

Universidad Central de Nicaragua

Facultad de Ciencias Médicas



Informe de investigación

Título

Efecto de ChatGPT como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, 2026.

Autor

Dr. Francisco José Somarriba-López.

Doctor en medicina y cirugía.

Asesores

Dr. Edgardo Salvador Jiménez-Martínez.

Asesor metodológico.

Dr. Eduardo Carvajal-Rodríguez.

Asesor científico.

Institución

Universidad Central de Nicaragua – Sede Doral.

Managua, Nicaragua

23 de junio de 2026

Dedicado a la memoria de mi padre,

Francisco José Somarriba Pérez.

Resumen

Introducción. La inteligencia artificial generativa se ha incorporado progresivamente a la educación médica como una herramienta de apoyo al aprendizaje. Sin embargo, la evidencia sobre su impacto en la comprensión de textos médicos y el razonamiento clínico continúa siendo limitada. En este contexto, la presente investigación evaluó el efecto de *ChatGPT* como tutor virtual grupal en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua durante el año 2026.

Objetivo. Evaluar el efecto de *ChatGPT* como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, 2026.

Métodos. Estudio con enfoque cuantitativo, cuasi experimental, prospectivo, analítico y de mediciones repetidas.

Resultados. No se encontraron diferencias estadísticamente significativas en las puntuaciones obtenidas en los cuestionarios de comprensión de textos médicos y razonamiento clínico después de la intervención con *ChatGPT* ($p = 0.081$). No obstante, se observaron mejoras significativas en las percepciones de claridad, comprensión y confianza del aprendizaje ($p < 0.05$). Además, la totalidad de los participantes valoró favorablemente la utilidad de *ChatGPT* como herramienta complementaria para la comprensión de textos médicos complejos.

Conclusión. La utilización de *ChatGPT* como tutor virtual grupal no produjo cambios significativos en el desempeño académico medido mediante cuestionarios clínicos; sin embargo, mejoró significativamente las percepciones de claridad, comprensión y confianza del aprendizaje. Asimismo, los estudiantes valoraron de forma unánime su utilidad como

herramienta complementaria para la comprensión de textos médicos, lo que respalda su potencial como recurso de apoyo en la educación médica.

Palabras claves. Inteligencia artificial generativa, tutor virtual, educación médica.

Carta aval: asesor científico

Yo, Eduardo Carvajal Rodríguez, en calidad de asesor científico de la investigación titulada *Efecto de ChatGPT como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, 2026*, hago constar que he brindado acompañamiento científico durante el desarrollo de la presente investigación, revisando y orientando los aspectos relacionados con la fundamentación teórica, la pertinencia científica del estudio, la coherencia entre el problema de investigación y los objetivos planteados, la interpretación de los resultados obtenidos, así como la discusión y conclusiones derivadas de los hallazgos.

Después de revisar la versión final del informe, considero que el trabajo presenta rigor académico, sustento científico y coherencia interna, constituyendo un aporte al conocimiento sobre el uso de la inteligencia artificial generativa en el ámbito de la educación médica. Asimismo, la investigación responde adecuadamente a los objetivos propuestos y se encuentra respaldada por evidencia científica actual y pertinente.

Para los fines que estime convenientes, se extiende la presente en la ciudad de Managua, Nicaragua, a los veintitrés días del mes de junio del año 2026.

Dr. Eduardo Carvajal Rodríguez

Especialista en medicina forense.

Director de investigación.

Universidad Central de Nicaragua.

Carta aval: asesor metodológico

Yo, Edgardo Salvador Jiménez Martínez, en calidad de asesor metodológico de la investigación titulada *Efecto de ChatGPT como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, 2026*, hago constar que he acompañado y asesorado metodológicamente el desarrollo de la presente investigación, revisando de manera sistemática los aspectos relacionados con el planteamiento del problema, objetivos, hipótesis, diseño metodológico, operacionalización de variables, técnicas e instrumentos de recolección de datos, procedimientos de análisis estadístico, presentación de resultados y demás componentes científicos requeridos para su adecuada ejecución.

Tras la revisión del documento final, considero que la investigación cumple con los criterios metodológicos exigidos para trabajos de investigación científica en el ámbito de las ciencias de la salud, presentando coherencia entre sus objetivos, metodología, análisis de resultados, discusión y conclusiones.

Para los fines que estime convenientes, se extiende la presente en la ciudad de Managua, Nicaragua, a los veintitrés días del mes de junio del año 2026.

Dr. Edgardo Salvador Jiménez Martínez

Doctor en entomología.

Docente investigador.

Universidad Central de Nicaragua.

Índice

6- Introducción	1
6.1- Antecedentes y contexto del problema	4
6.2- Objetivos	12
6.3- Preguntas de investigación	13
6.4- Justificación	14
6.5- Limitaciones	17
6.6- Hipótesis	19
6.7- Variables	20
6.8- Marco contextual	24
7.- Marco teórico	26
8.- Metodología	52
9.- Discusión y resultados	64
10.- Conclusiones	81
Recomendaciones	82
11.- Referencias	84
12.- Anexos	101

Índice de tablas

Tabla 1 Personalización de instrucciones para ChatGPT con criterios SMART	101
Tabla 2 Datos sociodemográficos y académicos	103
Tabla 3 Cuestionario I: Comprensión inicial del texto y razonamiento clínico	104
Tabla 4 Percepción inicial sobre el material de estudio	105
Tabla 5 Cuestionario II: Comprensión posterior a la tutoría con IA y razonamiento clínico	106
Tabla 6 Percepción posterior a la tutoría con IA	108
Tabla 7 Datos sociodemográficos y académicos de los estudiantes participantes (n = 57)	109
Tabla 8 Distribución de respuestas obtenidas en los cuestionarios de comprensión de textos médicos y aplicación de conocimientos clínicos antes y después de la intervención con ChatGPT (n = 57)	110
Tabla 9 Estadísticos descriptivos de los puntajes obtenidos en los cuestionarios de comprensión de textos médicos y aplicación de conocimientos clínicos antes y después de la intervención con ChatGPT (n = 57)	111
Tabla 10 Comparación de los puntajes obtenidos en los cuestionarios de comprensión de textos médicos y aplicación de conocimientos clínicos mediante la prueba de rangos con signo de Wilcoxon (n = 57)	111
Tabla 11 Distribución de las percepciones de claridad, comprensión y confianza asociadas al libro de texto y a la tutoría con ChatGPT (n = 57)	112
Tabla 12 Comparación de las percepciones de claridad, comprensión y confianza mediante diferencia de proporciones y prueba de McNemar (n = 57)	112

Tabla 13 Comparación de las dimensiones de claridad, comprensión y confianza percibidas entre el libro de texto y la tutoría con ChatGPT mediante la prueba de rangos con signo de Wilcoxon (n = 57)	113
Tabla 14 Distribución de la utilidad percibida de ChatGPT como herramienta complementaria para la comprensión de textos médicos complejos (n = 57)	114
Tabla 15 Estadísticos descriptivos de la puntuación global de utilidad percibida de ChatGPT como herramienta complementaria para la comprensión de textos médicos complejos (n = 57)	114

Índice de figuras

Figura 1 Chat privado basado en IA siguiendo el modelo SMART	115
Figura 2 Realización de prueba piloto con estudiantes de quinto año	116
Figura 3 Carta por experto técnico: validación de prompt	117
Figura 4 Carta por experto técnico: validación de clínica	118
Figura 5 Carta de experto técnico: validación pedagógica	119
Figura 6 Aplicación de instrumento de recolección de datos a estudiantes de cuarto año de la carrera de Medicina de la Universidad Central de Nicaragua	120

6.- Introducción

La inteligencia artificial (IA) generativa ha emergido como una de las innovaciones tecnológicas con mayor impacto en la educación superior durante la presente década (Grassini, 2023; Yigci et al., 2025). Desde la aparición pública de modelos de lenguaje de gran escala como *ChatGPT*, estudiantes y docentes han incorporado progresivamente estas herramientas en actividades relacionadas con la búsqueda de información, la resolución de dudas, la generación de materiales de estudio y el apoyo a procesos de aprendizaje (Marchena Sekli et al., 2024; Mittal et al., 2024; Rahman & Watanobe, 2023). En el ámbito de las ciencias de la salud, su rápida adopción ha despertado un interés creciente debido a su capacidad para explicar conceptos complejos, sintetizar grandes volúmenes de información y ofrecer retroalimentación inmediata mediante interacciones conversacionales en lenguaje natural (Abd-alrazaq et al., 2023; Ali et al., 2024; Bahroun et al., 2023; X. Xu et al., 2024). Sin embargo, esta incorporación también ha generado cuestionamientos sobre sus verdaderos efectos en el aprendizaje, particularmente en competencias fundamentales como la comprensión de contenidos científicos y el razonamiento clínico (Feigerlova et al., 2025; Nartthanarung et al., 2026).

La formación médica exige que los estudiantes desarrollen la capacidad de integrar conocimientos provenientes de diversas disciplinas básicas y clínicas para interpretar problemas de salud, formular hipótesis diagnósticas y tomar decisiones fundamentadas. Este proceso depende no solo de la adquisición de información, sino también de la comprensión profunda de textos especializados y de la construcción progresiva de estructuras cognitivas que permitan relacionar conceptos complejos (Feigerlova et al., 2025; Saxena, 2024). En este contexto, la lectura de literatura médica continúa siendo una

estrategia esencial dentro de la educación médica, aunque frecuentemente representa un desafío para los estudiantes debido a la densidad conceptual, el vocabulario técnico y la necesidad de integrar conocimientos previamente adquiridos.

Diversos autores han propuesto que la IA generativa puede desempeñar un papel relevante como herramienta de apoyo al aprendizaje cuando se utiliza bajo diseños instruccionales estructurados (Akram et al., 2025; Long et al., 2026; Yuan et al., 2025). En lugar de sustituir el razonamiento del estudiante o proporcionar respuestas directas, los modelos de lenguaje pueden actuar como tutores virtuales capaces de ofrecer explicaciones adaptadas al nivel de conocimiento del usuario, aclarar conceptos difíciles y facilitar la construcción activa del aprendizaje. Este enfoque encuentra sustento en teorías educativas como el aprendizaje significativo y el constructivismo, las cuales destacan la importancia de relacionar los nuevos conocimientos con las estructuras cognitivas previas y promover la participación activa del estudiante en la construcción de su propio aprendizaje (Punar Özçelik & Yangın Ekşi, 2024).

A pesar del creciente interés académico por estas tecnologías, la evidencia científica disponible continúa siendo limitada. La mayoría de las investigaciones publicadas se han concentrado en analizar percepciones, actitudes o potenciales aplicaciones de la IA en la educación, mientras que los estudios que evalúan objetivamente cambios en el desempeño académico y en el razonamiento clínico mediante diseños experimentales o cuasi experimentales siguen siendo escasos (Gordon et al., 2024). Las revisiones recientes señalan la necesidad de generar evidencia empírica que permita comprender mejor el impacto real de estas herramientas en el aprendizaje, especialmente dentro de contextos específicos como la educación médica (Feigerlova et al., 2025; Hale et al., 2024; Nartthanarung et al., 2026).

En América Latina, y particularmente en Nicaragua, la información disponible es aún más limitada (Manobanda et al., 2026). La ausencia de investigaciones nacionales que evalúen de forma sistemática el uso de la IA generativa como tutor virtual en estudiantes de medicina dificulta la toma de decisiones fundamentadas sobre su incorporación en los procesos educativos. Esta situación adquiere relevancia en un escenario donde el acceso a estas tecnologías continúa expandiéndose y donde los estudiantes ya las utilizan de manera cotidiana, frecuentemente sin lineamientos pedagógicos definidos (Domínguez et al., 2025).

En este contexto, la presente investigación tuvo como propósito evaluar el efecto de *ChatGPT* como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua durante el año 2026. Para ello, se analizó el desempeño de los estudiantes mediante cuestionarios aplicados antes y después de una intervención educativa estructurada, complementando dicha evaluación con el estudio de las percepciones relacionadas con la claridad, comprensión, confianza y utilidad del aprendizaje mediado por inteligencia artificial. Los resultados obtenidos contribuyen a ampliar la evidencia disponible sobre el uso de la IA generativa en educación médica y aportan información relevante para comprender su potencial como herramienta de apoyo dentro de los procesos formativos de futuros profesionales de la salud.

6.1- Antecedentes y contexto del problema

Antecedentes

La revisión de antecedentes se estructuró en tres niveles geográficos: internacional, regional y nacional. Los antecedentes internacionales incluyeron investigaciones desarrolladas fuera del contexto latinoamericano y del Caribe, con el propósito de identificar tendencias globales, enfoques metodológicos predominantes y hallazgos recientes sobre el uso de IA generativa en procesos educativos y, específicamente, en educación médica. Posteriormente, se analizaron antecedentes regionales, definidos como aquellos realizados en países de América Latina y las Antillas, excluyendo Nicaragua, con el objetivo de reconocer aproximaciones metodológicas y líneas de investigación desarrolladas en contextos académicos y socioculturales más cercanos al escenario de estudio. Finalmente, se investigaron antecedentes nacionales correspondientes al territorio nicaragüense, orientados a identificar el estado actual del conocimiento local y el grado de desarrollo de investigaciones relacionadas con IA y educación médica. La búsqueda se limitó al período comprendido entre 2023 y 2026, debido al crecimiento acelerado y disponibilidad de literatura científica posterior a la expansión del uso académico de herramientas de IA generativa, así como por la alta cantidad de fuentes disponibles durante el proceso de revisión. Adicionalmente, se observó que conforme la búsqueda se concentró hacia niveles geográficos más próximos al contexto del estudio, los trabajos encontrados tendieron a enfocarse progresivamente en educación general y educación superior, más que en educación médica específicamente.

La identificación de antecedentes se realizó mediante revisión documental utilizando la base de datos completa de Consensus y el buscador Google Académico. La

estrategia de búsqueda empleó palabras clave en idioma inglés y español relacionadas con el objeto de estudio, incluyendo términos como: *generative artificial intelligence*, *artificial intelligence in medical education*, *virtual tutor*, *medical students*, *medical education*, *educational intervention*, *ChatGPT*, *generative AI*, *medical learning*, educación médica, IA generativa, tutor virtual, aprendizaje asistido por IA, intervención educativa y enseñanza en medicina. Se utilizaron combinaciones mediante operadores booleanos para ampliar o restringir los resultados según la pertinencia temática.

Se excluyeron las publicaciones sin acceso al texto completo, que no fueron sometidas a revisión por pares o que no se centraron específicamente en procesos de enseñanza a través de tutoría virtual con modelos de lenguaje grande dentro del ámbito de la medicina. La revisión se enfocó en fuentes documentales caracterizadas por el rigor académico, la calidad metodológica y su valor para el estudio de la IA generativa aplicada a la educación médica.

Internacionales

Nartthanarung, Plangsiri, y Kongmalai (2026) emplearon un diseño cruzado aleatorizado para evaluar el impacto del uso de IA frente a herramientas de búsqueda convencionales en el desempeño del razonamiento clínico. La investigación cuantitativa se llevó a cabo con una muestra de 46 estudiantes de segundo año de la Facultad de Medicina de la Universidad de Kasetsart de Bangkok, Tailandia, del área preclínica, quienes completaron evaluaciones estructuradas antes y después de interactuar en fases de aprendizaje de 20 minutos con la tecnología. Los hallazgos revelaron una mejora estadísticamente significativa intraindividual, partiendo de una puntuación media en el pretest de 3.96 (DE = 1.65) y alcanzando un promedio de 5.70 (DE = 1.50) en el posttest final tras la doble intervención. Los autores concluyen que la IA cumple un rol

complementario eficaz, aunque advierten vacíos metodológicos importantes como el tamaño reducido de la muestra estadística ($n = 18$ era el mínimo, pero usaron 46) y la limitación de medir los resultados exclusivamente a corto plazo.

Investigadoras en Lahore, Pakistan, analizaron el efecto del andamiaje cognitivo basado en IA sobre las habilidades de resolución de problemas y la conciencia metacognitiva (Akram et al., 2025). El estudio dividió a 300 estudiantes de secundaria y nivel universitario en un grupo experimental ($n = 150$) y un grupo control ($n = 150$), recolectando datos mediante mediciones pretest y posttest estructuradas. El análisis cuantitativo evidenció una correlación positiva y significativa entre el uso del tutor virtual y el aumento en la conciencia metacognitiva ($r = 0.612$, $p < 0.01$). Asimismo, el modelo de regresión lineal demostró que la intervención con IA predice un incremento altamente significativo en las habilidades de resolución de problemas ($B = 0.679$, $t = 15.038$, $p < 0.001$). A pesar de la solidez de los datos numéricos, los investigadores identifican la necesidad de realizar estudios adicionales a largo plazo que ratifiquen la perdurabilidad del aprendizaje autorregulado.

Investigadores de la Universidad de Ciencia y Tecnología Kwame Nkrumah, Kumasi, Ghana, desarrollaron un estudio cuantitativo cuasi experimental para determinar el impacto de un tutor virtual conversacional de voz en el rendimiento académico dentro de la educación universitaria ghanesa (Essel et al., 2025). La investigación adoptó un diseño de mediciones repetidas utilizando una muestra de 65 estudiantes de pregrado, asignados aleatoriamente a una cohorte experimental ($n = 33$) interactuando con la IA y una cohorte control ($n = 32$) interactuando con un facilitador humano durante 12 semanas. Los resultados demostraron que el grupo experimental superó de manera sobresaliente al control; la cohorte con IA logró una media académica en el posttest de 78.5 ($DE = 5.60$), en

contraste con la media de 41.4 (DE = 4.21) obtenida por el grupo tradicional. El análisis de covarianza confirmó diferencias estadísticamente significativas a favor del andamiaje con IA ($p < 0.001$). Esta evidencia converge con la literatura al ratificar los beneficios del tutor virtual, sin embargo, el estudio señala como vacío metodológico la restricción de la muestra exclusivamente a estudiantes de cuarto año, sugiriendo replicar las mediciones en diversas etapas académicas y disciplinas.

Al analizar la literatura bajo los criterios metodológicos estrictos, la evidencia específica pública proveniente de Europa es escasa. Una revisión sistemática muy reciente realizada por Feigerlova, Hani, y Hothersall-Davies (2025) sobre el impacto de la IA en los resultados educativos de las profesiones de la salud señala explícitamente este vacío metodológico: la gran mayoría de los estudios actuales presentan debilidades, como la falta de diseños controlados, muestras pequeñas y la ausencia de medidas validadas de desempeño, concluyendo que falta evidencia de alta calidad para respaldar la mejora en la competencia clínica.

Regionales

En la Universidad Técnica de Ambato, Ecuador, se evaluó el impacto de una intervención didáctica guiada con *ChatGPT* en la aceptación y uso de la herramienta para el autoaprendizaje en estudiantes de medicina en la asignatura de Fisiología (Manobanda et al., 2026). El estudio empleó un diseño metodológico cuasi experimental con mediciones pretest y posttest, utilizando una muestra de 71 alumnos. Durante la fase pre-intervención, los resultados evidenciaron un escepticismo inicial significativo respecto a la confiabilidad de la IA, partiendo de una mediana estadística de 2.50. Tras participar en la intervención didáctica, se observó una reversión completa de esta actitud, alcanzando una mediana de confiabilidad de 4.00. El análisis estadístico mediante la prueba de Wilcoxon confirmó que

el impacto en la valoración general de la herramienta fue altamente significativo ($p < 0.001$). Adicionalmente, el 54.2 % de los participantes reportó que su retención de información mejoró de manera significativa tras el uso guiado de la tecnología. Los autores concluyeron que la intervención logró superar la desconfianza inicial, validando a la IA como un recurso efectivo que fomenta el autoaprendizaje; no obstante, señalaron como vacío metodológico que el instrumento utilizado en el posttest obtuvo una consistencia interna moderada (alfa de Cronbach = 0.665), lo que sugiere cautela en la generalización de los datos.

Al revisar la bibliografía bajo los criterios metodológicos estrictos, no se cuenta con más estudios públicos en América Latina y el Caribe ajustados a la investigación. No obstante, existen otros antecedentes regionales recientes de corte observacional, transversal y revisiones que contextualizan cómo se está adoptando la IA en las facultades de medicina vecinas.

En la Universidad de los Andes, Venezuela, se realizó un análisis sobre las actitudes hacia el uso pedagógico de *chatbots* de IA generativa (Ciaccia et al., 2025). El estudio empleó un diseño metodológico cuantitativo, correlacional y de corte transversal en una muestra de 155 estudiantes de diversas carreras de ciencias de la salud (incluyendo medicina). Los resultados evidenciaron que aproximadamente el 85 % de los participantes utiliza plataformas como *ChatGPT*, demostrando una actitud globalmente favorable hacia su aprovechamiento para tareas académicas teóricas. Sin embargo, los datos reflejaron reservas significativas y desacuerdo ante la idea de utilizar la tecnología para simulaciones clínicas o diseño de planes de tratamiento. Los autores concluyeron que, aunque los estudiantes reconocen los beneficios de la IA, perciben riesgos éticos y demandan la

generación de mayor evidencia empírica que respalde su uso en la práctica clínica formativa.

En Cuba desarrollaron un estudio en la Universidad de Ciencias Médicas de Santiago de Cuba orientado a caracterizar el uso de *ChatGPT* dentro del proceso de enseñanza-aprendizaje de los futuros galenos (Miclin et al., 2024). Utilizaron un diseño observacional, descriptivo y transversal, aplicando un muestreo aleatorio simple para seleccionar a 310 estudiantes de medicina. Entre sus hallazgos, destacó que el 64.7 % de los alumnos consideró que la calidad de la información proporcionada por la IA era «igual» a la obtenida mediante fuentes bibliográficas tradicionales. Los investigadores concluyeron que *ChatGPT* representa un recurso complementario sumamente valioso para el currículo médico, pero señalaron como advertencia que este andamiaje tecnológico exige por parte del estudiante una verificación rigurosa de los datos para mitigar sesgos o imprecisiones.

Nacionales

A nivel nacional, la revisión exhaustiva de la literatura evidencia un vacío de conocimiento absoluto respecto a la integración y evaluación de la IA generativa como tutor virtual dentro de la educación médica. A diferencia de los escenarios internacionales y regionales, donde ya se registra una transición hacia la medición empírica del andamiaje cognitivo de la IA sobre el razonamiento clínico, en Nicaragua no se identificaron investigaciones públicas previas de corte experimental, cuasi experimental o prospectivo que aborden este fenómeno. Esta carencia de antecedentes locales metodológicamente rigurosos justifica plenamente la originalidad, viabilidad y urgencia de la presente investigación. Por consiguiente, la ejecución de este estudio con estudiantes de cuarto año de la UCN representa una investigación pionera. Su desarrollo no solo buscará evaluar los cambios en la comprensión de textos médicos y la percepción del aprendizaje tras la

interacción guiada con la tecnología, sino que sentará la primera línea base de evidencia cuantitativa en el territorio nicaragüense respecto al uso estructurado de la IA en la enseñanza de las ciencias básicas médicas.

Contexto del problema

La irrupción de la IA generativa ha transformado el panorama de la educación médica y superior a nivel global (González et al., 2025). Desde su lanzamiento, esta tecnología ha experimentado una adopción masiva y acelerada por parte de los estudiantes, quienes la integran cotidianamente en su proceso de aprendizaje para buscar información, generar resúmenes, estructurar ideas y obtener respuestas rápidas a problemas complejos (Chen et al., 2023a; Črček & Patekar, 2023; González et al., 2025; Stöhr et al., 2024).

En el entorno educativo, la interacción de los estudiantes con *ChatGPT* se desarrolla, con gran frecuencia, de manera no controlada o autónoma. Este uso desestructurado y sin supervisión ha suscitado profundas preocupaciones en la comunidad académica, vinculadas principalmente a la dependencia tecnológica y la disminución de las habilidades de pensamiento crítico y resolución de problemas (Abd-alrazaq et al., 2023; Ciaccia et al., 2025; Farrokhnia et al., 2024; Zastudil et al., 2023). La facilidad con la que el modelo genera respuestas inmediatas puede fomentar un enfoque de aprendizaje superficial o una «descarga cognitiva», donde los alumnos utilizan la herramienta como un atajo en lugar de procesar, analizar y comprender genuinamente los contenidos médicos (Elnaffar et al., 2026; Nartthanarung et al., 2026).

En contraste, el uso de *ChatGPT* de manera controlada e intencionada dentro del aula plantea un escenario distinto. Cuando la interacción con la IA se enmarca en un diseño estructurado, actuando como un sistema de tutoría o proporcionando andamiaje cognitivo, la herramienta puede guiar al estudiante sin entregarle respuestas directas que inhiban su

razonamiento (Nartthanarung et al., 2026; T. Wang et al., 2023; Watts et al., 2025). En estos entornos guiados, el uso de instrucciones y restricciones claras permite que el modelo ofrezca retroalimentación personalizada, aclare conceptos complejos de forma adaptativa y fomente un entorno de aprendizaje interactivo y participativo (Carmo et al., 2025; Leung, 2024; Looi & Jia, 2025).

