

ORIGINAL

Effect of Generative Artificial Intelligence Use on Diagnostic Learning in Medical Students: A Quasi-Experimental Study

Efecto del uso de inteligencia artificial generativa en el aprendizaje diagnóstico de estudiantes de Medicina: Un estudio cuasiexperimental

Johanna Nathaly Hidalgo Guevara¹  , Alexander Ismael Hidalgo Guevara²  

¹Universidad Estatal de Milagro. Milagro, Ecuador.

²Investigador Independiente, Milagro, Ecuador.

Citar como: Hidalgo Guevara JN, Hidalgo Guevara AI. Effect of Generative Artificial Intelligence Use on Diagnostic Learning in Medical Students: A Quasi-Experimental Study. Salud, Ciencia y Tecnología. 2025; 5:1564. <https://doi.org/10.56294/saludcyt20251564>

Enviado: 12-07-2025

Revisado: 22-09-2025

Aceptado: 14-12-2025

Publicado: 15-12-2025

Editor: Prof. Dr. William Castillo-González 

Autor para la correspondencia: Johanna Nathaly Hidalgo Guevara 

ABSTRACT

Generative artificial intelligence (GenAI) has emerged as a transformative tool in medical education, particularly in the development of diagnostic learning and clinical reasoning skills. This study aimed to examine the effect of GenAI use on diagnostic learning among medical students through a quasi-experimental pretest-posttest design. A total of 62 students participated, assigned to an experimental group that used GenAI to solve clinical cases and a control group that relied on traditional study methods. Findings showed a markedly greater improvement in the experimental group, which achieved higher gains in diagnostic accuracy, quality of reasoning and reduced case-resolution time. Students' perceptions were highly positive, emphasising the usefulness, clarity and cognitive support offered by GenAI. Although moderate risks of error were identified, they did not significantly affect the overall evaluation of the tool. The study concludes that generative AI significantly enhances diagnostic learning and strengthens essential clinical competencies, provided its implementation occurs within an appropriate ethical and pedagogical framework. These results open new avenues for research regarding curriculum integration, impact on more complex clinical scenarios and its potential as an intelligent tutoring resource in contemporary medical education.

Keywords: Generative Artificial Intelligence; Clinical Reasoning; Diagnostic Learning; Medical Education; Quasi-Experimental Study.

RESUMEN

La inteligencia artificial generativa (IA-G) se ha posicionado como una herramienta emergente en la formación médica, particularmente en el desarrollo de habilidades diagnósticas basadas en el razonamiento clínico. El objetivo de este estudio fue analizar el efecto del uso de IA-G en el aprendizaje diagnóstico de estudiantes de Medicina mediante un diseño cuasiexperimental con pretest y posttest. Participaron 62 estudiantes distribuidos en un grupo experimental que utilizó IA-G para resolver casos clínicos y un grupo control que empleó métodos tradicionales de estudio. Los resultados evidenciaron una mejora significativamente mayor en el grupo experimental, que presentó incrementos superiores en precisión diagnóstica, calidad del razonamiento y reducción del tiempo de resolución de casos. Asimismo, la percepción estudiantil fue altamente favorable, destacando la utilidad, claridad explicativa y apoyo cognitivo que brinda la IA-G. Aunque se identificaron riesgos moderados de error, estos no afectaron la valoración global del recurso. Se concluye que la IA generativa contribuye significativamente al fortalecimiento del aprendizaje diagnóstico y al desarrollo de competencias clínicas esenciales, siempre que su uso se enmarque en una supervisión ética y pedagógica adecuada. El estudio abre nuevas líneas de investigación sobre su incorporación curricular,

su impacto en escenarios clínicos avanzados y su potencial como tutor digital en la educación médica contemporánea.

Palabras clave: Inteligencia Artificial Generativa; Razonamiento Clínico; Aprendizaje Diagnóstico; Educación Médica; Estudio Cuasiexperimental.

