declaran que no existe conflicto de intereses en relación con la investigación presentada, la autoría del manuscrito ni la publicación del presente artículo.
Declaración de financiamiento La presente investigación no recibió financiamiento específico de agencias públicas, comerciales o de organizaciones sin fines de lucro. En caso de existir financiamiento institucional o externo, este deberá ser declarado explícitamente por los autores en esta sección.
Declaración del editor El editor responsable certifica que el proceso editorial del presente artículo se desarrolló conforme a los principios de integridad científica, transparencia y buenas prácticas editoriales. El manuscrito fue sometido a un proceso de evaluación mediante revisión por pares doble ciego, garantizando la confidencialidad de la identidad de los autores y revisores durante todo el proceso de dictamen académico. Asimismo, el editor declara que el artículo cumple con los criterios científicos, metodológicos y éticos establecidos por la revista.
Declaración de los revisores Los revisores externos que participaron en la evaluación del presente manuscrito declaran haber realizado el proceso de revisión de manera objetiva, independiente y confidencial. Asimismo, manifiestan que no mantienen conflictos de interés con los autores ni con la investigación evaluada, y que sus observaciones y recomendaciones se fundamentan exclusivamente en criterios científicos, metodológicos y académicos.
Declaración ética de la investigación Los autores declaran que la investigación se desarrolló respetando los principios éticos de la investigación científica, garantizando la confidencialidad de los datos y el respeto a los participantes del estudio. En los casos en que la investigación involucre seres humanos, los procedimientos deben ajustarse a los principios éticos establecidos en la Declaración de Helsinki y a las normativas institucionales correspondientes.
Declaración sobre el uso de inteligencia artificial Los autores declaran que el uso de herramientas de inteligencia artificial, en caso de haberse utilizado durante el proceso de investigación o redacción del manuscrito, se realizó únicamente como apoyo técnico para mejorar la claridad del lenguaje o el análisis de información, manteniendo siempre la responsabilidad intelectual sobre el contenido del artículo. Las herramientas de inteligencia artificial no fueron utilizadas como autoras del manuscrito ni sustituyen la responsabilidad académica de los investigadores.
Disponibilidad de datos Los datos que respaldan los resultados de esta investigación estarán disponibles previa solicitud razonable al autor de correspondencia, respetando las normas éticas y de confidencialidad establecidas por la investigación.