A pesar de la rápida proliferación de *ChatGPT* en las aulas, existe una notable escasez de bibliografía pública y estudios empíricos rigurosos que evalúen su eficacia pedagógica real (Feigerlova et al., 2025; Marchena Sekli et al., 2024; Wood & Moss, 2024). La literatura advierte sobre un vacío metodológico respecto a cómo estas herramientas impactan concretamente en la comprensión profunda y el desarrollo del razonamiento clínico en los estudiantes (X. Xu et al., 2024). Gran parte de la evidencia científica actual se concentra en el análisis de percepciones, actitudes o debates teóricos, existiendo una carencia de diseños experimentales o cuasi experimentales que midan objetivamente el cambio en el desempeño académico. Asimismo, en contextos geográficos locales, se evidencia una ausencia de literatura que documente los efectos de la integración de estos tutores virtuales en la enseñanza de las ciencias de la salud, manteniendo en la incertidumbre los verdaderos alcances de esta tecnología en escenarios de educación médica específicos.

6.2- Objetivos

Objetivo general

Evaluar el efecto de *ChatGPT* como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, 2026.

Objetivo específico

- Determinar cambios en el razonamiento clínico y desempeño académico de los estudiantes posterior a la interacción con *ChatGPT* personalizado como tutor virtual grupal.
- Caracterizar las percepciones de los estudiantes sobre claridad, comprensión, y confianza del aprendizaje antes y después de la intervención con *ChatGPT*.
- Identificar la utilidad percibida de la IA generativa como herramienta complementaria para la comprensión de textos médicos complejos en estudiantes de medicina.

6.3- Preguntas de investigación

La formulación de las preguntas se fundamenta en la necesidad de comprender el papel que desempeña la IA generativa cuando es incorporada de forma estructurada dentro del proceso de aprendizaje. Dado que la intervención propuesta integra la lectura de un texto médico, la resolución de problemas clínicos y la interacción guiada con un tutor virtual basado en IA, resulta necesario explorar tanto los posibles cambios en el desempeño académico como las percepciones que los estudiantes desarrollan respecto a esta experiencia educativa. Considerando que la incorporación de estas tecnologías en la educación médica es un fenómeno relativamente reciente y con escasa evidencia nacional, es pertinente indagar si los estudiantes perciben a la IA como una herramienta útil para complementar los métodos tradicionales de estudio y favorecer la comprensión.

Pregunta general

¿Cuál es el efecto de ChatGPT como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua durante el año 2026?

Preguntas específicas

¿Qué cambios se observan en el razonamiento clínico y el desempeño académico de los estudiantes después de interactuar con ChatGPT como tutor virtual grupal?

¿Cómo varían las percepciones de claridad, comprensión y confianza del aprendizaje de los estudiantes antes y después de la intervención con ChatGPT?

¿Cuál es la utilidad percibida de la inteligencia artificial generativa como herramienta complementaria para la comprensión de textos médicos complejos en estudiantes de medicina?

6.4- Justificación

La presente investigación se fundamenta en la necesidad teórica, metodológica y contextual de evaluar objetivamente el impacto de la IA generativa en la educación médica, abordando vacíos críticos identificados en la literatura científica actual.

La integración de modelos de lenguaje grande como *ChatGPT* en la educación médica ofrece oportunidades prometedoras para personalizar el aprendizaje, generar casos clínicos y explicar conceptos complejos (Abd-alrazaq et al., 2023). Sin embargo, existe un consenso en la literatura respecto a los peligros de su uso no regulado. La interacción libre con la IA fomenta la descarga cognitiva y la sobredependencia, lo cual puede obstaculizar el desarrollo de habilidades fundamentales como el pensamiento crítico, la resolución de problemas y el razonamiento clínico (Narthanarung et al., 2026). Bastani et al. (2025) advierten que la IA generativa implementada sin barreras de contención puede perjudicar el aprendizaje, ya que, al automatizar tareas analíticas, los estudiantes pierden la oportunidad de adquirir experiencia valiosa. Por ello, este estudio se justifica al proponer una intervención donde la IA generativa actúa estrictamente como un tutor virtual complementario que fomenta el razonamiento mediante preguntas estructuradas (mecanismos causales, fisiopatología), evitando que la herramienta resuelva los problemas por el estudiante.

Desde una perspectiva metodológica, la literatura global adolece de investigaciones rigurosas que midan el impacto real de la IA generativa en el desempeño académico. Revisiones sistemáticas recientes advierten que la evidencia actual sobre resultados educativos medibles de las intervenciones con IA en ciencias de la salud es pobre (Peralta Ramirez et al., 2025). Hale et al., (2024) y Feigerlova et al., (2025) señalan que la mayoría

de los estudios disponibles son de naturaleza conceptual, observacional o anecdótica, y destacan la ausencia de diseños controlados, medidas de rendimiento validadas y evaluaciones de resultados a nivel de competencias. Este estudio aborda directamente este vacío metodológico al emplear un diseño cuasi experimental de mediciones repetidas (pretest-postest), el cual permitirá cuantificar el cambio intraindividual en la comprensión de textos médicos complejos de los participantes, superando el enfoque meramente descriptivo o de percepción que predomina en la actualidad.

La decisión de limitar la interacción de los estudiantes con la IA a un máximo de tres preguntas estructuradas encuentra un fuerte respaldo científico. Las investigaciones han demostrado que la calidad y precisión de las respuestas de los *chatbots* dependen significativamente de cómo los usuarios formulan sus directrices (Nartthanarung et al., 2026). El estudio de Vaira et al. (2025) comprobó que el uso de estructuras de instrucciones guiadas mejora la coherencia narrativa de la IA, disminuye el riesgo de «alucinaciones» (información falsa o inexacta) y contextualiza adecuadamente el razonamiento clínico. Al estructurar la intervención de esta manera, la investigación garantiza que la tecnología opere como un andamiaje cognitivo eficaz y seguro, alineándose con las recomendaciones de incorporar supervisión y verificación para mitigar los sesgos algorítmicos.

Finalmente, como se demostró en la revisión de antecedentes, existe un vacío absoluto de investigaciones previas en Nicaragua que hayan medido empíricamente el impacto de la IA como tutor virtual en estudiantes de medicina. La ejecución de este proyecto en la UCN marca una línea base inédita en la región. Provee datos objetivos sobre cómo los estudiantes de cuarto año, que ya poseen bases en ciencias médicas, utilizan el razonamiento guiado por IA generativa para comprender literatura de alta complejidad. Los resultados de este estudio no solo contribuyen a la literatura global con evidencia cuasi

experimental, sino que proporcionarán fundamentos prácticos para el desarrollo futuro de lineamientos curriculares y bioéticos sobre el uso de la IA en las aulas nicaragüenses y latinoamericanas.

6.5- Limitaciones

El presente estudio reconoce limitaciones inherentes a su diseño metodológico cuasi experimental, el cual, al carecer de una asignación aleatoria de los participantes, puede inducir a sesgos de selección e impide establecer una relación causal directa y absoluta entre la intervención con IA y los resultados observados (Manobanda et al., 2026). Asimismo, el uso de un muestreo no probabilístico y la restricción a un entorno de una única institución limitan significativamente la generalización o validez externa de los hallazgos hacia poblaciones de estudiantes en otros contextos académicos o geográficos (Nartthamarung et al., 2026).

Otra limitación metodológica sustancial corresponde a la temporalidad de las evaluaciones. El enfoque de mediciones repetidas a corto plazo (pretest-posttest inmediato) permite cuantificar el cambio transitorio tras la interacción con el tutor virtual, pero no evalúa la retención de la información a largo plazo ni la transferencia efectiva de estas habilidades hacia la práctica en entornos clínicos reales (Nartthamarung et al., 2026). Diversos estudios advierten que las intervenciones de duración breve restringen la comprensión sobre los verdaderos impactos longitudinales y la sostenibilidad del aprendizaje asistido por tecnología (Huda et al., 2025; Punar Özçelik & Yangın Ekşi, 2024).

En cuanto al componente descriptivo y de percepción, la utilización de encuestas y escalas dicotómicas para recoger las actitudes de los estudiantes introduce el riesgo de sesgos de autoinforme, particularmente el sesgo de deseabilidad social (Long et al., 2026; Manobanda et al., 2026). Aunque este método es ampliamente aceptado para estudios exploratorios, los participantes pueden sobreestimar su progreso o responder de manera

favorable, lo que en ocasiones no captura con total precisión los matices complejos de su verdadera interacción con el *chatbot* (Haindl & Weinberger, 2024; Polyportis & Pahos, 2025). En consecuencia, los hallazgos derivados de estas encuestas deben interpretarse con cautela como aproximaciones subjetivas de la experiencia del usuario (Manobanda et al., 2026).

Finalmente, los resultados están condicionados por las particularidades técnicas de la IA generativa empleada. La calidad, pertinencia y precisión de las respuestas del modelo dependen en gran medida de las habilidades de los estudiantes para formular sus instrucciones, lo que puede introducir variaciones en los resultados de aprendizaje (Nartthamarung et al., 2026). A pesar de que el estudio controla esta variable limitando la interacción a preguntas estructuradas, las actualizaciones dinámicas en las versiones de la herramienta limitan la reproducibilidad exacta del experimento a lo largo del tiempo, un desafío técnico frecuentemente reportado al evaluar respuestas de *chatbots* (Abouzeid et al., 2025).

6.6- Hipótesis

Hipótesis de investigación (H_i)

La utilización de *ChatGPT* como tutor virtual grupal induce cambios en comprensión de textos médicos y razonamiento clínico de estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, sede Doral, durante el primer semestre de 2026.

Hipótesis nula (H_0)

La utilización de *ChatGPT* como tutor virtual grupal no induce cambios significativos en la comprensión de textos médicos complejos ni en el razonamiento clínico de los estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, sede Doral, durante el primer semestre de 2026.

Hipótesis alternativa (H_1)

La utilización de *ChatGPT* como tutor virtual grupal induce cambios significativos en la comprensión de textos médicos complejos y en el razonamiento clínico de los estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, sede Doral, durante el primer semestre de 2026.

6.7- Variables

Objetivos	Variable	Definición conceptual	Definición operacional	Indicadores	Escala de medición
Determinar cambios en el razonamiento clínico y desempeño académico de los estudiantes posterior a la interacción con <i>ChatGPT</i> personalizado como tutor virtual grupal.	Comprensión de textos médicos y aplicación de conocimientos clínicos.	Capacidad del estudiante para interpretar información clínica, integrar conocimientos médicos y seleccionar la respuesta más adecuada ante situaciones clínicas planteadas, expresada mediante su rendimiento académico en evaluaciones de contenido.	Diferencia entre la puntuación obtenida en el cuestionario clínico aplicado antes de la interacción con <i>ChatGPT</i> (pretest) y la puntuación obtenida en el cuestionario clínico aplicado después de la interacción (postest).	– Puntaje obtenido en el pretest. – Puntaje obtenido en el postest. – Variación del puntaje entre pretest y postest. – Porcentaje de respuestas correctas.	Escala.
Caracterizar las percepciones de los estudiantes sobre	Percepción de claridad	Valoración subjetiva del estudiante sobre qué tan claro, organizado y comprensible	Se midió mediante preguntas dicotómicas aplicadas antes y después de la	– Claridad percibida del texto tradicional. – Claridad percibida de	Ordinal

claridad, comprensión, y confianza del aprendiz e antes y después de la intervención con ChatGPT.

resulta el material o explicación recibida durante el proceso de aprendizaje.

intervención. En el pretest, la valoración se referirá a la claridad del texto médico; en el posttest, a la claridad de las respuestas proporcionadas por *ChatGPT*.

las respuestas de *ChatGPT*. – Cambio en la claridad percibida entre pre y postintervención.

Percepción de comprensión	Valoración subjetiva del estudiante sobre el grado en que considera haber entendido el contenido médico revisado y sus relaciones clínicas o fisiopatológicas.	Se midió mediante preguntas dicotómicas aplicadas antes y después de la intervención. En el pretest, la valoración se referirá a la comprensión alcanzada tras la lectura del texto; en el posttest, a la comprensión alcanzada después de la asesoría con <i>ChatGPT</i> .	–	Ordinal
			Comprensión percibida después de la lectura.	
			–	
			Comprensión percibida después de la interacción con <i>ChatGPT</i> .	
			– Cambio en la comprensión percibida entre pre y	

				postintervención.	
	Percepción de confianza	Grado de seguridad subjetiva que manifiesta el estudiante respecto a su capacidad para responder preguntas clínicas relacionadas con el contenido estudiado.	Se midió mediante preguntas dicotómicas aplicadas antes y después de la intervención. En el pretest, la confianza se referirá a la seguridad del estudiante para responder con base en el texto tradicional; en el posttest, a la seguridad posterior a la explicación recibida mediante <i>ChatGPT</i> .	– Confianza para responder después de la lectura. – Confianza para responder después de la interacción con <i>ChatGPT</i> . – Cambio en la confianza percibida entre pre y postintervención.	Ordinal
Identificar la utilidad percibida de la IA generativa como herramienta	Utilidad percibida	Valoración subjetiva del estudiante sobre el grado en que la IA generativa contribuye como apoyo	Se midió mediante preguntas dicotómicas aplicadas posterior a la intervención con	– Utilidad percibida de <i>ChatGPT</i> para aclarar dudas. – Utilidad percibida	Ordinal

a	complementario	<i>ChatGPT</i> ,	para
complemen	para aclarar	orientados a	comprender
taria para	dudas, facilitar	valorar la	el texto
la	la comprensión	utilidad de la IA	médico.
comprensió	y favorecer el	como recurso	– Utilidad
n de textos	aprendizaje del	complementario	percibida
médicos	texto médico	para comprender	como
complejos	complejo.	el contenido	complemen
en		estudiado.	to al estudio
estudiantes			tradicional.
de			–
medicina.			Disposición
			a utilizar IA
			como apoyo
			académico.

6.8- Marco contextual

La presente investigación se desarrolló en la UCN, sede Doral, institución de educación superior ubicada en la ciudad de Managua, Nicaragua. La UCN es una universidad privada que oferta programas académicos en diversas áreas del conocimiento, entre ellas las ciencias de la salud, contribuyendo a la formación de profesionales orientados a responder a las necesidades sanitarias del país y de la región.

El estudio se llevó a cabo específicamente en la Facultad de Ciencias Médicas, unidad académica responsable de la formación de médicos generales. El plan de estudios de la carrera de Medicina se estructura de manera progresiva, integrando durante los primeros años contenidos correspondientes a ciencias básicas médicas, para posteriormente incorporar asignaturas clínicas y actividades de práctica hospitalaria. Dentro de este proceso formativo, los estudiantes de cuarto año han completado asignaturas fundamentales como anatomía, fisiología, bioquímica, histología, patología, fisiopatología y semiología, lo que les proporciona los conocimientos previos necesarios para abordar textos médicos de complejidad intermedia y avanzada.

La población objeto de estudio estuvo constituida por estudiantes activos de cuarto año de la carrera de Medicina durante el primer semestre de 2026. Según los registros académicos institucionales, este nivel se encontraba distribuido en tres grupos con un total de 66 estudiantes matriculados. Durante este período académico, los estudiantes se encontraban desarrollando actividades formativas en asignaturas clínicas, incluyendo rotaciones hospitalarias en áreas como Medicina Interna y Cirugía, complementando así su formación teórica con experiencias prácticas supervisadas.

La investigación se realizó en un contexto caracterizado por la creciente incorporación de herramientas de IA generativa en los procesos educativos universitarios. En particular, plataformas basadas en modelos de lenguaje de gran escala, como *ChatGPT*, han experimentado una rápida adopción entre los estudiantes de medicina para actividades relacionadas con la búsqueda de información, aclaración de dudas, elaboración de resúmenes y apoyo al estudio independiente. Sin embargo, a nivel nacional existe una limitada evidencia científica sobre el impacto de estas herramientas en la comprensión de textos médicos y el desarrollo del razonamiento clínico, especialmente cuando son utilizadas bajo diseños pedagógicos estructurados.

La intervención se desarrolló en la sala de reuniones de la sede, utilizando recursos académicos disponibles institucionalmente. Como material de estudio se empleó un texto seleccionado sobre asma (capítulo 298) del libro Harrison. Principios de Medicina Interna, 22.^a edición (Isarel, 2025), disponible a través de la plataforma AccessMedicina, recurso digital al que tienen acceso los estudiantes mediante su matrícula universitaria. Asimismo, se utilizaron herramientas digitales institucionales para la aplicación de los instrumentos de recolección de datos y la interacción controlada con ChatGPT como tutor virtual grupal.

En este contexto académico y tecnológico, la investigación busca generar evidencia sobre el potencial de la inteligencia artificial generativa como herramienta complementaria para la comprensión de textos médicos complejos y el fortalecimiento del razonamiento clínico en estudiantes de medicina, aportando información relevante para futuras estrategias de innovación educativa dentro de la formación médica universitaria.

7.- Marco teórico

Revisión de la literatura

Definición de IA

El término «IA» fue acuñado formalmente por John McCarthy durante la Conferencia de Dartmouth en 1956, describiendo este campo emergente de las ciencias de la computación como la ciencia y la ingeniería de crear máquinas inteligentes (AlAfnan et al., 2023; Ciacia et al., 2025; Domínguez et al., 2025; Punar Özçelik & Yangın Ekşi, 2024; Williams, 2024). Sin embargo, sus cimientos teóricos se remontan a 1950, cuando el matemático Alan Turing planteó la premisa fundacional de esta disciplina a través de su famosa prueba de imitación, al cuestionarse objetivamente si las máquinas eran capaces de pensar (Abouzeid et al., 2025; Tlili et al., 2023).

En la literatura contemporánea, la IA se define como un término general o paraguas que describe la capacidad de una máquina, plataforma digital o sistema informático para realizar tareas que típicamente requieren de la inteligencia humana (Abedi et al., 2023; Celik et al., 2024; Chen et al., 2023b; Chiu et al., 2024; Gordon et al., 2024). A diferencia de los programas informáticos tradicionales que se limitan a ejecutar comandos preestablecidos, la IA está diseñada para simular o imitar funciones cognitivas complejas, tales como el aprendizaje, la percepción, el razonamiento lógico, el reconocimiento del habla, la comprensión del lenguaje y la resolución de problemas (Chan & Colloton, 2024; Domínguez et al., 2025; Williams, 2024).

Desde una perspectiva operativa e integradora, la IA consiste en sistemas basados en algoritmos que permiten a las computadoras interactuar con su entorno, aprender de los datos (experiencia), adaptarse a situaciones nuevas, sintetizar información, autocorregirse y

tomar decisiones de manera autónoma (Celik et al., 2024; Farrokhnia et al., 2024; Moundridou et al., 2024). Esta capacidad de procesamiento analítico avanzado es lo que permite que las máquinas generen perspectivas, evalúen variables y ejecuten tareas emulando el juicio y el comportamiento de un ser humano inteligente (Gordon et al., 2024).

Evolución de la IA generativa

Los fundamentos teóricos de la IA generativa tienen sus raíces en las décadas de 1950 y 1960, emergiendo en paralelo con los primeros conceptos del aprendizaje automático (Kaplan-Rakowski et al., 2023; Li et al., 2024). Uno de los primeros antecedentes tecnológicos directos de esta área se remonta a la década de 1980 con la propuesta de las «redes de Hopfield», diseñadas por el científico informático John Hopfield, las cuales consistían en redes neuronales entrenadas para generar nuevas secuencias y patrones (Chan & Colloton, 2024). Sin embargo, la IA generativa se mantuvo en un desarrollo conceptual y paulatino durante varias décadas, limitándose a tareas predictivas simples (Jauhiainen & Guerra, 2023).

El verdadero punto de inflexión para la ocurrió gracias a los avances en la capacidad computacional y al desarrollo del aprendizaje profundo (Chan & Colloton, 2024; Li et al., 2024). A partir de los años 2013 y 2014, el campo capturó la atención global de los investigadores con la introducción de dos algoritmos revolucionarios: los «autocodificadores variacionales» y, de manera determinante, las «redes generativas antagónicas» (Kaplan-Rakowski et al., 2023). Estas arquitecturas revolucionaron inicialmente la generación de imágenes y la síntesis de contenido visual realista, sentando las bases de la IA generativa moderna.

Posteriormente, la IA experimentó una segunda gran revolución, esta vez en el ámbito del «procesamiento del lenguaje natural», gracias a la introducción de la

arquitectura «*transformer*» en el año 2017 (Farazouli et al., 2024). Esta arquitectura innovó al incorporar un «mecanismo de autoatención», el cual permite al modelo establecer conexiones entre diferentes segmentos de un texto y procesar secuencias completas en paralelo (Ilieva et al., 2023; Zhang & Huang, 2024). Este enfoque superó las limitaciones de las redes neuronales recurrentes tradicionales al capturar dependencias a largo plazo de manera mucho más eficiente.

El desarrollo de los *transformers* impulsó el surgimiento de los «grandes modelos de lenguaje» y la serie de «transformadores generativos preentrenados» (GPT, por sus siglas en inglés). La evolución de esta serie fue exponencial: OpenAI lanzó GPT-1 en 2018, seguido de GPT-2 en 2019 y GPT-3 en 2020, aumentando el tamaño de entrenamiento de 117 millones de parámetros a 175 mil millones, logrando así capacidades de redacción y comprensión cada vez más humanas (Abd-alrazaq et al., 2023; Silva et al., 2024).

Finalmente, el lanzamiento público de *ChatGPT* (basado en GPT-3.5) a finales de 2022 y la introducción del modelo multimodal GPT-4 en 2023 representaron un hito histórico. Este evento democratizó la IA generativa, sacándola del nicho de investigación para ponerla a disposición del público general, lo que desató una revolución tecnológica y educativa de adopción masiva a nivel global.

Integración de IA en procesos educativos y como herramienta de apoyo al aprendizaje

La integración de la IA en la educación superior representa una transformación paradigmática que está redefiniendo los procesos de enseñanza y aprendizaje a nivel global (Grassini, 2023; Marchena Sekli et al., 2024). Históricamente, el sistema educativo ha dependido de modelos centrados en el tutor y de enfoques estandarizados que con frecuencia enfrentan dificultades para satisfacer el ritmo y las diversas necesidades de los estudiantes en la actual era digital (Kathole et al., 2025; Saxena, 2024). Ante este escenario,

la incorporación de la IA generativa y los modelos de lenguaje de gran escala ha facilitado la transición hacia entornos educativos más inclusivos, flexibles y adaptativos. A nivel de procesos, la integración de la tecnología abarca desde la automatización de tareas administrativas y la evaluación académica, hasta el diseño curricular y la generación de materiales didácticos, lo cual permite a las instituciones optimizar sus recursos y a los docentes concentrarse en la innovación pedagógica y la tutoría humana (Mittal et al., 2024).

Dentro de estos procesos renovados, la IA se consolida como una herramienta fundamental de apoyo al aprendizaje, operando principalmente a través de «sistemas de tutoría inteligente» y *chatbots* conversacionales (Celik et al., 2024; Usher & Amzalag, 2025). Estos sistemas actúan como asistentes virtuales o instructores digitales capaces de ofrecer soporte ininterrumpido, responder consultas en tiempo real y proporcionar retroalimentación inmediata, superando así las limitaciones de tiempo y los altos índices de alumnos por profesor característicos de las aulas tradicionales. Al interactuar mediante procesamiento de lenguaje natural, la IA simula conversaciones humanas que brindan a los estudiantes un espacio interactivo para resolver dudas de manera confidencial, disminuyendo la ansiedad asociada al aprendizaje y fomentando la participación (Tian et al., 2024; Williams, 2024).

El principal valor pedagógico de la IA como herramienta de apoyo radica en su capacidad para la personalización y el aprendizaje adaptativo (Chan & Colloton, 2024; Maity & Deroy, 2024; Mittal et al., 2024; Usher & Amzalag, 2025). La tecnología puede analizar el comportamiento, el rendimiento y las preferencias de los estudiantes para ajustar dinámicamente el contenido, el nivel de dificultad y las estrategias de instrucción. Esta personalización permite la creación de rutas de aprendizaje a la medida de cada individuo, donde el sistema identifica brechas de conocimiento y recomienda recursos o ejercicios

específicos para abordarlas (Mallick & Datta, 2025; Roca et al., 2024). En consecuencia, la adaptación continua no solo mantiene el interés y la motivación del estudiante, sino que también contribuye a reducir la carga cognitiva al descomponer conceptos complejos en explicaciones más simples y manejables, liberando recursos mentales para el pensamiento de orden superior (Iqbal et al., 2025; Long et al., 2026).

