

Inteligencia artificial y razonamiento clínico en educación médica: una revisión sistemática

Artificial intelligence and clinical reasoning in medical education: a systematic review

 Rafael Emiliano Sulca Quispe¹

 Juan Miguel Purilla Janto²

 María del Rosario Jáuregui-Figueroa³

 María del Rosario Figueroa Bejarano De
Jauregui⁴

¹Universidad Nacional
Federico Villarreal. Lima,
Perú
Correo:
rsulca@unfv.edu.pe

²Universidad Nacional
Federico Villarreal. Lima,
Perú
Correo:
jpurilla@unfv.edu.pe

³Universidad Nacional
Federico Villarreal. Lima,
Perú
Correo:
mjaureguif@unfv.edu.pe

⁴Universidad Nacional
Federico Villarreal. Lima,
Perú
Correo:
mfigueroab@unfv.edu.pe

Resumen

Contexto: La inteligencia artificial ha transformado los escenarios de aprendizaje en ciencias de la salud y ha generado un debate sobre su papel en la formación del razonamiento clínico. **Objetivo:** Analizar el papel de la inteligencia artificial en el desarrollo del razonamiento clínico en la educación médica. **Método:** Se realizó una revisión sistemática siguiendo la guía PRISMA 2020. Se seleccionaron 22 publicaciones, incluidos estudios primarios, revisiones y textos de reflexión, publicados entre 2021 y 2026. **Resultados:** El metaanálisis incluido de [Takita et al. \(2025\)](#) reportó una precisión diagnóstica global del 52.1 % en 83 estudios. En [Gordon et al. \(2024\)](#), el 6.8 % (13/191) de los artículos/innovaciones se centró en enseñar o evaluar razonamiento clínico. [Arcos Huaman et al. \(2025\)](#) reportaron un 31 % de barreras regulatorias en diagnóstico cardiovascular latinoamericano, sin evaluar su prevalencia curricular. **Conclusiones:** La inteligencia artificial constituye un recurso pedagógico complementario valioso, cuya integración curricular exige supervisión docente activa, marcos éticos explícitos y validación empírica continua. **Palabras clave:** Educación médica; Formación médica; Inteligencia artificial; Modelos de lenguaje; Razonamiento clínico

Abstract

Context: Artificial intelligence has transformed learning environments in the health sciences and prompted debate about its role in developing clinical reasoning. **Objective:** To analyze the role of artificial intelligence in developing clinical reasoning in medical education. **Methods:** A systematic review followed PRISMA 2020 guidelines. Twenty-two publications, including primary studies, reviews, and reflective papers, published between 2021 and 2026 were selected. **Results:** The included meta-analysis by [Takita et al. \(2025\)](#) reported an overall diagnostic accuracy of 52.1% across 83 studies. In [Gordon et al. \(2024\)](#), 6.8% (13/191) of articles/innovations focused on teaching or assessing clinical reasoning. [Arcos Huaman et al. \(2025\)](#) reported 31% regulatory barriers in Latin American cardiovascular diagnosis, without assessing their curricular prevalence. **Conclusions:** Artificial intelligence is a valuable complementary educational resource whose curricular integration requires active faculty supervision, explicit ethical frameworks, and ongoing empirical validation.

Keywords: Medical education; Medical training; Artificial intelligence; Language models; Clinical reasoning

Cómo citar:

Sulca Quispe, R. E., Purilla Janto, J. M., Jáuregui-Figueroa, M. R., & Figueroa Bejarano de Jauregui, M. R. (2026). Inteligencia artificial y razonamiento clínico en la educación médica: una revisión sistemática. *Revista de Educación Simón Rodríguez*, 6(13), 1-18.
<https://doi.org/10.62319/simonrodriguez.v.6i13.239>

Introducción

El razonamiento clínico constituye la competencia nuclear que la educación médica debe desarrollar en sus estudiantes para transformar datos anamnésticos, hallazgos exploratorios y resultados de exámenes auxiliares en hipótesis diagnósticas pertinentes y decisiones terapéuticas oportunas. Tradicionalmente, la formación de esta competencia se ha sustentado en dos modelos cognitivos complementarios: el razonamiento analítico o hipotético-deductivo, empleado ante casos complejos o poco frecuentes, y el reconocimiento de patrones, consolidado mediante la exposición repetida a presentaciones típicas de la práctica médica. Ambos procesos requieren estrategias pedagógicas deliberadas, práctica supervisada y retroalimentación experta a lo largo del currículo, lo que sitúa al razonamiento clínico como un aprendizaje que se afina progresivamente con la experiencia y la reflexión (Pelaccia et al., 2011).

Desde una perspectiva internacional, el desarrollo acelerado de la inteligencia artificial, particularmente de los modelos de lenguaje de gran escala como ChatGPT, GPT-4, DeepSeek o Gemini, ha renovado el interés de las escuelas de medicina por comprender en qué medida estas herramientas pueden integrarse a los procesos de enseñanza del razonamiento clínico. Además, se ha documentado que estos modelos son capaces de aprobar exámenes de licenciatura médica y de resolver retos diagnósticos complejos con niveles de precisión que, en ciertos escenarios, igualan o superan a médicos no especialistas. Esta evidencia ha sido tomada como punto de partida para explorar su incorporación como recurso didáctico, aunque persiste la incertidumbre sobre su verdadero valor formativo (Joseph et al., 2025).

Las teorías cognitivas que sustentan el razonamiento clínico han evolucionado hacia modelos integradores que reconocen la interacción entre procesos intuitivos y analíticos. La teoría de los dos sistemas distingue entre un Sistema 1, rápido y automático, vinculado al reconocimiento de patrones, y un Sistema 2, lento y deliberado, asociado al razonamiento hipotético-deductivo. La regulación metacognitiva entre ambos sistemas determina la calidad del juicio clínico, así como la susceptibilidad a sesgos cognitivos como el anclaje, la disponibilidad o el sesgo de confirmación, los cuales comprometen la precisión diagnóstica y deben ser objeto de entrenamiento explícito durante la formación médica (Corrao y Argano, 2022).

En el contexto latinoamericano, la incorporación de la inteligencia artificial a los currículos de formación médica enfrenta brechas estructurales de diversa índole. La literatura especializada identifica limitaciones en infraestructura tecnológica, escasa formación docente en competencias digitales, ausencia de marcos regulatorios homogéneos y desigualdades de acceso entre instituciones públicas y privadas. Estas particularidades regionales justifican una revisión que no solo sintetice la evidencia internacional en el ámbito médico, sino que también visibilice el estado de adopción y los desafíos propios de la educación médica en la región (Chávez-Martínez y Ragacini, 2025).

Como antecedente histórico, la revisión de literatura de 2011–2017 examinó tendencias formativas relacionadas con tecnologías avanzadas e inteligencia artificial. La revisión integradora de la literatura identifica propuestas para orientar los programas para incluir

competencias digitales, razonamiento clínico asistido por tecnología y ética de la inteligencia artificial, con el propósito de preparar profesionales capaces de ejercer en entornos altamente digitalizados sin renunciar al componente humano de la atención (Han et al., 2019).

