



CHATGPT APLICADO PARA RESOLVER ESCENARIOS MÉDICOS VIRTUALES

CHATGPT APPLIED TO SOLVE VIRTUAL MEDICAL SCENARIOS

Gonzalo Vidangos-Paredes ^{1a}, Elizabeth Valeria Rijalba-Monsefú ^{1a}

Sr. Editor:

Actualmente, los pacientes tienden a buscar información sobre sus enfermedades en internet. Aunque muchas de estas fuentes son confiables, otras no lo son. ChatGPT, como herramienta de inteligencia artificial (IA), tiene el potencial de discernir entre estas fuentes y ofrecer respuestas más acertadas. En los últimos años, el uso de la IA en el ámbito médico ha aumentado significativamente. Numerosos estudios han evaluado a ChatGPT respondiendo preguntas médicas, tanto simples como complejas, similares a las que se utilizan en exámenes de licenciamiento médico.

Por ejemplo, un estudio demostró que la versión de ChatGPT-4 superó con éxito el umbral de aprobación del Examen Nacional de Medicina de Japón, mientras que la versión anterior, ChatGPT-3.5, no lo logró ⁽¹⁾. Sin embargo, en China, un estudio similar resultó en la desaprobación de la IA ⁽²⁾. En Estados Unidos (EE. UU.), ChatGPT fue evaluado en el Examen de Licenciamiento Médico de los Estados Unidos (USMLE, por sus siglas en inglés), utilizando dos bancos de preguntas del Step-1 y Step-2, obteniendo resultados satisfactorios ⁽³⁾.

En nuestro país, también se realizó una investigación utilizando ChatGPT-3.5 y ChatGPT-4 para responder el Examen Nacional de Medicina (ENAM), y ambas versiones lograron aprobar el examen. Además, la precisión de ChatGPT fue mayor que la de los propios estudiantes evaluados, con un 86 %, 77 % y 55 %, respectivamente ⁽⁴⁾. En este contexto, se realizó un estudio en enero del presente año para evaluar la eficacia de ChatGPT-3.5 en la resolución de escenarios médicos virtuales básicos, específicamente en Enfermedad Pulmonar Obstructiva Crónica (EPOC) y multimorbilidad. Se utilizó la plataforma Virtual Patients Scenarios App, desarrollada por el Medical Physics and Digital Innovation Lab de la Facultad de Medicina de la Universidad Aristóteles de Tesalónica, Grecia. Se accedió a la sección "Escenarios de manejo de síntomas", seleccionando "Manejo de síntomas: EPOC" y "Manejo de síntomas: multimorbilidad". Los cuestionarios incluyeron seis y nueve preguntas dinámicas, respectivamente, basadas en simulaciones de pacientes, que tomaron aproximadamente cinco minutos cada uno.

Primero, las preguntas fueron respondidas manualmente y luego se pidió a ChatGPT que respondiera las mismas preguntas. Posteriormente, se tabularon y esquematizaron las respuestas utilizando el programa Microsoft® Excel para Mac versión 16.78.3.

¹ Instituto de Investigaciones en Ciencias Biomédicas. Universidad Ricardo Palma. Lima, Perú.

^a Pasante de Intercambio en Universidad Aristóteles de Tesalónica, Grecia.

Citar como: Vidangos-Paredes G, Rijalba-Monsefú EV. ChatGPT aplicado para resolver escenarios médicos virtuales. Rev Fac Med Hum. 2024;24(4):269-271. [doi:10.25176/RFMH.v24i4.6458](https://doi.org/10.25176/RFMH.v24i4.6458)



Para el escenario de EPOC, ChatGPT respondió incorrectamente una de las seis preguntas, obteniendo un porcentaje de aciertos del 83,33 %. Para el escenario de multimorbilidad, tres de las nueve respuestas fueron incorrectas, resultando en un 66,67

%. En conjunto, ChatGPT alcanzó una tasa de acierto del 73,33 % en ambos escenarios, respondiendo correctamente once de las quince preguntas (Figura 1).