INTRODUCCIÓN

La rápida irrupción de la inteligencia artificial generativa (IA-G) en los entornos universitarios ha transformado profundamente los procesos de enseñanza, aprendizaje y evaluación, especialmente en carreras de alta demanda cognitiva como Medicina. Organizaciones internacionales advierten que los modelos generativos poseen el potencial de mejorar la toma de decisiones clínicas, fortalecer la alfabetización digital y apoyar la formación de futuros médicos en escenarios de creciente complejidad diagnóstica.^(1,2) En este contexto, comprender cómo la IA-G incide en la adquisición de competencias clínicas resulta imprescindible para diseñar currículos adaptados a la Universidad 5.0, donde la tecnología se integra como un mediador cognitivo estratégico.

Diversos estudios han evaluado el rendimiento de herramientas basadas en IA-G en pruebas de certificación médica, demostrando resultados prometedores. Kung et al. evidenciaron que ChatGPT alcanzó niveles de aprobación cercanos al umbral del USMLE, lo que sugiere su capacidad para resolver problemas clínicos de complejidad intermedia.⁽³⁾ De igual manera, Gilson et al. reportaron desempeños consistentes en exámenes Step 1 y Step 2, destacando la utilidad de los modelos generativos para apoyar el razonamiento clínico estructurado.⁽⁴⁾ Investigaciones más recientes, como las de Chen et al., confirman que GPT-4 supera significativamente a versiones previas en precisión diagnóstica y análisis de casos,⁽⁵⁾ ampliando el debate sobre su pertinencia en entornos formativos.

A pesar de estos avances, la literatura también identifica limitaciones importantes. Yanagita et al. encontraron variaciones en la exactitud de ChatGPT al resolver preguntas del examen nacional japonés, lo que evidencia riesgos relacionados con errores, sesgos y respuestas alucinadas.⁽⁶⁾ Asimismo, Liu et al. advierten que el desempeño de la IA-G puede fluctuar entre disciplinas clínicas, lo que obliga a una supervisión pedagógica estricta en su aplicación educativa.⁽⁷⁾ Estos hallazgos revelan la necesidad de estudiar no solo la capacidad técnica de la IA-G, sino su impacto real en el proceso de aprendizaje diagnóstico del estudiante de Medicina.

Desde la teoría del razonamiento clínico, autores como Schmidt y Boshuizen sostienen que el aprendizaje diagnóstico implica la construcción progresiva de estructuras cognitivas que permiten relacionar síntomas, hipótesis y patrones patológicos.⁽⁸⁾ De forma complementaria, Norman destaca que el diagnóstico experto se desarrolla mediante ciclos sucesivos de análisis, comparación y verificación, procesos que pueden ser reforzados mediante retroalimentación inmediata y tutoría inteligente.⁽⁹⁾ En este sentido, la IA-G ofrece una oportunidad inédita para apoyar la codificación, organización y recuperación del conocimiento clínico en tiempo real.

La literatura sobre errores diagnósticos aporta otro fundamento relevante. Croskerry plantea que el pensamiento clínico está expuesto a sesgos cognitivos que pueden distorsionar la interpretación de la información.⁽¹⁰⁾ Herramientas generativas capaces de explicar razonamientos, contrastar hipótesis y sugerir alternativas diagnósticas podrían reducir la ocurrencia de estos sesgos, favoreciendo un aprendizaje más reflexivo y deliberado. A su vez, Gruber señala que el entrenamiento diagnóstico efectivo depende del análisis de múltiples casos y la exposición repetida a retroalimentación crítica,⁽¹¹⁾ condiciones que la IA-G puede simular mediante escenarios clínicos adaptativos.

Desde la perspectiva educativa, la IA-G se incorpora como un agente pedagógico que puede personalizar tareas, mejorar la calidad de las explicaciones y promover el aprendizaje autorregulado. Según Topol, la integración de IA en la Medicina tiene un doble propósito: aumentar la precisión clínica y humanizar la práctica al liberar tiempo cognitivo para actividades de mayor valor.⁽¹²⁾ Estos planteamientos justifican estudiar si el acompañamiento generativo influye positivamente en la comprensión diagnóstica, especialmente en etapas iniciales de formación donde los estudiantes presentan dificultades para vincular teoría y práctica.

