

Entre la eficiencia académica y el razonamiento clínico: perfiles funcionales de uso de inteligencia artificial en estudiantes de medicina

LUIS MIGUEL DELGADO ARIAS¹, DANIEL ALBA OSPINA², MARIA ALEJANDRA LÓPEZ RÍOS³.

Recibido para publicación: 14-04-2026. Versión corregida: 30-09-2026. Aprobado para publicación: 07-10-2026.

Modelo de citación:

Delgado Arias L.M., Alba Ospina D., López Ríos M.A. **Entre la eficiencia académica y el razonamiento clínico: perfiles funcionales de uso de inteligencia artificial en estudiantes de medicina.** Arch Med (Manizales). 2026;26(2). <https://doi.org/10.30554/archmed.26.2.5667.2026>

Resumen

Introducción: La inteligencia artificial (IA) generativa se ha incorporado rápidamente a la educación médica, especialmente en tareas de estudio y productividad académica. Sin embargo, sigue siendo incierto si su uso se extiende de manera equivalente a procesos de mayor complejidad, como el razonamiento clínico. **Objetivo:** Caracterizar y comparar los perfiles funcionales de uso de IA orientados a la eficiencia académica y al razonamiento clínico en estudiantes de medicina. **Método:** Estudio transversal en 290 estudiantes de medicina. Se aplicó un cuestionario autoadministrado y se construyeron dos índices proporcionales de uso de IA: uno orientado a la eficiencia académica y el otro al razonamiento clínico. Se realizaron análisis descriptivos, prueba *t* para muestras pareadas y ANOVA de Welch con comparaciones post hoc. **Resultados:** La edad media fue $20,4 \pm 2,66$ años y el promedio académico $3,99 \pm 0,54$. El 71,3 % reportó alta frecuencia de uso de IA. El uso orientado a la eficiencia académica fue mayor que el orientado al razonamiento clínico ($0,66 \pm 0,22$ vs. $0,46 \pm 0,34$), con diferencia media de 0,20 (IC 95 %: 0,17–0,24; $p < 0,001$). Se observaron diferencias según el semestre académico. **Conclusiones:** El uso de IA en estudiantes de medicina se orienta principalmente a la eficiencia académica, con variaciones según la etapa de formación.

Palabras clave: Inteligencia artificial; educación médica; estudiantes de medicina; razonamiento clínico.

- 1 Magíster en Educación (Enseñanza). Médico. Universidad de Manizales, Manizales, Caldas, Colombia. ORCID: <https://orcid.org/0000-0001-5978-2017>. Correo electrónico: ldelgado@umanizales.edu.co. Autor de correspondencia.
- 2 Magíster en Ciencias Biomédicas con énfasis en Fisiología. Revive centro de entrenamiento, Manizales, Caldas, Colombia. ORCID: <https://orcid.org/0009-0000-2378-4619>. Correo electrónico: dalba@umanizales.edu.co
- 3 Magister en Proyectos para el Desarrollo Integral de la Infancia la Adolescencia y la Juventud. Universidad de Manizales, Manizales, Colombia. ORCID: <https://orcid.org/0009-0000-2129-8676>. Correo electrónico: psicologamariaalejandra@gmail.com

Between academic efficiency and clinical reasoning: functional profiles of artificial intelligence use in medical students.

Abstract

Introduction: Generative artificial intelligence has been rapidly integrated into medical education, particularly in study tasks and academic productivity. However, it remains unclear whether its use extends equally to more complex processes, such as clinical reasoning. **Objective:** To characterize and compare the functional profiles of AI use oriented toward academic efficiency and clinical reasoning among medical students. **Method:** A cross-sectional study involving 290 medical students. A self-administered questionnaire was administered, and two proportional indices of AI use were constructed: one oriented toward academic efficiency and the other toward clinical reasoning. Descriptive analyses, a paired t-test, and Welch's ANOVA with post hoc comparisons were performed. **Results:** The mean age was 20.4 ± 2.66 years and the GPA was 3.99 ± 0.54 . 71.3% reported high frequency of AI use. Use oriented toward academic efficiency was higher than that oriented toward clinical reasoning (0.66 ± 0.22 vs. 0.46 ± 0.34), with a mean difference of 0.20 (95% CI: 0.17–0.24; $p < 0.001$). Differences were observed according to academic semester. **Conclusions:** AI use among medical students is primarily oriented toward academic efficiency, with variations depending on the stage of training.