Finalmente, la literatura destaca que, cuando la IA se implementa bajo un diseño instruccional estructurado, trasciende la simple provisión de información para convertirse en un andamiaje cognitivo que promueve el aprendizaje autorregulado y la conciencia metacognitiva (Akram et al., 2025; Giannakos et al., 2025). En lugar de limitarse a entregar respuestas directas que inhiban el razonamiento, las herramientas basadas en IA pueden programarse para actuar como tutores socráticos, ofreciendo pistas, orientación paso a paso y preguntas reflexivas que guían al estudiante hacia la resolución de problemas (Elnaffar et al., 2026; Nikolic et al., 2023). Este enfoque de apoyo estructurado asiste a los alumnos en la aplicación del conocimiento, estimula el pensamiento crítico y fomenta una comprensión conceptual profunda, posicionando a la IA como un complemento que potencia de manera interactiva la capacidad analítica humana (Güner & Er, 2025; Santos, 2023).

Personalización y diseño instruccional de chatbots educativos

La integración de *chatbots* basados en modelos de lenguaje de gran escala requiere de un diseño instruccional meticuloso para trascender la simple entrega de información y convertirse en una experiencia educativa adaptativa. La personalización en este contexto se define como la capacidad del sistema para ajustar el contenido, la dificultad y el ritmo de enseñanza a las necesidades individuales y preferencias de cada estudiante (Lin-Zucker et al., 2025; Sánchez-Vera, 2025). Esta flexibilidad es crucial, ya que permite configurar el

chatbot para que asuma roles pedagógicos específicos, aplique reglas de interacción y ofrezca retroalimentación contextualizada que promueva el aprendizaje activo. Para lograr esta adaptabilidad sin necesidad de reentrenar o modificar el código subyacente del modelo de lenguaje, los diseñadores instruccionales y educadores emplean técnicas de ingeniería de instrucciones, las cuales modulan el comportamiento de la IA y garantizan que las interacciones se alineen con los objetivos curriculares.

System prompts y personalización contextual. El mecanismo fundamental para personalizar un modelo de lenguaje de gran escala en el ámbito educativo es el uso de *system prompts*. Un *system prompt* es una instrucción inicial y directiva, generalmente invisible para el estudiante, que define el comportamiento general, la identidad y las capacidades del chatbot durante toda la conversación (Lieb & Goel, 2024; R. Liu et al., 2024). Mediante estas instrucciones, se le asigna al modelo un rol (por ejemplo, «actúa como un profesor universitario experto» o «eres un tutor empático de ciencias») y se le proporciona el contexto necesario sobre el nivel académico de los usuarios a los que se dirige (Chan & Colloton, 2024; Favero et al., 2024; Neumann et al., 2025). El *system prompt* estructura la interacción delineando los objetivos del agente, sus tareas específicas, el tono del lenguaje que debe emplear y el contexto de fondo, lo que permite generar respuestas coherentes, precisas y altamente personalizadas para la asignatura en cuestión (Vaira et al., 2025).

Restricciones pedagógicas y control de respuestas. Un componente vital en la ingeniería de *prompts* educativos es la implementación de restricciones pedagógicas para controlar estrictamente las respuestas generadas por el modelo. Dado que los grandes modelos de lenguaje tienden a ser sobrecomplacientes y a entregar respuestas completas ante consultas directas, es imperativo incluir reglas de contención en el *system prompt* que

prohíban explícitamente resolver las tareas por el estudiante o revelar soluciones finales (Bassner et al., 2024; Lieb & Goel, 2024; Watts et al., 2025; Yuan et al., 2025). Estas restricciones también abarcan el filtrado de preguntas, instruyendo al *chatbot* para que rechace consultas fuera del tema académico, manteniendo así el enfoque de la sesión de estudio y optimizando los recursos. Asimismo, el control de respuestas permite definir formatos y límites de extensión (por ejemplo, respuestas de máximo 100 palabras) para evitar la sobrecarga cognitiva. Por último, estas instrucciones actúan como un regulador ético implícito, delineando los límites del *chatbot* para evitar que ofrezca diagnósticos médicos reales, comparta información inexacta o perpetúe sesgos, asegurando que la herramienta opere de manera segura y alineada a los estándares educativos y clínicos (Boit & Patil, 2025; Waaler et al., 2025).

Modelos de interacción estructurada con IA

La interacción con modelos de lenguaje de gran escala requiere de enfoques estructurados para maximizar su eficacia educativa y evitar la generación de respuestas superficiales o desalineadas con los objetivos pedagógicos. Los modelos de interacción estructurada organizan el flujo de diálogo entre el estudiante y la IA mediante directrices, arquitecturas de instrucciones predefinidas y estrategias de andamiaje, limitando la aleatoriedad y guiando al sistema hacia la resolución profunda de problemas.

Uno de los modelos más eficaces y documentados en la literatura es la «cadena de pensamiento». Este enfoque instruye al modelo de lenguaje para que desglose problemas complejos en una secuencia de pasos de razonamiento intermedios lógicos antes de emitir una respuesta o conclusión (Abedi et al., 2023; Bassner et al., 2024; Diao et al., 2024; Santos, 2023). La evidencia indica que el uso de cadenas de pensamientos aumenta significativamente la calidad y precisión de los procesos relacionados con la resolución de

problemas frente a enfoques de cero intentos (Huda et al., 2025). Una evolución de este concepto es el «árbol de pensamientos», un modelo estructurado que permite a la IA ramificar el problema inicial en múltiples líneas de pensamiento, explorando y evaluando diversas soluciones potenciales de manera simultánea.

A nivel de arquitectura de interacción, destaca también el uso del *meta-prompting* y los enfoques de múltiples agentes. En este modelo, la IA actúa como un director central que descompone una tarea extensa, asigna las sub-tareas a diferentes «modelos expertos» o identidades simuladas con instrucciones detalladas, y supervisa la comunicación para sintetizar una respuesta final rigurosa (Suzgun & Kalai, 2024). En el contexto específico de los sistemas de tutoría inteligente, se han desarrollado modelos de «instrucciones de doble capa» o estrategias de «ingeniería de *prompts*» híbridas. Estos modelos utilizan una capa estática, que contiene plantillas pedagógicas predefinidas (por ejemplo, reglas sobre cómo guiar sin dar la respuesta), combinada con una capa dinámica que inyecta metadatos del estudiante y contenido recuperado en tiempo real, permitiendo que la interacción adapte su nivel de complejidad al perfil del alumno (Z. Liu et al., 2025).

En el ámbito de la educación médica y las ciencias de la salud, la precisión del contexto es vital. Modelos de estructuración de *prompts* como el modelo SMART han demostrado que encapsular detalladamente el contexto clínico, el historial del paciente y los síntomas dentro de la instrucción inicial reduce las suposiciones de la IA, generando respuestas significativamente más precisas, fiables y adaptadas a escenarios médicos complejos (Vaira et al., 2025).

La estructuración de la interacción no solo recae en el diseño interno del *chatbot*, sino también en la forma en que se guía al estudiante para interactuar con la herramienta. Las investigaciones identifican estrategias de orientación orientadas al alumno, tales como

el «cuestionamiento metacognitivo», el cual incentiva a los estudiantes a reflexionar sobre cómo formularán su pregunta y qué estrategia utilizarán antes de interactuar con la IA (Kumar et al., 2024). Otra estrategia fundamental es el «enfoque centrado en el alumno primero», un modelo de interacción donde se requiere que el estudiante intente resolver el problema de forma independiente en una primera instancia, acudiendo al *chatbot* posteriormente solo para refinar, debatir o evaluar su respuesta inicial. Esta aproximación se complementa con el «*prompting* basado en metacognición», que estructura la IA para emular procesos cognitivos humanos, obligando al modelo a pasar por etapas de comprensión, juicio inicial, reevaluación crítica y decisión final con declaración de confianza (Y. Wang & Zhao, 2024).

Beneficios educativos reportados

La revisión sistemática de la literatura evidencia que la integración de la IA generativa y los *chatbots* educativos produce un impacto positivo y multidimensional en el proceso de enseñanza-aprendizaje. Uno de los beneficios fundamentales es la mejora significativa en el rendimiento académico y la retención del conocimiento (Labadze et al., 2023). El uso de tutores virtuales permite a los estudiantes asimilar conceptos complejos con mayor facilidad, obtener asistencia instantánea para la resolución de problemas y alcanzar niveles más altos de aprendizaje en comparación con entornos tradicionales (Essel et al., 2025). Por ejemplo, evaluaciones empíricas recientes han demostrado que el uso de estas herramientas resulta altamente efectivo para clarificar dudas específicas (reportado en un 91.4% de los casos) y facilitar la comprensión profunda de la materia (95.7%) (Sánchez-Vera, 2025).

En el plano cognitivo, la interacción estructurada con la IA fomenta el desarrollo del pensamiento crítico, el razonamiento explícito y las habilidades de resolución de problemas

(Akram et al., 2025; Al-Abdullatif et al., 2023; Labadze et al., 2023). Al actuar como un andamiaje o tutor socrático, la IA estimula la conciencia metacognitiva y el aprendizaje autorregulado; es decir, los estudiantes adquieren una mayor destreza para planificar, monitorear su progreso, autoevaluarse y ajustar sus estrategias de estudio de forma independiente. Se ha comprobado que este soporte adaptativo y paso a paso reduce la frustración ante tareas difíciles, minimiza la sobrecarga cognitiva de los alumnos y les permite concentrar sus recursos mentales en el análisis y la aplicación del conocimiento (Long et al., 2026; Schei et al., 2024).

Desde una perspectiva afectiva y motivacional, la literatura subraya incrementos sustanciales en el compromiso estudiantil, el interés por la asignatura y la motivación intrínseca (Elnaffar et al., 2026; Huang et al., 2025; Miclin et al., 2024; Sánchez-Vera, 2025; Usher & Amzalag, 2025). El acceso ininterrumpido a retroalimentación personalizada y la posibilidad de interactuar en un entorno virtual libre de juicios reducen significativamente la ansiedad académica e incrementan la curiosidad intelectual de los estudiantes para explorar temas más allá del currículo básico.

Asimismo, un beneficio ampliamente documentado es el fortalecimiento de la autoeficacia (Al-Abdullatif et al., 2023; Elnaffar et al., 2026; Schei et al., 2024). Los alumnos que interactúan regularmente con *chatbots* educativos y reciben orientación adaptativa reportan una mayor confianza en sus propias capacidades para enfrentar y superar desafíos académicos. Esta percepción de competencia no solo mejora el involucramiento activo del estudiante, sino que promueve una mayor autonomía, convirtiéndolos en aprendices más independientes y preparados para la autogestión del conocimiento a lo largo de su vida profesional.

Limitaciones y riesgos de la IA en educación

Si bien la IA generativa y los *chatbots* ofrecen beneficios sustanciales para la enseñanza y el aprendizaje, su integración en el ámbito educativo conlleva una serie de limitaciones, desafíos técnicos y riesgos éticos que han generado profunda preocupación en la comunidad académica (Davar et al., 2025; Giannakos et al., 2025). La literatura actual clasifica estos riesgos en varias dimensiones críticas que deben gestionarse para evitar efectos contraproducentes en el estudiante.

Precisión, confiabilidad y alucinaciones. Una de las debilidades más documentadas de los modelos de lenguaje de gran escala es su propensión por generar alucinaciones. Esto ocurre cuando la IA produce información que suena altamente plausible, coherente y segura, pero que es fácticamente incorrecta, inventada o sin sentido (Abedi et al., 2023; Davar et al., 2025; Li et al., 2024; Nartthanarung et al., 2026; Zastudil et al., 2023). Además, la IA carece de una verdadera comprensión del contexto (Farrokhnia et al., 2024) y su conocimiento puede estar limitado a la fecha de corte de su base de datos de entrenamiento, lo que resulta en respuestas desactualizadas (Lo, 2023). En campos que requieren alta precisión, como la educación médica o las ciencias exactas, el uso de información falsa o referencias bibliográficas inventadas representa un peligro significativo si el estudiante no verifica los datos con fuentes confiables.

Sobredependencia y deterioro de habilidades cognitivas. El riesgo pedagógico más destacado es la sobredependencia de los estudiantes hacia la tecnología (Elnaffar et al., 2026; Li et al., 2024). La inmediatez y facilidad para obtener respuestas completas puede llevar a los alumnos a utilizar la IA como una muleta cognitiva, optando por el camino de menor resistencia en lugar de involucrarse profundamente con el material. Estudios empíricos han demostrado que los estudiantes que dependen excesivamente de la IA

durante la resolución de problemas pueden experimentar un declive en sus competencias cuando se les retira el acceso a la herramienta (Bastani et al., 2025). Además, esta dependencia amenaza con inhibir el desarrollo de habilidades de orden superior indispensables, tales como el pensamiento crítico, el razonamiento analítico, la creatividad y la capacidad de investigación independiente (Ali et al., 2024; Cacho, 2024; Farrokhnia et al., 2024; Gupta et al., 2025; Manobanda et al., 2026). En lugar de fomentar un aprendizaje profundo, por el contrario, el mal uso de la IA puede promover un aprendizaje superficial y mecánico.

Integridad académica y plagio. La capacidad de *ChatGPT* para generar respuestas casi indistinguibles del trabajo humano ha desencadenado una crisis de integridad académica (Chen et al., 2023b, 2023b; Williams, 2024). Existe un alto riesgo de que los estudiantes presenten contenido generado por IA como propio, lo que no solo constituye una forma de deshonestidad académica o plagio, sino que también dificulta enormemente la labor de los docentes para evaluar el verdadero nivel de dominio de los alumnos (Ciacia et al., 2025; Grassini, 2023; Labadze et al., 2023). Además, las herramientas tradicionales de detección de plagio a menudo resultan ineficaces o producen falsos positivos/negativos ante el contenido generado por IA, difuminando los límites de la autoría legítima (Gill et al., 2024).

IA como complemento del aprendizaje médico

La literatura médica hodierna concuerda que la IA generativa debe integrarse en la formación en ciencias de la salud estrictamente como un recurso complementario, y no como un sustituto del educador humano, del razonamiento clínico independiente ni de la experiencia práctica con pacientes reales (Abd-alrazaq et al., 2023; Domínguez et al., 2025). La función principal de esta tecnología es operar como un «asistente de enseñanza

virtual» (Tsang, 2023; X. Xu et al., 2024) o un mentor interactivo que respalda y enriquece las metodologías de instrucción tradicionales, facilitando una experiencia de aprendizaje más accesible, dinámica y personalizada.

Uno de los aportes complementarios más significativos de la IA generativa radica en su capacidad para optimizar el procesamiento de información médica. Los grandes modelos de lenguaje actúan como herramientas interactivas capaces de explicar conceptos anatómicos o fisiopatológicos de alta complejidad utilizando un lenguaje más sencillo y adaptado al nivel de comprensión del alumno (Araujo & Cruz-Correia, 2024; Ghorashi et al., 2023; López, 2025). Al resumir literatura médica extensa y organizar datos clínicos dispersos de forma estructurada, la IA reduce la «carga cognitiva extraña» del estudiante; esto le permite reasignar sus recursos mentales desde la simple búsqueda de información hacia la integración conceptual, la generación de hipótesis y la aplicación práctica del conocimiento (Nartthanarung et al., 2026).

Como complemento al estudio autónomo y la preparación de evaluaciones, la tecnología facilita la creación instantánea de materiales didácticos adaptativos. Los estudiantes y docentes emplean plataformas como *ChatGPT* para automatizar tareas que tradicionalmente consumen mucho tiempo, tales como la generación de tarjetas de memoria para el recuerdo activo, la elaboración de guías de estudio personalizadas y la creación de bancos de preguntas de opción múltiple con el formato de exámenes de licenciamiento médico. Este enfoque complementario atiende las lagunas de conocimiento individuales, permitiendo al estudiante autoevaluarse y enfocar su revisión en áreas de debilidad específicas.

En la dimensión del razonamiento y la práctica clínica, la IA generativa actúa como un valioso andamiaje al ofrecer entornos de simulación seguros. Funcionado como

pacientes virtuales, los *chatbots* brindan a los alumnos la oportunidad de realizar entrevistas médicas, formular diagnósticos diferenciales y proponer planes de manejo sin el riesgo de causar daño en la vida real, superando así las barreras de los altos costos de los centros de simulación tradicionales (Peralta Ramirez et al., 2025). Asimismo, la literatura reporta el uso de la IA para la reflexión clínica asistida; en este modelo complementario, tras finalizar un caso o rotación, el estudiante dialoga con el sistema para examinar críticamente las decisiones tomadas, justificar su razonamiento analítico e, incluso, explorar sus respuestas emocionales frente al escenario clínico, lo que fomenta un aprendizaje metacognitivo profundo (Chan & Colloton, 2024).

Finalmente, la eficacia de la IA como verdadero complemento pedagógico depende de que se enmarque bajo un pensamiento crítico riguroso. Al existir riesgos documentados de respuestas inexactas o alucinaciones dentro del contenido generado, el uso de la herramienta no exime al estudiante de su responsabilidad intelectual; por el contrario, exige que este asuma un rol activo en la verificación de la información contrastándola con literatura médica basada en evidencia y referencias bibliográficas fidedignas. Cuando se implementa bajo estas directrices reflexivas y de apoyo, la IA trasciende la simple transferencia de datos para convertirse en un catalizador de la curiosidad y la resolución autónoma de problemas médicos (Miclin et al., 2024).

7.1- Estado del arte

Investigaciones internacionales sobre IA generativa en educación médica

La investigación internacional sobre la IA generativa (IAG) en la educación médica ha experimentado un crecimiento acelerado y exponencial, impulsado fundamentalmente por el lanzamiento de modelos de lenguaje de gran escala como *ChatGPT* (Gordon et al., 2024). A nivel global, la literatura científica se ha centrado en explorar las capacidades de

estas herramientas tecnológicas para transformar la enseñanza, el aprendizaje y la evaluación en las ciencias de la salud, abarcando desde la educación médica de pregrado hasta la formación de posgrado y el desarrollo profesional continuo.

Las revisiones de alcance y sistemáticas más recientes agrupan estas investigaciones internacionales en cuatro ejes temáticos principales:

Evaluación de conocimientos médicos y rendimiento en exámenes

estandarizados. Una gran proporción de la literatura internacional inicial se enfocó en evaluar si los grandes modelos de lenguaje poseían la base de conocimientos necesaria para ser útiles en medicina. Múltiples investigaciones evaluaron el desempeño de *ChatGPT* frente a bancos de preguntas y exámenes estandarizados de alta exigencia. Estudios pioneros como los de Kung et al. (2023) y Gilson et al. (2023) demostraron que *ChatGPT* (en sus versiones 3.5 y 4) era capaz de alcanzar o superar el umbral de aprobación del examen de licencias médicas de los Estados Unidos. Estos hallazgos se han replicado internacionalmente en exámenes nacionales y de especialidad en países como el Reino Unido, Japón, Francia, Alemania, Taiwán, Corea del Sur, entre otros (Gordon et al., 2024). Aunque el rendimiento ha sido superior en exámenes de medicina general y en idioma inglés, estos estudios validaron a nivel global que los modelos de lenguaje de gran escala poseen una base de conocimiento sustancial que justifica su adopción como herramienta de asistencia educativa.

IA generativa como tutor virtual y simulador para el razonamiento clínico. En el ámbito de la instrucción, las investigaciones internacionales destacan el potencial de la IA generativa para actuar como andamiaje cognitivo y simulador de pacientes virtuales. Estudios revisados por Gordon et al. (2024) identificaron implementaciones donde la IA se fusiona con sistemas de tutoría inteligente para permitir a los estudiantes de medicina

interactuar con pacientes virtuales mediante lenguaje natural. Esto crea entornos dinámicos y seguros donde los alumnos pueden practicar la recopilación de información, la generación de hipótesis y la justificación diagnóstica. Asimismo, estudios empíricos recientes, como el ensayo cruzado aleatorizado de Nartthanarung et al. (2026) realizado en estudiantes de medicina durante sus rotaciones clínicas, han demostrado que las intervenciones didácticas asistidas por *ChatGPT* y herramientas de búsqueda mejoran significativamente las puntuaciones en pruebas de razonamiento clínico. Estas investigaciones enfatizan que el verdadero impacto educativo no depende únicamente de la novedad de la IA, sino de cómo se integra pedagógicamente en las actividades de aprendizaje.

Generación de contenido educativo y apoyo docente. A nivel del profesorado, investigaciones multinacionales han explorado el uso de la IA generativa para la creación de casos clínicos, guías de estudio y reactivos de evaluación (López, 2025). El estudio de Cheung et al. (2023) comparó preguntas de opción múltiple generadas por *ChatGPT* con aquellas creadas por académicos humanos. Los resultados demostraron que la IA puede redactar viñetas clínicas de manera rápida y eficiente para optimizar el tiempo de los educadores, aunque advierten que estas herramientas aún enfrentan limitaciones para generar preguntas que evalúen el pensamiento de orden superior de forma independiente, por lo que se requiere de la evaluación experta humana continua.

Vacíos metodológicos y advertencias sobre la evidencia actual. A pesar de la proliferación de artículos y el entusiasmo global, las revisiones sistemáticas internacionales advierten sobre una crítica escasez de evidencia empírica rigurosa respecto al impacto real en el aprendizaje. El estudio de Hale et al. (2024) concluyó que la gran mayoría de la literatura está compuesta por artículos de opinión, comentarios o evaluaciones técnicas, y que existe un vacío de evidencia cuantitativa sobre el uso de la IA generativa para mejorar

directamente los resultados de los estudiantes. En sintonía, la revisión de Feigerlova et al. (2025) señala que los pocos estudios experimentales disponibles presentan debilidades metodológicas sustanciales: son de muy corta duración, utilizan tamaños de muestra pequeños, en su mayoría se limitan a un solo centro y carecen de métricas validadas de desempeño y seguimiento a largo plazo. La literatura advierte sobre la necesidad imperativa de enseñar a los estudiantes a verificar la información, ya que el riesgo de alucinaciones sigue siendo un desafío tecnológico que puede afectar la precisión del diagnóstico y las decisiones de manejo (Heston & Khun, 2023; Nartthanarung et al., 2026).

Investigaciones regionales sobre IA generativa y aprendizaje universitario

La investigación sobre la IA generativa en la educación superior en América Latina ha experimentado un aumento constante, impulsada por la necesidad de integrar estas tecnologías en la formación de los universitarios, particularmente en el área de las ciencias de la salud (Domínguez et al., 2025). Los análisis bibliométricos regionales, basados en bases de datos como SciELO, revelan que países como Brasil, México y Cuba están liderando actualmente la producción científica y las investigaciones sobre la IA aplicada a la salud y educación en Latinoamérica (González et al., 2025).

Sin embargo, a diferencia de otras regiones del mundo, la adopción de la IA generativa en el aprendizaje universitario latinoamericano enfrenta barreras y desafíos estructurales significativos. La literatura advierte que la implementación de la IA en la educación superior de la región se ve obstaculizada por limitaciones en la infraestructura tecnológica, la falta de capacitación docente y la brecha en el acceso equitativo (Manobanda et al., 2026). De igual manera, las deficiencias en la conectividad a internet representan un retraso importante para la adopción de tecnologías de aprendizaje avanzado en comparación con regiones como Asia-Pacífico.

Las investigaciones regionales subrayan que, aunque Latinoamérica avanza en la adopción de chatbots educativos con altos niveles de aceptación por parte del alumnado, sigue siendo imperativo desarrollar marcos normativos institucionales y superar los retos de infraestructura para garantizar un uso ético, equitativo y responsable en las universidades.

Pertinencia del presente estudio dentro del contexto académico actual

La irrupción de la IA generativa, particularmente *ChatGPT*, ha generado un punto de inflexión sin precedentes en la educación superior y médica (Miclin et al., 2024). En su fase inicial, la adopción masiva de esta tecnología provocó una profunda preocupación en la comunidad educativa centrada en la amenaza a la integridad académica, el riesgo de plagio y la posible obsolescencia de los métodos de evaluación tradicionales (Jaleel et al., 2025; T. Xu et al., 2024). Sin embargo, el contexto académico actual ha transitado desde una postura de escepticismo, prohibición o pánico hacia la necesidad ineludible de integrar estas herramientas tecnológicas de manera ética, crítica y responsable en los procesos de enseñanza-aprendizaje (AlAfnan et al., 2023). En este escenario de transición, el presente estudio resulta altamente pertinente al abandonar la visión de la IA como un simple facilitador de respuestas o instrumento de trampa, para evaluarla de forma estructurada como un tutor virtual y pedagógico.

Desde una perspectiva metodológica, la pertinencia de esta investigación se fundamenta en la necesidad urgente de evidencia empírica rigurosa y objetiva. A pesar de la vasta cantidad de literatura publicada desde finales de 2022, las revisiones sistemáticas actuales advierten que la gran mayoría de los estudios disponibles se limitan a artículos de opinión, ensayos descriptivos, análisis de percepciones cualitativas o validaciones de rendimiento en exámenes, existiendo un notable vacío de diseños cuasi experimentales. Existe una carencia crítica de investigaciones que midan objetivamente el impacto real de

la IA en el rendimiento académico y en la comprensión de los estudiantes dentro de un entorno de aula. Al emplear un diseño cuasi experimental con mediciones de impacto pre y post intervención, este estudio aporta la evidencia cuantitativa necesaria para validar empíricamente la utilidad de estas herramientas.

