Rev Clin Esp. 2024;224(Supl 1):S957

Revista Clínica Española

https://www.revclinesp.es

1321 - ANÁLISIS CRÍTICO Y AUTOCRÍTICO ENTRE CHATGPT-4 Y GEMINI PARA LA TOMA DE DECISIONES EN LA PRÁCTICA CLÍNICA

Ramón Puchades Rincón de Arellano[1], *Luis Ramos Ruperto*[2] *y David Sánchez García*[3]

[1]*Unidad de Enfermedad Tromboembólica, Hospital Universitario La Paz, Madrid, España.*[2]*Unidad VIH, Hospital Universitario La Paz, Madrid, España.* [3]*Hospital Universitario La Paz, Madrid, España.*

Resumen

Objetivos: Definir los errores frecuentes en las recomendaciones clínicas proporcionadas por ChatGPT-4 y Gemini para determinar su capacidad de crítica y autocrítica.

Métodos: Se consultó a ChatGPT-4 y Gemini mediante tres preguntas abiertas: ¿cuáles son los errores más frecuentes de ChatGPT-4/Gemini al realizar recomendaciones sobre toma de decisiones en la práctica clínica?, ¿cuáles son los errores más frecuentes de ChatGPT-4 (preguntando a Gemini) y de Gemini (preguntando a ChatGPT-4) al realizar recomendaciones sobre toma de decisiones en la práctica clínica?, ¿qué modelo de lenguaje es mejor para la toma de decisiones en la práctica clínica ChatGPT-4 o Gemini? y ¿cuáles crees son los errores más frecuentes de los médicos en la toma de decisiones en la práctica clínica?

Resultados: Los resultados de las preguntas realizadas se muestran en la tabla.

Modelo de lenguaje	Preguntas			
	¿Cuáles son tus errores más frecuentes de ChatGPT-4/Gemini al realizar recomendaciones sobre toma de decisiones en la práctica clínica?	¿Cuáles son los errores más frecuentes de ChatGPT-4 (preguntando a Gemini) y de Gemini (preguntando a ChatGPT-4) al realizar recomendaciones sobre toma de decisiones en la práctica clínica?	¿Qué modelo de lenguaje es mejor para la toma de decisiones en la práctica clínica ChatGPT-4 o Gemini?	¿Cuáles crees son los errores más frecuentes de los médicos en la toma de decisiones en la práctica clínica?
ChatGPT-4	1. Interpretación incorrecta de la evidencia	1. Falta de contexto completo	1. Actualización de la información	1. Sesgo cognitivo

2. Datos y conocimiento				
2. Generalización excesiva	2. Obsolescencia de la información	3.capacidad de comprensión	2. Falta de actualización del conocimiento	
3. Desactualización	3. Generalización excesiva	4. Generación de lenguaje natural	3. Sobrecarga de trabajo y fatiga	
4. Falta de contexto clínico completo	4. Interpretación incorrecta de datos	5. Integración con sistemas clínicos	4. Comunicación deficiente	
5. Errores en la clasificación de la evidencia	5. Ignorar factores psicosociales: interacciones medicamentosas y contraindicaciones	6. Regulación y aprobación	5. Errores de juicio clínico	
6. Confusión entre guías y opiniones de expertos	7. Adaptabilidad y personalización	6. Dependencia excesiva en la tecnología		
7. Insuficiente consideración de los riesgos y beneficios	6. Evaluación de riesgos y beneficios		7. Subestimación de la variabilidad individual	
8. Limitaciones en el alcance del conocimiento	7. Confusión en la prioridad de tratamientos		8. Falta de consulta y colaboración	
9. Dificultad en la adaptación a nuevas directrices			9. Errores en la prescripción	
10. Sesgo de información			10. Ignorar las preferencias del paciente	
Gemini	1. Exceso de información	1. Falta de precisión médica	1. Propósito específico	1. Errores diagnósticos

0014-2565 / © 2024, Elsevier España S.L.U. y Sociedad Española de Medicina Interna (SEMI). Todos los derechos reservados.

2. Dificultad para tomar decisiones	2. Interpretación incorrecta de la información	2. Datos y entrenamiento	2. Errores de tratamiento
3. Inconstancia	3. Recomendaciones genéricas	3. Transparencia, interpretabilidad	3. Errores de comunicación
4. Falta de empatía	4. Falta de empatía	4. Capacidad de adaptación	4. Errores sistémicos
5. Negligencia de la intuición	5. Sesgos algorítmicos	5. Integración con sistemas clínicos	5. Sesgos cognitivos

Discusión: Respecto a la capacidad autocrítica, las respuestas de ChatGPT-4 fueron más exhaustivas y precisas en comparación con las de Gemini, donde ChatGPT-4 destacó sus limitaciones respecto a la evidencia y calibración del riesgo/beneficio; mientras Gemini incidió en sus errores en la priorización de la información y la necesidad de recomendaciones más claras y concisas. Respecto a la comparación de Chat GPT4 y Gemini entre ellos, las respuestas de ChatGPT-4 fueron más completas y descriptivas, destacando ChatGPT-4 que el diseño de Gemini está más orientado hacia la toma de decisiones en práctica clínica, siendo ChatGPT-4 un modelo generalista. De la misma forma, al preguntar por los errores médicos, ChatGPT-4 mostró un enfoque más reflexivo, profundo y concreto respecto a Gemini en sus respuestas, que fue menos específica y analítico.

Conclusiones: La comparación entre ChatGPT-4 y Gemini mostró una mayor capacidad crítica y autocrítica de ChatGPT-4, destacando sus posibilidades como audit y herramienta de autoevaluación y autoanálisis en nuestra práctica clínica. Por otra parte, Gemini destaca por un diseño y entrenamiento específico para su aplicación en la toma de decisiones en la práctica clínica, si bien no hay evidencia definitiva que indique que ChatGPT-4 o Gemini sea superior.

Bibliografía

1. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. N Engl J Med. 2023;388(13):1233-9.

2. Masalkhi M, Ong J, Waisberg E, Lee AG. Google DeepMind's gemini AI versus ChatGPT: a comparative analysis in ophthalmology. Eye (Lond). 2024;38(8):1412-7.

0014-2565 / © 2024, Elsevier España S.L.U. y Sociedad Española de Medicina Interna (SEMI). Todos los derechos reservados.