



El pulso de la Inteligencia Artificial y la alfabetización digital en Medicina: Nuevas herramientas, viejos desafíos

The Pulse of Artificial Intelligence and Digital Literacy in Medicine: New tools, old challenges

Sr. editor:

Tras leer con interés el debate suscitado por los dos artículos del equipo del Dr. Guillén-López ⁽¹⁻²⁾, coincidimos con los autores en la importancia de que los alumnos conozcan las distintas plataformas para la búsqueda bibliográfica ⁽¹⁾, así como sus puntos fuertes y debilidades, pues no todas ofrecen la misma funcionalidad y acceso a contenidos.

Si los estudiantes no tienen una formación sólida en el método científico, el resto de la comunidad médica se debilita, ya que tenderán a buscar información en fuentes no contrastadas o sin una revisión por pares, o simplemente no se sentirán seguros a la hora de compartir los casos de interés que encuentren en su devenir profesional. La habilidad de buscar y evaluar literatura médica de forma crítica es esencial para todas las especialidades médicas, ya que afecta directamente la toma de decisiones clínicas y la calidad del cuidado del paciente ⁽³⁾, que siempre han de estar basados en procedimientos sustentados en evidencias demostrables ⁽⁴⁾. Asimismo, los estudiantes deben desarrollar su juicio crítico y aprender a distinguir las fuentes confiables de aquellas que pueden estar sesgadas por intereses comerciales.

Los artículos del equipo del Dr. Guillén-López nos han hecho reflexionar sobre el impacto de la reciente aparición de las herramientas de acceso a la información basadas en Large Language Models (LLM), como GPT-4 de OpenAI o Bard de Google. Los LLM han demostrado ser útiles en muchos campos del conocimiento al interactuar en lenguaje natural. En las simulaciones de pruebas académicas médicas han destacado con calificaciones extremadamente altas (especialmente en inglés). Poco después de ponerse a disposición del público, chatGPT (también llamado

GPT-3.5) alcanzó puntuaciones cercanas al criterio de admisión (53,61% -58,78%) en simulaciones de las tres etapas del examen United States Medical Licensing Examination (USMLE) ⁽⁵⁾. La actual versión GPT-4 de OpenAI obtiene calificaciones muy superiores (86,7%) en pruebas semejantes ⁽⁶⁾, sin necesidad de un entrenamiento específico. Muy por delante de otros LLM especialmente diseñados para el ámbito médico (como el modelo Med-PaLM y Flan-PaLM 540B de Google). El fácil acceso vía web de estas herramientas y la aparente empatía que muestran hacia los médicos y pacientes ⁽⁷⁾, las han convertido en herramientas atractivas a las que preguntar acerca de todo tipo de dolencias y tratamientos, hasta el punto de ser reconocido su potencial para mejorar la educación médica ⁽⁵⁾.

Sin embargo, es importante tener en cuenta los riesgos que estas nuevas herramientas pueden presentar para los estudiantes de medicina. Una limitación clave es su incapacidad actual para citar las fuentes originales de los textos que generan. Esta falta de atribución puede dificultar la verificación de la exactitud y la fiabilidad de la información proporcionada por los LLM, lo que es especialmente preocupante en nuestro contexto, en el que la inexactitud de la información puede tener consecuencias significativas para la salud del paciente.

Además, los LLM también pueden sufrir de “alucinaciones”, un fenómeno por el que generan información que parece plausible pero que no está respaldada por los datos en los que fueron entrenados. Estas alucinaciones pueden ser particularmente problemáticas en los casos en que el conjunto de entrenamiento no tiene suficiente información sobre el tema de interés ⁽⁸⁾.

La alfabetización informacional y digital en medicina es un campo en rápida evolución. Por ello coincidimos plenamente con los autores de los trabajos originales ⁽¹⁻²⁾ en la necesidad de estudiar y contrastar las fuentes bibliográficas, pues la información clínica ofrecida por la nueva generación de herramientas basadas en inteligencia artificial mejorará con el tiempo, pero también puede inducir a error o carecer de precisión. Al fin y al cabo, es un médico humano quien está detrás de todos los algoritmos y sistemas informáticos, por tanto, las decisiones clínicas deben estar siempre ponderadas por el análisis cuidadoso de las evidencias, el espíritu crítico, la experiencia e, incluso, la intuición.

Rebeca Tenajas ^{1,a} , **David Miraut** ^{2,b} 

Correspondencia:

David Miraut

GMV. Calle Grisolia 4. 28760 Tres Cantos. España.

e-mail: dmiraut@gmv.com

REFERENCIAS BIBLIOGRÁFICAS

1. Guillén-López OB, Álvarez-Mayorga JH, Calle-Jacinto de Guillén DE. Fuentes de información utilizados por estudiantes de pregrado de Medicina. *Rev Med Hered.* 2023; 34(2): 107-9. doi:10.20453/rmh.v34i3.4550
2. Guillén-López OB, Álvarez-Mayorga JH, Calle-Jacinto de Guillén DE. Percepción de los estudiantes de pregrado de Medicina sobre sus habilidades para la investigación. *Rev Med Hered.* 2021; 32(4): 257-9. doi: <https://doi.org/10.20453/rmh.v32i4.4124>
3. Cabello López JB. Lectura crítica de la evidencia clínica. Elsevier Health Sciences; 2021. p 296.
4. Cook D, Rennie D, Guyatt G, Meade M. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice, Third Edition.* McGraw Hill Education; 2015. p 544.
5. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS digit. health.* 2023; 2(2): 1-12. doi:10.1371/journal.pdig.0000198
6. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375.* 2023; 1-35. doi:10.48550/arXiv.2303.13375
7. Ayers JW, Poliak A, Dredze M, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med.* 2023; 183(6): 589–596. doi:10.1001/jamainternmed.2023.1838
8. McCoy T, Pavlick E, Linzen T. “Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference”, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* 2019; pp. 3428-3448. doi:10.18653/v1/P19-1334

Recibido: 15/08/2023

Aceptado: 07/09/2023

¹ Departamento de Medicina de Familia, Centro de Sud de Arroyomolinos, España.

² Advanced Healthcare Technologies Department. GMV.

^a Doctora en Medicina, Máster en Medicina Clínica.

^b PhD.