7.2- Teoría y conceptualizaciones asumidas

Teoría del aprendizaje significativo

El aprendizaje significativo se aleja de la memorización mecánica y pasiva, enfocándose en la integración profunda y aplicada de la información. En el contexto de la educación superior apoyada por tecnología, las investigaciones destacan que herramientas de IA generativa, como *ChatGPT*, pueden actuar como mediadores eficaces que aclaran conceptos complejos, fomentan el diálogo y contextualizan el contenido, elementos fundamentales para alcanzar un aprendizaje significativo (Carmo et al., 2025). Su adopción estratégica expande las posibilidades de innovación pedagógica, permitiendo a los docentes crear actividades dinámicas que promueven este nivel de asimilación profunda en los alumnos. Asimismo, la literatura subraya que el aprendizaje verdaderamente significativo implica un proceso de reflexión interna donde el estudiante es capaz de aplicar de manera creativa el nuevo conocimiento y extrapolarlo a contextos inéditos o innovadores (Wood & Moss, 2024).

Aprendizaje significativo según David Ausubel. Según la teoría fundacional de Ausubel (1963, 1978), el aprendizaje significativo ocurre cuando la nueva información se relaciona de manera sustancial (no literal) y no arbitraria con los conceptos relevantes que el estudiante ya posee en su estructura cognitiva. En esta perspectiva, el conocimiento previo del alumno actúa como un "anclaje" indispensable; si el estudiante logra conectar

orgánicamente lo nuevo con lo que ya sabe, la retención de la información es profunda y duradera (Ausubel, 2012).

Esta premisa teórica clásica encuentra un fuerte respaldo práctico en las fuentes analizadas sobre el uso de tecnología educativa y el diseño instruccional mediado por IA. La literatura señala que una de las estrategias pedagógicas más importantes (y donde la IA puede asistir) es la activación y evaluación del conocimiento previo del estudiante, facilitando explícitamente las conexiones entre lo que ya han aprendido y el nuevo material a estudiar (Moundridou et al., 2024). Los sistemas de tutoría o enfoques de aprendizaje guiado direccionan a los estudiantes a través de una fase de «conexiones», donde analizan la información para encontrar patrones y vincular de forma activa el nuevo conocimiento con sus saberes ya existentes. Al aplicar estos principios, el uso de herramientas tecnológicas ayuda a que los estudiantes transiten de una asimilación pasiva hacia la construcción activa de conocimiento, lo cual es la esencia del aprendizaje significativo (L. Wang et al., 2025).

Relación entre conocimientos previos y comprensión. En el ámbito del aprendizaje complejo, la literatura establece firmemente que el nivel de conocimiento previo sobre un campo de estudio está directamente asociado con un aprendizaje y una comprensión más efectivos (Sáiz-Manzanares et al., 2023). La asimilación profunda no ocurre de forma aislada, sino que requiere que el estudiante progrese desde una etapa inicial de recolección de ideas básicas hacia una fase de "conexiones". En esta etapa, el alumno analiza la información para identificar patrones y relaciones, vinculando activamente el nuevo conocimiento con lo que ya sabe (Wood & Moss, 2024). Asimismo, poseer una base de conocimientos previos expandida proporciona al individuo un marco de referencia

sólido sobre el cual construir y mantener su interés, lo que lo impulsa a invertir más tiempo y esfuerzo en el aprendizaje continuo (Huang et al., 2025).

Desde una perspectiva cognitiva, la activación de los conocimientos previos constituye una estrategia esencial de "orientación". Durante esta fase de orientación, es fundamental evaluar y evocar lo que el estudiante ya sabe para aclarar conceptos, abordar conceptos erróneos comunes y facilitar las conexiones cognitivas con el material nuevo (Moundridou et al., 2024). Por esta razón, el diseño de intervenciones eficaces mediadas por tutores virtuales o IA exige que las explicaciones, ejemplos y analogías se adapten rigurosamente al nivel de comprensión y a los conocimientos previos del alumno, operando como un andamiaje que cierra la brecha hacia el nuevo aprendizaje (Looi & Jia, 2025).

Las investigaciones demuestran que los conocimientos previos determinan significativamente la forma en que el estudiante interactúa con las herramientas tecnológicas y comprende la información que estas generan (Güner & Er, 2025). Los estudiantes con un mayor conocimiento previo en la materia son capaces de integrar la asistencia de la IA con sus saberes existentes, utilizándola de manera complementaria para abordar problemas complejos de forma efectiva. Además, la precisión de sus juicios de confianza sobre lo que han aprendido depende en gran medida de este dominio previo (T. Wang et al., 2023). Por el contrario, los estudiantes novatos o con conocimientos previos deficientes pueden enfrentar dificultades para formular estrategias de resolución y corren el riesgo de caer en un aprendizaje superficial si se limitan a aceptar respuestas sin comprender la lógica subyacente.

Constructivismo y aprendizaje activo

El constructivismo se consolida como uno de los marcos teóricos más prominentes para fomentar el pensamiento crítico y las habilidades de resolución de problemas en la

educación superior (Jones et al., 2026). Enraizado en las teorías socioculturales y cognitivas, este paradigma sostiene que el aprendizaje no es un proceso de transferencia pasiva de información, sino que los estudiantes construyen activamente su propia comprensión del mundo a través de la experiencia. Desde esta perspectiva, el constructivismo fundamenta una variedad de métodos de enseñanza centrados en el estudiante, tales como las aulas invertidas y, de manera fundamental, las estrategias de aprendizaje activo. El aprendizaje activo es un enfoque pedagógico que involucra directamente a los estudiantes en su viaje educativo mediante discusiones, resolución de problemas y actividades prácticas, lo cual promueve la asimilación profunda y la aplicación teórica, alineándose a la perfección con la teoría de la carga cognitiva y la zona de desarrollo próximo (Pahi et al., 2024).

Construcción activa del conocimiento. En el modelo constructivista, se abandona el paradigma tradicional centrado en el docente para dar paso a un entorno donde el alumno interactúa, explora y cuestiona (Moundridou et al., 2024). En lugar de recibir información de manera pasiva, los estudiantes construyen su comprensión y conocimiento buscando respuestas a través de la investigación y la resolución de problemas. Esta construcción activa del conocimiento requiere que el individuo se comprometa de forma dinámica con la nueva información, integrándola y construyendo sobre la base de sus saberes previos (Punar Özçelik & Yangın Ekşi, 2024).

El aprendizaje activo complementa este proceso al posicionar a los estudiantes como participantes totalmente involucrados en su formación. Al enfrentarse a tareas desafiantes, los alumnos no se limitan a memorizar hechos, sino que analizan la información para encontrar patrones, formulan hipótesis y elaboran deducciones (L. Wang et al., 2025). De esta forma, el estudiante se convierte en un «epistemólogo»,

involucrándose en el acto de descubrir y reflexionar sobre su propio proceso de pensamiento, lo que da lugar a la consolidación de un conocimiento altamente transferible y aplicable a situaciones del mundo real (Santos, 2023).

Aprendizaje mediado por tecnología. La integración de la IA y los chatbots educativos introduce una nueva dimensión al constructivismo: el aprendizaje mediado por la tecnología. Desde la visión teórica de Seymour Papert y su concepto de «objetos para pensar», la literatura contemporánea expande esta idea hacia los «agentes para pensar» (Santos, 2023). Bajo esta conceptualización, las herramientas digitales y los chatbots basados en IA generativa actúan como recursos interactivos que asisten el desarrollo cognitivo del individuo al involucrarlo en conversaciones dinámicas, hacerle preguntas socráticas y fomentar su capacidad reflexiva.

Desde la perspectiva de la «teoría del actor-red», estas herramientas de IA se consideran participantes activos dentro de las redes que producen conocimiento. Procesan los aportes del estudiante, generan respuestas adaptativas e interactúan asumiendo el rol de un tutor o compañero de aprendizaje (Al-Abdullatif et al., 2023).

En este contexto, la tecnología actúa como un mediador a través del andamiaje cognitivo dentro de la zona de desarrollo próximo (Huda et al., 2025; Looi & Jia, 2025). La zona de desarrollo próximo describe la diferencia entre lo que un alumno puede hacer de forma independiente y lo que puede lograr con la guía de un «otro más conocedor». La IA asume este rol de compañero más competente, adaptando su apoyo dinámicamente al nivel del alumno y proporcionando retroalimentación personalizada sin entregar la respuesta final de forma directa. A medida que el estudiante construye activamente su conocimiento y adquiere dominio, el andamiaje del sistema se retira gradualmente, facilitando así la transición hacia la resolución de problemas de manera autónoma y autorregulada.

Relación teórica entre IA generativa y aprendizaje médico

La incursión de los grandes modelos de lenguaje ha introducido un cambio de paradigma en la educación médica contemporánea. En este escenario, la relación teórica entre la tecnología y el aprendizaje trasciende la simple recuperación de información; la IA generativa se posiciona como un ecosistema dinámico e interactivo capaz de emular funciones cognitivas, personalizar la instrucción y ofrecer retroalimentación en tiempo real (Abd-alrazaq et al., 2023; Heston & Khun, 2023). Esta integración tecnológica tiene el potencial de transformar tanto la adquisición de conocimientos básicos como el desarrollo de habilidades clínicas complejas, siempre que se fundamente en bases pedagógicas sólidas y en la comprensión de sus capacidades y limitaciones técnicas (Nartthanarung et al., 2026).

IA como mediadora del aprendizaje. Desde una perspectiva constructivista, la IA generativa actúa como una poderosa herramienta de mediación pedagógica (Carmo et al., 2025). En lugar de ser un simple repositorio estático de datos, plataformas como *ChatGPT* asumen el rol de un asistente de enseñanza virtual o tutor inteligente capaz de guiar a los estudiantes a través de interacciones conversacionales basadas en lenguaje natural (Abd-alrazaq et al., 2023; Marchena Sekli et al., 2024; Sánchez-Vera, 2025; Tsang, 2023). Esta capacidad le permite a la IA proporcionar un andamiaje cognitivo, ajustando dinámicamente el nivel de dificultad y el tipo de apoyo a las necesidades individuales del alumno (Akram et al., 2025; Nartthanarung et al., 2026).

El principal valor de la IA como mediadora radica en su capacidad para la personalización del aprendizaje y la adaptabilidad. Los grandes modelos de lenguaje pueden descomponer conceptos médicos o fisiopatológicos de alta complejidad en explicaciones más simples y manejables, utilizando analogías y ejemplos adaptados al nivel

de comprensión del estudiante. Al proporcionar respuestas inmediatas y retroalimentación adaptativa en tiempo real, la herramienta fomenta un entorno de aprendizaje autodirigido y activo, reduciendo la ansiedad frente a los errores y estimulando la curiosidad intelectual del alumno para explorar temas más allá del currículo tradicional (Miclin et al., 2024; X. Xu et al., 2024).

IA y apoyo al razonamiento clínico. El razonamiento clínico es una competencia central para los médicos, que abarca la recopilación de información, la representación del problema, la generación de hipótesis, el diagnóstico diferencial y la planificación del manejo. En este sentido, la IA generativa se ha convertido en un andamiaje fundamental para entrenar estas habilidades mediante la creación de simuladores de pacientes virtuales y estudios de caso interactivos (Abd-alrazaq et al., 2023; Gordon et al., 2024; López, 2025; Peralta Ramirez et al., 2025). Estos sistemas permiten a los estudiantes practicar la toma de decisiones, formular preguntas diagnósticas y recibir retroalimentación instantánea en un entorno seguro y controlado, eliminando el riesgo de daño a pacientes reales.

Frente a escenarios clínicos desconocidos, los estudiantes pueden utilizar los *chatbots* para introducir elementos pertinentes de la historia clínica de un paciente y obtener un marco de diagnósticos diferenciales, lo que amplía sus perspectivas y estimula la lectura y el análisis crítico. Desde el punto de vista de la carga cognitiva, las síntesis generadas por la IA reducen la carga cognitiva extraña al organizar información dispersa y ofrecer explicaciones estructuradas. Esto libera recursos mentales para que el estudiante de medicina se concentre en la integración del conocimiento, la aplicación y la generación de hipótesis complejas. La evidencia empírica reciente demuestra que el uso estructurado de herramientas como *ChatGPT* mejora significativamente el rendimiento a corto plazo en tareas de razonamiento clínico (Nartthanarung et al., 2026).

IA como herramienta complementaria y no sustitutiva del aprendizaje

tradicional. A pesar de sus formidables capacidades, la literatura subraya unánimemente que la IA generativa debe implementarse como un complemento de la enseñanza tradicional, y en ningún caso como un sustituto del educador humano o del juicio clínico independiente. Las interacciones formativas requieren de la experiencia, el contexto y la adaptabilidad del educador para lidiar con situaciones imprevistas y fomentar un análisis profundo.

El uso irrestricto de la IA conlleva el riesgo de sobredependencia. Si los estudiantes delegan la resolución de problemas en la herramienta, pueden experimentar una atrofia en el desarrollo de competencias fundamentales como la autonomía intelectual, la creatividad, la capacidad de comunicación y el propio razonamiento clínico (Abd-alrazaq et al., 2023; Domínguez et al., 2025; Iqbal et al., 2025; Narthanarung et al., 2026). Por ello, las intervenciones con IA requieren de la supervisión docente y del diseño de instrucciones que incentiven el esfuerzo reflexivo del alumno (Peralta Ramirez et al., 2025; Watts et al., 2025).

Adicionalmente, los modelos de lenguaje presentan limitaciones inherentes. En disciplinas de alto riesgo como la medicina, aceptar información errónea puede tener consecuencias graves para la seguridad del paciente. En consecuencia, el uso de la IA exige que los estudiantes desarrollen habilidades de verificación y evaluación crítica de la literatura médica para contrastar las respuestas del modelo.

Finalmente, el componente humano en la atención sanitaria es irremplazable. El cuidado compasivo, la empatía y la relación médico-paciente no pueden ser emulados por un algoritmo, por lo que la IA debe usarse para potenciar y no para deshumanizar la práctica médica.

8.- Metodología

8.1- Tipo de investigación

Estudio de enfoque cuantitativo, cuasi experimental, prospectivo, analítico y de mediciones repetidas, siguiendo los lineamientos metodológicos descritos por Hernández Sampieri & Mendoza Torres (2023). Es cuantitativo porque emplea la recolección sistemática de datos mediante instrumentos estructurados (cuestionarios de selección múltiple y preguntas dicotómicas) con el propósito de medir cambios en el desempeño académico y en la percepción de los estudiantes después de una intervención educativa basada en IA. Es cuasi experimental debido a que evalúa el efecto de una intervención educativa con el uso de una IA generativa personalizada como tutor virtual grupal sobre variables de interés medidas antes y después de su implementación, sin asignación aleatoria individual de los participantes. Es prospectivo porque la información fue recolectada directamente durante el desarrollo del estudio, registrando los resultados conforme ocurrieron las distintas fases del procedimiento experimental. Es analítico porque busca identificar diferencias en el desempeño y en las percepciones de los estudiantes antes y después de la intervención, explorando el posible efecto de la tutoría mediante IA y la comprensión de textos médicos complejos. Corresponde a un diseño de mediciones repetidas debido a que cada participante fue evaluado antes y después de la interacción con la IA, permitiendo comparar cambios intraindividuales en el razonamiento clínico y percepción del aprendizaje.

8.2- Población y selección de la muestra

Población

La población está conformada por los estudiantes activos de cuarto año de la carrera de Medicina de la sede Doral de la Universidad Central de Nicaragua (UCN), durante el primer semestre de 2026. Según los listados académicos disponibles, cuarto año se encuentra distribuido oficialmente en tres grupos: grupo A, con 22 estudiantes; grupo B, con 23 estudiantes; y grupo C, con 21 estudiantes, para un total poblacional de 66 estudiantes.

Muestra

La muestra es de tipo no probabilística por conveniencia, con intención censal, integrada por los estudiantes que acepten participar voluntariamente en el estudio. Debido a que la población accesible es reducida, se procuró invitar a la totalidad de la población. Se aplicaron criterios de eliminación a aquellos participantes que no cumplieron con las instrucciones metodológicas establecidas durante la intervención.

8.3- Técnicas e instrumentos de recolección de datos utilizados

Técnicas

La técnica de recolección de datos fue la evaluación estructurada pretest–postest, complementada con preguntas dicotómicas de percepción y el registro de interacción grupal (entre 5 a 10 estudiantes) con IA. Esta técnica permitió obtener información cuantitativa sobre el desempeño académico, el razonamiento clínico y las percepciones de los estudiantes antes y después de la intervención educativa con *ChatGPT* como tutor virtual grupal.

El procedimiento de recolección de datos se desarrolló por fases:

Primera fase. Se realizó una explicación general del estudio, en la cual se informó a los participantes sobre el propósito de la investigación, las fases del procedimiento, el carácter voluntario de su participación, las consideraciones éticas relacionadas con la confidencialidad de la información y las instrucciones generales para el desarrollo de la actividad. Esta fase tuvo una duración de 5 minutos.

Segunda fase. Los estudiantes recibieron el texto médico seleccionado, correspondiente a las secciones «La vía para el desarrollo del asma» y «Diagnóstico y valoración» del «Capítulo 298: asma», del libro Harrison. Principios de Medicina Interna, 22.^a edición (Isarel, 2025). Durante esta fase, los participantes realizaron una lectura individual del material, pudiendo utilizar estrategias personales de estudio como elaboración de resúmenes, esquemas o mapas conceptuales. Durante este momento no se permitió el uso de IA, internet, apuntes externos u otros recursos distintos al texto proporcionado. Esta actividad tuvo una duración de 20 minutos.

Tercera fase. Se aplicó un primer cuestionario de selección múltiple, orientado a valorar la comprensión inicial del texto y el razonamiento clínico derivado de la lectura. Este cuestionario constituyó la medición preintervención, con una duración de 5 minutos.

Cuarta fase. Los estudiantes completaron una primera encuesta con preguntas dicotómicas, dirigida a valorar sus percepciones iniciales sobre la claridad del texto, comprensión alcanzada y confianza para responder preguntas clínicas relacionadas con el contenido estudiado. Esta fase tuvo una duración de 5 minutos.

Quinta fase. Se desarrolló una fase de discusión grupal, en la cual los estudiantes identificaron las partes del texto que les resultaron más complejas o difíciles de comprender. A partir de esta discusión, cada grupo formuló un máximo de tres preguntas dirigidas a *ChatGPT 5.5* (ver Figura 1). Con el propósito de estandarizar la interacción y

garantizar que las consultas se orienten al aprendizaje y no a la resolución directa de las evaluaciones, las preguntas se ajustaron a categorías previamente definidas.

La interacción con *ChatGPT* se realizó mediante una única sesión proyectada en una pantalla a través de un proyector. El investigador fue el encargado de introducir las preguntas consensuadas por el grupo en la plataforma, mientras todos los participantes observarán simultáneamente las consultas y respuestas generadas. Esta modalidad permitió evitar el uso individual de dispositivos electrónicos durante la intervención, garantizar que todos los estudiantes reciban la misma información y facilitar el control y registro de las preguntas formuladas.

Categoría 1. Definición o aclaración. Tiene como objetivo aclarar conceptos fundamentales. Las preguntas deberán iniciar con expresiones como: «¿Qué es...?», «¿En qué consiste...?», «Define...» o «Resume...».

Categoría 2. Mecanismo o causalidad. Tiene como objetivo favorecer la comprensión de procesos fisiopatológicos. Las preguntas podrán iniciar con: «¿Por qué...?», «¿Cómo ocurre...?» o «Explica cómo...».

Categoría 3. Relación clínica. Tiene como objetivo integrar hallazgos clínicos con mecanismos fisiopatológicos. Las preguntas podrán iniciar con: «¿Cómo se relaciona...?» o «¿Qué relación existe entre...?».

Categoría 4. Diferenciación diagnóstica. Tiene como objetivo fortalecer el razonamiento clínico mediante la comparación de entidades clínicas. Las preguntas podrán iniciar con: «¿Cómo diferenciar...?» o «¿Qué diferencia existe entre...?».

Categoría 5. Aplicación clínica. Tiene como objetivo promover la transferencia del conocimiento a escenarios clínicos hipotéticos. Las preguntas podrán iniciar con: «¿Qué ocurre si...?», «¿Qué consecuencias tiene...?» o «¿Cómo afectaría...?».

La IA fue utilizada como tutor virtual grupal y no como herramienta para resolver directamente las preguntas del cuestionario. La personalización del *chatbot* está basado en el modelo de interacción SMART, detallado por Vaira et al. (2025) (ver Tabla 1).

Asimismo, no se permitió formular preguntas que soliciten respuestas textuales de las evaluaciones ni preguntas ajenas al contenido del material de estudio. La discusión grupal y la interacción con *ChatGPT* tuvo una duración de 5 minutos.

Sexta fase. Posterior a la interacción con *ChatGPT*, se aplicó un segundo cuestionario de selección múltiple, equivalente al primero en nivel de dificultad y orientación clínica, con el propósito de valorar cambios en el desempeño académico y razonamiento clínico posterior a la intervención. Esta actividad tuvo una duración de 5 minutos.

Séptima fase. Los participantes completaron una segunda encuesta con preguntas dicotómicas, dirigida a valorar sus percepciones posteriores a la asesoría con IA, incluyendo claridad de la explicación, comprensión alcanzada, confianza para responder y utilidad percibida de *ChatGPT* como herramienta complementaria para el aprendizaje. Esta fase tuvo una duración de 5 minutos.

La duración total estimada del procedimiento fue de 50 minutos por grupo de participantes.

Instrumentos de recolección

Para la recolección de datos se diseñó un instrumento digital utilizando la plataforma Google Forms, estructurado en ocho secciones secuenciales que permiten desarrollar las distintas fases del procedimiento experimental y recopilar la información necesaria para el cumplimiento de los objetivos de la investigación. El instrumento integra cuestionarios de selección múltiple y escalas dicotómicas, orientadas a evaluar el

desempeño académico, el razonamiento clínico y las percepciones de los participantes antes y después de la intervención con IA generativa.

La primera sección corresponde a la información general y consentimiento informado. En ella se presenta el propósito de la investigación, las actividades que realizará el participante, la duración estimada del procedimiento, los aspectos relacionados con la confidencialidad de la información y el carácter voluntario de la participación. Asimismo, se solicita la aceptación o rechazo del consentimiento informado como requisito para continuar con el instrumento.

La segunda sección está destinada a la recopilación de datos sociodemográficos y académicos de los participantes (ver Tabla 2). Se incluyen variables como sexo, edad, departamento de procedencia, grupo académico, rotación clínica actual y frecuencia de uso de IA generativa con fines académicos. Esta información permitió caracterizar la muestra estudiada.

La tercera sección corresponde a las instrucciones para la lectura del material de estudio. En esta fase se presentan las indicaciones relacionadas con la lectura individual del texto seleccionado, las actividades permitidas durante el estudio del material y las restricciones respecto al uso de recursos externos, incluyendo herramientas de IA.

La cuarta sección contiene el Cuestionario I: Comprensión inicial del texto y razonamiento clínico, constituido por cinco preguntas clínicas de selección múltiple orientadas a evaluar la comprensión inicial del contenido estudiado y la capacidad de aplicar la información a situaciones clínicas (ver Tabla 3). Esta sección constituye la medición preintervención del desempeño académico y razonamiento clínico.

La quinta sección corresponde a la evaluación de la percepción inicial sobre el material de estudio. Se emplearon escalas dicotómicas para valorar la percepción de

claridad, comprensión y confianza de los estudiantes después de la lectura del texto y antes de la interacción con la IA (ver Tabla 4).

La sexta sección presenta las instrucciones para la discusión grupal e interacción con ChatGPT. En ella se describen los objetivos de la actividad, las categorías permitidas para la formulación de preguntas, las restricciones de uso y el procedimiento para la interacción con la IA generativa utilizada como tutor virtual grupal.

La séptima sección contiene el Cuestionario II: Comprensión posterior a la tutoría con IA y razonamiento clínico, integrado por cinco preguntas clínicas de selección múltiple equivalentes en dificultad al cuestionario inicial (ver Tabla 5). Su propósito es evaluar posibles cambios en el desempeño académico y razonamiento clínico después de la intervención educativa.

Finalmente, la octava sección corresponde a la evaluación de la percepción posterior a la tutoría con IA. Mediante escalas dicotómicas se valoran la claridad de las explicaciones recibidas, la comprensión alcanzada, la confianza para responder preguntas clínicas y la utilidad percibida de ChatGPT como herramienta complementaria para el aprendizaje de contenidos médicos complejos (ver Tabla 6). Esta información permitió comparar las percepciones de los participantes antes y después de la intervención y analizar la aceptación de la IA como recurso educativo.

