EL MERCURIO

Fecha: 15/04/2026 Audiencia: 320.543 Sección: CONTRAPORTADA
Vpe: $8.792.035 Tirada: 126.654 Frecuencia: 0
Vpe pág: $20.570.976 Difusión: 126.654
Vpe portada: $20.570.976 Ocupación: 42,74%

Investigaciones pusieron a prueba modelos como ChatGPT y Gemini para orientación médica

Los chatbots fallan como consejeros de salud: cometen errores y usan datos de foros y redes sociales

Aunque pueden acertar cuando tienen amplia información del usuario, suelen dar respuestas inexactas e incompletas, concluyen estudios. Además, se expresan con confianza aunque estén equivocados.

JANINA MARCANO

Los chatbots de IA generativa, como ChatGPT, se han convertido en un consultor para múltiples temas, y la salud es uno. De hecho, en América Latina, ChatGPT fue la aplicación más utilizada en 2025 para consultar sobre temas de salud, desplazando incluso a páginas web de centros médicos, según un estudio del Instituto Tecnológico de Buenos Aires.

Así, a medida que más personas recurren a estos sistemas para consultar síntomas o buscar orientación en salud —y a que también han comenzado a utilizarse como apoyo en la atención médica—, han surgido investigaciones científicas que ponen a prueba su desempeño.

Esta semana se publicaron los resultados de dos estudios que no son alentadores: aunque pueden entregar respuestas correctas cuando cuentan con suficiente información, estas herramientas cometen errores, usan fuentes cuestionables y responden con seguridad incluso cuando se equivocan.

"Todos los chatbots dan respuestas diferentes a las mismas preguntas de salud, basadas en sus datos de entrenamiento", comenta a "El Mercurio" Nicholas Tiller, coautor de una investigación liderada por el Instituto Lundquist para la Innovación Biomédica, organización afiliada a la U. de California.

Tiller y su equipo analizaron cinco chatbots populares y gratuitos —Gemini, DeepSeek, Meta AI, ChatGPT y Grok—, poniéndolos a prueba con preguntas comunes en categorías como cáncer y vacunas.

La principal conclusión es que la información que proporcionan es "inexacta e incompleta" y que "la mitad de las respuestas son problemáticas", ya que pueden inducir "a los usuarios a tratamientos ineficaces o con consecuencias para la salud". El estudio se publicó en la revista BMJ Open.

Tiller afirma que los modelos de lenguaje evaluados "no son buenos revisando la evidencia científica ni haciendo juicios, porque generan resultados infiriendo patrones estadísticos a partir de sus datos de entrenamiento y prediciendo secuencias de palabras. No razonan ni sopesan la evidencia".

El trabajo también concluyó que los chatbots utilizan datos de foros de preguntas y respuestas y de redes sociales, y que el contenido científico que emplean se limita a estudios gratuitos, los que representan entre el 30% y el 50% de las investigaciones publicadas.

Ricardo Seguel, profesor de la Facultad de Ingeniería y Ciencias de la UAI, explica que "los modelos son entrenados con fuentes muy diversas. Cuando uno les consulta, buscan patrones y generan respuestas coherentes. El problema es que no necesariamente distinguen cuál es la más adecuada".

Otro aspecto que relevaron los autores es que las respuestas de los modelos se "expresan con confianza y seguridad". "De 250 preguntas, solo hubo dos que se negaron a responder, ambas por parte de Meta AI", cuenta Tiller.

Y añade: "Son como tontos expertos, porque incluso cuando no tienen la información correcta, actúan como si la tuvieran, y el usuario interpreta su confianza como credibilidad. Es preocupante".

Denis Parra, investigador del Instituto Milenio en Ingeniería e Inteligencia Artificial para la Salud (IHealth), señala que uno de los grandes desafíos de estos sistemas es que "puedan distinguir cuándo realmente no tienen información suficiente para responder. Hoy eso no está resuelto".

Otra investigación, publicada el lunes en JAMA Network Open —y realizada por investigadores de la red de hospitales Mass General Brigham, asociada a la U. de Harvard—, concluye que la IA aún no está preparada para tomar decisiones médicas.

Réplica apresurada

Según los resultados, aunque los chatbots de IA aciertan en el diagnóstico en más del 90% de los casos cuando disponen de información clínica completa, muestran deficiencias importantes cuando deben trabajar con datos limitados, razonar o hacer "un diagnóstico diferencial", es decir, identificar la causa de los síntomas de un paciente, distinguiéndola de otras con manifestaciones similares.

"Por ejemplo, el dolor torácico podría reflejar reflujo, neumonía, embolia pulmonar o infarto. Un buen médico mantiene abiertas las posibilidades peligrosas, pero lo que observamos en los modelos de IA es que suelen llegar demasiado rápido a una única respuesta, lo que podría deteriorar al paciente o impulsar procedimientos innecesarios", dijo a este diario Marc Succi, investigador del Hospital General de Massachusetts Brigham y autor principal del estudio.

El equipo evaluó 21 de los modelos más avanzados —entre ellos, GPT-5, DeepSeek y Gemini—, pidiéndoles que actuaran como médicos en distintos escenarios.

Sobre esto, Parra comenta: "Los médicos saben cuándo detenerse, pedir más información y no apresurarse". Los modelos de lenguaje, en cambio, añade, "no tienen ese proceso deliberativo, porque su entrenamiento está orientado a siempre producir una salida".

El trabajo constató, eso sí, que los modelos más recientes superan a los más antiguos, lo que demuestra que siguen mejorando.

"Como usuario, actualmente no utilizaría un chatbot para orientación médica, pero eso no significa que las cosas no puedan cambiar a futuro", plantea Tiller.

Recomendación

Jaime de los Hoyos, jefe del Departamento de Informática Biomédica de Clínica Alemana, plantea que el uso de chatbots en salud debería limitarse, por ahora, a la información, por ejemplo, para entender mejor una enfermedad, pero no a la toma de decisiones. "Son una forma de aprender y acceder a información, pero todavía no reemplazan el juicio clínico", afirma. Los entrevistados coinciden en que no es aconsejable usarla para autodiagnosticarse, decidir tratamientos, o definir si acudir o no al médico.

FREEPIK/CREATIVE COMMONS

"La postura y hasta cómo el paciente entra caminando a consulta es información clínica para tomar decisiones de salud", plantea el médico Jaime de los Hoyos, sobre cómo la comunicación con la IA puede pasar por alto datos al dar un diagnóstico.