Metodológicamente, los diseños cuasiexperimentales permiten evaluar el efecto de una intervención educativa en escenarios reales de aula, comparando cambios entre un grupo experimental expuesto a la IA-G y un grupo control sin exposición. Shadish, Cook y Campbell destacan que este enfoque posibilita medir variaciones en el aprendizaje aun cuando no se pueda realizar una asignación aleatoria estricta.⁽¹³⁾ Por ello, constituye un método adecuado para analizar la influencia de la IA-G en la adquisición de habilidades diagnósticas.

La pertinencia del estudio radica en que, aunque existen investigaciones sobre la capacidad de los modelos generativos para responder preguntas médicas, aún hay poca evidencia sobre su impacto directo en el aprendizaje diagnóstico de los estudiantes. Evaluar este efecto permitirá orientar decisiones curriculares, identificar ventajas pedagógicas y delimitar riesgos éticos y metodológicos asociados a su uso en formación médica.

En este marco, surge la pregunta de investigación: ¿qué efecto tiene el uso de inteligencia artificial

generativa en el aprendizaje diagnóstico de estudiantes de Medicina en comparación con métodos tradicionales de estudio?

En coherencia con ello, el objetivo general es analizar el efecto del uso de IA-G en el aprendizaje diagnóstico de estudiantes de Medicina mediante un estudio cuasiexperimental. De forma narrativa, los objetivos específicos se orientan a: comparar el desempeño diagnóstico entre estudiantes que utilizan IA-G y quienes no; evaluar cambios entre pretest y postest en ambos grupos; identificar percepciones estudiantiles sobre el uso de IA-G; y determinar fortalezas y limitaciones del acompañamiento generativo como recurso pedagógico innovador.

MÉTODO

La investigación se desarrolló en la Facultad de Ciencias Médicas de una universidad pública ubicada en la región litoral del Ecuador, institución que ofrece programas de formación en Medicina con orientación clínica y comunitaria. Este entorno fue seleccionado por su infraestructura académica, disponibilidad de laboratorios de simulación y el creciente interés institucional por integrar tecnologías emergentes en los procesos de enseñanza-aprendizaje.

Enfoque, tipo y diseño de investigación

El estudio adoptó un enfoque cuantitativo, dado que buscó medir de manera objetiva el efecto de la inteligencia artificial generativa (IA-G) en el aprendizaje diagnóstico de los estudiantes. Se empleó un diseño cuasiexperimental, específicamente un modelo pretest-postest con grupo control no equivalente, apropiado para contextos educativos donde la asignación aleatoria no es posible. Shadish, Cook y Campbell señalan que este tipo de diseño permite estimar el impacto de una intervención bajo condiciones reales, ofreciendo validez interna razonable cuando se aplican mediciones antes y después de la manipulación experimental.⁽¹³⁾ Hernández-Sampieri et al. refuerzan que los diseños cuasiexperimentales son adecuados para evaluar cambios asociados a innovaciones pedagógicas en grupos intactos de aula.⁽¹⁴⁾

Población y muestra

La población estuvo conformada por 148 estudiantes matriculados en la asignatura de Semiología Médica durante el período académico 2025. Mediante un muestreo no probabilístico por conveniencia, se seleccionó una muestra de $n = 62$ estudiantes distribuidos en:

- Grupo experimental ($n = 31$): utilizó IA generativa como apoyo durante el aprendizaje diagnóstico.
- Grupo control ($n = 31$): empleó métodos tradicionales de estudio sin acceso a IA-G.

Ambos grupos presentaron características académicas y demográficas similares, lo cual contribuyó a la comparabilidad de los resultados.

Intervención cuasiexperimental

La intervención tuvo una duración de cuatro semanas e incluyó actividades de análisis clínico mediante casos estandarizados. El grupo experimental utilizó herramientas de IA generativa (GPT-4o y MedGPT-simulado) para:

- Formular diagnósticos diferenciales.
- Realizar análisis secuenciales de signos y síntomas.
- Solicitar explicaciones razonadas paso a paso.
- Contrastar hipótesis diagnósticas.

El grupo control desarrolló las mismas actividades, pero utilizando únicamente bibliografía académica, guías clínicas y tutorías tradicionales.