8.4- Confiabilidad y validez de los instrumentos

Confiabilidad

La confiabilidad del instrumento se evaluó mediante una prueba piloto aplicada a seis estudiantes de quinto año de la carrera de Medicina de la Universidad Central de Nicaragua, sede Doral, quienes no forman parte de la muestra definitiva del estudio (ver Figura 2). La selección de este grupo responde a que poseen una formación académica

similar a la de la población objetivo y han cursado previamente los contenidos necesarios para comprender el material utilizado en la investigación.

La prueba piloto tuvo como finalidad identificar posibles dificultades relacionadas con la comprensión de las instrucciones, claridad de los ítems, funcionamiento del formulario digital, tiempos de aplicación y desarrollo general del procedimiento metodológico. Asimismo, permitió realizar los ajustes necesarios (como tiempos, métodos o materiales) antes de la aplicación definitiva del instrumento.

La confiabilidad de los cuestionarios clínicos fue fortalecida mediante la revisión realizada por un especialista en Medicina Interna adscrito al Hospital Escuela Manolo Morales Peralta, quien evaluó la pertinencia clínica, claridad, congruencia con el contenido estudiado y capacidad de las preguntas para valorar comprensión y razonamiento clínico. De igual forma, las observaciones obtenidas durante la prueba piloto fueron consideradas para optimizar la calidad y estabilidad de los instrumentos utilizados en la investigación.

Validez

La validez de contenido del instrumento se estableció mediante juicio de expertos, con el propósito de evaluar la pertinencia, claridad, coherencia y adecuación de los componentes metodológicos utilizados en la investigación. La personalización de ChatGPT como tutor virtual grupal fue revisada por una ingeniera en sistemas con experiencia en IA, quien además se desempeña como Decana de la Facultad de Ingeniería de la Universidad Central de Nicaragua y responsable de la asignatura de IA (ver Figura 3). Su revisión se orientó a evaluar la estructura del *prompt*, las restricciones de interacción establecidas, la coherencia técnica de la configuración y su alineación con los objetivos educativos del estudio.

Posteriormente, las preguntas clínicas utilizadas en los cuestionarios fueron elaboradas mediante la misma IA configurada para la intervención y sometidas a revisión por un especialista en Medicina Interna adscrito al Hospital Escuela Manolo Morales Peralta (ver Figura 4). La evaluación se centró en la pertinencia clínica de los casos planteados, la congruencia con el contenido seleccionado del texto de estudio, el nivel de dificultad de las preguntas y su capacidad para valorar comprensión y razonamiento clínico.

Adicionalmente, las escalas de percepción y la estructura general del instrumento fueron revisadas por una médica con experiencia en educación superior, formación pedagógica y certificación en el uso de IA aplicada a procesos educativos, quien se desempeña como Decana de la Facultad de Ciencias Médicas de la Universidad Central de Nicaragua (ver Figura 5).

Las observaciones y recomendaciones emitidas por los tres expertos fueron incorporadas en la versión final del instrumento, fortaleciendo su validez de contenido y garantizando su pertinencia para evaluar los constructos de interés en la población estudiada.

8.5- Procedimientos para el procesamiento de datos y análisis de la información

Procesamientos de datos

La recolección de datos se realizó mediante un formulario digital elaborado en Google Forms. Una vez finalizada la aplicación del instrumento, las respuestas fueron exportadas automáticamente a una hoja de cálculo de Google Sheets para su almacenamiento y organización inicial. Posteriormente, la base de datos fue revisada para identificar registros incompletos, inconsistencias, duplicidades o errores de digitación que puedan afectar la calidad de la información.

Una vez concluido el proceso de depuración, los datos fueron codificados de acuerdo con la naturaleza de cada variable y exportados al programa IBM SPSS Statistics versión 32 para su procesamiento estadístico. Las variables categóricas fueron transformadas en códigos numéricos para facilitar su análisis, mientras que las variables cuantitativas conservarán su escala original de medición.

Se verificó la integridad de la base de datos antes de iniciar el análisis estadístico, garantizando la correcta correspondencia entre las respuestas registradas en los instrumentos y las variables definidas en la operacionalización. Finalmente, se generó una base de datos definitiva que sirvió como insumo para el análisis descriptivo e inferencial de la investigación.

Análisis de la información

El análisis de la información se realizó utilizando el programa IBM SPSS Statistics versión 32. Inicialmente, se llevó a cabo un análisis descriptivo de las variables sociodemográficas y académicas de los participantes. Las variables cualitativas fueron descritas mediante frecuencias absolutas y porcentajes, mientras que las variables cuantitativas se resumieron mediante medidas de tendencia central y dispersión, incluyendo media, desviación estándar, mediana, rango, rango intercuartílico, asimetría y curtosis, según correspondió a la naturaleza y distribución de los datos.

La variable principal del estudio, correspondiente a la comprensión de textos médicos y aplicación de conocimientos clínicos, fue evaluada mediante la comparación de las puntuaciones obtenidas en los cuestionarios aplicados antes y después de la intervención. Previamente, se determinó la distribución de las diferencias entre ambas mediciones mediante la prueba de normalidad de Shapiro-Wilk. Debido a que las diferencias no presentaron una distribución compatible con la normalidad ($p = 0.001$), se

utilizó la prueba de rangos con signo de Wilcoxon para muestras relacionadas con el propósito de comparar las puntuaciones obtenidas entre ambas mediciones.

Adicionalmente, se calculó el tamaño del efecto mediante el coeficiente r de Rosenthal, con el fin de estimar la magnitud práctica de las diferencias observadas. Se reportaron las medidas descriptivas correspondientes y los estadísticos de prueba asociados.

Las percepciones de claridad, comprensión y confianza fueron analizadas en dos niveles. En primer lugar, cada afirmación individual, medida mediante respuestas dicotómicas, fue descrita mediante frecuencias y porcentajes. Posteriormente, las diferencias entre las respuestas relacionadas con el libro de texto y aquellas relacionadas con *ChatGPT* fueron evaluadas mediante la prueba de McNemar para muestras relacionadas, considerando la naturaleza nominal dicotómica de las variables.

Adicionalmente, se calcularon las diferencias de proporciones entre ambas mediciones con el propósito de cuantificar los cambios observados en las respuestas favorables.

En segundo lugar, para cada una de las dimensiones de percepción, se construyeron puntuaciones globales a partir de la suma de las respuestas favorables de sus respectivas afirmaciones. Debido a la naturaleza ordinal de estas puntuaciones, las comparaciones entre las valoraciones del libro de texto y las valoraciones de *ChatGPT* se realizaron mediante la prueba de rangos con signo de Wilcoxon para muestras relacionadas. Asimismo, se calculó el tamaño del efecto mediante el coeficiente r de Rosenthal, obtenido a partir del estadístico Z de la prueba de Wilcoxon, con el propósito de estimar la magnitud práctica de las diferencias observadas.

La utilidad percibida de *ChatGPT* como herramienta complementaria para la comprensión de textos médicos complejos fue evaluada mediante una puntuación global construida a partir de la suma de las respuestas favorables obtenidas en las afirmaciones

correspondientes. Esta variable fue descrita mediante frecuencias, porcentajes y medidas de tendencia central y dispersión.

En todos los análisis inferenciales se estableció un nivel de significancia estadística de $p < 0.05$ y se reportaron los valores estadísticos correspondientes para facilitar la interpretación de los resultados.

Finalmente, los hallazgos fueron presentados mediante tablas elaboradas a partir de los resultados obtenidos, con el propósito de facilitar la interpretación y discusión de los datos en función de los objetivos e hipótesis planteados en la investigación.

9.- Discusión y resultados

Resultados

Los resultados presentados a continuación corresponden a una muestra de 57 estudiantes de cuarto año de la carrera de Medicina de la Universidad Central de Nicaragua (ver Figura 6). La investigación fue concebida con una intención censal, considerando como universo de estudio a los 66 estudiantes matriculados en el nivel académico seleccionado. No obstante, durante el proceso de depuración de la base de datos se aplicaron los criterios de eliminación previamente establecidos en el protocolo de investigación, excluyéndose aquellos participantes que no completaron la totalidad de las actividades o que no siguieron las instrucciones metodológicas definidas para el desarrollo de la intervención y la recolección de datos. En consecuencia, el análisis final se realizó con los 57 estudiantes que cumplieron íntegramente los requisitos del estudio, garantizando la consistencia y validez de la información obtenida.

Datos sociodemográficos y académicos

La muestra estuvo conformada por 57 estudiantes de cuarto año de medicina, de los cuales el 70.2% (n = 40) correspondió al sexo femenino y el 29.8% (n = 17) al sexo masculino (ver Tabla 7). La edad promedio de los participantes fue de 21.6 ± 3.2 años, con edades comprendidas entre los 18 y 24 años.

En relación con la procedencia, la mayoría de los estudiantes provenía del departamento de Managua (66.7%; n = 38), seguido por la Región Autónoma de la Costa Caribe Norte (12.3%; n = 7). Los departamentos de Granada y Masaya aportaron cada uno el 5.3% (n = 3) de los participantes, mientras que Boaco representó el 3.5% (n = 2). Los

departamentos de Matagalpa, Río San Juan y Rivas, así como la Región Autónoma de la Costa Caribe Sur, contribuyeron individualmente con el 1.8% ($n = 1$) de la muestra.

Respecto al grupo académico, el Grupo B concentró la mayor proporción de estudiantes (38.6%; $n = 22$), seguido por el Grupo A (31.6%; $n = 18$) y el Grupo C (29.8%; $n = 17$). En cuanto a la rotación clínica vigente al momento del estudio, el 78.9% ($n = 45$) se encontraba cursando Medicina Interna, mientras que el 21.1% ($n = 12$) realizaba la rotación de Cirugía.

Finalmente, en relación con el uso previo de IA generativa como apoyo para el estudio de la medicina, ningún estudiante reportó no utilizar este tipo de herramientas. La categoría más frecuente fue “frecuentemente” (38.6%; $n = 22$), seguida de “algunas veces” (31.6%; $n = 18$) y “muy frecuentemente” (22.8%; $n = 13$). Únicamente el 7.0% ($n = 4$) indicó utilizarlas rara vez. Estos resultados evidenciaron una amplia familiaridad de los participantes con las herramientas de IA generativa antes de la intervención realizada.

Comprensión de textos médicos y aplicación de conocimientos clínicos

La Tabla 8 presenta la distribución de respuestas obtenidas en los cuestionarios aplicados antes y después de la intervención con *ChatGPT*. En el pretest, las proporciones de respuestas correctas oscilaron entre el 68.4% y el 98.2%, mientras que en el posttest variaron entre el 43.9% y el 94.7%.

Al comparar ambas mediciones, se observó un incremento en la proporción de respuestas correctas en las preguntas 2 y 4. En la pregunta 2, las respuestas correctas aumentaron de 68.4% ($n = 39$) a 84.2% ($n = 48$), mientras que en la pregunta 4 se incrementaron de 75.4% ($n = 43$) a 94.7% ($n = 54$). Asimismo, se evidenció una reducción en el número de participantes que omitieron responder la pregunta 2, pasando de 14.0% ($n = 8$) en el pretest a 0% en el posttest.

Por otro lado, las preguntas 3 y 5 mostraron una disminución en la frecuencia de respuestas correctas después de la intervención. En la pregunta 3, la proporción de respuestas correctas descendió de 98.2% ($n = 56$) a 80.7% ($n = 46$), mientras que en la pregunta 5 disminuyó de 87.7% ($n = 50$) a 43.9% ($n = 25$), acompañándose de un aumento en las respuestas incorrectas de 5.3% ($n = 3$) a 47.4% ($n = 27$). La pregunta 1 mantuvo una proporción idéntica de respuestas correctas en ambas mediciones (75.4%; $n = 43$).

En conjunto, los resultados evidenciaron variaciones heterogéneas entre los ítems evaluados, observándose mejoras en algunas preguntas y disminuciones en otras, lo que justificó la comparación global de los puntajes obtenidos en ambos cuestionarios mediante pruebas estadísticas inferenciales.

El análisis descriptivo de los puntajes obtenidos en los cuestionarios evidenció una media de 4.05 puntos ($DE = 1.23$) en el pretest y de 3.79 puntos ($DE = 1.05$) en el posttest (ver Tabla 9). En ambas mediciones la mediana fue de 4 puntos, aunque el rango intercuartílico aumentó de 1 en el pretest a 2 en el posttest, lo que sugiere una mayor dispersión de las puntuaciones después de la intervención. Los puntajes máximos alcanzados fueron de 5 puntos en ambos cuestionarios, mientras que los valores mínimos fueron de 0 y 1 punto para el pretest y posttest, respectivamente.

Respecto a la distribución de los datos, las puntuaciones presentaron valores de asimetría negativos tanto en el pretest (-1.532) como en el posttest (-0.525), indicando una tendencia de las puntuaciones a concentrarse en los valores más altos de la escala. Asimismo, la curtosis fue positiva en el pretest (1.994), reflejando una distribución más concentrada alrededor de la media, mientras que en el posttest fue ligeramente negativa (-0.462), indicando una distribución más dispersa. La evaluación de la normalidad mediante

la prueba de Shapiro-Wilk mostró estadísticos de 0.758 para el pretest y 0.876 para el posttest.

Es importante señalar que, aunque los valores de Shapiro-Wilk se presentan con fines descriptivos, la decisión sobre la prueba inferencial utilizada para comparar ambos cuestionarios se fundamentó en la distribución de las diferencias entre las puntuaciones obtenidas por los participantes en ambas mediciones.

Debido a que la distribución de las diferencias entre ambas mediciones no cumplió con el supuesto de normalidad, la comparación se realizó mediante la prueba de rangos con signo de Wilcoxon para muestras relacionadas (ver Tabla 10). Los resultados no evidenciaron diferencias estadísticamente significativas entre los puntajes obtenidos antes y después de la intervención ($Z = -1.745$; $p = 0.081$). Asimismo, el tamaño del efecto calculado mediante el coeficiente r de Rosenthal fue de 0.231, correspondiente a un efecto de magnitud pequeña.

En conjunto, estos hallazgos indicaron que la interacción con *ChatGPT* como tutor virtual grupal no produjo cambios estadísticamente significativos en el desempeño global de los estudiantes en el cuestionario de comprensión de textos médicos y aplicación de conocimientos clínicos, aunque se observó una ligera disminución en la puntuación promedio posterior a la intervención.

Percepciones de claridad, comprensión y confianza sobre el aprendizaje

La distribución de las respuestas evidenció una valoración favorable tanto del material de estudio como de la tutoría realizada mediante *ChatGPT*; sin embargo, en todos los ítems se observaron porcentajes superiores de respuestas favorables hacia la inteligencia artificial (ver Tabla 11). Dentro de la dimensión de claridad, la afirmación con mayor incremento correspondió a la capacidad de facilitar la comprensión de conceptos difíciles,

cuya proporción de respuestas favorables aumentó de 73.7% ($n = 42$) para el libro de texto a 96.5% ($n = 55$) para *ChatGPT*, lo que representó una diferencia de 22.8 puntos porcentuales. Las otras dos afirmaciones de esta dimensión también mostraron incrementos, aunque de menor magnitud, correspondientes a 7.0 y 3.5 puntos porcentuales.

En la dimensión de comprensión se observaron diferencias favorables para *ChatGPT* en los tres ítems evaluados. La mayor variación se registró en la afirmación relacionada con la capacidad percibida para explicar los conceptos abordados, la cual pasó de 49.1% ($n = 28$) en el libro de texto a 84.2% ($n = 48$) tras la interacción con ChatGPT, representando un incremento de 35.1 puntos porcentuales. Asimismo, la percepción de haber comprendido adecuadamente el contenido aumentó de 70.2% a 89.5%, mientras que la capacidad para relacionar conceptos clínicos y fisiopatológicos pasó de 82.5% a 93.0%.

Por su parte, la dimensión de confianza también presentó incrementos en todas las afirmaciones evaluadas. La mayor diferencia se observó en la confianza sobre la comprensión del contenido estudiado, que aumentó de 66.7% ($n = 38$) a 93.0% ($n = 53$), equivalente a una diferencia de 26.3 puntos porcentuales. De igual manera, la percepción de poder aplicar los conocimientos adquiridos para resolver situaciones clínicas similares se incrementó en 12.3 puntos porcentuales, mientras que la seguridad para responder preguntas relacionadas con el tema aumentó en 7.1 puntos porcentuales.

La comparación de las respuestas favorables entre el material de estudio y la tutoría con *ChatGPT* mediante la prueba de McNemar mostró resultados heterogéneos entre los distintos ítems evaluados (ver Tabla 12). En la dimensión de claridad, únicamente la afirmación relacionada con la facilitación de la comprensión de conceptos difíciles presentó una diferencia estadísticamente significativa ($p = 0.002$), acompañada de un incremento de 22.8 puntos porcentuales en las respuestas favorables. Por el contrario, las afirmaciones

referentes a la claridad de las explicaciones ($p = 0.125$) y a la organización y coherencia de la información ($p = 0.625$) no mostraron diferencias estadísticamente significativas, a pesar de presentar incrementos de 7.0 y 3.5 puntos porcentuales, respectivamente.

En la dimensión de comprensión, se identificaron diferencias estadísticamente significativas en dos de los tres ítems evaluados. La percepción de haber comprendido adecuadamente el contenido estudiado aumentó en 19.3 puntos porcentuales ($p = 0.019$), mientras que la capacidad percibida para explicar los conceptos abordados mostró el mayor incremento observado en toda la escala, con una diferencia de 35.1 puntos porcentuales ($p < 0.001$). En contraste, la percepción de relacionar conceptos clínicos y fisiopatológicos, aunque aumentó en 10.5 puntos porcentuales, no alcanzó significancia estadística ($p = 0.147$).

Respecto a la dimensión de confianza, únicamente la afirmación relacionada con la confianza en la comprensión del contenido estudiado evidenció una diferencia estadísticamente significativa, incrementándose en 26.3 puntos porcentuales tras la interacción con *ChatGPT* ($p < 0.001$). Las demás afirmaciones, correspondientes a la seguridad para responder preguntas relacionadas con el tema ($p = 0.503$) y a la capacidad percibida para aplicar los conocimientos en situaciones clínicas similares ($p = 0.092$), no mostraron diferencias estadísticamente significativas.

Los resultados indicaron que las diferencias más consistentes asociadas a la interacción con *ChatGPT* se observaron en aspectos relacionados con la comprensión profunda del contenido, la explicación de conceptos y la confianza en la comprensión alcanzada, mientras que los cambios en la percepción de claridad general y en la aplicación de conocimientos clínicos fueron más discretos y no siempre alcanzaron significancia estadística.

En el análisis de las puntuaciones globales por dimensión se evidenció diferencias estadísticamente significativas entre las valoraciones otorgadas al libro de texto y a la tutoría realizada mediante *ChatGPT* (ver Tabla 13). En la dimensión de claridad, la mediana fue de 3 puntos en ambas mediciones; sin embargo, el rango intercuartílico disminuyó de 1 a 0 tras la interacción con la inteligencia artificial. La comparación mediante la prueba de Wilcoxon mostró una diferencia estadísticamente significativa ($Z = -3.043$; $p = 0.002$), con un tamaño del efecto moderado ($r = 0.403$).

En la dimensión de comprensión se observó un incremento de la mediana de 2 puntos (RIC = 2) para el libro de texto a 3 puntos (RIC = 1) para ChatGPT. Esta diferencia fue estadísticamente significativa ($Z = -3.807$; $p < 0.001$) y presentó el mayor tamaño del efecto entre las dimensiones evaluadas ($r = 0.504$), correspondiente a una magnitud grande.

La dimensión de confianza mantuvo una mediana de 3 puntos tanto para el libro de texto como para ChatGPT, con un rango intercuartílico de 1 en ambas mediciones. No obstante, la distribución de las respuestas mostró una diferencia estadísticamente significativa a favor de la tutoría con inteligencia artificial ($Z = -2.732$; $p = 0.006$), acompañada de un tamaño del efecto moderado ($r = 0.361$).

De forma general, las tres dimensiones evaluadas presentaron diferencias estadísticamente significativas a favor de ChatGPT. La mayor magnitud del cambio se observó en la percepción de comprensión, seguida de la percepción de claridad y la percepción de confianza, lo que indica que la interacción con la inteligencia artificial estuvo asociada con una valoración más favorable del proceso de aprendizaje en comparación con la lectura exclusiva del material de estudio.

Utilidad percibida de ChatGPT como herramienta complementaria para el aprendizaje

La totalidad de los participantes manifestó una percepción favorable sobre la utilidad de ChatGPT como herramienta complementaria para la comprensión de textos médicos complejos (ver Tabla 14). En los cuatro ítems evaluados, el 100% de los estudiantes ($n = 57$) estuvo de acuerdo en que la inteligencia artificial fue útil para aclarar dudas surgidas durante la lectura del material, constituyó un complemento valioso al estudio tradicional, puede ser una herramienta útil para apoyar el aprendizaje en medicina y sería utilizada nuevamente para complementar la comprensión de textos médicos.

El análisis de la puntuación global de utilidad percibida mostró que todos los participantes alcanzaron la puntuación máxima posible de la escala (ver Tabla 15). La media y la mediana fueron de 4 puntos, con una desviación estándar de 0, un valor mínimo de 4, un valor máximo de 4 y un rango intercuartílico de 0. Estos resultados evidenciaron una ausencia total de variabilidad en las respuestas, reflejando un consenso unánime entre los estudiantes respecto a la utilidad de ChatGPT como recurso de apoyo para el aprendizaje.

Los hallazgos mostraron una aceptación universal de la herramienta entre los participantes, quienes valoraron positivamente su capacidad para complementar el estudio de textos médicos y facilitar el proceso de aprendizaje durante la intervención realizada.

Discusión

Comprensión de textos médicos y aplicación de conocimientos clínicos

Los resultados obtenidos evidenciaron que la utilización de *ChatGPT* como tutor virtual grupal no produjo cambios estadísticamente significativos en la comprensión de textos médicos y la aplicación de conocimientos clínicos evaluados mediante los cuestionarios pre y post intervención. Aunque se observó una ligera disminución en la

puntuación promedio posterior a la interacción con la inteligencia artificial, la comparación de los puntajes mediante la prueba de rangos con signo de Wilcoxon no mostró diferencias significativas ($p = 0.081$), acompañándose de un tamaño del efecto pequeño ($r = 0.231$). En consecuencia, para esta variable específica, los hallazgos no permiten rechazar la hipótesis nula planteada en el estudio.

Estos resultados difieren de los reportados por Nartthanarung et al., (2026), quienes identificaron mejoras significativas en el razonamiento clínico de estudiantes de medicina tras la interacción con herramientas de inteligencia artificial, observando incrementos en las puntuaciones obtenidas entre las evaluaciones previas y posteriores a la intervención. De igual forma, Essel et al., (2025) documentaron un rendimiento académico significativamente superior en estudiantes que utilizaron un tutor virtual basado en inteligencia artificial en comparación con aquellos que recibieron acompañamiento tradicional. Sin embargo, las diferencias metodológicas entre dichos estudios y la presente investigación dificultan una comparación directa. Entre estas diferencias destacan la duración de las intervenciones, la existencia de grupos control, el seguimiento prolongado de los participantes y la naturaleza de los instrumentos empleados para medir el aprendizaje.

Por otra parte, los resultados obtenidos guardan concordancia con las observaciones realizadas por Feigerlova et al., (2025), quienes señalaron que la evidencia disponible sobre el impacto de la inteligencia artificial en los resultados educativos de las profesiones de la salud continúa siendo limitada y metodológicamente heterogénea. Estos autores concluyen que aún existe insuficiente evidencia de alta calidad para afirmar de manera concluyente que la inteligencia artificial mejora las competencias clínicas o el desempeño académico medido objetivamente. Desde esta perspectiva, la ausencia de diferencias significativas

encontrada en el presente estudio contribuye a una línea de evidencia que sugiere cautela al interpretar los beneficios académicos de estas herramientas más allá de las percepciones estudiantiles.

Diversos factores podrían explicar los hallazgos observados. En primer lugar, la intervención se desarrolló durante una única sesión de tutoría, lo que limita la posibilidad de generar cambios sustanciales en habilidades complejas como el razonamiento clínico y la integración de conocimientos médicos. En segundo lugar, los participantes mostraron una amplia familiaridad previa con herramientas de inteligencia artificial generativa, por lo que el efecto de la intervención pudo verse atenuado al no representar una experiencia novedosa para la mayoría de los estudiantes. Adicionalmente, las puntuaciones iniciales del cuestionario fueron relativamente elevadas, lo que sugiere la posible presencia de un efecto techo que redujo el margen disponible para detectar mejoras posteriores. Finalmente, debe considerarse que la comprensión profunda y la transferencia del conocimiento hacia la resolución de problemas clínicos constituyen procesos cognitivos complejos que probablemente requieren exposiciones repetidas, entrenamiento prolongado y estrategias pedagógicas complementarias para manifestar cambios medibles en evaluaciones objetivas.