Keywords: Artificial intelligence; medical education; medical students; clinical reasoning.

Entre a eficiência acadêmica e o raciocínio clínico: perfis funcionais do uso da inteligência artificial em estudantes de medicina

Resumo

Introdução: A inteligência artificial generativa tem vindo a ser rapidamente incorporada na educação médica, especialmente em tarefas de estudo e produtividade académica. No entanto, continua a ser incerto se a sua utilização se estende de forma equivalente a processos de maior complexidade, como o raciocínio clínico. **Objetivo:** Caracterizar e comparar os perfis funcionais de utilização da inteligência artificial orientados para a eficiência académica e para o raciocínio clínico em estudantes de medicina. **Método:** Estudo transversal em 290 estudantes de medicina. Foi aplicado um questionário autoaplicável e foram construídos dois índices proporcionais de utilização da IA: um orientado para a eficiência académica e outro para o raciocínio clínico. Foram realizadas análises descritivas, o teste t para amostras emparelhadas e a ANOVA de Welch com comparações post hoc. **Resultados:** A idade média foi de $20,4 \pm 2,66$ anos e a média académica de $3,99 \pm 0,54$. 71,3% relataram elevada frequência de utilização da IA. O uso orientado para a eficiência académica foi superior ao orientado para o raciocínio clínico ($0,66 \pm 0,22$ vs. $0,46 \pm 0,34$), com uma diferença média de 0,20 (IC95%:

0,17–0,24; $p < 0,001$). Foram observadas diferenças de acordo com o semestre letivo.

Conclusões: O uso da IA em estudantes de medicina é orientado principalmente para a eficiência acadêmica, com variações de acordo com a fase da formação.

Palavras-chave: Inteligência artificial; educação médica; estudantes de medicina; raciocínio clínico.

Introducción

La expansión de los modelos de lenguaje de gran escala y sus interfaces conversacionales ha cambiado en pocos años la forma de estudiar, redactar y resolver dudas en educación médica. Revisiones recientes describen oportunidades relevantes en el aprendizaje autodirigido, generación de materiales, apoyo a escritura e investigación, pero también riesgos formativos que pueden interferir en la validez de la evaluación y la integridad académica si no se enseñan límites, verificación y declaración transparente de uso. Un factor que aceleró el interés fue el desempeño de ChatGPT en tareas de examen: se reportó rendimiento cercano al umbral de aprobación en el USMLE [1], junto con explicaciones que “suenan” coherentes, aunque no siempre sean correctas, lo que refuerza su atractivo como tutor informal [2].

En el panorama internacional, la evidencia empírica basada en encuestas transversales sugiere adopción amplia y un uso predominantemente orientado a la productividad académica. En Estados Unidos se documentó el uso frecuente y preferencia por ChatGPT frente a algunos recursos tradicionales en escenarios específicos, motivado por el ahorro de tiempo y la clarificación de contenidos [3]. En otros entornos académicos, los estudiantes reportan utilidad para explicar conceptos, resumir y apoyar tareas escritas, junto con preocupaciones recurrentes por desinformación, plagio y falta de formación formal en IA y ética; además, se observa discrepancia entre el uso real de herramientas tipo chat y la enseñanza estructurada de IA/ética en el currículo [4,5]. La adopción también se describe en poblaciones

fuera de Norteamérica, con patrones de utilidad académica coexistiendo con inquietudes por confiabilidad y prácticas indebidas [6].

El punto crítico para la educación médica no es si la IA se usa, sino cómo se integra funcionalmente según el tipo de tarea y el nivel de riesgo. La ventaja de la IA generativa para tareas académicas (explicar, resumir, generar preguntas) no se traduce automáticamente a actividades que exigen razonamiento clínico con incertidumbre, priorización y toma de decisiones [7]. En un ensayo clínico aleatorizado, el acceso a un modelo de lenguaje a gran escala como apoyo diagnóstico no mejoró significativamente el desempeño de razonamiento diagnóstico de médicos frente a recursos convencionales, lo que sugiere que la simple disponibilidad de la herramienta no garantiza mejor proceso cognitivo ni mejores resultados [8]. En evaluaciones sobre casos complejos, se ha observado que modelos como GPT-4 pueden incluir el diagnóstico correcto en una parte importante de los diferenciales, pero con fallas frecuentes al priorizar el diagnóstico más probable, lo que limita su papel a apoyo complementario y no sustituto del juicio clínico [9]. Además, evaluaciones recientes que desagregan “etapas del cuidado” indican variabilidad del desempeño entre la recomendación de exámenes, la decisión diagnóstica y la planificación terapéutica, con degradación en tareas de mayor complejidad y consecuencias potenciales si el usuario sobreconfía en salidas plausibles [10].