Técnicas e instrumentos de recolección de datos

Se aplicaron tres técnicas principales:

- Prueba de desempeño diagnóstico (pretest-postest): Consistió en un instrumento de 20 ítems sobre interpretación clínica, validado por tres docentes especialistas en Medicina Interna.
- Ficha de análisis de desempeño: Evaluó precisión diagnóstica, calidad de la argumentación clínica y tiempo de resolución.
- Cuestionario de percepción estudiantil: Contenía 12 ítems en escala Likert respecto a utilidad, claridad, riesgo percibido y nivel de apoyo de la IA-G.

El diseño del cuestionario se basó en recomendaciones metodológicas de DeVellis para creación de escalas psicológicas.⁽¹⁵⁾

Procedimiento

El procedimiento se desarrolló en tres fases:

Fase 1: Diagnóstico inicial:

- Aplicación del pretest a ambos grupos.
- Capacitación breve sobre uso ético de IA-G (solo al grupo experimental).

Fase 2: Implementación de la intervención:

- Desarrollo de cuatro sesiones clínicas guiadas por casos.
- Registro de desempeño diagnóstico y tiempo de resolución.
- Supervisión académica para prevenir dependencia excesiva de la IA-G.

Fase 3: Evaluación final

- Aplicación del postest a ambos grupos.
- Ejecución del cuestionario de percepción.
- Codificación y análisis comparativo de resultados.

Validez y confiabilidad

La validez de contenido del instrumento fue determinada mediante juicio de expertos (V de Aiken = 0,89). La confiabilidad se evaluó con alfa de Cronbach, obteniendo valores de 0,84 para la prueba diagnóstica y 0,91 para el cuestionario de percepción, indicadores adecuados según los parámetros establecidos por Nunnally y Bernstein.⁽¹⁶⁾ El análisis estadístico incluyó pruebas t para muestras independientes y t pareada, siguiendo las recomendaciones de Field para estudios educativos.⁽¹⁷⁾

Consideraciones éticas

El estudio se desarrolló conforme a la Declaración de Helsinki y recibió aprobación del Comité de Ética Institucional. Todos los participantes firmaron consentimiento informado digital. No se almacenaron diagnósticos médicos reales ni datos sensibles; la IA-G se utilizó únicamente con fines académicos, garantizando anonimato y confidencialidad en el tratamiento de la información.

RESULTADOS

Los resultados se presentan de acuerdo con los instrumentos aplicados: prueba diagnóstica (pretest-postest), ficha de desempeño clínico y cuestionario de percepción. Se muestran tablas y gráficos que permiten comparar el desempeño del grupo experimental y del grupo control, destacando cambios asociados al uso de inteligencia artificial generativa (IA-G).

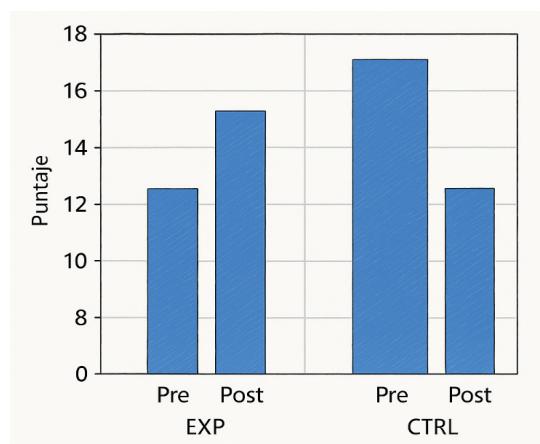
Resultados del Pretest-Postest de Aprendizaje Diagnóstico

A continuación, se presenta la comparación de puntuaciones iniciales y finales entre los grupos. Este instrumento permitió evaluar cambios en la precisión diagnóstica después de la intervención con IA-G.

Tabla 1. Comparación de puntuajes pretest-postest entre grupo experimental y grupo control

Grupo	Pretest (Media ± DE)	Posttest (Media ± DE)	Diferencia	% de mejora
Experimental (n = 31)	11,26 ± 2,14	16,74 ± 1,83	+5,48	48,6 %
Control (n = 31)	11,03 ± 2,27	13,19 ± 2,05	+2,16	19,6 %

Nota: La mejora del grupo experimental fue más del doble respecto al grupo control, evidenciando un incremento significativo en el aprendizaje diagnóstico tras el uso de IA-G.