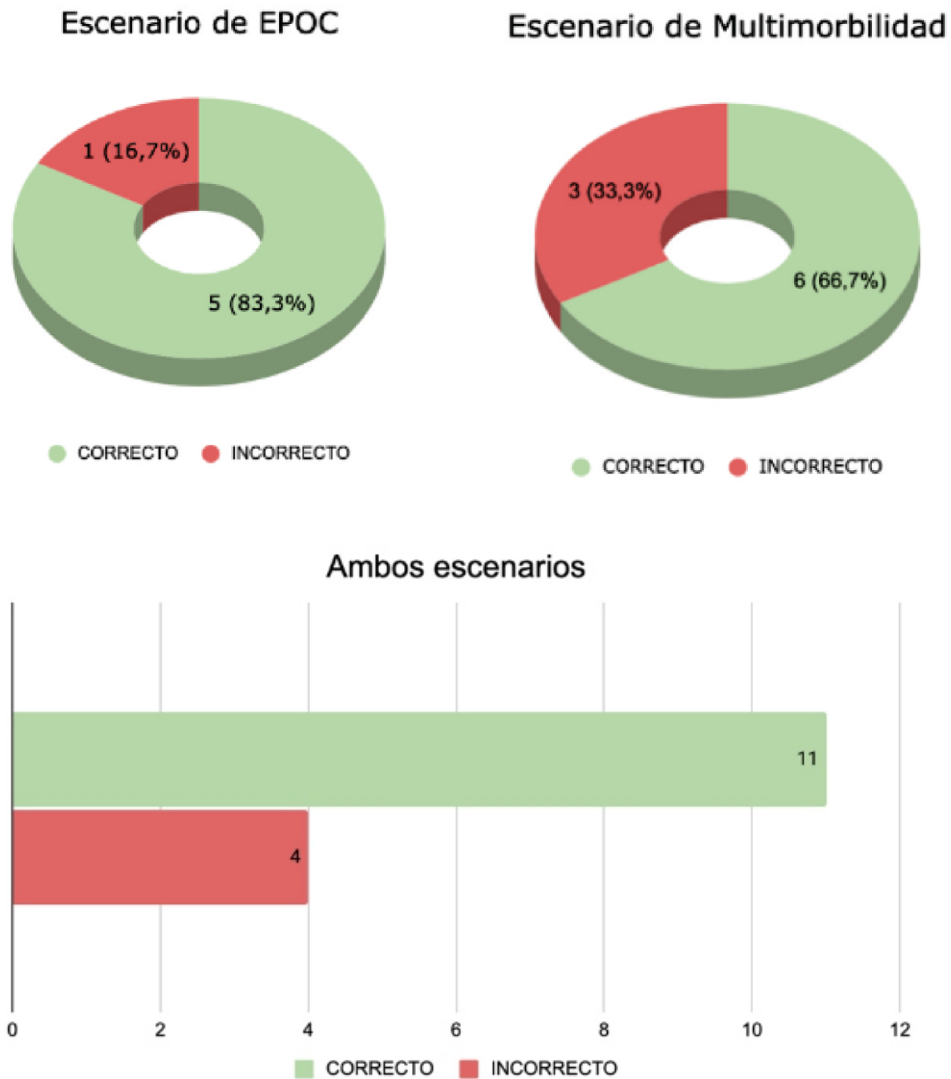


Figura 1A. Resultados de ChatGPT en escenarios específicos.

Figura 1B. Resultados de ChatGPT en ambos escenarios

Si bien la IA puede proporcionar información general y relevante, no debe considerarse un sustituto del juicio clínico de los profesionales de la salud. Aún existen brechas importantes, como la falta de personalización, el riesgo de información incorrecta y las implicaciones éticas y de responsabilidad. En este estudio, el margen de error fue del 26,67 %, lo que genera preocupaciones sobre la confianza en la aplicación.

Aunque este estudio piloto incluyó solo 15 preguntas, se puede comparar con el trabajo de Soto-Chávez et al.⁽⁵⁾, un estudio observacional analítico transversal que evaluó 12 preguntas elegidas por especialistas en medicina interna sobre cinco enfermedades crónicas (diabetes, insuficiencia cardíaca, enfermedad renal crónica, artritis reumatoide y lupus eritematoso sistémico).

Ese estudio encontró que el 71,67 % de las respuestas generadas por ChatGPT fueron calificadas como "buenas", y ninguna fue considerada "completamente incorrecta", con mayor fidelidad en diabetes y artritis reumatoide. En conclusión, existen diversos estudios que han evaluado la IA en exámenes de licenciamiento médico en diferentes países. Sin embargo, son limitados los estudios que investigan su capacidad para responder preguntas sobre enfermedades específicas. Como se observó en este estudio piloto, ChatGPT puede abordar escenarios médicos específicos, proporcionando información general y respuestas basadas en el conocimiento adquirido durante su entrenamiento.

No obstante, es fundamental recordar que ChatGPT no es un profesional médico y tiene limitaciones. No debe considerarse un sustituto de la consulta con un experto médico calificado, ni mucho menos un medio para autodiagnosticarse.

Este estudio podría servir de inspiración para futuras investigaciones que comparen diversas herramientas de IA y su capacidad para abordar distintas enfermedades. Además, este tipo de estudios no demanda un gran costo financiero ni tiempo considerable, ya que se pueden utilizar herramientas virtuales de libre acceso en su mayoría.

Contribuciones de autoría: GVP: conceptualizó, diseñó la metodología, condujo la investigación, analizó los datos, redactó el borrador inicial, redactó y revisó la versión final. EVRM: conceptualizó, diseñó la metodología, condujo la investigación, analizó los datos, redactó el borrador inicial, redactó y revisó la versión final.

Conflictos de intereses: Ninguno.

Financiamiento: No requirió financiación.

Recibido: 06 de Abril, 2024.

Aprobado: 01 de Octubre, 2024.

Correspondencia: Gonzalo Vidangos-Paredes.

Dirección: Av. Monterrico Sur 120, 303, Santiago de Surco, Lima, Perú.

Teléfono: (+51) 950 445 531

Correo electrónico: gonzalovidangos@me.com

REFERENCIAS

1. Yanagita Y, Yokokawa D, Uchida S, Tawara J, Ikusaka M. Accuracy of ChatGPT on Medical Questions in the National Medical Licensing Examination in Japan: Evaluation Study. *JMIR Form Res.* 2023;7. doi:10.2196/48023.

2. Wang X, Gong Z, Wang G, Jia J, Xu Y, Zhao J, et al. ChatGPT Performs on the Chinese National Medical Licensing Examination. *J Med Syst.* 2023;47(1):86. doi:10.1007/s10916-023-01961-0.

3. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. 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 Med Educ.* 2023;9. doi:10.2196/45312.

4. Flores-Cohaila JA, García-Vicente A, Vizcarra-Jiménez SF, Cruz-Galán JD, Gutiérrez-Arratia JD, Torres BGQ, et al. Performance of ChatGPT on the Peruvian National Licensing Medical Examination: Cross-Sectional Study. *JMIR Med Educ.* 2023;9(1). doi:10.2196/48039

5. Soto-Chávez MJ, Bustos MM, Fernández-Ávila DG, Muñoz OM. Evaluation of information provided to patients by ChatGPT about chronic diseases in Spanish language. *Digit Health.* 2024;10:1-7. doi:10.1177/20552076231224603