En educación médica se ha planteado que el uso temprano y no guiado de chatbots generativos puede funcionar como un precursor de automatización, al entrenar hábitos de de-

legación cognitiva sin estándares explícitos de verificación, trazabilidad y límites de uso [11]. Además, el debate sobre integridad académica se intensificó porque la IA generativa facilita la producción de texto con apariencia experta y aumenta el riesgo de plagio, elaboración de referencias y externalización opaca del trabajo intelectual; por ello, se ha advertido que los errores y la invención de información pueden erosionar el profesionalismo y la integridad si no se construyen barreras educativas y normativas [12]. Estas preocupaciones se amplifican en contextos con menor capacidad institucional para políticas de evaluación y supervisión, donde se han descrito tensiones adicionales por inequidad, privacidad y sobredependencia [13].

La respuesta educativa más robusta no es la prohibición general, sino generar alfabetización en IA y rediseño curricular basado en competencias. En este sentido, se han propuesto dominios y competencias para el uso efectivo y ético de herramientas de IA en salud (conocimiento básico, implicaciones sociales, éticas, evaluación basada en evidencia, integración en flujos de trabajo y aprendizaje continuo), útiles como marco para la enseñanza y la evaluación [14]. También se han desarrollado currículos mediante consenso, por ejemplo, con recursos como el método Delphi, para establecer contenidos esenciales de IA en pregrados médicos [15,7,16].

A pesar de la rápida adopción de la IA, aún no se comprende cómo su uso varía entre el simple apoyo administrativo y el desarrollo del juicio clínico. Esta brecha de conocimiento es crítica, ya que el uso no supervisado podría fomentar una dependencia cognitiva prematura. Por lo tanto, este estudio tiene como objetivo caracterizar y comparar los perfiles funcionales de uso de IA orientados a la eficiencia académica frente al razonamiento clínico en estudiantes de medicina. Los hallazgos buscan orientar la integración curricular de estas herramientas bajo un marco de uso responsable y profesional.

Materiales y método

Se utilizó un muestreo no probabilístico por conveniencia. La invitación a participar se realizó a través de canales institucionales y/o redes de comunicación estudiantil, compartiendo un enlace al cuestionario. La participación fue anónima y no se ofrecieron incentivos.

Instrumento y variables

Se aplicó un cuestionario autoadministrado diseñado para evaluar.

Construcción de índices de uso de IA

Con el fin de caracterizar los perfiles funcionales de uso de IA, se construyeron dos índices compuestos a partir de las actividades académicas reportadas por los estudiantes.

El índice de uso orientado a la eficiencia académica se construyó sumando seis actividades relacionadas con el apoyo instrumental al estudio: explicar temas no comprendidos, resumir textos o guías, resolver preguntas tipo examen, ordenar ideas de un texto, generar apuntes y preparar presentaciones.

El índice de uso orientado al razonamiento clínico se construyó sumando cuatro actividades relacionadas con procesos clínicos: elaboración de historias clínicas, análisis de historias clínicas, discusión de casos clínicos y propuesta de tratamiento.

Cada actividad fue codificada de forma dicotómica (1 = uso reportado; 0 = no uso). El valor de cada índice correspondió a la suma de las actividades reportadas por cada estudiante. Debido a que ambos índices estaban compuestos por diferente número de actividades, los puntajes fueron estandarizados dividiendo el total obtenido entre el número máximo posible de actividades de cada escala. De esta manera, se obtuvieron proporciones de uso con rango entre 0 y 1, lo que permitió realizar comparaciones directas entre ambos perfiles.

Con el fin de evaluar la orientación relativa del uso de IA entre actividades académicas y clínicas, se construyó un índice de orientación del uso de IA, calculado como la diferencia entre el puntaje proporcional del índice de uso orientado a eficiencia académica y el puntaje proporcional del índice de uso orientado a razonamiento clínico. Este índice presentó un rango teórico entre -1 y $+1$. Los valores positivos indican una mayor orientación hacia usos académicos de la IA, los valores cercanos a cero reflejan un uso relativamente equilibrado entre actividades académicas y clínicas, y los valores negativos indican una mayor orientación hacia actividades relacionadas con razonamiento clínico.