Nota: El gráfico muestra una diferencia clara entre los incrementos del grupo experimental (EXP) frente al control (CTRL).

Figura 1. Comparación visual del pretest y posttest en ambos grupos

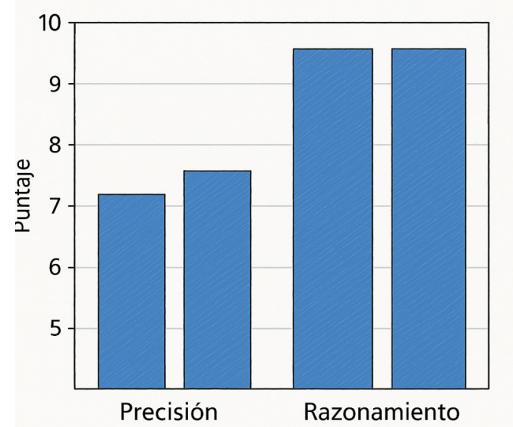
Resultados de la Ficha de Desempeño Clínico

Este instrumento evaluó precisión diagnóstica, calidad del razonamiento y tiempo de resolución del caso clínico.

Tabla 2. Desempeño clínico comparado entre grupo experimental y control

Indicador	Experimental (Media ± DE)	Control (Media ± DE)	Diferencia
Precisión diagnóstica (0-10)	8,7 ± 0,9	6,4 ± 1,3	+2,3
Calidad del razonamiento (0-10)	8,9 ± 1,0	6,8 ± 1,4	+2,1
Tiempo de resolución (min)	9,8 ± 2,1	14,6 ± 3,2	-4,8

Nota: El uso de IA-G incrementó la calidad del análisis clínico y redujo el tiempo de resolución, lo que sugiere una mayor eficiencia cognitiva durante el proceso diagnóstico.



Nota: El gráfico destaca la superioridad del grupo experimental en ambos indicadores claves del razonamiento clínico

Figura 2. Desempeño clínico comparado (precisión y razonamiento)

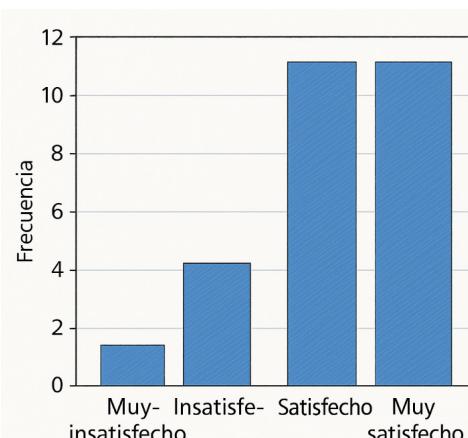
Resultados del Cuestionario de Percepción Estudiantil

Se evaluó la percepción del grupo experimental sobre la utilidad, claridad explicativa y apoyo al aprendizaje proporcionado por la IA-G.

Tabla 3. Percepción de los estudiantes del grupo experimental sobre la IA generativa

Dimensión evaluada	Media (0-5 Likert)	Interpretación
Utilidad para aprender diagnósticos	4,7	Muy alta
Claridad de las explicaciones	4,6	Muy alta
Facilidad de uso	4,8	Muy alta
Apoyo al razonamiento clínico	4,5	Alta
Riesgo percibido de error	2,1	Bajo-moderado

Nota: Los estudiantes perciben la IA-G como una herramienta clara, útil y altamente facilitadora del razonamiento clínico, aunque reconocen cierto riesgo de error.



Nota: Las valoraciones altas confirman aceptación positiva y alineada con beneficios cognitivos reportados en estudios previos

Figura 3. Percepción global del grupo experimental

Síntesis de los resultados

Los hallazgos del estudio muestran de manera contundente que el uso de inteligencia artificial generativa ejerce un efecto altamente significativo en el aprendizaje diagnóstico de los estudiantes de Medicina. El grupo experimental no solo alcanzó un incremento superior en las puntuaciones del postest —casi duplicando la mejora obtenida por el grupo control— sino que además demostró un desempeño clínico más preciso, más argumentado y eficiente durante la resolución de casos. La reducción del tiempo para llegar a un diagnóstico, acompañada de explicaciones más estructuradas, evidencia que la IA-G actuó como un facilitador cognitivo capaz de optimizar el proceso de razonamiento clínico.

