

Viaje al corazón de las palabras

Inteligencia artificial y lenguaje médico (y IV)



El mes pasado¹ les mostraba varios ejemplos del modo en el que un bot de conversación basado en inteligencia artificial (IA) generativa, ChatGPT (en adelante, GePeTo o Gepeto), es capaz de desenvolverse con una pericia nada desdeñable en el terreno de la comunicación médico-paciente. Era de esperar también que la IA superara claramente a la inteligencia humana en la interpretación de sonidos auscultatorios, electrocardiogramas y técnicas de diagnóstico por la imagen (véanse los datos publicados en *Lancet Oncology*²).

Pero ¿y ejercer la medicina? ¿Podemos imaginar un «doctor Gepeto» al que acudan los pacientes en busca del diagnóstico más probable, una orientación pronóstica o las mejores pautas de tratamiento? A juzgar por los dos ejemplos siguientes, no parece ya ciencia ficción.

EL EMPÁTICO DOCTOR GEPETO

En un artículo publicado en abril de 2023 en la revista *JAMA Internal Medicine*^{3,4}, los investigadores compararon las respuestas ofrecidas por médicos de carne y hueso en una red social de consultas médicas («Ask a Doctor» de Reddit)⁵ con las que da GePeTo. Seleccionaron de manera aleatoria 195 preguntas planteadas por los usuarios y respondidas por médicos reales entre octubre y diciembre de 2022; a finales de diciembre, les hicieron esas mismas preguntas a GePeTo, y pidieron luego a un grupo integrado por tres profesionales sanitarios que evaluaran con sendas escalas de cinco puntos las respuestas en cuanto a calidad de la información proporcionada (muy mala, mala, aceptable, buena o muy buena) y en cuanto a empatía (nada empática, ligeramente empática, moderadamente empática, empática o muy empática). Con ambos criterios, GePeTo superó ampliamente a los facultativos humanos.

La proporción de respuestas evaluadas como de calidad «buena» o «muy buena» fue más de tres veces superior con GePeTo (79%) que con los médicos de carne y hueso (22%). Y en lo tocante a la empatía, la diferencia fue todavía mayor: el 45,1% de las respuestas de GePeTo se consideraron «empáticas» o «muy empáticas», frente a un escuálido 4,6% de las respuestas humanas. Aunque también es preciso tener en cuenta que los médicos, en general, tenemos fama de ser uno de los colectivos profesionales menos empáticos que uno pueda echarse en cara.

GEPETO APRUEBA SOBRODAMENTE EL EXAMEN MIR ESTADOUNIDENSE

El UMSLE (*United States Medical Licensing Examination*) es un examen nacional obligatorio que debe aprobar todo graduado en medicina para poder especializarse o ejercer esta disciplina en los Estados Unidos, más o menos equivalente, pues, al examen MIR (médico interno residente) en España o al ENARM (examen nacional de aspirantes a residencias médicas) en México.

A mediados de marzo de 2023, un grupo de investigadores² de Microsoft y OpenAI pidieron a la versión mejorada de GePeTo (GPT-4, lanzado por la tecnológica californiana OpenAI apenas unos días antes) que respondiera un examen UMSLE oficial, y los resultados obtenidos por el doctor GePeTo fueron espectaculares: sin ningún tipo de capacitación ni entrenamiento específicos en tareas médicas, superó en más de 20 puntos la nota de corte necesaria para aprobar el examen.

No solo respondió correctamente la mayor parte de las preguntas de elección múltiple, acertó incluso preguntas complejas que iban acompañadas de imágenes o fotografías (que GePeTo no podía «ver», pues en ese momento era todavía un modelo programado únicamente para leer texto escrito).

Ya no es utópico ni inimaginable un futuro próximo en el que GePeTo pueda emitir diagnósticos y pronósticos, pautar tratamientos y citar para revisión a un paciente. Lo que sí es seguro es que desde ya mismo podrá usarse con fines docentes. En el artículo de *arXiv*, los autores aportan un ejemplo sumamente ilustrativo. Cuando piden al bot de conversación que haga un diagnóstico diferencial, su respuesta es impecable: razona el diagnóstico probable y explica los motivos por los que descarta las demás posibilidades. Si así se le solicita, es capaz de explicar con detalle a un alumno que haya seleccionado una respuesta incorrecta por qué esta lo es; también, de indicar cuál pudo haber sido el razonamiento de ese alumno para seleccionar una respuesta errónea, e incluso, de escribir una nota de corrección para el estudiante con el estilo de una crítica comprensiva y constructiva.