Procedimiento de recolección de datos

Los datos se recolectaron mediante un formulario electrónico autoaplicado. Antes de iniciar el cuestionario se presentó información sobre el objetivo del estudio, voluntariedad de la participación y confidencialidad de los datos. El consentimiento informado se obtuvo de manera electrónica.

Consideraciones éticas

El estudio se condujo conforme a los principios éticos para investigación con seres humanos. La participación fue voluntaria, anónima y sin recolección de datos identificables. El protocolo se clasificó como investigación sin riesgo por lo que no se solicitó aval del Comité de Ética.

Análisis estadístico

Se realizó un análisis descriptivo de las variables de interés. Las variables continuas se resumieron mediante medias y desviaciones estándar, y las variables categóricas mediante frecuencias absolutas y relativas.

La distribución de los índices de uso orientado a la eficiencia académica y al razonamiento clínico, se exploró gráficamente mediante dia-

gramas de violín. Para comparar ambos perfiles de uso dentro de los mismos participantes, se empleó una prueba t para muestras pareadas, utilizando los puntajes proporcionales de ambos índices. Se estimó la diferencia media, su intervalo de confianza del 95 % (IC 95 %) y el tamaño del efecto mediante d de Cohen.

Posteriormente, para evaluar si la orientación del uso de IA variaba según el semestre académico, se comparó el índice de orientación entre grupos mediante ANOVA de Welch, considerando la posible heterogeneidad de varianzas entre semestres. Las comparaciones múltiples se realizaron mediante la prueba post hoc de Games-Howell. Se consideró estadísticamente significativo un valor de $p < 0,05$.

Los análisis se realizaron utilizando el software Jamovi (versión 2.6.4).

Resultados

Características generales de la muestra

Se incluyeron 290 estudiantes de medicina. La edad media fue de $20,4 \pm 2,66$ años, y el promedio académico global fue de $3,99 \pm 0,54$ en escala de 1 a 5. Entre los estudiantes que reportaron uso de herramientas de IA con fines académicos, el 71,3 % presentó alta frecuencia de uso.

En términos de los perfiles funcionales construidos, la proporción media de uso de IA orientado a la eficiencia académica fue de $0,66 \pm 0,22$, mientras que la proporción media de uso orientado al razonamiento clínico fue de $0,46 \pm 0,34$ (Tabla 1).

Distribución de los perfiles de uso de IA

La distribución de los índices de uso de IA orientados a la eficiencia académica y al razonamiento clínico se presentan en la figuras 1 y 2, respectivamente. El índice de uso orientado a la eficiencia académica mostró

una mayor concentración de valores en rangos intermedios-altos, mientras que el índice de uso orientado al razonamiento clínico presentó una distribución más amplia y una mayor dispersión de valores.

Tabla 1. Características generales de la muestra y perfiles proporcionales de uso de IA.

Variable	Resultado
Edad (media $\pm$ DE)	20,4 $\pm$ 2,66
Promedio académico (media $\pm$ DE)	3,99 $\pm$ 0,54
Alta frecuencia de uso de IA, n (%)	71,3%
Uso orientado a eficiencia académica (proporcional), media $\pm$ DE	0,66 $\pm$ 0,22
Uso orientado a razonamiento clínico (proporcional), media $\pm$ DE	0,46 $\pm$ 0,34

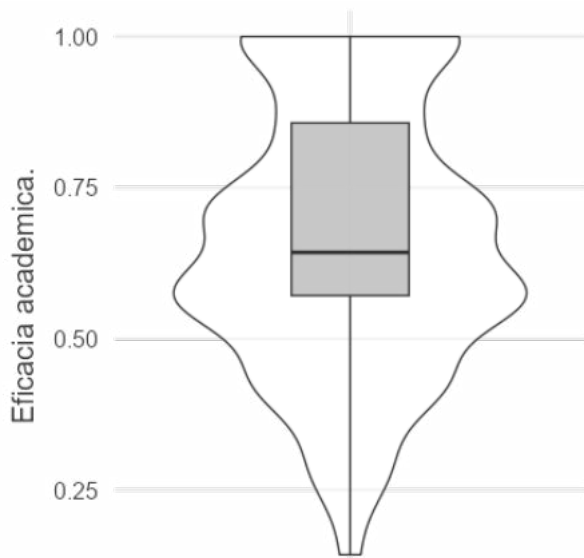


Figura 1. Distribución del índice de uso de IA orientado a la eficiencia académica.