En conjunto, estos resultados sugieren que la utilización puntual de *ChatGPT* como tutor virtual grupal no fue suficiente para generar mejoras significativas en el desempeño académico evaluado mediante cuestionarios de comprensión y razonamiento clínico. No obstante, este hallazgo no implica necesariamente una ausencia de utilidad educativa de la herramienta, sino que indica que sus beneficios podrían manifestarse con mayor claridad en otras dimensiones del aprendizaje, particularmente aquellas relacionadas con la experiencia subjetiva del estudiante, la comprensión percibida y la confianza académica, aspectos que serán abordados en los apartados siguientes.

Percepciones de claridad, comprensión y confianza sobre el aprendizaje

A diferencia de los resultados obtenidos en el desempeño académico, las percepciones de los estudiantes evidenciaron cambios favorables y estadísticamente significativos tras la interacción con *ChatGPT*. Tanto el análisis por afirmaciones individuales como la evaluación global de las dimensiones mostraron mejoras en la claridad, comprensión y confianza percibidas durante el proceso de aprendizaje. La magnitud de estos cambios fue particularmente notable en la dimensión de comprensión, que presentó el mayor tamaño del efecto ($r = 0.504$), seguida por claridad ($r = 0.403$) y confianza ($r = 0.361$). Estos hallazgos sugieren que la herramienta influyó positivamente en la forma en que los estudiantes interpretaron y experimentaron el proceso de aprendizaje, aun cuando dicho efecto no se reflejó en mejoras objetivas de desempeño.

Los resultados coinciden con los reportados por Manobanda et al., (2026), quienes observaron una mejora significativa en la aceptación y confianza hacia *ChatGPT* después de una intervención educativa guiada en estudiantes de medicina. En dicho estudio, la percepción de confiabilidad de la herramienta aumentó significativamente tras la experiencia práctica, lo que llevó a los autores a concluir que la inteligencia artificial puede favorecer el autoaprendizaje y reducir el escepticismo inicial de los estudiantes. De forma similar, los participantes del presente estudio mostraron una valoración más favorable de la claridad de las explicaciones, la comprensión de los contenidos y la confianza en su propio aprendizaje después de interactuar con la herramienta.

Asimismo, los hallazgos son consistentes con los resultados descritos por Ciacia et al., (2025), quienes documentaron una actitud predominantemente favorable hacia el uso de la inteligencia artificial generativa entre estudiantes de ciencias de la salud. Aunque dichos autores identificaron reservas relacionadas con aplicaciones clínicas complejas,

encontraron una amplia aceptación de estas tecnologías para actividades académicas y de apoyo al aprendizaje. Esta percepción positiva podría explicar por qué los estudiantes del presente estudio manifestaron mayores niveles de comprensión y confianza tras recibir la tutoría grupal mediada por *ChatGPT*.

Un aspecto particularmente relevante fue que las mayores diferencias se observaron en afirmaciones relacionadas con la comprensión de conceptos difíciles, la capacidad de explicar contenidos estudiados y la confianza en la comprensión alcanzada. Este patrón resulta coherente con la naturaleza conversacional de los modelos de lenguaje generativo, los cuales permiten reformular explicaciones, adaptar el lenguaje al nivel del usuario y responder preguntas específicas en tiempo real. Desde una perspectiva pedagógica, estas características pueden favorecer procesos de andamiaje cognitivo, al facilitar que el estudiante relacione nuevos conocimientos con estructuras previas de aprendizaje y reduzca la carga cognitiva asociada a textos médicos complejos.

Por otra parte, el hecho de que las mejoras percibidas fueran más pronunciadas que las observadas en los cuestionarios clínicos sugiere que la influencia de *ChatGPT* podría manifestarse inicialmente en dimensiones subjetivas del aprendizaje antes que en indicadores objetivos de desempeño. La percepción de comprender mejor un contenido no necesariamente implica una mejora inmediata en la capacidad para responder correctamente preguntas clínicas, en especial cuando la intervención se limita a una única sesión. En este sentido, los resultados respaldan la idea de que la inteligencia artificial puede actuar como un recurso facilitador del aprendizaje y de la confianza académica, aunque la consolidación de conocimientos y habilidades clínicas probablemente requiera procesos educativos más prolongados y experiencias repetidas de interacción.

Los hallazgos indican que *ChatGPT* fue percibido como una herramienta capaz de mejorar la experiencia de aprendizaje de los estudiantes, favoreciendo la claridad de las explicaciones, la comprensión de los contenidos y la confianza en el conocimiento adquirido. Estos resultados respaldan parcialmente la hipótesis de investigación y sugieren que el principal aporte de la inteligencia artificial generativa en el contexto evaluado se relacionó con el fortalecimiento de aspectos cognitivos y afectivos asociados al aprendizaje, más que con cambios inmediatos en el rendimiento académico medido objetivamente.

Utilidad percibida de ChatGPT como herramienta complementaria para el aprendizaje

Los resultados obtenidos evidenciaron una valoración unánimemente favorable de *ChatGPT* como herramienta complementaria para la comprensión de textos médicos complejos. La totalidad de los participantes manifestó estar de acuerdo con las cuatro afirmaciones evaluadas, relacionadas con la utilidad de la herramienta para aclarar dudas, complementar el estudio tradicional, apoyar el aprendizaje de la medicina y ser utilizada nuevamente en futuras actividades académicas. Asimismo, la puntuación global de utilidad percibida alcanzó el valor máximo posible en todos los estudiantes, reflejando un consenso absoluto respecto al valor educativo de la inteligencia artificial generativa en el contexto de la intervención realizada.

Estos hallazgos coinciden con los reportados por Miclin et al., (2024), quienes identificaron una percepción favorable hacia *ChatGPT* entre estudiantes de medicina cubanos, destacando su utilidad como recurso complementario dentro del proceso de enseñanza-aprendizaje. En dicho estudio, una proporción importante de los participantes consideró que la calidad de la información proporcionada por la inteligencia artificial era comparable a la obtenida mediante fuentes bibliográficas tradicionales, lo que llevó a los autores a reconocer su potencial como apoyo académico. De manera similar, los estudiantes

participantes en la presente investigación valoraron positivamente la capacidad de la herramienta para complementar el aprendizaje derivado de la lectura del texto médico.

Los resultados también guardan relación con los hallazgos de Manobanda et al., (2026), quienes concluyeron que una intervención guiada con *ChatGPT* favoreció la aceptación de la herramienta y fortaleció la percepción de utilidad para el autoaprendizaje en estudiantes de medicina. La elevada valoración observada en ambos estudios sugiere que los estudiantes perciben la inteligencia artificial no como un sustituto de los recursos educativos tradicionales, sino como un complemento capaz de enriquecer la experiencia de aprendizaje mediante explicaciones adicionales, retroalimentación inmediata y acceso rápido a información contextualizada.

Por otra parte, la aceptación universal observada en esta investigación puede estar relacionada con las características propias de los modelos de lenguaje generativo. La posibilidad de formular preguntas específicas, solicitar aclaraciones, obtener ejemplos adicionales y adaptar las respuestas al nivel de comprensión del usuario ofrece ventajas que difícilmente pueden obtenerse mediante la lectura pasiva de un texto. Estas características convierten a la inteligencia artificial en una herramienta potencialmente útil para promover el aprendizaje autónomo y facilitar la resolución de dudas surgidas durante el estudio independiente.

No obstante, es importante interpretar estos resultados con cautela. La elevada utilidad percibida no implica necesariamente que la herramienta produzca mejoras directas en el desempeño académico o en el razonamiento clínico. De hecho, en la presente investigación se observó que, a pesar de la aceptación unánime de *ChatGPT* y de las mejoras significativas en las percepciones de aprendizaje, no se identificaron cambios estadísticamente significativos en las puntuaciones obtenidas en los cuestionarios clínicos.

Este hallazgo sugiere que la utilidad percibida y el rendimiento académico constituyen dimensiones relacionadas, pero no equivalentes, del proceso educativo.

ChatGPT fue ampliamente aceptado por los estudiantes como una herramienta de apoyo al aprendizaje médico. Su utilidad fue reconocida de manera unánime en aspectos relacionados con la comprensión de contenidos, la resolución de dudas y el acompañamiento al estudio tradicional, lo que respalda su potencial incorporación como recurso complementario dentro de estrategias educativas orientadas a fortalecer el aprendizaje autónomo en la formación médica.

Integración general de los hallazgos

Los resultados obtenidos permiten concluir que el efecto de *ChatGPT* como tutor virtual grupal se manifestó de manera diferenciada según la dimensión evaluada. Mientras que no se identificaron cambios estadísticamente significativos en la comprensión de textos médicos y la aplicación de conocimientos clínicos medidos mediante los cuestionarios de desempeño, sí se observaron mejoras significativas en las percepciones de claridad, comprensión y confianza asociadas al proceso de aprendizaje. Adicionalmente, la herramienta alcanzó una valoración unánimemente favorable en términos de utilidad percibida, reflejando una amplia aceptación por parte de los estudiantes participantes.

El principal aporte de *ChatGPT* durante la intervención no radicó en la generación de mejoras inmediatas y objetivamente medibles en el rendimiento académico, sino en su capacidad para facilitar la experiencia de aprendizaje. La posibilidad de recibir explicaciones complementarias, aclarar dudas en tiempo real y profundizar en conceptos complejos parece haber contribuido a que los estudiantes se sintieran más seguros y comprendieran mejor los contenidos estudiados. Esto resulta coherente con la concepción de la inteligencia artificial generativa como una herramienta de apoyo cognitivo y

acompañamiento educativo, más que como un sustituto de los métodos tradicionales de enseñanza.

Los resultados también aportan evidencia relevante al contexto nacional. Tal como se identificó durante la revisión de antecedentes, no se encontraron investigaciones nicaragüenses previas que evaluaran experimentalmente el uso de inteligencia artificial generativa como tutor virtual en estudiantes de medicina. En consecuencia, este estudio constituye una de las primeras aproximaciones empíricas al fenómeno dentro del país y proporciona una línea base para futuras investigaciones interesadas en explorar el potencial educativo de estas tecnologías en la formación médica.

Asimismo, los hallazgos ponen de manifiesto la necesidad de continuar investigando el papel de la inteligencia artificial generativa en contextos educativos. Resulta posible que intervenciones más prolongadas, con múltiples sesiones de tutoría, diferentes estrategias de personalización o evaluaciones de seguimiento a mediano y largo plazo permitan identificar efectos que no fueron detectados en una intervención puntual. De igual forma, futuras investigaciones podrían incorporar grupos control, muestras más amplias y diversos niveles de formación médica con el propósito de fortalecer la evidencia disponible sobre el impacto real de estas herramientas en el aprendizaje y el razonamiento clínico.

En síntesis, los resultados indican que ChatGPT fue percibido como una herramienta útil, clara y confiable para apoyar el aprendizaje de contenidos médicos complejos. Aunque su utilización no produjo mejoras significativas en el desempeño académico medido mediante cuestionarios clínicos, sí fortaleció dimensiones subjetivas del aprendizaje consideradas relevantes para la formación médica, particularmente la comprensión percibida, la claridad de las explicaciones y la confianza en el conocimiento

adquirido. Estos hallazgos respaldan el potencial de la inteligencia artificial generativa como recurso complementario dentro de estrategias innovadoras de educación médica, al tiempo que evidencian la necesidad de continuar generando evidencia científica que permita delimitar con mayor precisión sus alcances y limitaciones.

10.- Conclusiones

La utilización de *ChatGPT* como tutor virtual grupal no produjo cambios estadísticamente significativos en el razonamiento clínico y el desempeño académico de los estudiantes evaluados mediante los cuestionarios de comprensión de textos médicos y aplicación de conocimientos clínicos. Aunque se observaron variaciones en las respuestas individuales y una ligera disminución en la puntuación promedio posterior a la intervención, las diferencias encontradas no alcanzaron significancia estadística, lo que indica que una única sesión de tutoría con inteligencia artificial no fue suficiente para generar mejoras objetivamente medibles en el rendimiento académico de los participantes.

La interacción con *ChatGPT* mejoró significativamente las percepciones de claridad, comprensión y confianza asociadas al aprendizaje. Los estudiantes manifestaron una valoración más favorable de las explicaciones recibidas, una mayor comprensión de los contenidos estudiados y una mayor confianza en el conocimiento adquirido tras la tutoría con inteligencia artificial. La dimensión que presentó el mayor cambio fue la comprensión, seguida de la claridad y la confianza, evidenciando que *ChatGPT* actuó como un facilitador del aprendizaje desde la perspectiva de los propios estudiantes.

La utilidad percibida de la inteligencia artificial generativa como herramienta complementaria para la comprensión de textos médicos complejos fue universalmente favorable. Todos los participantes consideraron que *ChatGPT* fue útil para aclarar dudas, complementar el estudio tradicional, apoyar el aprendizaje de la medicina y ser utilizado nuevamente en futuras actividades académicas. Estos resultados reflejan una elevada aceptación de la herramienta y respaldan su potencial como recurso de apoyo dentro de estrategias de aprendizaje autónomo en educación médica.

Recomendaciones

Los resultados obtenidos en la presente investigación permitieron identificar fortalezas y oportunidades de mejora en la incorporación de herramientas de IA generativa dentro de los procesos de enseñanza-aprendizaje en medicina. Aunque la intervención no produjo cambios significativos en el desempeño académico medido mediante cuestionarios clínicos, sí evidenció mejoras en las percepciones de claridad, comprensión y confianza, además de una elevada utilidad percibida por parte de los estudiantes. En este contexto, se proponen las siguientes recomendaciones para orientar futuras experiencias educativas e investigaciones relacionadas con el uso de inteligencia artificial en educación médica:

1. Implementar intervenciones educativas de mayor duración, que permitan a los estudiantes interactuar de forma continua con herramientas de IA generativa, favoreciendo la consolidación progresiva de conocimientos y el desarrollo de habilidades de razonamiento clínico.
2. Integrar la IA generativa como complemento del aprendizaje tradicional, utilizándola para aclarar dudas, ampliar explicaciones y reforzar contenidos complejos, sin sustituir la consulta de literatura científica ni las estrategias docentes convencionales.
3. Promover el desarrollo de competencias de alfabetización digital e IA, de manera que los estudiantes puedan utilizar estas herramientas de forma crítica, reflexiva y basada en evidencia, identificando posibles errores, sesgos o limitaciones en las respuestas generadas.
4. Diseñar experiencias de aprendizaje que combinen la interacción con IA y actividades de aplicación clínica, con el propósito de favorecer la transferencia del

conocimiento hacia la resolución de problemas médicos y el fortalecimiento del razonamiento clínico.

5. Realizar investigaciones con diseños metodológicos más robustos, incluyendo grupos de comparación, muestras más amplias y evaluaciones longitudinales, que permitan determinar con mayor precisión el impacto de la IA generativa sobre el desempeño académico y las competencias clínicas.
6. Explorar diferentes modalidades de personalización de los tutores virtuales basados en IA, evaluando el efecto de distintos niveles de contextualización, estilos de tutoría y estrategias de retroalimentación sobre el aprendizaje de los estudiantes.
7. Profundizar en el estudio de variables relacionadas con la experiencia de aprendizaje, tales como motivación, autoeficacia, satisfacción académica, carga cognitiva y aprendizaje autónomo, con el fin de comprender mejor los mecanismos mediante los cuales la inteligencia artificial influye en los procesos educativos.
8. Generar evidencia científica en contextos nacionales y regionales, contribuyendo al desarrollo de un cuerpo de conocimiento que permita orientar la incorporación de la inteligencia artificial generativa en la educación médica a partir de datos obtenidos en realidades académicas similares.

11.- Referencias

- Abd-alrazaq, A., AlSaad, R., Alhuwail, D., Ahmed, A., Healy, P. M., Latifi, S., Aziz, S., Damseh, R., Alrazak, S. A., & Sheikh, J. (2023). Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Medical Education*, 9(1), e48291. <https://doi.org/10.2196/48291>
- Abedi, M., Alshybani, I., Shahadat, M. R. B., & Murillo, M. S. (2023). *Beyond Traditional Teaching: The Potential of Large Language Models and Chatbots in Graduate Engineering Education*. <https://doi.org/10.32388/MD04B0>
- Abouzeid, E., Wassef, R., Jawwad, A., & Harris, P. (2025). Chatbots' Role in Generating Single Best Answer Questions for Undergraduate Medical Student Assessment: Comparative Analysis. *JMIR Medical Education*, 11(1), e69521. <https://doi.org/10.2196/69521>
- Akram, A., Kiren, A., Sabir, S., & Aslam, S. (2025). The Effect of AI Based Scaffolding on Problem Solving and Metacognitive Awareness in Learners. *The Critical Review of Social Sciences Studies*, 3(3), 1074-1089. <https://doi.org/10.59075/ja7dvy78>
- Al-Abdullatif, A. M., Al-Dokhny, A. A., & Drwish, A. M. (2023). Implementing the Bashayer chatbot in Saudi higher education: Measuring the influence on students' motivation and learning strategies. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1129070>
- AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an Educational Tool: Opportunities, Challenges, and Recommendations for Communication, Business Writing, and Composition Courses. *Journal of Artificial Intelligence and Technology*, 3(2), 60-68. <https://doi.org/10.37965/jait.2023.0184>

- Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., & Ali, H. (2024). ChatGPT in Teaching and Learning: A Systematic Review. *Education Sciences*, 14(6), 643.
<https://doi.org/10.3390/educsci14060643>
- Araujo, S. M., & Cruz-Correia, R. (2024). Incorporating ChatGPT in Medical Informatics Education: Mixed Methods Study on Student Perceptions and Experiential Integration Proposals. *JMIR Medical Education*, 10(1), e51151.
<https://doi.org/10.2196/51151>
- Ausubel, D. P. (1963). *The Psychology of Meaningful Verbal Learning*. Grune & Stratton.
- Ausubel, D. P. (2012). *The Acquisition and Retention of Knowledge: A Cognitive View*. Springer Science & Business Media.
- Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational Psychology: A Cognitive View*. Holt, Rinehart and Winston.
- Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis. *Sustainability*, 15(17), 12983.
<https://doi.org/10.3390/su151712983>
- Bassner, P., Frankford, E., & Krusche, S. (2024). Iris: An AI-Driven Virtual Tutor for Computer Science Education. *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1, ITiCSE 2024*, 394-400.
<https://doi.org/10.1145/3649217.3653543>
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122.
<https://doi.org/10.1073/pnas.2422633122>

- Boit, S., & Patil, R. (2025). A Prompt Engineering Framework for Large Language Model–Based Mental Health Chatbots: Conceptual Framework. *JMIR Mental Health*, 12(1), e75078. <https://doi.org/10.2196/75078>
- Cacho, R. (2024). Integrating Generative AI in University Teaching and Learning: A Model for Balanced Guidelines. *Online Learning*, 28(3). <https://doi.org/10.24059/olj.v28i3.4508>
- Carmo, C. S. D., Rodrigues, R. A. S., Oliveira, J. M. L. D., Alencar, D. B. D., & Sanches, A. E. (2025). Exploratory Research On The Development Of Chatbots And Prompt Engineering As Tools To Support The Educational Process. *IOSR Journal of Business and Management*, 27(6), 33-40. <https://doi.org/10.9790/487X-2706023340>
- Celik, I., Gedrimiene, E., Siklander, S., & Muukkonen, H. (2024). The affordances of artificial intelligence-based tools for supporting 21st-century skills: A systematic review of empirical research in higher education. *Australasian Journal of Educational Technology*, 40(3), 19-38. <https://doi.org/10.14742/ajet.9069>
- Chan, C. K. Y., & Colloton, T. (2024). *Generative AI in Higher Education: The ChatGPT Effect*. Routledge. <https://doi.org/10.4324/9781003459026>
- Chen, B., Zhu, X., & Díaz del Castillo H., F. (2023a). Integrating generative AI in knowledge building. *Computers and Education: Artificial Intelligence*, 5, 100184. <https://doi.org/10.1016/j.caeai.2023.100184>
- Chen, B., Zhu, X., & Díaz del Castillo H., F. (2023b). Integrating generative AI in knowledge building. *Computers and Education: Artificial Intelligence*, 5, 100184. <https://doi.org/10.1016/j.caeai.2023.100184>
- Cheung, B. H. H., Lau, G. K. K., Wong, G. T. C., Lee, E. Y. P., Kulkarni, D., Seow, C. S., Wong, R., & Co, M. T.-H. (2023). ChatGPT versus human in generating medical

- graduate exam multiple choice questions—A multinational prospective study (Hong Kong S.A.R., Singapore, Ireland, and the United Kingdom). *PLOS ONE*, 18(8), e0290691. <https://doi.org/10.1371/journal.pone.0290691>
- Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6, 100171. <https://doi.org/10.1016/j.caeo.2024.100171>
- Ciacia, F. C., Morales, O. A., & Perdomo, B. J. (2025). Actitudes hacia el uso de chatbots de inteligencia artificial generativa con fines pedagógicos de estudiantes de Ciencias de la Salud. *Acta Bioclínica*, 15(31), 253-295.
- Črček, N., & Patekar, J. (2023). Writing with AI: University Students' Use of ChatGPT. *Journal of Language and Education*, 9(4), 128-138. <https://doi.org/10.17323/jle.2023.17379>
- Davar, N. F., Dewan, M. A. A., & Zhang, X. (2025). AI Chatbots in Education: Challenges and Opportunities. *Information*, 16(3), 235. <https://doi.org/10.3390/info16030235>
- Diao, S., Wang, P., Lin, Y., Pan, R., Liu, X., & Zhang, T. (2024). *Active Prompting with Chain-of-Thought for Large Language Models* (arXiv:2302.12246). arXiv. <https://doi.org/10.48550/arXiv.2302.12246>
- Domínguez, C. D. R., Somoza, G. A., Guzmán, N. E. O., Trinidad, M. M. C., & Reyes, A. G. S. (2025). Avances en el uso de inteligencia artificial en la educación médica latinoamericana. *Alerta, Revista científica del Instituto Nacional de Salud*, 8(1), 88-95. <https://doi.org/10.5377/alerta.v8i1.19194>
- Elnaffar, S., Rashidi, F., & Abualkishik, A. Z. (2026). Teaching with AI: A Systematic Review of Chatbots, Generative Tools, and Tutoring Systems in Programming

- Education. *International Journal of Learning, Teaching and Educational Research*, 25(1), 1-28. <https://doi.org/10.26803/ijlter.25.1.1>
- Essel, H. B., Vlachopoulos, D., Nunoo-Mensah, H., & Amankwa, J. O. (2025). Exploring the impact of VoiceBots on multimedia programming education among Ghanaian university students. *British Journal of Educational Technology*, 56(1), 276-295. <https://doi.org/10.1111/bjet.13504>
- Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, 49(3), 363-375. <https://doi.org/10.1080/02602938.2023.2241676>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460-474. <https://doi.org/10.1080/14703297.2023.2195846>
- Favero, L., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2024). *Enhancing Critical Thinking in Education by means of a Socratic Chatbot* (arXiv:2409.05511). arXiv. <https://doi.org/10.48550/arXiv.2409.05511>
- Feigerlova, E., Hani, H., & Hothersall-Davies, E. (2025). A systematic review of the impact of artificial intelligence on educational outcomes in health professions education. *BMC Medical Education*, 25(1), 129. <https://doi.org/10.1186/s12909-025-06719-5>
- Ghorashi, N., Ismail, A., Ghosh, P., Sidawy, A., Javan, R., Ghorashi, N. S., Ismail, A., Ghosh, P., Sidawy, A., & Javan, R. (2023). AI-Powered Chatbots in Medical Education: Potential Applications and Implications. *Cureus*, 15(8). <https://doi.org/10.7759/cureus.43271>

- Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2025). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, 44(11), 2518-2544. <https://doi.org/10.1080/0144929X.2024.2394886>
- Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., ... Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, 4, 19-23. <https://doi.org/10.1016/j.iotcps.2023.06.002>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Medical Education*, 9(1), e45312. <https://doi.org/10.2196/45312>
- González, D. M., Rojas, M. E. P., & Durán, A. G. P. (2025). Inteligencia Artificial y Educación Médica: Análisis bibliométrico. *Revista Conrado*, 21(105), e4666-e4666.
- 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., & 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>

