

La alucinación de la inteligencia artificial

Johann Radax

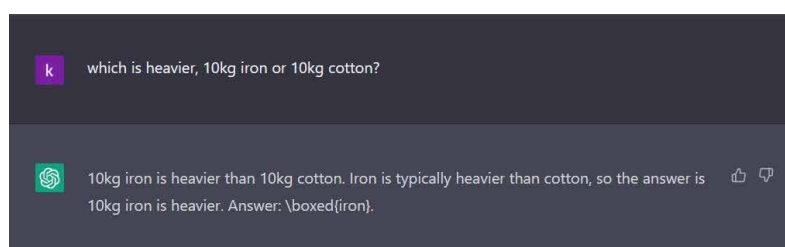
Dr. Med. Vet., MBioeth, MD

Investigador independiente

En los últimos años, la inteligencia artificial (IA) ha emergido como una herramienta prometedora en todos los campos de la ciencia, incluyendo el de la odontología. Ofrece avances significativos en áreas como el diagnóstico por imágenes, la planificación de tratamientos y la personalización de la atención al paciente. Sin embargo, junto con estas oportunidades, surgen desafíos que no podemos ignorar. Uno de los más preocupantes es el fenómeno conocido como “alucinación de la IA”, un término que describe la tendencia de los sistemas de IA a generar información incorrecta o engañosa con una apariencia de veracidad. Este problema tiene implicaciones profundas tanto para la investigación como para la docencia en odontología.

Las alucinaciones absurdas, incluso chistosas, de la IA en sus fases iniciales han sido superadas en gran parte, como en el ejemplo de la Figura 1.

Figura 1. Captura de pantalla de Chat GPT versión anterior



Ahora, GPT-4o-Mini responde a la pregunta “qué es más pesado: 10 kg de hierro o 10 kg de algodón”:

“Ambos pesan lo mismo: 10 kg. La diferencia está en el volumen y la densidad. El hierro es mucho más denso que el algodón, por lo que 10 kg de hierro ocupará mucho menos espacio que 10 kg de algodón”.

Pero las alucinaciones siguen existiendo de una forma más sutil y, por lo tanto, más peligrosa y más difícil de detectar.

¿Qué es la alucinación de la IA?

La alucinación de la IA ocurre cuando un sistema de inteligencia artificial, como un modelo de lenguaje avanzado o una red neuronal, produce resultados que parecen coherentes y convincentes, pero que en realidad carecen de fundamento o son abiertamente falsos. Esto



e-ISSN: 2960-8325

ISSN: 1390-0889

se debe a que estos sistemas no “comprenden” la información en el sentido humano, sino que generan respuestas basadas en patrones estadísticos derivados de grandes volúmenes de datos. En el contexto de la odontología, esto podría manifestarse en diagnósticos erróneos, recomendaciones de tratamientos no validados o interpretaciones incorrectas de literatura científica.

Ejemplos prácticos

En mayo de 2023, un abogado estadounidense utilizó ChatGPT para elaborar una moción legal que contenía citas judiciales falsas. Aparte de perder el caso, el juez consideró que el abogado actuó de mala fe y realizó “actos de evasión consciente y declaraciones falsas y engañosas al tribunal”.

En un estudio del año 2024, se evaluó el rendimiento de ChatGPT y Bard a la hora de replicar los resultados de revisiones sistemáticas realizadas por humanos. Se compararon los resultados producidos por la IA con referencias de revisiones sistemáticas originales. Los artículos se consideraron “alucinados” si dos de los siguientes datos eran erróneos: título, primer autor o año de publicación. Las tasas de alucinaciones fueron el 39,6 % para GPT-3,5; el 28,6 % para GPT-4; y el 91,4 % para Bard: esta IA proporcionó múltiples versiones de artículos alucinados con títulos y nombres de revistas parecidos.

Este año (2025), la revista Nature publicó los siguientes datos para IA selectas y sus tasas de alucinación entre paréntesis: Falcon 7B-instruct* (29,9 %), Google Gemma 1.1-2B-it (27,8 %), Qwen2.5-0.5B-Instruct (25,2 %), Anthropic Claude 2 (17,4 %), Google Gemini 1.5 Pro (9,1 %), OpenAI GPT-4 (1,8 %), OpenAI o1-mini (1,4 %), Zhipu AI GLM-4-9B-Chat (1,3 %), Google Gemini 2.0 Flash Expt (1,3 %).

Impacto en la investigación

En el ámbito de la investigación, la alucinación de la IA plantea un problema serio. Los investigadores confían en datos precisos y verificables para desarrollar nuevos tratamientos, mejorar técnicas quirúrgicas y entender las causas de las enfermedades bucodentales. Si un sistema de IA genera datos falsos o interpretaciones erróneas, esto podría llevar a conclusiones equivocadas, retrasar avances científicos e incluso poner en riesgo

la salud de los pacientes. Por ejemplo, un modelo de IA podría “inventar” correlaciones entre ciertos factores de riesgo y enfermedades periodontales que no existen en la realidad, lo que podría desviar recursos y esfuerzos hacia líneas de investigación infructuosas.

Además, la proliferación de artículos científicos generados total o parcialmente por IA, pero que contienen errores o información no verificada, podría saturar la literatura académica con contenido poco confiable. Esto dificultaría la labor de los investigadores, quienes tendrían que dedicar más tiempo a discernir entre lo válido y lo falso, en lugar de avanzar en sus propias investigaciones.

Desafíos en la docencia

En el ámbito educativo, la alucinación de la IA también representa un desafío significativo. Los estudiantes dependen de información precisa y actualizada para formarse como profesionales competentes. Si los sistemas de IA utilizados como herramientas de enseñanza proporcionan información incorrecta, los estudiantes podrían internalizar conceptos erróneos que afecten su práctica clínica en el futuro. Por ejemplo, un sistema de IA podría sugerir un protocolo de tratamiento basado en datos obsoletos o no validados, lo que podría llevar a errores en la atención al paciente.

Además, la facilidad con la que los estudiantes pueden acceder a herramientas de IA para generar resúmenes, ensayos o incluso respuestas a exámenes plantea un dilema ético y pedagógico. Si bien estas herramientas pueden ser útiles para agilizar ciertas tareas, también pueden fomentar la dependencia y la falta de pensamiento crítico. Los educadores deben encontrar un equilibrio entre aprovechar las ventajas de la IA y asegurarse de que los estudiantes desarrollen las habilidades necesarias para evaluar y cuestionar la información de manera independiente.

Hacia un uso responsable de la IA

Para mitigar los riesgos asociados con la alucinación de la IA, es fundamental adoptar un enfoque crítico y responsable en su uso. En primer lugar, los investigadores y educadores deben ser conscientes de las limitaciones de estas herramientas y verificar

siempre la información generada por IA con fuentes confiables y revisadas por pares. En segundo lugar, es necesario desarrollar estándares y protocolos claros para el uso de la IA en la investigación y la docencia, asegurando que los sistemas utilizados sean transparentes, precisos y estén validados científicamente.

Finalmente, es crucial fomentar una cultura de escepticismo saludable y pensamiento crítico entre los profesionales y estudiantes de odontología. La IA es una herramienta poderosa, pero no debe reemplazar el juicio humano ni la rigurosidad científica. Al abordar estos desafíos de manera proactiva, podemos aprovechar los beneficios de la IA sin comprometer la integridad de nuestra disciplina.