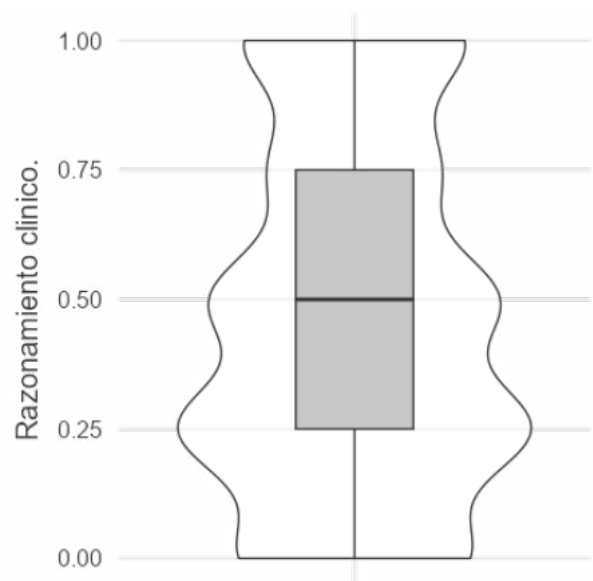


Figura 2. Distribución del índice de uso de IA orientado al razonamiento clínico.

Comparación entre uso académico y uso clínico de IA

La proporción media de uso de IA orientada a LA eficiencia académica fue significativamente mayor que la relacionada con actividades de razonamiento clínico. La diferencia media entre ambos índices fue de 0,20 (IC 95 %: 0,17–0,24), resultado que fue estadísticamente significativo ($t(289) = 11,5$; $p < 0,001$). El tamaño del efecto fue moderado-alto (d de Cohen = 0,68), lo que indica una predominancia consistente del uso instrumental de la IA para tareas académicas en comparación con su utilización en actividades vinculadas al razonamiento clínico (ver Tabla 2).

Tabla 2. Comparación entre perfiles de uso de IA orientados a la eficiencia académica y al razonamiento clínico.

Perfil de uso de IA	Media	DE	Rango
Uso orientado a la eficiencia académica (proporcional)	0,664	0,219	0,14–1,00
Uso orientado al razonamiento clínico (proporcional)	0,461	0,339	0,00–1,00

Orientación del uso de IA según semestre académico

La variación del índice de orientación del uso de IA según el semestre académico se presenta en la Figura 3. En todos los grupos, el índice presentó valores positivos, lo que indica un predominio general del uso de IA orientado a la eficiencia académica sobre su utilización en actividades vinculadas al razonamiento clínico.

No obstante, se observaron diferencias significativas entre semestres (ANOVA de Welch: $F(4,126) = 3,50$; $p = 0,010$). En particular, los estudiantes de 7–8 semestre mostraron valores significativamente menores del índice en comparación con los de 5–6 semestre (diferencia de medias = 0,19; $p = 0,013$), lo que sugiere una integración relativamente mayor de la IA en tareas relacionadas con el razonamiento clínico en esta etapa de la formación.

En conjunto, estos hallazgos muestran que el uso de IA en estudiantes de medicina se orienta predominantemente hacia tareas de eficiencia académica, aunque este patrón presenta variaciones según la etapa de formación y la exposición a actividades clínicas.

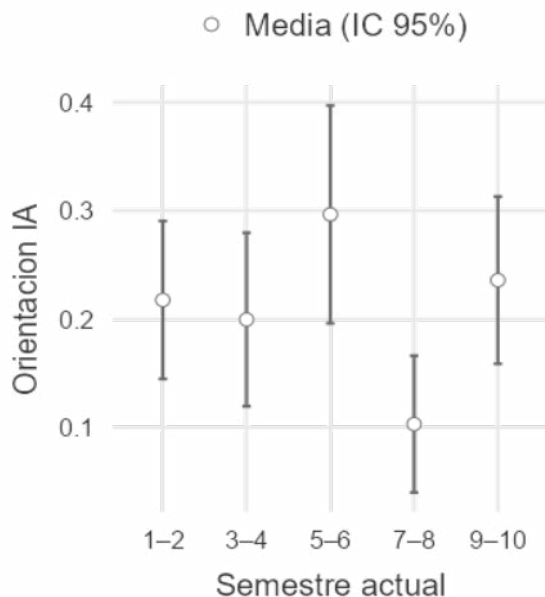


Figura 3. Índice de orientación del uso de IA según el semestre académico.