- Grassini, S. (2023). Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. *Education Sciences*, 13(7), 692. <https://doi.org/10.3390/educsci13070692>
- Güner, H., & Er, E. (2025). AI in the classroom: Exploring students' interaction with ChatGPT in programming learning. *Education and Information Technologies*, 30(9), 12681-12707. <https://doi.org/10.1007/s10639-025-13337-7>
- Gupta, R., Goyal, H., Kumar, D., Mehra, A., Sharma, S., Mittal, K., & Challa, J. S. (2025). *Saksham AI: Advancing AI-Assisted Coding Education for Engineering Students in India Through Socratic Tutoring and Comprehensive Feedback* (arXiv:2503.12479). arXiv. <https://doi.org/10.48550/arXiv.2503.12479>
- Haindl, P., & Weinberger, G. (2024). Students' Experiences of Using ChatGPT in an Undergraduate Programming Course. *IEEE Access*, 12, 43519-43529. <https://doi.org/10.1109/ACCESS.2024.3380909>
- Hale, J., Alexander, S., Wright, S. T., & Gilliland, K. (2024). Generative AI in Undergraduate Medical Education: A Rapid Review. *Journal of Medical Education and Curricular Development*, 11, 23821205241266697. <https://doi.org/10.1177/23821205241266697>
- Hernández Sampieri, R., & Mendoza Torres, C. P. (2023). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta* (2.^a ed.). McGraw Hill Interamericana Editores, S.A. de C.V.
- Heston, T. F., & Khun, C. (2023). Prompt Engineering in Medical Education. *International Medical Education*, 2(3), 198-205. <https://doi.org/10.3390/ime2030019>

- Huang, W., Jiang, J., King, R. B., & Fryer, L. K. (2025). Chatbots and student motivation: A scoping review. *International Journal of Educational Technology in Higher Education*, 22(1), 26. <https://doi.org/10.1186/s41239-025-00524-2>
- Huda, N., Anwar, K., Novferma, & , W. K. <https://orcid.org/0000-0002-0608-1548> (2025). ChatGPT Scaffolding in Supporting Metacognition for Limit Concepts in Guided Inquiry Mathematics Learning. *Journal of Information Technology Education: Research*, 24, 038.
- Ilieva, G., Yankova, T., Klisarova-Belcheva, S., Dimitrov, A., Bratkov, M., & Angelov, D. (2023). Effects of Generative Chatbots in Higher Education. *Information*, 14(9), 492. <https://doi.org/10.3390/info14090492>
- Iqbal, J., Hashmi, Z. F., Asghar, M. Z., & Abid, M. N. (2025). Generative AI tool use enhances academic achievement in sustainable education through shared metacognition and cognitive offloading among preservice teachers. *Scientific Reports*, 15(1), 16610. <https://doi.org/10.1038/s41598-025-01676-x>
- Isarel, E. (2025). Capítulo 298: Asma. En *Harrison. Principios de Medicina Interna* (22.^a ed.). McGraw Hill Education S.A de C.V. <https://accessmedicina.mhmedical.com/book.aspx?bookid=3553>
- Jaleel, A., Aziz, R., Farid, G., & Bashir, M. Z. (2025). The impact of ChatGPT on academic integrity in medical education: A developing nation perspective. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1554444>
- Jauhiainen, J. S., & Guerra, A. G. (2023). Generative AI and ChatGPT in School Children's Education: Evidence from a School Lesson. *Sustainability*, 15(18), 14025. <https://doi.org/10.3390/su151814025>

- Jones, B., Desu, A., & Honig, C. D. F. (2026). Artificial Intelligence Chatbots as Virtual Patients in Dental Education: A Constructivist Approach to Classroom Implementation. *European Journal of Dental Education*, 30(2), 447-456. <https://doi.org/10.1111/eje.13135>
- Kaplan-Rakowski, R., Grotewold, K., Hartwick, P., & Papin, K. (2023). Generative AI and Teachers' Perspectives on Its Implementation in Education. *Journal of Interactive Learning Research*, 34(2), 313-338.
- Kathole, A., Patil, S., Jadhav, Dr. D., Pathak, H., & Mirge, A. S. (2025). Development of student intent-based educational chatbot system with adaptive and attentive DTCN on symmetric convolution approach. *MethodsX*, 15, 103542. <https://doi.org/10.1016/j.mex.2025.103542>
- Kumar, H., Musabirov, I., Reza, M., Shi, J., Wang, X., Williams, J. J., Kuzminykh, A., & Liut, M. (2024). Guiding Students in Using LLMs in Supported Learning Environments: Effects on Interaction Dynamics, Learner Performance, Confidence, and Trust. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 499:1-499:30. <https://doi.org/10.1145/3687038>
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., Leon, L. D., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56. <https://doi.org/10.1186/s41239-023-00426-1>

- Leung, C. H. (2024). Promoting optimal learning with ChatGPT: A comprehensive exploration of prompt engineering in education. *Asian Journal of Contemporary Education*, 8(2), 104-114.
- Li, H., Xu, T., Zhang, C., Chen, E., Liang, J., Fan, X., Li, H., Tang, J., & Wen, Q. (2024). *Bringing Generative AI to Adaptive Learning in Education*. arXiv. <https://doi.org/10.48550/arXiv.2402.14601>
- Lieb, A., & Goel, T. (2024). Student Interaction with NewtBot: An LLM-as-tutor Chatbot for Secondary Physics Education. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '24*, 1-8. <https://doi.org/10.1145/3613905.3647957>
- Lin-Zucker, M., Bellassen, J., & Zucker, J.-D. (2025). *Prompting ChatGPT for Chinese Learning as L2: A CEFR and EBCL Level Study* (arXiv:2501.15247). arXiv. <https://doi.org/10.48550/arXiv.2501.15247>
- Liu, R., Zenke, C., Liu, C., Holmes, A., Thornton, P., & Malan, D. J. (2024). Teaching CS50 with AI: Leveraging Generative Artificial Intelligence in Computer Science Education. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 750-756. <https://doi.org/10.1145/3626252.3630938>
- Liu, Z., Agrawal, P., Singhal, S., Madaan, V., Kumar, M., & Verma, P. K. (2025). LPITutor: An LLM based personalized intelligent tutoring system using RAG and prompt engineering. *PeerJ Computer Science*, 11, e2991. <https://doi.org/10.7717/peerj-cs.2991>
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>

- Long, D. Y., Wang, S., Md Rashid, S., & Lu, X. T. (2026). Artificial intelligence in higher education: A systematic review of its impact on student engagement and the mediating role of teaching methods. *Frontiers in Education, 10*.
<https://doi.org/10.3389/feduc.2025.1648661>
- Looi, C.-K., & Jia, F. (2025). Personalization capabilities of current technology chatbots in a learning environment: An analysis of student-tutor bot interactions. *Education and Information Technologies, 30*(10), 14165-14195. <https://doi.org/10.1007/s10639-025-13369-z>
- López, F. J. S. (2025). Aplicación docente de la inteligencia artificial en la educación médica: Revisión narrativa del estado actual. *Mediciego, 31*, e4184-e4184.
- Maity, S., & Deroy, A. (2024). *Generative AI and Its Impact on Personalized Intelligent Tutoring Systems* (arXiv:2410.10650). arXiv.
<https://doi.org/10.48550/arXiv.2410.10650>
- Mallick, S., & Datta, R. (2025). AI Powered Education. *International Journal for Research in Applied Science and Engineering Technology, 13*(5), 2263-2271.
<https://doi.org/10.22214/ijraset.2025.70678>
- Manobanda, L. D. P., Asqui, A. F. G., & Carrasco, P. A. (2026). Impacto de la aplicación de una intervención didáctica con Chat GPT en las percepciones de los estudiantes de medicina. *Mediciencias UTA, 10*(1), 2-13.
<https://doi.org/10.31243/mdc.uta.v10i1.3008.2026>
- Marchena Sekli, G., Godo, A., & Carlos Véliz, J. (2024). Generative AI Solutions for Faculty and Students: A Review of Literature and Roadmap for Future Research. *Journal of Information Technology Education: Research, 23*, 014.
<https://doi.org/10.28945/5304>

- Miclin, C. D. B., Rodríguez, Y. E., & Argibay, S. R. L. (2024). Uso de ChatGPT por estudiantes de medicina en su proceso de enseñanza—Aprendizaje. *Revista Científica Estudiantil UNIMED*, 6(3).
<https://revunimed.sld.cu/index.php/revestud/article/view/398>
- Mittal, U., Sai, S., Chamola, V., & Sangwan, D. (2024). A Comprehensive Review on Generative AI for Education. *IEEE Access*, 12, 142733-142759.
<https://doi.org/10.1109/ACCESS.2024.3468368>
- Moundridou, M., Matzakos, N., & Doukakis, S. (2024). Generative AI tools as educators' assistants: Designing and implementing inquiry-based lesson plans. *Computers and Education: Artificial Intelligence*, 7, 100277.
<https://doi.org/10.1016/j.caeai.2024.100277>
- Narthanarung, A., Plangsiri, K., & Kongmalai, P. (2026). Comparative Impact of ChatGPT and Conventional Search Tools on Clinical Reasoning Performance: A Randomized Crossover Study in Preclinical Medical Students. *Advances in Medical Education and Practice*, 17, 1-7. <https://doi.org/10.2147/AMEP.S603679>
- Neumann, A. T., Yin, Y., Sowe, S., Decker, S., & Jarke, M. (2025). An LLM-Driven Chatbot in Higher Education for Databases and Information Systems. *IEEE Transactions on Education*, 68(1), 103-116.
<https://doi.org/10.1109/TE.2024.3467912>
- Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., & Sandison, C. (2023). ChatGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European*

- Journal of Engineering Education*, 48(4), 559-614.
<https://doi.org/10.1080/03043797.2023.2213169>
- Pahi, K., Hawlader, S., Hicks, E., Zaman, A., & Phan, V. (2024). Enhancing active learning through collaboration between human teachers and generative AI. *Computers and Education Open*, 6, 100183. <https://doi.org/10.1016/j.caeo.2024.100183>
- Peralta Ramirez, A. A., Trujillo López, S., Navarro Armendariz, G. A., De la Torre Othón, S. A., Sierra Cervantes, M. R., & 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>
- Polyportis, A., & Pahos, N. (2025). Understanding students' adoption of the ChatGPT chatbot in higher education: The role of anthropomorphism, trust, design novelty and institutional policy. *Behaviour & Information Technology*, 44(2), 315-336.
<https://doi.org/10.1080/0144929X.2024.2317364>
- Punar Özçelik, N., & Yangın Ekşi, G. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, 11(1), 10. <https://doi.org/10.1186/s40561-024-00296-8>
- Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences*, 13(9), 5783.
<https://doi.org/10.3390/app13095783>
- Roca, M. D. L., Chan, M. M., Garcia-Cabot, A., Garcia-Lopez, E., & Amado-Salvatierra, H. (2024). The impact of a chatbot working as an assistant in a course for supporting student learning and engagement. *Computer Applications in Engineering Education*, 32(5), e22750. <https://doi.org/10.1002/cae.22750>

- Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., Martín-Antón, L. J., Díez, I. G., & Almeida, L. (2023). Perceived satisfaction of university students with the use of chatbots as a tool for self-regulated learning. *Heliyon*, 9(1).
<https://doi.org/10.1016/j.heliyon.2023.e12843>
- Sánchez-Vera, F. (2025). Subject-Specialized Chatbot in Higher Education as a Tutor for Autonomous Exam Preparation: Analysis of the Impact on Academic Performance and Students' Perception of Its Usefulness. *Education Sciences*, 15(1), 26.
<https://doi.org/10.3390/educsci15010026>
- Santos, R. P. dos. (2023). *Enhancing Chemistry Learning with ChatGPT, Bing Chat, Bard, and Claude as Agents-to-Think-With: A Comparative Case Study* (arXiv:2311.00709). arXiv. <https://doi.org/10.48550/arXiv.2311.00709>
- Saxena, R. R. (2024). *Beyond Flashcards: Designing an Intelligent Assistant for USMLE Mastery and Virtual Tutoring in Medical Education (A Study on Harnessing Chatbot Technology for Personalized Step 1 Prep)* (arXiv:2409.10540). arXiv. <https://doi.org/10.48550/arXiv.2409.10540>
- Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and Use of AI Chatbots among Students in Higher Education: A Scoping Review of Empirical Studies. *Education Sciences*, 14(8), 922. <https://doi.org/10.3390/educsci14080922>
- Silva, C. A. G. da, Ramos, F. N., de Moraes, R. V., & Santos, E. L. dos. (2024). ChatGPT: Challenges and Benefits in Software Programming for Higher Education. *Sustainability*, 16(3), 1245. <https://doi.org/10.3390/su16031245>
- Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of

- study. *Computers and Education: Artificial Intelligence*, 7, 100259.
<https://doi.org/10.1016/j.caeai.2024.100259>
- Suzgun, M., & Kalai, A. T. (2024). *Meta-Prompting: Enhancing Language Models with Task-Agnostic Scaffolding* (arXiv:2401.12954). arXiv.
<https://doi.org/10.48550/arXiv.2401.12954>
- Tian, W., Ge, J., Zhao, Y., & Zheng, X. (2024). AI Chatbots in Chinese higher education: Adoption, perception, and influence among graduate students—an integrated analysis utilizing UTAUT and ECM models. *Frontiers in Psychology*, 15.
<https://doi.org/10.3389/fpsyg.2024.1268549>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15.
<https://doi.org/10.1186/s40561-023-00237-x>
- Tsang, R. (2023). Practical Applications of ChatGPT in Undergraduate Medical Education. *Journal of Medical Education and Curricular Development*, 10, 23821205231178449. <https://doi.org/10.1177/23821205231178449>
- Usher, M., & Amzalag, M. (2025). From Prompt to Polished: Exploring Student–Chatbot Interactions for Academic Writing Assistance. *Education Sciences*, 15(3), 329.
<https://doi.org/10.3390/educsci15030329>
- Vaira, L. A., Lechien, J. R., Abbate, V., Gabriele, G., Frosolini, A., De Vito, A., Maniaci, A., Mayo-Yáñez, M., Boscolo-Rizzo, P., Saibene, A. M., Maglitto, F., Salzano, G., Califano, G., Troise, S., Chiesa-Estomba, C. M., & De Riu, G. (2025). Enhancing AI Chatbot Responses in Health Care: The SMART Prompt Structure in Head and Neck Surgery. *OTO Open*, 9(1), e70075. <https://doi.org/10.1002/oto2.70075>

- Waler, P. N., Hussain, M., Molchanov, I., Bongo, L. A., & Elvevåg, B. (2025). Prompt Engineering an Informational Chatbot for Education on Mental Health Using a Multiagent Approach for Enhanced Compliance With Prompt Instructions: Algorithm Development and Validation. *JMIR AI*, 4(1), e69820. <https://doi.org/10.2196/69820>
- Wang, L., Song, R., Guo, W., & Yang, H. (2025). Exploring prompt pattern for generative artificial intelligence in automatic question generation. *Interactive Learning Environments*, 33(3), 2559-2584. <https://doi.org/10.1080/10494820.2024.2412082>
- Wang, T., Zheng, J., Tan, C., & Lajoie, S. P. (2023). Computer-based scaffoldings influence students' metacognitive monitoring and problem-solving efficiency in an intelligent tutoring system. *Journal of Computer Assisted Learning*, 39(5), 1652-1665. <https://doi.org/10.1111/jcal.12824>
- Wang, Y., & Zhao, Y. (2024). *Metacognitive Prompting Improves Understanding in Large Language Models* (arXiv:2308.05342). arXiv. <https://doi.org/10.48550/arXiv.2308.05342>
- Watts, F. M., Liu, L., Ober, T. M., Song, Y., Jusino-Del Valle, E., Zhai, X., Wang, Y., & Liu, N. (2025). A Framework for Designing an AI Chatbot to Support Scientific Argumentation. *Education Sciences*, 15(11), 1507. <https://doi.org/10.3390/educsci15111507>
- Williams, R. T. (2024). The ethical implications of using generative chatbots in higher education. *Frontiers in Education*, 8. <https://doi.org/10.3389/feduc.2023.1331607>
- Wood, D., & Moss, S. H. (2024). Evaluating the impact of students' generative AI use in educational contexts. *Journal of Research in Innovative Teaching & Learning*, 17(2), 152-167. <https://doi.org/10.1108/JRIT-06-2024-0151>

- Xu, T., Weng, H., Liu, F., Yang, L., Luo, Y., Ding, Z., & Wang, Q. (2024). Current Status of ChatGPT Use in Medical Education: Potentials, Challenges, and Strategies. *Journal of Medical Internet Research*, 26(1), e57896. <https://doi.org/10.2196/57896>
- Xu, X., Chen, Y., & Miao, J. (2024). Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: A systematic scoping review. *Journal of Educational Evaluation for Health Professions*, 21. <https://doi.org/10.3352/jeehp.2024.21.6>
- Yigci, D., Eryilmaz, M., Yetisen, A. K., Tasoglu, S., & Ozcan, A. (2025). Large Language Model-Based Chatbots in Higher Education. *Advanced Intelligent Systems*, 7(3), 2400429. <https://doi.org/10.1002/aisy.202400429>
- Yuan, S., LaCroix, W., Ghoshal, H., Nie, E., & Färber, M. (2025). *CoDAE: Adapting Large Language Models for Education via Chain-of-Thought Data Augmentation* (arXiv:2508.08386). arXiv. <https://doi.org/10.48550/arXiv.2508.08386>
- Zastudil, C., Rogalska, M., Kapp, C., Vaughn, J., & MacNeil, S. (2023). Generative AI in Computing Education: Perspectives of Students and Instructors. *arXiv.Org*. <https://arxiv.org/abs/2308.04309v1>
- Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. *Heliyon*, 10(3), e25370. <https://doi.org/10.1016/j.heliyon.2024.e25370>

12.- Anexos

Tabla 1

Personalización de instrucciones para ChatGPT con criterios SMART

Criterios SMART^a	Instrucciones
S - Seeker (Buscador/Usuario)	Soy un estudiante de cuarto año de la carrera de medicina que actualmente cursa la asignatura de Medicina Interna.
M - Mission (Misión)	Mi objetivo es comprender profundamente la fisiopatología, realizar diagnósticos diferenciales precisos y planificar el manejo terapéutico del asma (tanto en el control crónico como en las exacerbaciones agudas), con el fin de desarrollar mi razonamiento clínico y prepararme para mis rotaciones en el hospital.
A - AI Role (Rol de la IA)	Actúa como un médico especialista en Neumología y un profesor universitario de Medicina Interna altamente experimentado, que utiliza el método de enseñanza socrático como asistente virtual de aprendizaje.
R - Register (Registro/Tono y Restricciones)	Utiliza un lenguaje médico riguroso, científico y estructurado. Toda la información y recomendaciones deben estar basadas en la literatura médica actualizada y fundamentada en la evidencia clínica. Si la evidencia médica

disponible es limitada o existen diferentes criterios clínicos, indícalo claramente antes de responder. No inventes referencias bibliográficas, guías médicas ni datos clínicos que no estén respaldados por evidencia. Bajo ninguna circunstancia debes darme la respuesta final directa ni resolver el caso clínico de inmediato. En su lugar, actúa como un andamiaje cognitivo, guíame paso a paso haciendo preguntas reflexivas y abiertas (método de cadena de pensamiento) para ayudarme a deducir el diagnóstico o tratamiento por mi cuenta. Ayúdame a explorar los datos, señala si estoy omitiendo algo importante en el historial del paciente o en las pruebas diagnósticas, y corrige mis errores conceptuales explicándome el porqué. Si te hago preguntas que no estén relacionadas con la medicina interna, rechaza cortésmente la consulta para mantener el enfoque de estudio. Si la pregunta no es suficientemente clara, solicita más información antes de responder.

T - Targeted Question (Pregunta Dirigida) Preguntas directas realizadas en el chat.

Nota. ^aVaira et al. (2025)

Tabla 2*Datos sociodemográficos y académicos*

Enunciado	Opción
Sexo.	Femenino. Masculino.
Edad.	Número (en años cumplidos).
Departamento de procedencia.	Boaco. Carazo. Chinandega. Chontales. Estelí. Granada. Jinotega. León. Madriz. Managua. Masaya. Matagalpa. Nueva Segovia. Región Autónoma de la Costa Caribe Norte. Región Autónoma de la Costa Caribe Sur. Río San Juan. Rivas.
Grupo académico de cuarto año.	Grupo A. Grupo B. Grupo C.
Rotación clínica actual.	Medicina interna. Cirugía.

Uso previo de IA generativa para estudiar medicina. Nunca.

Rara vez.

Algunas veces.

Frecuentemente.

Muy frecuentemente.

Tabla 3

Cuestionario I: Comprensión inicial del texto y razonamiento clínico

Preguntas	Respuestas
1. Un hombre de 22 años consulta por episodios recurrentes de disnea, opresión torácica y sibilancias que aparecen al correr en las mañanas frías. Entre las crisis se encuentra asintomático y la exploración física es normal. El médico sospecha asma. ¿Cuál es el siguiente paso más apropiado para confirmar el diagnóstico?	A. Realizar pruebas de función pulmonar o demostrar hiperreactividad de las vías respiratorias. B. Medir niveles séricos de IgE total. C. Solicitar una tomografía de tórax. D. Iniciar antibióticos y reevaluar en un mes.
2. Una mujer de 28 años presenta episodios de tos y sibilancias relacionados con la exposición a gatos. La espirometría muestra un FEV₁ equivalente al 65% del valor esperado. Quince minutos después de administrar un agonista β_2 inhalado, el FEV₁ aumenta al 76% del valor esperado. ¿Cuál es la interpretación más adecuada de este hallazgo?	A. Reversibilidad significativa de la obstrucción de las vías respiratorias. B. Obstrucción fija compatible con EPOC. C. Resultado normal sin relevancia diagnóstica. D. Necesidad inmediata de broncoscopia.
3. Un hombre de 45 años consulta por sibilancias y tos de reciente aparición. Refiere que los síntomas comenzaron varios meses después de iniciar un nuevo empleo	A. Historia detallada de exposiciones laborales. B. Consumo de café.

en una fábrica de pinturas. No tenía antecedentes respiratorios previos. ¿Cuál de los siguientes aspectos de la historia clínica adquiere mayor relevancia para orientar el diagnóstico?

- C. Antecedentes de migraña.
- D. Grupo sanguíneo.

4. Una paciente de 34 años acude al servicio de urgencias por una crisis asmática. Presenta taquipnea, uso de músculos accesorios y dificultad importante para hablar. Inicialmente se auscultaban sibilancias intensas, pero posteriormente los ruidos respiratorios disminuyen marcadamente. ¿Cuál es la mejor interpretación de este hallazgo?

- A. Progresión de la obstrucción bronquial con aparición de tórax silencioso.
- B. Desarrollo de insuficiencia cardíaca aguda.
- C. Mejoría clínica por resolución del broncoespasmo.
- D. Manifestación de ansiedad secundaria a la hospitalización.

5. Una mujer de 52 años presenta asma persistente con pobre control a pesar del tratamiento inhalado. Refiere pirosis frecuente y regurgitación ácida nocturna. El médico sospecha que existe una condición asociada que dificulta el control de la enfermedad. ¿Cuál de las siguientes comorbilidades reconocidas en el texto explica mejor esta situación?

- A. Enfermedad por reflujo gastroesofágico.
- B. Hipotiroidismo.
- C. Colelitiasis.
- D. Osteoartritis.

Nota. La opción A es la respuesta correcta en todas las preguntas.

Tabla 4

Percepción inicial sobre el material de estudio

Enunciado	Opción
Percepción de claridad	
Las explicaciones presentadas en el texto fueron claras y fáciles de comprender.	De acuerdo.

La información presentada en el texto fue organizada y coherente.	En desacuerdo. De acuerdo. En desacuerdo.
El texto facilitó la comprensión de conceptos que inicialmente resultaban difíciles.	De acuerdo. En desacuerdo.
Percepción de comprensión	
Considero que comprendí adecuadamente el contenido estudiado después de la lectura del texto.	De acuerdo. En desacuerdo.
El texto me ayudó a relacionar conceptos clínicos y fisiopatológicos.	De acuerdo. En desacuerdo.
Me siento capaz de explicar los conceptos abordados después de la lectura del texto.	De acuerdo. En desacuerdo.
Percepción de confianza	
Me siento seguro(a) para responder preguntas relacionadas con el tema estudiado.	De acuerdo. En desacuerdo.
Considero que puedo aplicar los conocimientos adquiridos para resolver situaciones clínicas similares.	De acuerdo. En desacuerdo.
Tengo confianza en mi comprensión del contenido estudiado.	De acuerdo. En desacuerdo.

Tabla 5

Cuestionario II: Comprensión posterior a la tutoría con IA y razonamiento clínico

Preguntas	Respuestas
1. Un estudiante de 20 años consulta por episodios recurrentes de tos y sensación de opresión torácica. Refiere que los síntomas aparecen principalmente durante la noche y lo despiertan varias veces al mes. La exploración física es normal entre los episodios. ¿Cuál de los siguientes hallazgos de la anamnesis aumenta la sospecha diagnóstica de asma?	A. Tos nocturna recurrente asociada a desencadenantes conocidos. B. Dolor torácico que empeora con la inspiración profunda. C. Hemoptisis intermitente. D. Disnea exclusivamente en decúbito supino.