La irrupción de la inteligencia artificial generativa ha ampliado el espectro de oportunidades y desafíos para la educación médica, al habilitar herramientas conversacionales capaces de generar casos clínicos, retroalimentación automatizada y simulaciones de pacientes. Para Preiksaitis y Rose (2023), este campo identifica beneficios potenciales en el aprendizaje personalizado y en la formación del razonamiento clínico, así como riesgos asociados a la precisión, la equidad y la integridad académica, lo que aconseja un abordaje cauto y orientado por marcos éticos explícitos antes de su implementación curricular masiva.

La definición de competencias específicas en inteligencia artificial para el pregrado médico constituye un avance relevante para ordenar la integración curricular. Las propuestas recientes sugieren estructurar el aprendizaje en torno a dominios como el uso seguro de herramientas, la interpretación crítica de resultados, el conocimiento de limitaciones y sesgos, y la valoración ética de las aplicaciones. Estos lineamientos ofrecen a las facultades de medicina un mapa operativo para incorporar la inteligencia artificial sin reducirla a una herramienta instrumental, sino como objeto de aprendizaje y de razonamiento profesional (Kim et al., 2025).

En el caso específico de la región, una revisión de 13 artículos examina tendencias y desafíos de la inteligencia artificial en educación médica (Ruiz Orizaga et al., 2024). La distribución de experiencias entre instituciones públicas y privadas y el alcance de las políticas universitarias regionales requieren documentación específica. Este panorama configura un vacío de conocimiento respecto a las condiciones pedagógicas, éticas y de infraestructura necesarias para que la integración de la inteligencia artificial sea equitativa y pertinente para el contexto latinoamericano. En consecuencia, el presente estudio tuvo como objetivo analizar el papel de la inteligencia artificial en el desarrollo del razonamiento clínico en la educación médica. De ahí que la pregunta que guía el artículo sea: ¿de qué manera la inteligencia artificial influye en el desarrollo del razonamiento clínico en la educación médica?

Método

La presente investigación corresponde a una revisión sistemática desarrollada con referencia a la declaración PRISMA 2020 como guía de reporte (Page et al., 2021), orientada a transparentar la identificación, selección y síntesis de la evidencia científica disponible sobre el uso de la inteligencia artificial en el desarrollo del razonamiento clínico dentro de la educación médica. El enfoque adoptado fue cualitativo y descriptivo, orientado a la integración interpretativa de hallazgos provenientes de estudios primarios, revisiones y textos de reflexión incluidos en la Tabla 1 publicados en el período 2021-2026.

Criterios de selección y bases de datos

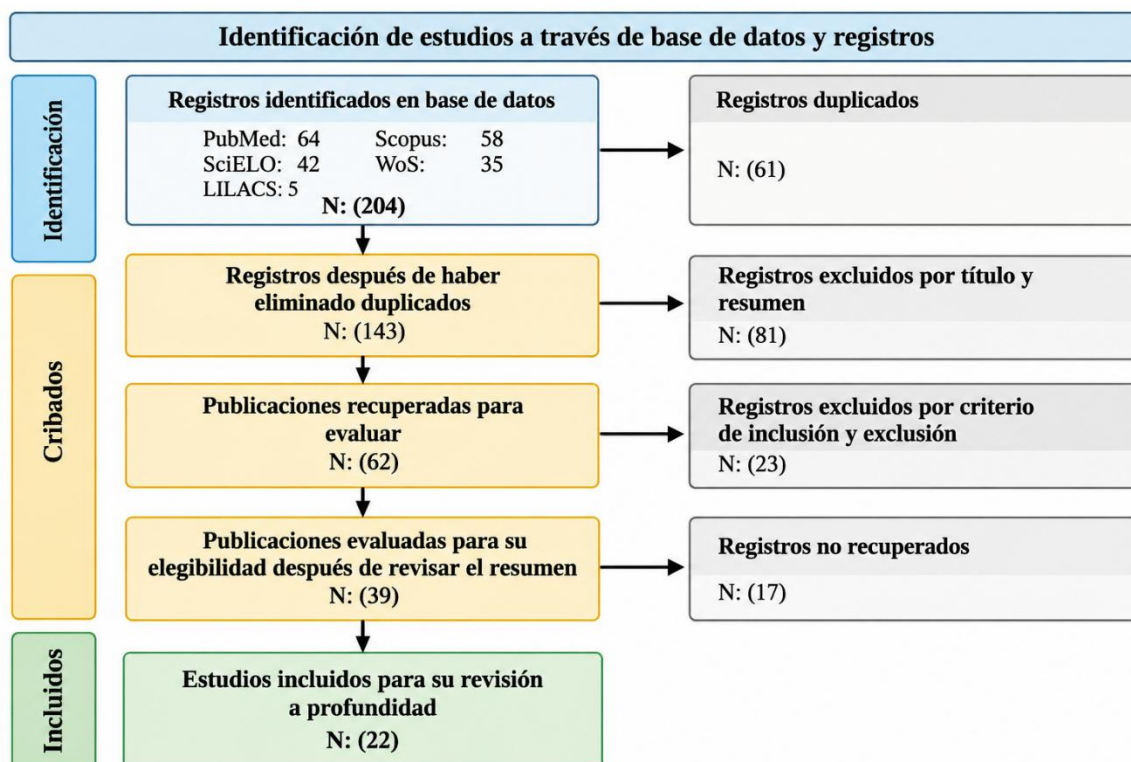
Los criterios de elegibilidad se definieron con el fin de asegurar pertinencia temática, calidad metodológica y actualidad. Se incluyeron estudios empíricos, revisiones sistemáticas,

revisiones de alcance, meta-análisis y estudios comparativos sobre el uso de inteligencia artificial, aprendizaje automático o modelos de lenguaje en la enseñanza, el entrenamiento o la evaluación del razonamiento clínico en la educación médica, publicados entre 2021 y 2026, en idioma español, inglés o portugués, con acceso libre al texto completo. Se declaró la exclusión de textos que no correspondieran a investigaciones empíricas, literatura no indexada y ponencias de eventos académicos sin arbitraje, estudios publicados en idiomas diferentes al español, inglés o portugués, documentos sin acceso libre al texto completo, editoriales, cartas al editor, opiniones sin evidencia empírica y resúmenes de congreso. La presencia de revisiones, reflexiones y actas en la Tabla 1 requiere reconciliar los criterios declarados con el registro real de inclusión, antes de considerar cerrado el corpus.

La búsqueda se realizó en PubMed, Scopus, SciELO, WoS y LILACS, complementada con la revisión de repositorios institucionales de universidades latinoamericanas y de actas de congresos especializados en educación médica. Los descriptores empleados fueron "educación médica", "inteligencia artificial", "razonamiento clínico", "modelos de lenguaje", "medical education", "large language models" y "clinical reasoning", combinados mediante los operadores booleanos AND y OR.