Discusión

Los hallazgos del presente estudio revelan una adopción predominante de IA orientada a la eficiencia académica sobre el razonamiento clínico en estudiantes de medicina, con un 71,3 % de los participantes reportando una alta frecuencia de uso. Esta marcada inclinación hacia tareas instrumentales y de productividad (0,66 $\pm$ 0,22) coincide con lo reportado por Sallam [17], quien identifica que la facilidad para sintetizar información y la gestión de tiempos son los principales motores de adopción en etapas tempranas de la formación médica. Sin embargo, la brecha significativa hallada entre el uso para la eficiencia y para el razonamiento clínico ($d = 0,68$; $p < 0,001$) sugiere que, aunque la IA está integrada en la rutina estudiantil, su aplicación en procesos cognitivos de orden superior sigue siendo limitada. Esta disparidad podría reflejar una falta de alfabetización en IA específica para el entorno clínico, lo que, según Masters [18], puede derivar en un uso superficial que no capitaliza el potencial de estas herramientas para el diagnóstico o el manejo de pacientes.

Un aspecto importante sería que los estudiantes emplean primero la IA donde obtienen ventajas inmediatas de tiempo y retroalimentación. Darvishi et al. [19] señalan que el uso de IA permite recibir correcciones inmediatas sobre diversos ejercicios, facilitando un proceso de aprendizaje continuo y autónomo al evitar la demora de las revisiones tradicionales. De igual modo, Araujo-Sandoval [20] halló que los estudiantes valoran la IA como herramienta de aprendizaje rápido, pese a cuestionar su confiabilidad. Estos estudios apoyan que la IA se adopta inicialmente para actividades de aprendizaje donde el beneficio es claro, dejando el razonamiento clínico bajo revisión humana crítica.

El razonamiento clínico exige pensamiento crítico y manejo de incertidumbre explícito. Los modelos generativos actuales, si bien pueden

ser convincentes, carecen de validación lógica interna. Estudios recientes lo reflejan. Gilson et al. [21] muestran que ChatGPT no supera consistentemente el examen para certificarse como médico en Estados Unidos, y Song et al. [22] destacan que su falta de razonamiento paso a paso limita la exactitud en casos complejos. En este contexto, los estudiantes perciben o aprenden que desviar el juicio clínico al modelo de IA es menos tolerable, lo que reduce su uso en situaciones diagnósticas exigentes [23].

Un aspecto crítico observado es la transición del perfil de uso según el nivel de formación académica. Los resultados indican que los estudiantes de semestres superiores (7-8) muestran una mayor integración de la IA en tareas de razonamiento clínico en comparación con sus pares de ciclos básicos e intermedios. Este fenómeno es consistente con la teoría de la carga cognitiva aplicada a la educación médica, donde la exposición a la práctica clínica real demanda herramientas de soporte para la toma de decisiones complejas [24]. Al comparar estos datos con investigaciones en contextos globales, se observa una tendencia similar donde la utilidad percibida de la IA muta de un soporte administrativo a uno de asistencia diagnóstica a medida que el estudiante adquiere competencias clínicas fundamentales.

En Latinoamérica, las barreras lingüísticas y de recursos pueden acentuar esta diferencia. Liu et al. [25] encontraron que ChatGPT puede aprobar exámenes médicos a nivel mundial, sugiriendo adaptación lingüística suficiente para tareas de estudio. Sin embargo, también advirtieron gran variabilidad según la especialidad, lo que indica limitaciones en dominios clínicos específicos. Además, en países en desarrollo se observa que la falta de infraestructura y formación en IA tiende a priorizar su uso para preparar exámenes o tareas escritas [26,27], reforzando el sesgo hacia actividades académicas.

El tamaño del efecto observado ($d \approx 0,68$) señala que la brecha no es insignificante. Las instituciones no deberían prohibir el uso de IA,

sino encaminarlo y regularlo pedagógicamente. Se debe fortalecer la alfabetización en IA en el currículo: enseñar a los estudiantes a verificar la información generada, citar correctamente y reconocer sesgos [28,29]. Para fomentar su uso en el razonamiento clínico, hay que diseñar actividades supervisadas; por ejemplo, casos clínicos simulados con prompts estructurados, donde el alumno deba analizar críticamente la respuesta de la IA [30]. Este enfoque ayuda a integrar la IA como complemento formativo, sin suplantar el juicio humano [31].