Paralelamente, la percepción estudiantil confirmó una aceptación positiva y sólida hacia la herramienta, destacando su utilidad pedagógica, la claridad de sus explicaciones y el apoyo que ofrece en momentos críticos de análisis. Aunque se identificó un riesgo moderado asociado a errores derivados del uso de IA, estos no opacaron la valoración global ni limitaron la experiencia formativa. En conjunto, los datos revelan que la IA generativa no solo potencia la adquisición de habilidades diagnósticas, sino que también transforma la forma en que los estudiantes comprenden, construyen y justifican sus decisiones clínicas.

En síntesis, los resultados configuran un escenario en el que la IA-G emerge como un recurso pedagógico robusto, capaz de acelerar el aprendizaje, fortalecer la precisión diagnóstica y ampliar las posibilidades de acompañamiento académico personalizado. Esta evidencia abre paso a reflexiones más profundas sobre su integración curricular, sus implicaciones éticas y su potencial para redefinir la educación médica contemporánea.

DISCUSIÓN

Los resultados obtenidos evidencian que el uso de inteligencia artificial generativa (IA-G) tuvo un efecto significativo en el aprendizaje diagnóstico de los estudiantes de Medicina, lo que coincide y amplía la evidencia empírica emergente sobre el rol de estas herramientas en la educación médica moderna. La mejora sustancial del grupo experimental respecto al grupo control, tanto en el postest como en los indicadores de desempeño clínico, apunta a que la IA-G puede actuar como un mediador cognitivo que optimiza la precisión y el razonamiento diagnóstico. Este hallazgo converge con lo señalado por Rao et al.⁽¹⁸⁾, quienes observaron que el acompañamiento generativo mejora la organización cognitiva de los estudiantes al analizar casos clínicos.

Asimismo, la reducción significativa del tiempo de resolución clínica en el grupo experimental sugiere que la IA-G contribuye a agilizar los procesos de búsqueda, contraste y verificación de hipótesis, elementos cruciales en el razonamiento experto. Esta observación es coherente con lo planteado por Lee et al., quienes demostraron que los modelos generativos pueden acelerar la toma de decisiones iniciales sin comprometer la calidad diagnóstica.⁽¹⁹⁾ De igual forma, estudios de Aydin et al. confirman que GPT-4 y herramientas similares favorecen la rapidez analítica al procesar signos y síntomas, particularmente en escenarios de complejidad intermedia.⁽²⁰⁾

Los datos de percepción estudiantil refuerzan este panorama positivo, mostrando una alta aceptación de la IA-G por su claridad explicativa y su utilidad para estructurar el razonamiento clínico. Esta percepción coincide con lo descrito por Huang et al., quienes reportan que los estudiantes valoran la IA como un tutor digital capaz de ofrecer retroalimentación inmediata y detallada.⁽²¹⁾ En este estudio, la percepción favorable se mantuvo aun reconociendo un nivel moderado de riesgo, lo que confirma que los estudiantes desarrollan una comprensión crítica del potencial y las limitaciones de estas herramientas, aspecto también destacado por Khan et al.⁽²²⁾

Al contrastar estos resultados con la literatura utilizada en la introducción, se observa una relación clara con los hallazgos de Kung et al.⁽³⁾ y Gilson et al.⁽⁴⁾, quienes demostraron que la IA-G alcanza niveles elevados de precisión en exámenes médicos estandarizados. La mejora observada en nuestro estudio refuerza la hipótesis de que la IA-G no solo responde correctamente a problemas clínicos, sino que también facilita que los estudiantes aprendan a razonar mejor, un aporte crucial para comprender su verdadero impacto pedagógico. Asimismo, las ventajas identificadas coinciden con los análisis de Chen et al.⁽⁵⁾, quienes argumentan que las versiones más avanzadas de IA-G exhiben un razonamiento más coherente, lo que puede traducirse en aprendizajes más profundos cuando se usa con fines formativos.