Estrategias de búsqueda y proceso de selección de estudios

El proceso de identificación proporcionó un total de 204 registros. Una vez eliminados 61 duplicados, la suma de registros fue de 143. A continuación, 81 artículos fueron excluidos a través de la lectura del título y el resumen, quedando 62 artículos para evaluación preliminar, y de estos últimos se excluyeron 23 artículos por no cumplir los criterios establecidos. Finalmente, 39 artículos fueron objeto de revisión, de los cuales 17 no pudieron recuperarse en texto completo y solo 22 cumplieron con los requisitos exigidos. Este proceso delimitó el conjunto analizado; la restricción de acceso no garantiza su representatividad y puede introducir sesgo de selección (ver Figura 1).

Figura 1. Diagrama de flujo para la selección de los artículos según PRISMA

Nota. Se conservan las cantidades de la figura original. Los rótulos de recuperación y elegibilidad y los motivos de las 23 exclusiones requieren reconstrucción desde el registro de selección; no se han inferido razones ni cantidades nuevas.

Marco ético y científico

La investigación se enmarca en los principios éticos de honestidad, transparencia y atribución adecuada de las fuentes. Por tratarse de una revisión sistemática de literatura publicada, no se requirió aprobación de comité de ética institucional, en tanto no se trabajó con datos primarios ni con seres humanos. Se respetaron los lineamientos de buena práctica científica en el manejo de citas y referencias, garantizando la trazabilidad de cada afirmación a su fuente original.

Análisis de contenido

El análisis de la información se realizó de forma descriptiva y cualitativa, organizando los hallazgos en cinco ejes temáticos orientados a la educación médica: (a) desempeño diagnóstico de la inteligencia artificial como sustento empírico de su uso formativo; (b) la inteligencia artificial como herramienta pedagógica, incluidos los simuladores virtuales, para el entrenamiento del razonamiento clínico; (c) la inteligencia artificial como apoyo durante la práctica supervisada; (d) riesgos de dependencia, deskilling e interpretabilidad; y (e) particularidades del contexto latinoamericano y peruano en la formación médica. Cada estudio fue codificado según autor, año, país, enfoque, diseño y principales hallazgos, lo que permitió construir las tablas síntesis y articular la discusión.

Resultados

Los estudios seleccionados evidencian un panorama heterogéneo sobre el uso de la inteligencia artificial en la formación del razonamiento clínico. A continuación, se presentan los hallazgos organizados en cinco ejes: desempeño diagnóstico, función pedagógica (incluidos los simuladores virtuales), apoyo en la práctica supervisada, riesgos formativos y contexto latinoamericano. En la Tabla 1 se sintetizan las características descriptivas de cada estudio.

Tabla 1. Descripción de los estudios seleccionados según autor, año, país, enfoque y revista

No	Autor (año)	País	Enfoque y tipo documental	Revista o actas
1	Arcos Huaman et al. (2025)	Perú	Revisión sistemática	Actas LACCEI
2	Cuestas (2023)	Argentina	Reflexión	Revista de la Facultad de Ciencias Médicas de Córdoba
3	Gaebe y van der Woerd (2025)	Internacional	Cuantitativo	PLoS ONE
4	Goh et al. (2024)	Estados Unidos	Cuantitativo	JAMA Network Open
5	Gonzales Aste (2025)	Perú	Reflexión clínica	Respirar
6	Hager et al. (2024)	Alemania	Cuantitativo	Nature Medicine
7	Kanjee et al. (2023)	Estados Unidos	Cuantitativo	JAMA
8	Gordon et al. (2024)	Internacional	Cualitativo	Medical Teacher
9	Wei et al. (2025)	Internacional	Revisión sistemática y metaanálisis	International Dental Journal
10	Savage et al. (2024)	Internacional	Cuantitativo	NPJ Digital Medicine
11	Ahsan (2025)	Internacional	Cualitativo	BMC Medical Education
12	Turney et al. (2026)	Internacional	Cualitativo	Advances in Medical Education and Practice
13	Takita et al. (2025)	Internacional	Cuantitativo	NPJ Digital Medicine
14	Siam et al. (2026)	EE. UU./ Asia/ Emiratos	Cuantitativo	Scientific Reports
15	Strong et al. (2023)	Estados Unidos	Cuantitativo	JAMA Internal Medicine
16	Zhong et al. (2025)	China	Cuantitativo	JMIR
17	Hernández Rincón et al. (2025)	Internacional	Cualitativo	BMC Medical Education
18	Shan et al. (2025)	China	Cuantitativo	JMIR Medical Informatics
19	Çiçek et al. (2025)	Turquía/ Internacional	Cuantitativo	Postgraduate Medical Journal

No	Autor (año)	País	Enfoque y tipo documental	Revista o actas
20	García-Torres et al. (2025)	España	Observacional transversal	JMIR Medical Education
21	Peralta Ramírez et al. (2025)	México	Mixto	Journal of CME
22	Sindhvananda et al. (2026)	Tailandia	Mixto	BMC Medical Education

Desempeño diagnóstico de la inteligencia artificial como sustento de su uso formativo

Los hallazgos confirman que los modelos de lenguaje de gran escala alcanzan niveles notables de precisión en tareas de conocimiento médico estructurado, lo que constituye el principal argumento empírico para explorar su incorporación en la enseñanza del razonamiento clínico. En este sentido al evaluarse cuatro modelos avanzados frente a preguntas del examen de licenciatura médica de los Estados Unidos, se encontró que dos de ellos superaron consistentemente a los demás, con DeepSeek alcanzando un 93 % de aciertos en Step 2 CK ([Siam et al., 2026](#)). El uso de preguntas públicas obliga a considerar su posible exposición durante el entrenamiento. Sin embargo, un meta-análisis reciente que comparó el desempeño diagnóstico de la inteligencia artificial generativa frente a médicos en 83 estudios reportó una precisión diagnóstica global de apenas el 52.1 %, sin diferencias estadísticamente significativas frente a médicos no expertos ($p = .93$), pero con menor desempeño frente a expertos ($p = .007$). La ausencia de significancia frente a no expertos no demuestra equivalencia, lo que matiza el entusiasmo inicial sobre su idoneidad como modelo de referencia para la enseñanza ([Takita et al., 2025](#)).

Otro estudio comparativo evaluó el diagnóstico de 121 enfermedades raras y mostró variaciones según el modelo, el idioma y el uso de recuperación aumentada y cadenas de razonamiento. Esta distinción resulta pedagógicamente relevante, pues advierte que aprobar un examen de opción múltiple no equivale a razonar frente a un caso real y que la evaluación en la formación médica no puede sustentarse únicamente en este tipo de pruebas estandarizadas ([Zhong et al., 2025](#)). Asimismo, una revisión sistemática con meta-análisis que comparó la precisión diagnóstica entre profesionales clínicos y modelos de lenguaje identificó que los modelos superaron a los humanos en 10 de 30 estudios (33.3 %). Esta proporción no demuestra por sí misma una ventaja humana específica en enfermedades raras ([Shan et al., 2025](#)).