Estos hallazgos han de interpretarse considerando ciertas limitaciones. Un caso puede ser la categoría uso académico versus uso clínico que puede solaparse en la práctica real, introduciendo subjetividad. Los modelos de IA evolucionan rápido: versiones futuras con entrenamiento en razonamiento médico podrían reducir las diferencias observadas. Los datos provienen de autorreportes de estudiantes, lo que implica posibles sesgos de deseabilidad. Sin embargo, la consistencia de estos resultados en varios estudios sugiere que el patrón es robusto.

Conclusión

En estudiantes de medicina el uso de IA se orienta predominantemente hacia la eficiencia académica más que hacia el razonamiento clínico, con variaciones según la etapa de formación. Estos hallazgos respaldan la necesidad de integrar alfabetización en IA y estrategias curriculares supervisadas que favorezcan un uso crítico, ético y clínicamente pertinente.

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.

Contribución de los autores

LMDA: Conceptualización y diseño del estudio, recolección de datos, análisis estadístico, revisión bibliográfica, redacción del manuscrito y aprobación final.

DAO: Diseño del estudio, análisis estadístico, revisión del manuscrito, redacción y aprobación final.

MALR: Diseño del estudio y del análisis estadístico, revisión del manuscrito, participación en la redacción y aprobación final.

Declaración de uso de IA

Los autores declaran que se emplearon herramientas de Inteligencia Artificial generativa exclusivamente para el apoyo en la revisión gramatical, la corrección de estilo y la optimización de la estructura sintáctica del manuscrito.

Referencias

1. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
2. Preiksaitis C, Rose C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. JMIR Med Educ. 2023;9:e48785. <https://doi.org/10.2196/48785>
3. Zhang JS, Yoon C, Williams DKA, Pinkas A. Exploring the usage of ChatGPT among medical students in the United States. J Med Educ Curric Dev. 2024;11:23821205241264695. <https://doi.org/10.1177/23821205241264695>
4. Weidener L, Fischer M. Artificial intelligence in medicine: cross-sectional study among medical students on application, education, and ethical aspects. JMIR Med Educ. 2024;10:e51247. <https://doi.org/10.2196/51247>
5. Moldt JA, Festl-Wietek T, Madany Mamlouk A, Nieselt K, Fuhl W, Herrmann-Werner A. Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. Med Educ Online. 2023;28(1):2182659. <https://doi.org/10.1080/10872981.2023.2182659>
6. Tangadulrat P, Sono S, Tangtrakulwanich B. Using ChatGPT for clinical practice and medical education: cross-sectional survey of medical students' and physicians' perceptions. JMIR Med Educ. 2023;9:e50658. <https://doi.org/10.2196/50658>
7. Levingston H, Anderson MC, Roni MA. From theory to practice: artificial intelligence (AI) literacy course for first-year medical students. Cureus. 2024;16(10):e70706. <https://doi.org/10.7759/cureus.70706>
8. Goh E, Gallo R, Hom J, He Z, Posada J, Peng L, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open. 2024;7(10):e2440969. <https://doi.org/10.1001/jamanetworkopen.2024.40969>
9. Ríos-Hoyo AR, Shan NL, Li A, Pearson AT, Pusztai L, Howard FM. Evaluation of large language models as a diagnostic aid for complex medical cases. Front Med (Lausanne). 2024;11:1380148. <https://doi.org/10.3389/fmed.2024.1380148>
10. Qiu P, Wu C, Liu S, Feng J, Wang J, Zhang Y, et al. Quantifying the reasoning abilities of LLMs on clinical cases. Nat Commun. 2025;16:9799. <https://doi.org/10.1038/s41467-025-64769-1>
11. Nguyen T. ChatGPT in medical education: a precursor for automation bias? JMIR Med Educ. 2024;10:e50174. <https://doi.org/10.2196/50174>
12. Currie GM. Academic integrity and artificial intelligence: is ChatGPT hype, hero or heresy? Semin Nucl Med. 2023;53(5):719-30. <https://doi.org/10.1053/j.semnuclmed.2023.04.008>
13. Jaleel A, Aziz RA, Farid GF, Bashir MZ. The impact of ChatGPT on academic integrity in medical education: a developing nation perspective. Front Educ. 2025;10:1554444. <https://doi.org/10.3389/educ.2025.1554444>
14. Russell RG, Lovett Novak L, Patel M, Garvey KV, Craig KJT, Jackson GP, et al. Competencies for the use of artificial intelligence-based tools by health care professionals. Acad Med. 2023;98(3):348-356. <https://doi.org/10.1097/ACM.0000000000004963>
15. Singla R, Pupic N, Ghaffarizadeh SA, Kim C, Hu R, Forster BB, et al. Developing a Canadian artificial intelligence medical curriculum using a Delphi study. npj Digit Med. 2024;7:323. <https://doi.org/10.1038/s41746-024-01307-1>
16. Oftring ZS, Deutsch K, Tolks D, Jungmann F, Kuhn S. Female medical students' perception of AI for academic and clinical use: a qualitative interview study. JMIR Med Educ. 2025;11:e65220. <https://doi.org/10.2196/65220>

17. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel). 2023;11(6):887. <https://doi.org/10.3390/healthcare11060887>
18. Masters K. Ethical use of artificial intelligence in health professions education: AMEE Guide N.º 158. Med Teach. 2023;45(6):574-84. <https://doi.org/10.1080/0142159X.2023.2186203>
19. Darvishi A, Khosravi H, Sadiq S, Gašević D, Siemens G. Impacto de la asistencia de IA en la autonomía del estudiante. Comput Educ. 2023;210:104967. <https://doi.org/10.1016/j.compedu.2023.104967>
20. Araujo-Sandoval O. El impacto de la inteligencia artificial en el aprendizaje activo. Diario Horizonte Nexus. 2024;2(4):43. <https://doi.org/10.70881/hnj/v2/n4/43>
21. Gilson A, Safranek C, Huang T, Socrates V, Chi L, Taylor R, et al. ¿Cómo se desempeña ChatGPT en el Examen de Licencia Médica de los Estados Unidos (USMLE)? Las implicaciones de los modelos de lenguaje grandes para la educación médica y la evaluación del conocimiento. JMIR Med Educ. 2023;9:e45312. <https://doi.org/10.2196/45312>
22. Song Z, Yue S, Zhang J. Pensar no es una ilusión: superando las limitaciones de los modelos de razonamiento mediante la ampliación de herramientas. [Preprint]. ArXiv 2025 jul 23 [citado 2026 Abr 13]. Disponible en: <https://doi.org/10.48550/arxiv.2507.17699>
23. Cherrez-Ojeda I, Gallardo-Bastidas JC, Robles-Velasco K, Osorio MF, Vélez León EM, León Velastegui M, et al. Understanding health care students' perceptions, beliefs, and attitudes toward AI-powered language models: cross-sectional study. JMIR Med Educ. 2024;10(1):e51757. <https://doi.org/10.2196/51757>
24. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ. 2023;9(1):e48291. <https://doi.org/10.2196/48291>
25. Liu M, Okuhara T, Chang X, Okada H, Kiuchi T. Rendimiento de ChatGPT en exámenes de licencia médica en países de todo el mundo: una revisión sistemática y protocolo de metaanálisis. PLoS One. 2024;19(10):e0312771. <https://doi.org/10.1371/journal.pone.0312771>
26. Kshetri N. Inteligencia artificial en países en desarrollo. IT Prof. 2020;22(4):63-8. <https://doi.org/10.1109/mitp.2019.2951851>
27. Oftring ZS, Deutsch K, Tols D, Jungmann F, Kuhn S. Novel blended learning on artificial intelligence for medical students: qualitative interview study. JMIR Med Educ. 2025 Jan 20;11:e65220. <https://doi.org/10.2196/65220>
28. Walter Y. Abrazando el futuro de la Inteligencia Artificial en el aula: la relevancia de la alfabetización en IA, la ingeniería de impulsos y el pensamiento crítico en la educación moderna. Int J Educ Technol High Educ. 2024;21(1):1-29. <https://doi.org/10.1186/s41239-024-00448-3>
29. McDonald N, Johri A, Ali A, Collier A. Inteligencia artificial generativa en la educación superior: evidencia de un análisis de políticas y directrices institucionales. Comput Human Behav Artif Hum. 2025;3:100121. <https://doi.org/10.1016/j.chbah.2025.100121>
30. Brügge E, Ricchizzi S, Arenbeck M, Keller MN, Schur L, Stummer W, et al. Large language models improve clinical decision making of medical students through patient simulation and structured feedback: a randomized controlled trial. BMC Med Educ. 2024;24(1):1391. <https://doi.org/10.1186/s12909-024-06399-7>
31. Thesen T, Park SH. A generative AI teaching assistant for personalized learning in medical education. npj Digit Med. 2025;8:627. <https://doi.org/10.1038/s41746-025-02022-1>