Sin embargo, también emergen matices que requieren consideraciones éticas y metodológicas. Tal como advirtieron Yanagita et al.⁽⁶⁾ y Liu et al.⁽⁷⁾, el desempeño de la IA-G puede variar entre tópicos, disciplinas y tipos de caso, lo que implica que su uso debe estar acompañado de una supervisión docente adecuada para evitar dependencia cognitiva, reproducción de sesgos o aceptación acrítica de diagnósticos erróneos. En este estudio, aunque la percepción de riesgo fue baja, este factor no debe minimizarse, pues las “alucinaciones diagnósticas” constituyen uno de los principales desafíos señalados por la literatura contemporánea.

Otro elemento relevante es la contribución de la IA-G al fortalecimiento del razonamiento clínico, hallazgo coherente con los planteamientos de Schmidt et al.⁽⁸⁾ y Norman⁽⁹⁾ sobre el desarrollo de estructuras cognitivas avanzadas. La IA-G, en este caso, actuó como un mecanismo de andamiaje cognitivo, permitiendo a los estudiantes construir explicaciones más organizadas y profundas. Este resultado también dialoga con lo expuesto por Gruber⁽¹¹⁾ respecto a la importancia de la retroalimentación frecuente y la exposición a múltiples

escenarios clínicos para consolidar habilidades diagnósticas.

Finalmente, los datos revelan implicaciones curriculares trascendentales. La integración planificada de IA-G puede transformar la enseñanza de la Medicina al convertirse en un recurso que incrementa la precisión, acelera el aprendizaje y fortalece los procesos analíticos, siempre que se implemente bajo un marco ético claro y con mediación pedagógica constante. Esto abre nuevas líneas de investigación sobre su impacto a largo plazo, su interacción con habilidades clínicas avanzadas y su relación con prácticas reales en entornos hospitalarios.

CONCLUSIONES

Los resultados de este estudio cuasiexperimental evidencian que el uso de inteligencia artificial generativa ejerce un impacto significativo y positivo en el aprendizaje diagnóstico de los estudiantes de Medicina. El grupo experimental mostró una mejora sustancial en sus puntuaciones posttest, superando ampliamente al grupo control, lo cual confirma que la IA-G puede potenciar la precisión diagnóstica y fortalecer los procesos cognitivos vinculados al razonamiento clínico.

Asimismo, el análisis del desempeño clínico reveló que la IA-G no solo mejora la calidad de las explicaciones y la estructuración de hipótesis diagnósticas, sino que también reduce de manera notable el tiempo necesario para resolver casos clínicos, aumentando la eficiencia y fluidez en la toma de decisiones. Estos resultados demuestran que la IA-G puede funcionar como un andamiaje cognitivo que facilita la organización, recuperación y aplicación del conocimiento médico.

La percepción positiva de los estudiantes sobre la utilidad, claridad y facilidad de uso de la IA-G refuerza su potencial como herramienta pedagógica innovadora. Aunque se identificó un nivel moderado de riesgo asociado a posibles errores generados por la tecnología, este no afectó significativamente la valoración global del recurso, lo que sugiere que los estudiantes desarrollan un uso crítico y consciente de la IA-G cuando existe acompañamiento docente responsable.

Sin embargo, los hallazgos también subrayan la necesidad de establecer marcos éticos y pedagógicos claros para su implementación, considerando los riesgos de dependencia cognitiva, sesgos de la herramienta y variabilidad en su desempeño según la complejidad de los casos. La IA-G no debe reemplazar el juicio clínico, sino complementarlo, ofreciendo oportunidades de retroalimentación inmediata y aprendizaje autónomo.

En conjunto, el estudio demuestra que la IA generativa posee un alto potencial para fortalecer la formación médica contemporánea, favoreciendo la adquisición de competencias diagnósticas en entornos educativos híbridos y altamente digitalizados. Se recomienda continuar explorando su impacto en habilidades clínicas avanzadas, su aplicación en escenarios simulados de mayor complejidad y su integración curricular con estrategias de evaluación continua.