La inteligencia artificial como herramienta pedagógica en la enseñanza del razonamiento clínico

Un cuerpo creciente de evidencia explora el papel de la inteligencia artificial como recurso pedagógico para los estudiantes de ciencias de la salud, como es el caso de la revisión de [Gordon et al. \(2024\)](#), que identificó 191 artículos/innovaciones dentro de 278 publicaciones sobre inteligencia artificial en educación médica e identificó que apenas 13 de ellos empleaban específicamente la inteligencia artificial para enseñar o evaluar el razonamiento clínico,

mediante simulaciones de casos, sistemas de retroalimentación automatizada y plataformas conversacionales tipo chatbot.

A su vez [Ahsan \(2025\)](#), concluye que estas herramientas funcionan como un andamiaje pedagógico potente para el rediseño curricular y la evaluación de competencias, aunque su integración no debe prescindir de la supervisión docente ni de la retroalimentación de profesionales experimentados. En la misma línea, un ensayo aleatorizado comparó la retroalimentación generada por un modelo de lenguaje con la retroalimentación experta para mejorar el razonamiento clínico de estudiantes de medicina, con 129 participantes y sin diferencias globales significativas inmediatas ni a diez días, aunque con ventaja de la retroalimentación experta en casos complicados; estos resultados no demuestran equivalencia formal ni beneficios duraderos ([Çiçek et al., 2025](#)).

La inteligencia artificial como apoyo durante la práctica supervisada

Un hallazgo relevante de esta revisión es que la disponibilidad de un modelo de lenguaje como herramienta de apoyo durante el proceso diagnóstico no garantiza, por sí sola, una mejora diagnóstica de médicos residentes y tratantes. El ensayo aleatorizado con viñetas clínicas realizado con médicos no encontró diferencias significativas en el diagnóstico de los profesionales que contaban con acceso a un modelo de lenguaje frente a quienes utilizaban recursos convencionales de búsqueda ([Goh et al., 2024](#)). Estos resultados sugieren que el principal cuello de botella no reside únicamente en la capacidad técnica del algoritmo, sino en la interacción humano-máquina: la forma en que los profesionales formulan consultas, interpretan respuestas y las integran en su proceso deliberativo. La transferencia de este hallazgo al aprendizaje de estudiantes de pregrado requiere evaluación específica.

La evidencia sobre estrategias de formulación de instrucciones indica que un uso más estructurado de la herramienta puede mejorar la precisión técnica de las respuestas obtenidas, sin medir directamente adquisición de razonamiento por estudiantes ([Gaebe y van der Woerd, 2025](#)), lo que reafirma la necesidad de que los programas de educación médica formen a sus estudiantes no solo en el uso de la inteligencia artificial, sino en su uso crítico y metodológicamente adecuado.

Riesgos formativos: dependencia, deskilling e interpretabilidad

Un eje crítico de los hallazgos se refiere al riesgo de que la exposición temprana y no supervisada a sistemas de inteligencia artificial reduzca la práctica activa del razonamiento clínico independiente en los estudiantes, fenómeno descrito en la literatura como deskilling o inhibición del desarrollo de habilidades. Una revisión sistemática reciente advierte que este efecto podría comprometer la capacidad de juicio profesional autónomo de los futuros médicos si los programas de formación no establecen estrategias pedagógicas que equilibren el uso de la tecnología con la práctica deliberada y supervisada ([Turney et al., 2026](#)).

A este riesgo se suma la limitación persistente de interpretabilidad de los modelos de lenguaje, frecuentemente descritos como cajas negras cuyo proceso interno de decisión no resulta transparente para el docente ni para el estudiante. También se evaluaron las limitaciones de los modelos de lenguaje en la toma de decisiones sobre casos y datos clínicos y propusieron

estrategias de mitigación centradas en la validación externa y la supervisión experta, aplicables al contexto formativo (Hager et al., 2024). En contraste, se ha demostrado que es posible inducir a un modelo avanzado, mediante instrucciones específicas, a generar explicaciones de sus respuestas; estas explicaciones no demuestran acceso fiel a su proceso interno, lo que abre una vía prometedora para que los docentes utilicen estas explicaciones como material didáctico y para que los estudiantes aprendan a evaluar la plausibilidad de las respuestas generadas (Savage et al., 2024).

Simuladores virtuales y aprendizaje basado en casos

El desarrollo de plataformas de pacientes virtuales impulsadas por inteligencia artificial ha abierto nuevas oportunidades para el entrenamiento del razonamiento clínico. Un estudio observacional transversal realizado con 156 estudiantes de psicología evaluó la percepción de interacciones con pacientes virtuales generados por inteligencia artificial y encontró que las variables del sistema, como el parámetro de temperatura del modelo, y las características de la interacción influyeron en la valoración de la experiencia, con implicaciones para el diseño de escenarios formativos (García-Torres et al., 2025). La mejora percibida no se acompañó de una diferencia significativa en el examen práctico frente a la cohorte anterior; su pertinencia para educación médica es indirecta. En la misma línea, un estudio sobre la simulación clínica con inteligencia artificial generativa exploró la creación de escenarios interactivos y la valoración estudiantil, evidenciando percepciones favorables hacia esta herramienta como complemento del entrenamiento tradicional (Peralta Ramírez et al., 2025).

La incorporación de pacientes simulados con retroalimentación automatizada se ha extendido a dominios clínicos específicos. Un estudio mixto evaluó una plataforma de paciente simulado con retroalimentación automática para el entrenamiento de la anamnesis en cefalea, con 25 estudiantes en Tailandia y mostrando mejoras en OSCE y CRI-HT en un diseño pre-post de grupo único; no se demostró superioridad frente a un control ni retención prolongada (Sindhvananda et al., 2026). La revisión sistemática sobre aprendizaje basado en problemas y casos con inteligencia artificial sintetizó evidencia sobre su efectividad en la adquisición de competencias clínicas, especialmente en etapas tempranas e intermedias de la formación, sin sustituir la función docente de supervisión (Wei et al., 2025).

El contexto latinoamericano y peruano en la formación médica

La incorporación de la inteligencia artificial a los currículos de formación médica en Latinoamérica presenta particularidades que no siempre son visibles en la literatura predominantemente anglosajona. Una revisión sistemática centrada en el diagnóstico de enfermedades cardiovasculares en la región identificó como principales obstáculos las barreras regulatorias, la falta de personal capacitado, los altos costos de implementación, la dificultad de validación clínica y la infraestructura insuficiente, factores del diagnóstico cardiovascular cuya posible repercusión docente requiere evaluación específica, además de señalar inequidades de acceso especialmente marcadas en zonas rurales (Arcos Huaman et al., 2025), publicado en actas de congreso.