Para rizar el rizo, GePeTo puede también reescribir el caso clínico de tal modo que, con el menor número de cambios posibles, la respuesta correcta pase a ser otra distinta que se le señale; es decir, es capaz de crear una hipotética situación contrafactual. No sé cómo lo ven ustedes: a mí me parece sencillamente espectacular.

* * *

¿Es perfecto ChatGPT? Desde luego que no; nadie lo es. Con lupa en mano, cabe señalar algunas muletillas (como el abuso de «realizar» y «utilizar») o el uso angloide de adjetivos posesivos con las partes del cuerpo (en lugar de artículos apoyados en un verbo construido con un pronombre personal, que es lo natural en español), pero eso mismo vemos a menudo en textos escritos por médicos españoles de carne y hueso. Es evidente también que ChatGPT se maneja mucho mejor en su lengua materna —el inglés— que en español; pero en esta (y muchas otras lenguas más) no lo hace nada mal. Cinco años atrás, muy pocos habrían vaticinado que en 2023 estaríamos hablando a este nivel con

una máquina. Si los datos que he recopilado para esta miniserie («Inteligencia artificial y lenguaje médico») corresponden al primer trimestre de 2023, ¿qué maravillas no podrá traernos la IA generativa dentro de otros cinco años más, allá por 2028?

Estoy convencido de que sí: los modelos de lenguaje se incorporarán a la actividad asistencial para facilitar la comunicación entre médico y paciente, y la IA se convertirá más pronto que tarde en un instrumento esencial de los profesionales sanitarios, de los profesionales de la lengua y de la población general.

No me pilló por sorpresa, pues, el pasado mes de enero, el nacimiento de una nueva revista médica interdisciplinar, *NEJM AI*⁶, que, bajo la dirección de Isaac (Zak) S. Kohane, catedrático y jefe del Departamento de Informática Biomédica de la Universidad de Harvard, explorará los avances, perspectivas, oportunidades y tendencias futuras de la IA en el ámbito de la medicina. La propia *New England Journal of Medicine* recopila también desde el año 2023, en un rincón de su ciber sitio llamado «AI in Medicine»⁷, los principales artículos de revisión sobre posibles aplicaciones prácticas de la IA a la asistencia sanitaria, que son muchas y cada vez serán más.

Fernando A. Navarro

Consejo Editorial, Revista Española de Cardiología

Obras de referencia recomendadas:

Diccionario de dudas y dificultades de traducción del inglés médico (3.ª edición), 2013-2024; en la plataforma Cosnautas disponible en www.cosnautas.com/es/catalogo/librorojo.

«Laboratorio del lenguaje» de *Diario Médico*, 2006-2024, disponible en www.diariomedico.com/opinion/fernando-navarro.html.

BIBLIOGRAFÍA

1. Navarro FA. Inteligencia artificial y lenguaje médico (III). *Rev Esp Cardiol*. 2024;77.
2. Lång K, Josefsson V, Larsson A-M, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *Lancet Oncol*. 2023;24:936–944.
3. Nori H, King N, McKinney SM, Carignan D, Horvitz E. «Capabilities of GPT-4 on medical challenge problems». *arXiv*. 2023. Disponible en <https://arxiv.org/pdf/2303.13375.pdf>. Consultado 12 Ago 2023..
4. 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 Int Med*. 2023;183:589–596.
5. Reddit AskDocs. Disponible en <https://www.reddit.com/r/AskDocs>. Consultado 12 Ago 2023.
6. NEJM AI. Disponible en <https://ai.nejm.org>. Consultado 12 Ago 2023.
7. VV. AA.. AI in Medicine. *New Eng J Med*. 2023. Disponible en www.nejm.org/aiin-medicine. Consultado 12 Ago 2023.