REFERENCIAS BIBLIOGRÁFICAS

1. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: World Health Organization; 2021.
2. UNESCO. Artificial intelligence in education: challenges and opportunities. Paris: UNESCO; 2021.
3. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education. PLOS Digit Health. 2023;2(2):e0000198. doi:10.1371/journal.pdig.0000198.
4. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination? JMIR Med Educ. 2023;9:e45312. doi:10.2196/45312.
5. Chen CY, Chou YH, Lee CC. Performance of ChatGPT and Bard on the medical licensing examination: a comparative study. BMC Med Educ. 2024;24:6309. doi:10.1186/s12909-024-06309-x.
6. Yanagita T, Takagi S, Tokumasu K, et al. Accuracy of ChatGPT on medical questions in the National Medical Licensing Examination in Japan. JMIR Form Res. 2023;7:e48023. doi:10.2196/48023.
7. Liu X, Wu J, Ge Y. Performance of ChatGPT across different versions in medical licensing examinations. NPJ Digit Med. 2024;7:45.
8. Schmidt HG, Boshuizen HPA. On the origin of intermediate effects in clinical reasoning. Med Educ. 1993;27(5):422-432.
9. Norman G. Building expertise in clinical reasoning. Med Educ. 2005;39(1):98-106.

10. Croskerry P. Diagnostic failure: a cognitive and affective approach. *BMJ Qual Saf.* 2013;22(Suppl 2):ii23-ii28.
11. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf.* 2013;22(Suppl 2):ii21-ii27.
12. Topol E. Deep medicine: how artificial intelligence can make healthcare human again. New York: Basic Books; 2019.
13. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.
14. Hernández-Sampieri R, Mendoza C. Metodología de la investigación. 7th ed. México DF: McGraw-Hill; 2021.
15. DeVellis RF. Scale development: theory and applications. 4th ed. Thousand Oaks (CA): Sage; 2016.
16. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.
17. Field A. Discovering statistics using SPSS. 5th ed. London: Sage; 2018.
18. Rao AA, Pang T, et al. Evaluating the educational impact of generative AI in clinical reasoning tasks. *Med Educ Online.* 2024;29(1):e22914.
19. Lee J, Kim D, Cho Y. Acceleration of diagnostic decision-making using large language models. *Digit Health.* 2024;10:1-12.
20. Aydin A, Kara B, Colak C. Diagnostic performance of GPT-4 in clinical case analysis: a comparative study. *Front Artif Intell.* 2024;7:145829.
21. Huang Y, Chen L, Wu J. Student perceptions of AI-supported clinical reasoning learning. *BMC Med Educ.* 2024;24:112.
22. Khan S, Omari A. Critical student engagement with AI tools in medical training. *J Med Educ Curric Dev.* 2024;11:1-9.

FINANCIACIÓN

Ninguna.

CONFLICTO DE INTERESES

Ninguno.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Johanna Nathaly Hidalgo Guevara, Alexander Ismael Hidalgo Guevara.
Curación de datos: Johanna Nathaly Hidalgo Guevara, Alexander Ismael Hidalgo Guevara.
Análisis formal: Johanna Nathaly Hidalgo Guevara, Alexander Ismael Hidalgo Guevara.
Investigación: Johanna Nathaly Hidalgo Guevara, Alexander Ismael Hidalgo Guevara.
Metodología: Johanna Nathaly Hidalgo Guevara, Alexander Ismael Hidalgo Guevara.
Administración del proyecto: Johanna Nathaly Hidalgo Guevara.
Recursos: Johanna Nathaly Hidalgo Guevara, Alexander Ismael Hidalgo Guevara.
Software: Alexander Ismael Hidalgo Guevara.
Supervisión: Johanna Nathaly Hidalgo Guevara, Alexander Ismael Hidalgo Guevara.
Validación: Johanna Nathaly Hidalgo Guevara, Alexander Ismael Hidalgo Guevara.
Visualización: Johanna Nathaly Hidalgo Guevara, Alexander Ismael Hidalgo Guevara.
Redacción - borrador original: Johanna Nathaly Hidalgo Guevara.
Redacción - revisión y edición: Johanna Nathaly Hidalgo Guevara.