En el caso peruano específico, se plantea, a propósito de la especialidad de neumología, un marco ético de siete principios para orientar la adopción de la inteligencia artificial en el sistema de salud y en la formación médica nacional, enfatizando la necesidad de equidad y contextualización para que los beneficios de estas tecnologías no se concentren únicamente en instituciones privadas de alta complejidad ni en facultades con mayores recursos ([Gonzales Aste, 2025](#)). La posibilidad de que las brechas tecnológicas se traduzcan en diferencias formativas entre facultades públicas y privadas constituye un riesgo potencial; esta reflexión clínica no evalúa comparativamente su adopción curricular.

Otros hallazgos de relevancia y respaldo

Según [Gordon et al. \(2024\)](#), de los 191 artículos/innovaciones sobre inteligencia artificial en educación médica identificados dentro de 278 publicaciones, solo una minoría se orientó a la enseñanza del razonamiento, por consiguiente, existe una brecha entre la disponibilidad tecnológica y su uso formativo específico. Por otra parte, es necesario distinguir entre el razonamiento clínico hipotético-deductivo humano y el procesamiento estadístico de la inteligencia artificial; es decir, la enseñanza del razonamiento humano no es equiparable al procesamiento algorítmico automatizado ([Cuestas, 2023](#)). En cuanto al planteamiento de un marco ético de principios para la adopción de la inteligencia artificial en la práctica y la formación médica peruana, la equidad y la contextualización constituyen condiciones indispensables para una adopción ética ([Gonzales Aste, 2025](#)), como propuesta clínica de neumología cuya aplicación docente es interpretativa.

La inteligencia artificial se integra en la educación médica para reforzar el razonamiento clínico, rediseñar los currículos y mejorar la evaluación; su incorporación curricular debe orientarse por criterios pedagógicos y éticos ([Ahsan, 2025](#)). Asimismo, se ha reportado el desempeño superior de un modelo avanzado en el diagnóstico de enfermedades raras, aunque con variabilidad según el idioma, por ello, los currículos multilingües deben considerar la variabilidad del desempeño de los modelos ([Zhong et al., 2025](#)).

En síntesis, la evidencia muestra que la inteligencia artificial alcanza alta precisión en conocimientos estructurados, pero su desempeño diagnóstico global es moderado, sin superioridad demostrada frente a médicos no expertos y con menor desempeño frente a expertos. Además, su uso pedagógico es prometedor mediante simuladores, chatbots y retroalimentación automatizada, aunque solo una pequeña minoría de artículos/innovaciones de la revisión de [Gordon et al. \(2024\)](#) abordó específicamente el razonamiento clínico. También persisten riesgos de dependencia, deskilling y opacidad interpretativa, donde la interacción humano-máquina y la formación en uso crítico resultan claves. En Latinoamérica y específicamente en Perú, barreras regulatorias, costos e inequidad limitan su adopción formativa equitativa.

Discusión

Los hallazgos de esta revisión muestran un cuadro matizado respecto al desempeño diagnóstico de la inteligencia artificial, pues algunos modelos alcanzan precisiones elevadas

en exámenes estandarizados, su rendimiento global en viñetas y evaluaciones retrospectivas no mostró diferencias significativas frente a médicos no expertos, aunque fue inferior al de expertos; estos resultados no equivalen a evidencia prospectiva de práctica clínica. Esta paradoja puede explicarse por la heterogeneidad metodológica de las evaluaciones y por la naturaleza misma del razonamiento clínico, que combina conocimiento factual con integración contextual, juicio probabilístico y consideración de preferencias del paciente. La evidencia disponible sugiere que la inteligencia artificial generativa ofrece un rendimiento competitivo en tareas acotadas, pero no reemplaza la complejidad cognitiva del razonamiento clínico experto (Takita et al., 2025).

La comparación con la práctica profesional humana revela además que puede existir un riesgo de sobreestimación de la precisión, asociado al sesgo de automatización. La revisión de barreras de implementación sanitaria no mide directamente aceptación acrítica por estudiantes. Esta tendencia es especialmente preocupante en formación médica, donde la verificación crítica constituye una dimensión central del juicio clínico. Las estrategias de mitigación propuestas incluyen la incorporación de mecanismos de transparencia, la supervisión experta y la validación externa de las respuestas generadas antes de su uso en decisiones formativas o clínicas (Ahmed et al., 2023).

Desde una perspectiva pedagógica, la integración de la inteligencia artificial en la educación médica exige superar la visión instrumental de la herramienta y avanzar hacia un enfoque centrado en el aprendizaje. La interpretación pedagógica de los estudios revisados sugiere que el valor formativo no reside en la mera exposición a la tecnología, sino en el diseño de experiencias de aprendizaje que articulen simulación, retroalimentación y reflexión metacognitiva. La teoría del andamiaje, ampliamente utilizada en educación médica, encuentra en la inteligencia artificial un recurso potencial para personalizar el aprendizaje, aunque su uso eficaz requiere formación docente específica y currículos explícitamente orientados a este fin (Borg et al., 2025). Esta prueba de concepto con 15 estudiantes en Suecia aporta percepciones sobre simulación robótica, sin demostrar una mejora general de aprendizaje ni validar por sí misma el andamiaje.

Además, la evidencia empírica muestra que la inteligencia artificial puede cumplir funciones complementarias en la formación del razonamiento clínico, como la generación de casos virtuales, la práctica de la anamnesis, el entrenamiento del razonamiento diferencial y la evaluación formativa. Sin embargo, la simple disponibilidad de estas herramientas no se traduce automáticamente en aprendizajes significativos; se requiere intencionalidad pedagógica, integración curricular coherente y desarrollo de competencias docentes para acompañar su uso. La literatura especializada advierte que sin esta intencionalidad, la herramienta puede quedar subutilizada o, peor aún, convertirse en un sustituto acrítico del esfuerzo cognitivo del estudiante (Kang y Ahn, 2025).

Uno de los hallazgos más preocupantes de la revisión es el riesgo de deskilling o inhibición del desarrollo de habilidades diagnósticas propias en estudiantes que emplean inteligencia artificial de manera temprana y no supervisada. Este fenómeno, ampliamente descrito en la literatura sobre tecnología y cognición, cobra especial relevancia en la formación médica,

donde la automatización de tareas cognitivas puede debilitar el desarrollo del razonamiento clínico autónomo. La revisión teórica recomienda prevenir este riesgo cuando la herramienta se incorpora con supervisión experta y como complemento, no como sustituto, de la práctica deliberada (Bin Abdulrahman y Hefny, 2026), sin constituir una prueba longitudinal de pérdida de habilidades.