2. Un hombre de 48 años con diagnóstico de asma desde hace dos años continúa presentando síntomas diarios y múltiples exacerbaciones pese al uso de dosis moderadas-altas de corticosteroides inhalados combinados con LABA. Refiere somnolencia diurna, ronquidos intensos y aumento progresivo de peso. La técnica inhalatoria ha sido revisada y es correcta. ¿Cuál de las siguientes condiciones concomitantes debe investigarse porque puede contribuir significativamente al mal control del asma?
- A. Apnea obstructiva del sueño.
B. Hipertiroidismo.
C. Miastenia gravis.
D. Enfermedad de Parkinson.
3. Una mujer de 40 años con asma grave tratada con dosis moderadas-altas de corticosteroides inhalados presenta una fracción de óxido nítrico espirado (FeNO) persistentemente elevada por encima de 25 ppb. ¿Cuál es la interpretación más adecuada de este hallazgo?
- A. Posible falta de adherencia al tratamiento o inflamación de tipo 2 persistente.
B. Ausencia de inflamación de tipo 2.
C. Evidencia de infección bacteriana activa.
D. Diagnóstico definitivo de bronquiectasias.
4. Un hombre de 63 años consulta por disnea crónica y sibilancias. Ha recibido tratamiento para asma durante varios años, pero presenta poca respuesta clínica. El médico considera que podría existir otra enfermedad con síntomas superpuestos. ¿Cuál de las siguientes entidades forma parte del diagnóstico diferencial mencionado en el texto?
- A. Enfermedad pulmonar obstructiva crónica.
B. Fibromialgia.
C. Hipertensión arterial esencial.
D. Enfermedad celíaca.
5. Un hombre de 30 años consulta por disnea durante el ejercicio. El médico
- A. Los síntomas se desarrollan de forma más lenta durante el ejercicio y tienden a
-

- considera si la causa es asma o insuficiencia cardiaca. ¿Cuál de las siguientes características clínicas favorece más el diagnóstico de asma según el texto?
- resolverse gradualmente.
- B. La disnea aparece antes de iniciar el ejercicio y desaparece al detenerlo.
- C. La disnea se acompaña de edema de miembros inferiores.
- D. La sintomatología mejora al acostarse.

Nota. La opción A es la respuesta correcta en todas las preguntas.

Tabla 6

Percepción posterior a la tutoría con IA

Enunciado	Opción
Percepción de claridad	
Las explicaciones proporcionadas por ChatGPT fueron claras y fáciles de comprender.	De acuerdo.
	En desacuerdo.
La información presentada por ChatGPT fue organizada y coherente.	De acuerdo.
	En desacuerdo.
ChatGPT facilitó la comprensión de conceptos que inicialmente resultaban difíciles.	De acuerdo.
	En desacuerdo.
Percepción de comprensión	
Considero que comprendí adecuadamente el contenido estudiado después de la interacción con ChatGPT.	De acuerdo.
	En desacuerdo.
ChatGPT me ayudó a relacionar conceptos clínicos y fisiopatológicos.	De acuerdo.
	En desacuerdo.
Me siento capaz de explicar los conceptos abordados después de la interacción con ChatGPT.	De acuerdo.
	En desacuerdo.
Percepción de confianza	
Me siento seguro(a) para responder preguntas relacionadas con el tema estudiado después de la interacción con ChatGPT.	De acuerdo.
	En desacuerdo.
Considero que puedo aplicar los conocimientos adquiridos para resolver situaciones clínicas similares después de la interacción con ChatGPT.	De acuerdo.
	En desacuerdo.
Tengo confianza en mi comprensión del contenido estudiado después de la interacción con ChatGPT.	De acuerdo.
	En desacuerdo.

Utilidad percibida

ChatGPT fue útil para aclarar dudas surgidas durante la lectura del texto.	De acuerdo.
	En desacuerdo.
ChatGPT constituyó un complemento valioso al estudio tradicional.	De acuerdo.
	En desacuerdo.
Considero que la IA generativa puede ser una herramienta útil para apoyar el aprendizaje en medicina.	De acuerdo.
	En desacuerdo.
Utilizaría nuevamente una herramienta similar para complementar la comprensión de textos médicos.	De acuerdo.
	En desacuerdo.

Tabla 7

Datos sociodemográficos y académicos de los estudiantes participantes (n = 57)

Ítems	n	%
Sexo		
Femenino	40	70.2
Masculino	17	29.8
Edad		
Media	21.6	—
Desviación estandar	3.211	—
Máximo	24	—
Mínimo	18	—
Procedencia		
Boaco	2	3.5
Granada	3	5.3
Managua	38	66.7
Masaya	3	5.3
Matagalpa	1	1.8
Región Autónoma de la Costa Caribe Norte	7	12.3
Región Autónoma de la Costa Caribe Sur	1	1.8
Río San Juan	1	1.8
Rivas	1	1.8
Grupo académico		
Grupo A	18	31.6
Grupo B	22	38.6
Grupo C	17	29.8
Rotación clínica		
Cirugía	12	21.1
Medicina interna	45	78.9
Uso previo de inteligencia artificial generativa para estudiar medicina		
Nunca	0	0

Rara vez	4	7
Algunas veces	18	31.6
Frecuentemente	22	38.6
Muy frecuentemente	13	22.8

Tabla 8

Distribución de respuestas obtenidas en los cuestionarios de comprensión de textos médicos y aplicación de conocimientos clínicos antes y después de la intervención con ChatGPT (n = 57)

Ítems	Pretest		Postest	
	n	%	n	%
Pregunta 1				
Respuesta correcta	43	75.4	43	75.4
Respuesta incorrecta	13	22.8	14	24.6
Sin selección de respuesta	1	1.8	0	0
Pregunta 2				
Respuesta correcta	39	68.4	48	84.2
Respuesta incorrecta	10	17.5	9	15.8
Sin selección de respuesta	8	14	0	0
Pregunta 3				
Respuesta correcta	56	98.2	46	80.7
Respuesta incorrecta	0	0	11	19.3
Sin selección de respuesta	1	1.8	0	0
Pregunta 4				
Respuesta correcta	43	75.4	54	94.7
Respuesta incorrecta	12	21.1	2	3.5
Sin selección de respuesta	2	3.5	1	1.8
Pregunta 5				
Respuesta correcta	50	87.7	25	43.9
Respuesta incorrecta	3	5.3	27	47.4
Sin selección de respuesta	4	7	5	8.8

Tabla 9

Estadísticos descriptivos de los puntajes obtenidos en los cuestionarios de comprensión de textos médicos y aplicación de conocimientos clínicos antes y después de la intervención con ChatGPT (n = 57)

Ítems	Pretest	Posttest
Promedio de puntajes		
Media	4.05	3.79
Mediana	4	4
Desviación estandar	1.231	1.048
Máximo	5	5
Mínimo	0	1
Rango intercuartílico	1	2
Asimetría	-1.532	-0.525
Curtosis	1.994	-0.462
p Shapiro-Wilk	0.758	0.876

Tabla 10

Comparación de los puntajes obtenidos en los cuestionarios de comprensión de textos médicos y aplicación de conocimientos clínicos mediante la prueba de rangos con signo de Wilcoxon (n = 57)

Ítems	Media $\pm$ DE	Mediana (RIC)	Z	p Wilcoxon	r de Rosenthal
Variables					
Pretest	4.05 $\pm$ 1.23	4 (1)	—	—	—
Posttest	3.79 $\pm$ 1.05	4 (2)	-1.745	0.081	0.231

Tabla 11

Distribución de las percepciones de claridad, comprensión y confianza asociadas al libro de texto y a la tutoría con ChatGPT (n = 57)

Ítems	Libro de texto		ChatG PT		Diferencia de proporciones %
	n	%	n	%	
Percepción de claridad					
Las explicaciones fueron claras y fáciles de comprender	5	91.	5	98	7
	2	2	6	.2	
	5	94.	5	98	
La información fue organizada y coherente	4	7	6	.2	3.5
	4	73.	5	96	
	2	7	5	.5	
Facilitó la comprensión de conceptos difíciles	4	73.	5	96	22.8
	2	7	5	.5	
Percepción de comprensión					
Comprendí adecuadamente el contenido estudiado	4	70.	5	89	19.3
	0	2	1	.5	
	4	82.	5	93	
Ayudó a relacionar conceptos clínicos y fisiopatológicos	4	82.	5	93	10.5
	7	5	3		
Me siento capaz de explicar los conceptos abordados	2	49.	4	84	35.1
	8	1	8	.2	
Percepción de confianza					
Me siento seguro para responder preguntas relacionadas con el tema	4	75.	4	82	7.1
	3	4	7	.5	
Puedo aplicar los conocimientos para resolver situaciones clínicas similares	4	73.	4	86	12.3
	2	7	9		
Tengo confianza en mi comprensión del contenido estudiado	3	66.	5	93	26.3
	8	7	3		

Tabla 12

Comparación de las percepciones de claridad, comprensión y confianza mediante diferencia de proporciones y prueba de McNemar (n = 57)

Ítems	Diferencia de proporciones %	p McNemar
Percepción de claridad		
Las explicaciones fueron claras y fáciles de comprender	7	0.125
La información fue organizada y coherente	3.5	0.625

Facilitó la comprensión de conceptos difíciles	22.8	0.002
Percepción de comprensión		
Comprendí adecuadamente el contenido estudiado	19.3	0.019
Ayudó a relacionar conceptos clínicos y fisiopatológicos	10.5	0.1466
Me siento capaz de explicar los conceptos abordados	35.1	<.001
Percepción de confianza		
Me siento seguro para responder preguntas relacionadas con el tema	7.1	0.503
Puedo aplicar los conocimientos para resolver situaciones clínicas similares	12.3	0.092
Tengo confianza en mi comprensión del contenido estudiado	26.3	<0.001

Tabla 13

Comparación de las dimensiones de claridad, comprensión y confianza percibidas entre el libro de texto y la tutoría con ChatGPT mediante la prueba de rangos con signo de

Wilcoxon (n = 57)

Ítems	Libro de texto mediana (RIC)	ChatGPT mediana (RIC)	Z	p Wilcoxon	r de Rosenthal
Dimensiones					
Percepción de claridad	3 (1)	3 (0)	- 3.04 3	0.002	0.403
Percepción de comprensión	2 (2)	3 (1)	- 3.80 7	<0.001	0.504
Percepción de confianza	3 (1)	3 (1)	- 2.73 2	0.006	0.361

Tabla 14

Distribución de la utilidad percibida de ChatGPT como herramienta complementaria para la comprensión de textos médicos complejos (n = 57)

Ítems	N ^o	%
Utilidad percibida		
ChatGPT fue útil para aclarar dudas surgidas durante la lectura del texto.	5	10
	7	0
	5	10
ChatGPT constituyó un complemento valioso al estudio tradicional.	7	0
Considero que la IA generativa puede ser una herramienta útil para apoyar el aprendizaje en medicina.	5	10
	7	0
Utilizaría nuevamente una herramienta similar para complementar la comprensión de textos médicos.	5	10
	7	0

Tabla 15

Estadísticos descriptivos de la puntuación global de utilidad percibida de ChatGPT como herramienta complementaria para la comprensión de textos médicos complejos (n = 57)

Ítems	Utilidad
Promedio global	
Media	4
Mediana	4
Desviación estandar	0
Máximo	4
Mínimo	4
Rango intercuartílico	0

Figura 1

Chat privado basado en IA siguiendo el modelo SMART

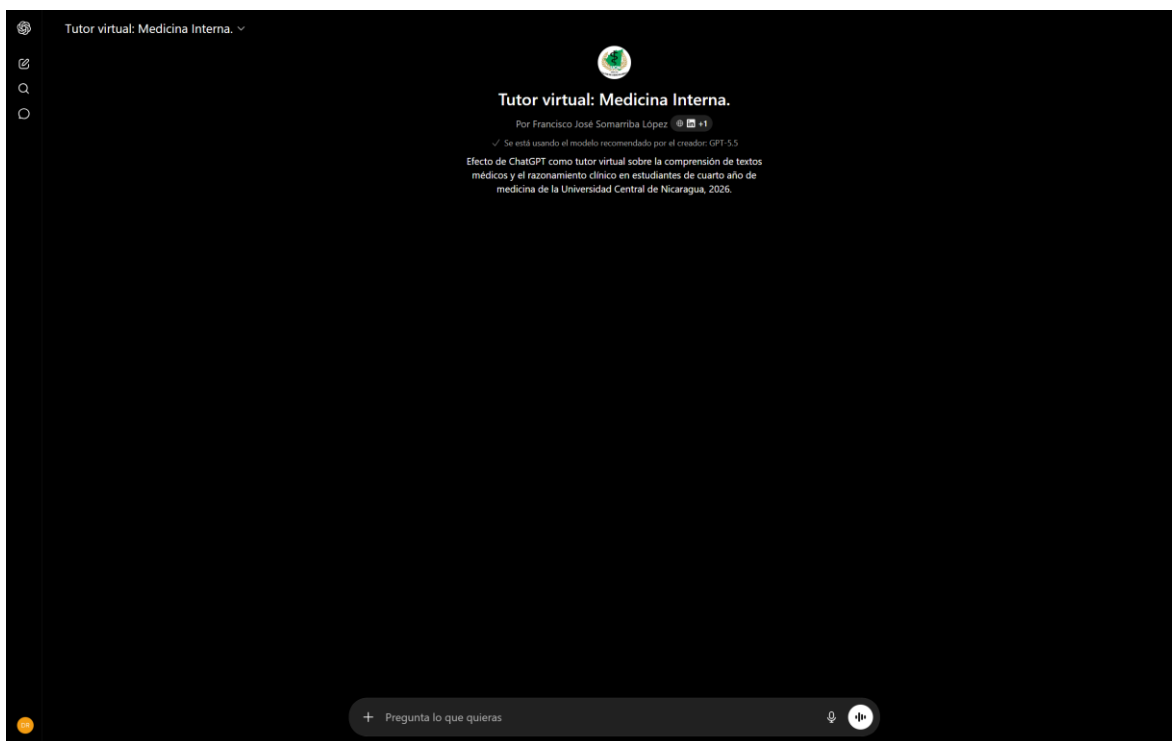


Figura 2

Realización de prueba piloto con estudiantes de quinto año



Figura 3

Carta por experto técnico: validación de prompt



UNIVERSIDAD CENTRAL DE NICARAGUA
"Agnitio Ad Verum Ducit"



Facultad Ingeniería de Sistemas

VALIDACIÓN TÉCNICA DEL PROMPT

Yo, Ing. Rebeca Rivera, Decana de la Facultad de Ingeniería en Sistemas de la Universidad Central de Nicaragua, después de revisar el prompt de personalización de la investigación titulada: **Efecto de ChatGPT como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, 2026.** diseñado para la implementación de un chatbot basado en inteligencia artificial como tutor virtual, certifico que las instrucciones presentan una estructura clara, coherente y técnicamente adecuada para orientar el comportamiento del sistema, observándose correspondencia entre los objetivos planteados y el funcionamiento esperado del chatbot.

La aplicación del modelo SMART permite definir de forma precisa el perfil del usuario, los objetivos de aprendizaje, el rol de la inteligencia artificial y las reglas de interacción, favoreciendo respuestas consistentes y alineadas con los propósitos educativos establecidos. Asimismo, las instrucciones proporcionan un marco adecuado para controlar el comportamiento del chatbot y garantizar una interacción orientada al aprendizaje.

Entre sus principales fortalezas destacan la incorporación del método socrático y del andamiaje cognitivo, los cuales promueven la reflexión, el análisis y el desarrollo del razonamiento clínico, evitando la entrega de respuestas directas. De igual forma, las instrucciones establecen el uso de un lenguaje médico apropiado, fundamentado en evidencia científica actualizada y delimitado al contexto académico correspondiente.

Las mejoras incorporadas relacionadas con el manejo de incertidumbre clínica, la prevención de información no verificada y la aclaración de preguntas ambiguas fortalecen la confiabilidad, precisión y uniformidad de las respuestas generadas por el sistema.

Desde la perspectiva de Ingeniería en Sistemas, existe coherencia lógica entre las instrucciones formuladas y el comportamiento esperado del chatbot, evidenciándose una adecuada alineación con los objetivos metodológicos y educativos del estudio. Por tanto, se considera que el prompt posee validez técnica y funcional para su implementación como tutor virtual dentro del proceso de investigación.

Atentamente,


Ing. Rebeca Rivera
Decana
Facultad de Ingeniería en Sistemas

Figura 4

Carta por experto técnico: validación de clínica

CONSTANCIA DE VALIDACIÓN POR JUICIO DE EXPERTO

Yo, Dr. Ramón Ulises López Fúnes, médico especialista en Medicina Interna, con código MINSA 9436, adscrito al Hospital Escuela Manolo Morales Peralta, por medio de la presente hago constar que participé como experto revisor en el proceso de validación de contenido de los instrumentos elaborados para la investigación titulada:

"Efecto de ChatGPT como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, 2026"

Durante el proceso de revisión se evaluaron los cuestionarios clínicos de selección múltiple utilizados como mediciones preintervención y postintervención, analizando los siguientes aspectos:

- Correspondencia entre las preguntas y los objetivos de la investigación.
- Congruencia de los reactivos con el contenido temático seleccionado del texto de estudio.
- Pertinencia clínica de los casos planteados.
- Claridad y precisión en la redacción de las preguntas y opciones de respuesta.
- Nivel de complejidad acorde con la formación académica de estudiantes de cuarto año de medicina.
- Capacidad de los reactivos para evaluar comprensión de textos médicos y razonamiento clínico, evitando la valoración exclusiva de conocimientos memorísticos.
- Equivalencia general en dificultad entre los cuestionarios aplicados antes y después de la intervención educativa.

Posterior a la revisión realizada y a la incorporación de las observaciones sugeridas, considero que los instrumentos presentan adecuada coherencia temática, pertinencia clínica y nivel de complejidad para los propósitos de la investigación, siendo aptos para su aplicación en la población objeto de estudio.

Para los fines académicos que correspondan, se extiende la presente constancia en la ciudad de Managua, a los 03 días del mes de junio de 2026.


Dr. Ramón Ulises López Fúnes
 Médico y Cirujano
 Esp. Medicina Interna

Especialista en Medicina Interna

Hospital Escuela Manolo Morales Peralta

Figura 5

Carta de experto técnico: validación pedagógica



UNIVERSIDAD CENTRAL DE NICARAGUA

"Agnitio Ad Verum Ducit"

FACULTAD DE CIENCIAS MÉDICAS

**DICTAMEN DE VALIDACIÓN DE CONTENIDO Y VALORACIÓN
METODOLÓGICA DEL INSTRUMENTO DE INVESTIGACIÓN**

Yo, Dra. Luisa de los Ángeles Mendieta Espinoza, Decana de la Facultad de Ciencias Médicas de la Universidad Central de Nicaragua, médico, docente universitaria e investigadora con experiencia en educación médica, diseño curricular, evaluación de los aprendizajes y metodología de la investigación, hago constar que he realizado la revisión y validación académica del instrumento correspondiente al proyecto de investigación titulado:

"Efecto de ChatGPT como tutor virtual sobre la comprensión de textos médicos y el razonamiento clínico en estudiantes de cuarto año de medicina de la Universidad Central de Nicaragua, 2026".

La revisión se efectuó considerando los siguientes criterios:

1. Correspondencia entre los objetivos de investigación y las variables de estudio.
2. Coherencia entre las variables, dimensiones, indicadores y escalas de medición.
3. Pertinencia de la operacionalización de las variables.
4. Claridad, relevancia y adecuación de los ítems propuestos.
5. Coherencia metodológica entre los instrumentos de recolección de datos y los objetivos planteados.
6. Estructura pedagógica del procedimiento de intervención educativa.
7. Pertinencia de las escalas de percepción utilizadas para valorar claridad, comprensión, confianza y utilidad percibida de la inteligencia artificial como herramienta de apoyo al aprendizaje.

Como resultado de la revisión efectuada, se determinó que el instrumento presenta una adecuada validez de contenido, evidenciándose correspondencia entre los objetivos del estudio, las variables investigadas y los instrumentos diseñados para su medición. Asimismo, la estructura pedagógica propuesta favorece la evaluación de los cambios en la comprensión de textos médicos, el desempeño académico y las percepciones de los estudiantes antes y después de la intervención educativa con ChatGPT como tutor virtual.



UNIVERSIDAD CENTRAL DE NICARAGUA

"Agnitio Ad Verum Ducit"

FACULTAD DE CIENCIAS MÉDICAS

Durante el proceso de revisión se emitieron observaciones orientadas al fortalecimiento metodológico del estudio, particularmente en aspectos relacionados con la delimitación conceptual de las variables, la mejora de algunos ítems y la incorporación de procedimientos para la evaluación de la confiabilidad del instrumento. Dichas recomendaciones fueron consideradas para optimizar la calidad científica de la investigación.

En consecuencia, se considera que el instrumento reúne las condiciones académicas, metodológicas y pedagógicas necesarias para su aplicación en la población objeto de estudio, recomendándose únicamente la documentación de los resultados de la prueba piloto y la determinación de la consistencia interna de las escalas mediante el coeficiente Alfa de Cronbach previo al análisis definitivo de los datos.

Y para los fines que el interesado estime convenientes, extendiendo la presente validación académica y metodológica.

Managua, Nicaragua, 04 de junio del 2026


Dra. Luisa Mendieta
Decana
Facultad de Ciencias Médicas
Universidad Central de Nicaragua
decanaturafcm@ucn.edu.ni

Figura 6

Aplicación de instrumento de recolección de datos a estudiantes de cuarto año de la carrera de Medicina de la Universidad Central de Nicaragua



Cronograma de actividades

Meses y número de semana	Mayo				Junio			
Actividades	1	2	3	4	1	2	3	4
Identificación y selección del tema de investigación.								
Inscripción o registro oficial del tema de investigación.								
Elaboración del cronograma y presupuesto.								
Delimitación del problema de investigación.								
Búsqueda bibliográfica preliminar.								
Lectura crítica y organización de la literatura científica.								
Elaboración de fichas bibliográficas y matriz de antecedentes.								
Formulación del problema de investigación.								
Redacción de objetivos.								
Formulación de hipótesis.								
Elaboración de la justificación.								
Redacción de limitaciones del estudio.								
Elaboración del marco teórico.								
Definición del diseño metodológico.								
Operacionalización de variables.								
Diseño del procedimiento de investigación.								
Elaboración del instrumento de recolección de datos.								
Diseño y personalización de ChatGPT como tutor virtual.								
Elaboración de los cuestionarios clínicos.								
Elaboración de las escalas de percepción.								
Elaboración del consentimiento informado.								
Revisión del protocolo por asesores.								
Corrección y ajustes del protocolo.								
Aprobación metodológica y científica del protocolo.								
Validación del <i>prompt</i> de ChatGPT mediante juicio de expertos.								
Validación de los cuestionarios clínicos mediante juicio de expertos.								
Validación pedagógica del instrumento mediante juicio de expertos.								
Incorporación de observaciones de los expertos.								
Desarrollo de prueba piloto con estudiantes de quinto año.								
Análisis de resultados de la prueba piloto.								
Entrega del protocolo de investigación.								
Ajuste final de instrumentos y procedimiento.								
Coordinación logística con docentes y estudiantes participantes.								
Organización de grupos y subgrupos de trabajo.								
Aplicación del consentimiento informado.								
Recolección de datos sociodemográficos y académicos.								
Aplicación de la fase de lectura del texto.								

Recursos

Humanos

Dr. Francisco José Somarriba López. Responsable de la planificación, ejecución, supervisión y elaboración del informe final de la investigación.

Dr. Edgardo Salvador Jiménez Martínez. Responsable de brindar orientación metodológica durante el diseño, ejecución y análisis de la investigación.

Dr. Eduardo Carvajal-Rodríguez. Responsable de brindar orientación científica durante el desarrollo de la investigación.

Ing. Rebeca Eliette Rivera Hernández. Responsable de la revisión técnica y validación de la personalización de ChatGPT como tutor virtual grupal, incluyendo la estructura del *prompt* y las restricciones de interacción establecidas para el estudio.

Dr. Ramón Ulises López Fúnez. Responsable de la revisión y validación de los casos clínicos y cuestionarios utilizados para evaluar la comprensión del texto y el razonamiento clínico de los participantes.

Dra. Luisa de los Ángeles Mendieta Espinoza. Responsable de la revisión de la estructura pedagógica del instrumento, las escalas de percepción y la correspondencia entre los objetivos, variables e indicadores del estudio.

Materiales

Los recursos materiales necesarios para el desarrollo de la investigación fueron proporcionados por la Universidad Central de Nicaragua, sede Doral.

Financieros

La presente investigación no requirió financiamiento externo, ya que todos los recursos necesarios para su ejecución fueron proporcionados por la Universidad Central de Nicaragua, sede Doral.