El concepto de descarga cognitiva, emparentado con el deskilling, describe la tendencia a transferir a la tecnología parte del esfuerzo mental necesario para resolver problemas. En educación médica, esto puede traducirse en una menor elaboración de las hipótesis diagnósticas, una reducción del razonamiento diferencial y una pérdida progresiva de la capacidad de juicio clínico independiente. Como recomendación pedagógica por evaluar, se propone, en consecuencia, que los currículos equilibren el uso de inteligencia artificial con prácticas que exijan el ejercicio activo del razonamiento, como los exámenes de respuesta libre, las discusiones clínicas supervisadas y la resolución de casos sin acceso a la herramienta. El estudio de Fehr et al. (2022) examina transparencia y confiabilidad de herramientas médicas, no la eficacia curricular de estas estrategias.

La limitada interpretabilidad de los modelos de lenguaje constituye un desafío central para su incorporación educativa. La metáfora de la caja negra describe la opacidad del proceso interno mediante el cual estos modelos generan sus respuestas, lo que dificulta su empleo como modelo explícito de razonamiento clínico para los estudiantes. Sin embargo, se muestra que estrategias como la formulación de instrucciones específicas pueden inducir explicaciones textuales de las respuestas, sin probar fidelidad al proceso interno, abriendo una vía para utilizar estas explicaciones como material didáctico y como objeto de evaluación crítica por parte de los estudiantes (Savage et al., 2024). El análisis filosófico de Da Silva (2023) aporta una perspectiva distinta sobre explicabilidad y justificación pública.

La transparencia algorítmica no es solo un imperativo técnico, sino un requisito ético y pedagógico. Los docentes deben ayudar a sus estudiantes a evaluar críticamente las explicaciones generadas y su coherencia clínica, qué limitaciones presenta y en qué casos su desempeño puede ser engañoso. La literatura propone que la formación médica incorpore la evaluación crítica de las explicaciones generadas por la inteligencia artificial como una competencia transversal, lo que contribuiría a desarrollar una actitud reflexiva ante la tecnología y a reducir el riesgo de aceptación acrítica (Kıyak et al., 2025), cuyo estudio evalúa retroalimentación quirúrgica automatizada, sin acreditar transparencia interna del modelo.

La alfabetización en inteligencia artificial emerge como una competencia fundamental para los futuros médicos, en la medida en que estas herramientas se incorporan progresivamente a la práctica clínica. Esta alfabetización excede el manejo instrumental de la herramienta e incluye el conocimiento de sus principios básicos, sus limitaciones, sus sesgos y las implicaciones éticas de su uso. La evidencia sugiere que la formación en esta competencia debe iniciarse tempranamente en el currículo y mantenerse a lo largo de toda la formación, con progresión desde el uso guiado hasta la evaluación crítica autónoma (Gaur et al., 2026).

La formación en el uso de inteligencia artificial debe distinguir entre habilidades técnicas de formulación de instrucciones y competencias más amplias de alfabetización, transferibles a

herramientas en evolución. La encuesta transversal de 118 estudiantes japoneses aporta percepciones e intenciones de uso (Tajima et al., 2025), sin comparar la eficacia curricular de ambas alternativas. Por ello, su importancia relativa requiere evaluación educativa específica.

La integración de la inteligencia artificial en la educación médica plantea desafíos éticos de diversa índole. La equidad en el acceso a las tecnologías, la protección de la privacidad de los datos de pacientes simulados o reales, la integridad académica frente al uso de herramientas generativas y la transparencia en la evaluación son dimensiones que requieren marcos normativos explícitos. La opinión de la red estadounidense de farmacia clínica advierte, desde un enfoque normativo, que la ausencia de regulación homogénea entre instituciones y países configura un escenario de desigualdad formativa, especialmente marcado en contextos de recursos limitados (Abdalla et al., 2025).

La incorporación de plataformas de inteligencia artificial en facultades de medicina con menores recursos podría reproducir brechas entre sistemas públicos y privados, regiones urbanas y rurales y países con distintos recursos. Como recomendación institucional, se propone orientar su adopción hacia la equidad mediante financiación compartida, acceso a herramientas validadas y formación docente. El ensayo PROSCA informa a pacientes sobre cáncer de próstata (Baumgärtner et al., 2024), por lo que no demuestra estas desigualdades curriculares. La posible traducción de brechas tecnológicas en brechas formativas y asistenciales requiere evidencia regional específica.

Esta revisión sistemática presenta limitaciones que deben reconocerse, como la heterogeneidad metodológica de los estudios incluidos, que abarcan desde ensayos aleatorizados hasta revisiones narrativas, lo que dificulta la síntesis cuantitativa y obliga a una integración predominantemente cualitativa. Además, la restricción a publicaciones en español, inglés o portugués, si bien responde a criterios de accesibilidad, puede haber excluido evidencia relevante publicada en otros idiomas, mientras que el predominio de literatura anglosajona y la escasez de estudios latinoamericanos limitan la transferencia directa de los hallazgos al contexto regional. Se añaden la restricción a acceso abierto, el posible solapamiento entre estudios primarios y revisiones, la heterogeneidad de versiones de modelos y la ausencia de medición directa de aprendizaje en los benchmarks. La inclusión de textos reflexivos debe reconciliarse con los criterios declarados y distinguirse de la evidencia empírica.

Conclusiones

La presente revisión evidenció que los modelos de lenguaje de gran escala alcanzaron niveles de precisión diagnóstica competitivos en escenarios estandarizados de evaluación, sin que los benchmarks demuestren por sí mismos mejora del aprendizaje. El ensayo con médicos residentes y tratantes no permite generalizar ese resultado a estudiantes de pregrado.

Asimismo, se reveló que la inteligencia artificial posee un potencial pedagógico relevante para el entrenamiento del razonamiento clínico, mediante la generación de casos simulados, la retroalimentación automatizada y las plataformas conversacionales, siempre que dicho uso

estuviera acompañado de supervisión docente y no sustituyera la práctica deliberada ni la exposición directa a la formación médica.

No obstante, persistieron riesgos formativos no resueltos vinculados a la dependencia tecnológica, la posible inhibición del desarrollo de habilidades diagnósticas propias y la limitada interpretabilidad de los procesos algorítmicos, que exigieron marcos de supervisión ética y validación empírica continua dentro de los programas de formación médica.

Los antecedentes latinoamericanos y peruanos, principalmente contextuales y reflexivos, sugirieron examinar si la integración de estas tecnologías en los currículos médicos enfrenta barreras estructurales de infraestructura, financiamiento, formación docente y regulación, lo que fundamenta la recomendación de desarrollar políticas educativas que promovieran una adopción equitativa, crítica y centrada en la seguridad del paciente.

Por último, se recomendó avanzar en la formulación de lineamientos curriculares explícitos para la incorporación de la inteligencia artificial en la formación médica regional, en el desarrollo de estudios empíricos controlados con seguimiento longitudinal y en la construcción de marcos éticos operativos que orienten su uso pedagógico de manera responsable y contextualizada.

Acerca de

Contribución de los autores: Los autores contribuyeron a la conceptualización del estudio, al desarrollo metodológico, al análisis e interpretación de los datos, a la redacción del manuscrito y la revisión crítica de su contenido intelectual.

Financiamiento: Los autores declaran que no recibieron financiamiento para esta investigación.

Conflicto de interés: Los autores declaran no tener conflicto de intereses.

Certificación ética: Por tratarse de una revisión sistemática de literatura científica publicada, el estudio no involucró directamente a participantes humanos ni recopiló datos personales. Se respetaron los principios de integridad académica, reconocimiento de las fuentes consultadas y uso responsable de la información científica. Por su naturaleza documental, no se requirió consentimiento informado.

Historia del artículo: Artículo recibido: 02 de julio 2026 | Arbitrado: 30 de julio 2026 | Aceptado: 03 de septiembre 2026 | Publicado: 12 de octubre 2026.

Referencias

Abdalla, M., Saad, M., Abazia, D., Advani, A., Alhammad, A. M., Bernknopf, A., Bitra, V., Blommel, M., Conn, K. M., Goldwire, M. A., Gosser, R. A., Iso, T., Johnson, S. T., Kier, K. L., Kostrzewa, A., Majerczyk, D., Munir, F., Phillips, J., Segovia, M., et al. (2025). Responsible use of artificial intelligence in health care: evidence, challenges, and best practices: an opinion of the drug information practice and research network of the American college of clinical pharmacy. *Journal of the American College of Clinical Pharmacy*, 8(12), 1333-1361. <https://doi.org/10.1002/jac5.70131>

- Ahmed, M. I., Spooner, B., Isherwood, J., Lane, M., Orrock, E. y Dennison, A. (2023). A systematic review of the barriers to the implementation of artificial intelligence in healthcare. *Cureus*, 15(10), e46454. <https://doi.org/10.7759/cureus.46454>
- Ahsan, Z. (2025). Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions. *BMC medical education*, 25, 1187. <https://doi.org/10.1186/s12909-025-07744-0>
- Arcos Huaman, J. D., Andia Ascama, C. A. y Cabrera Garcia, C. A. (2025). Systematic Review of Obstacles to the Implementation of Artificial Intelligence in the Diagnosis of Cardiovascular Diseases in Latin America. *Actas de la 23.ª Conferencia Internacional LACCEI*. <https://doi.org/10.18687/LACCEI2025.1.1.773>
- Baumgärtner, K., Byczkowski, M., Schmid, T., Muschko, M., Woessner, P., Gerlach, A., Bonekamp, D., Schlemmer, H.-P., Hohenfellner, M. y Görtz, M. (2024). Effectiveness of the medical chatbot PROSCA to inform patients about prostate cancer: results of a randomized controlled trial. *European urology open science*, 69, 80-88. <https://doi.org/10.1016/j.euros.2024.08.022>
- Bin Abdulrahman, K. A. y Hefny, M. A. (2026). Theoretical evolution of AI in medical education: models, frameworks, and future directions. *Frontiers in Education*, 11, 1796632. <https://doi.org/10.3389/feduc.2026.1796632>
- Borg, A., Georg, C., Jobs, B., Huss, V., Waldenlind, K., Ruiz, M., Edelbring, S., Skantze, G. y Parodis, I. (2025). Virtual patient simulations using social robotics combined with large language models for clinical reasoning training in medical education: mixed methods study. *Journal of Medical Internet Research*, 27, e63312. <https://doi.org/10.2196/63312>
- Chávez-Martínez, O. y Ragacini, L. A. (2025). Educación médica e inteligencia artificial: perspectivas y desafíos éticos. *Revista Médica del Instituto Mexicano del Seguro Social*, 63(5), e6736. <https://doi.org/10.5281/zenodo.16748310>
- Cheng, F., Yang, C. y Deng, R. (2026). The Evolution of Visualization Technologies in Healthcare: A Bibliometric Analysis of Studies Published from 1994 to 2025. *Information*, 17(3), 281. <https://doi.org/10.3390/info17030281>
- Çiçek, F. E., Ülker, M., Özer, M. y Kıyak, Y. S. (2025). ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: a randomized controlled trial. *Postgraduate Medical Journal*, 101(1195), 458-463. <https://doi.org/10.1093/postmj/qgae170>
- Corrao, S. y Argano, C. (2022). Rethinking clinical decision-making to improve clinical reasoning. *Frontiers in medicine*, 9, 900543. <https://doi.org/10.3389/fmed.2022.900543>
- Cuestas, E. (2023). El razonamiento clínico y la inteligencia artificial. *Revista de la Facultad de Ciencias Médicas de Córdoba*, 80(4), 306-310. <https://doi.org/10.31053/1853.0605.v80.n4.42544>
- Da Silva, M. (2023). Explainability, Public Reason, and Medical Artificial Intelligence. *Ethical Theory and Moral Practice*, 26(5), 743-762. <https://doi.org/10.1007/s10677-023-10390-4>
- Fehr, J., Jaramillo-Gutierrez, G., Oala, L., Gröschel, M. I., Bierwirth, M., Balachandran, P., Werneck-Leite, A. y Lippert, C. (2022). Piloting a survey-based assessment of transparency and trustworthiness with three medical AI tools. *Healthcare*, 10(10), 1923. <https://doi.org/10.3390/healthcare10101923>

- Gaebe, K. y van der Woerd, B. (2025). Evaluation of large language models as a diagnostic tool for medical learners and clinicians using advanced prompting techniques. *PLoS One*, 20(8), e0325803. <https://doi.org/10.1371/journal.pone.0325803>
- García-Torres, D., Fernández, C., Mira, J. J., Morales, A. y Vicente, M. A. (2025). Using AI-based virtual simulated patients for training in psychopathological interviewing: Cross-sectional observational study. *JMIR Medical Education*, 11, e78857. <https://doi.org/10.2196/78857>
- Gaur, A., Kirani Tan, J., Sridhar Rao, M., Fuad, M., Bhatti, T., Rishab Bharadwaj, H. y Mohamed Ahmed, K. A. H. (2026). AI in UK Medical Education: A Framework for Curriculum Reform. *JMIR Medical Education*, 12, e81953. <https://doi.org/10.2196/81953>
- Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J. A., Kanjee, Z., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, A., Olson, A. P. J., Rodman, A. y Chen, J. H. (2024). Large language model influence on diagnostic reasoning: A randomized clinical trial. *JAMA Network Open*, 7(10), e2440969. <https://doi.org/10.1001/jamanetworkopen.2024.40969>
- Gonzales Aste, M. (2025). Inteligencia artificial en neumología ¿aliada clínica o desafío ético en el Perú? *Respirar*, 17(4), 511-526. <https://doi.org/10.55720/respirar.17.4.12>
- Gordon, M., Daniel, M., Ajiboye, A., Uraiby, H., Xu, N. Y., Bartlett, R., Hanson, J., Haas, M., Spadafore, M., Grafton-Clarke, C., Gasiea, R. Y., Michie, C., Corral, J., Kwan, B., Dolmans, D. y Thammasitboon, S. (2024). A scoping review of artificial intelligence in medical education: BEME Guide No. 84. *Medical teacher*, 46(4), 446-470. <https://doi.org/10.1080/0142159X.2024.2314198>
- Hager, P., Jungmann, F., Holland, R., Bhagat, K., Hubrecht, I., Knauer, M., Vielhauer, J., Makowski, M., Braren, R., Kaissis, G. y Rueckert, D. (2024). Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nature medicine*, 30(9), 2613-2622. <https://doi.org/10.1038/s41591-024-03097-1>
- Han, E. R., Yeo, S., Kim, M. J., Lee, Y. H., Park, K. H. y Roh, H. (2019). Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. *BMC medical education*, 19, 460. <https://doi.org/10.1186/s12909-019-1891-5>
- Hernández Rincón, E. H., Jimenez, D., Chavarro Aguilar, L. A., Pérez Flórez, J. M., Romero Tapia, Á. E. y Jaimes Peñuela, C. L. (2025). Mapping the use of artificial intelligence in medical education: a scoping review. *BMC medical education*, 25, 526. <https://doi.org/10.1186/s12909-025-07089-8>
- Joseph, G., Bhatti, N., Mittal, R. y Bhatti, A. (2025). Current application and future prospects of artificial intelligence in healthcare and medical education: a review of literature. *Cureus*, 17(1), e77313. <https://doi.org/10.7759/cureus.77313>
- Kang, J. y Ahn, J. (2025). Technologies, opportunities, challenges, and future directions for integrating generative artificial intelligence into medical education: a narrative review. *Ewha medical journal*, 48(4), e53. <https://doi.org/10.12771/emj.2025.00787>
- Kanjee, Z., Crowe, B. y Rodman, A. (2023). Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. *JAMA*, 330(1), 78-80. <https://doi.org/10.1001/jama.2023.8288>

- Kim, D. H., Kang, Y. J. y Lee, Y. M. (2025). Twelve tips for developing and implementing AI curriculum for undergraduate medical education. *Medical education online*, 30(1), 2585637. <https://doi.org/10.1080/10872981.2025.2585637>
- Kıyak, Y. S., Emekli, E., İş Kara, T., Coşkun, Ö. y Budakoğlu, I. İ. (2025). AI teaches surgical diagnostic reasoning to medical students: evidence from an experiment using a fully automated, low-cost feedback system. *Journal of surgical education*, 82(10), 103639. <https://doi.org/10.1016/j.jsurg.2025.103639>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pelaccia, T., Tardif, J., Tribby, E. y Charlin, B. (2011). An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Medical education online*, 16(1), 5890. <https://doi.org/10.3402/meo.v16i0.5890>
- Peralta Ramírez, A. A., Trujillo López, S., Navarro Armendariz, G. A., De la Torre Othón, S. A., Sierra Cervantes, M. R. y Medina Aguirre, J. A. (2025). Clinical simulation with ChatGPT: a revolution in medical education? *Journal of CME*, 14(1), 2525615. <https://doi.org/10.1080/28338073.2025.2525615>
- Preiksaitis, C. y Rose, C. (2023). Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. *JMIR Medical Education*, 9, e48785. <https://doi.org/10.2196/48785>
- Ruiz Orizaga, G., Romero Pérez, D. Y., Romero Pérez, F. D., González Ibarra, E. A. y Durán Bautista, M. (2024). Inteligencia artificial en la educación médica: Tendencias y desafíos. *Ciencia y Reflexión*, 3(2), 594-605. <https://doi.org/10.70747/cr.v3i2.55>
- Savage, T., Nayak, A., Gallo, R., Rangan, E. y Chen, J. H. (2024). Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *NPJ Digital Medicine*, 7, 20. <https://doi.org/10.1038/s41746-024-01010-1>
- Shan, G., Chen, X., Wang, C., Liu, L., Gu, Y., Jiang, H. y Shi, T. (2025). Comparing diagnostic accuracy of clinical professionals and large language models: systematic review and meta-analysis. *JMIR medical informatics*, 13, e64963. <https://doi.org/10.2196/64963>
- Siam, K., Varela, A., Faruk, M. J. H., Cheng, J. Q., Gu, H., Al Maruf, A. y Aung, Z. (2026). Benchmarking large language models on the United States medical licensing examination for clinical reasoning and medical licensing scenarios. *Scientific Reports*, 16, 1387. <https://doi.org/10.1038/s41598-025-31010-4>
- Sindhvananda, K., Ruengwattanachot, K., Leksuwankun, S., Pongpitakmetha, T., Hiransuthikul, A., Surawattanawong, T., Anukoolwittaya, P. y Panjasriprakarn, P. (2026). AI-powered simulated patients with automated feedback for enhancing headache history-taking skills: a convergent mixed-methods study. *BMC Medical Education*, 26, 1151. <https://doi.org/10.1186/s12909-026-09511-1>
- Strong, E., DiGiammarino, A., Weng, Y., Kumar, A., Hosamani, P., Hom, J. y Chen, J. H. (2023). Chatbot vs medical student performance on free-response clinical reasoning examinations. *JAMA internal medicine*, 183(9), 1028-1030. <https://doi.org/10.1001/jamainternmed.2023.2909>

- Tajima, H., Kasai, H., Shikino, K., Shimizu, I. y Ito, S. (2025). Perceptions and intentions to use generative AI among first-year medical students in Japan: cross-sectional survey study. *JMIR Medical Education*, 11, e77552. <https://doi.org/10.2196/77552>
- Takita, H., Kabata, D., Walston, S. L., Tatekawa, H., Saito, K., Tsujimoto, Y., Miki, Y. y Ueda, D. (2025). A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians. *NPJ Digital Medicine*, 8, 175. <https://doi.org/10.1038/s41746-025-01543-z>
- Turney, J., Young, T. M., Chauhan, D. R., Beeharry, R. y Mahmud, M. (2026). AI use for medical students: Impact on clinical skill acquisition and retention. A systematic review. *Advances in Medical Education and Practice*, 17, 1–13. <https://doi.org/10.2147/AMEP.S583763>
- Wei, H., Dai, Y., Yuan, K., Li, K. Y., Hung, K. F., Hu, E. M., Lee, A. H. C., Chang, J. W. W., Zhang, C. y Li, X. (2025). AI-powered problem-and case-based learning in medical and dental education: a systematic review and meta-analysis. *International Dental Journal*, 75(4), 100858. <https://doi.org/10.1016/j.identj.2025.100858>
- Zhong, W., Liu, Y., Liu, Y., Yang, K., Gao, H., Yan, H., Hao, W., Yan, Y. y Yin, C. (2025). Performance of ChatGPT-4o and four open-source large language models in generating diagnoses based on China's rare disease catalog: comparative study. *Journal of Medical Internet Research*, 27, e69929. <https://doi.org/10.2196/69929>
-

Este artículo está bajo la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0).