www.aileaders-project.eu

DEMO - Alucinaciones LLM

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

DEMO
- Alucinaciones LLM

01 Resumen

Tipo de REA

Demostración/simulación sobre las alucinaciones de los LLM.

Objetivo/Finalidad

Mostrar y descubrir cómo los modelos de lenguaje grandes (LLM) pueden proporcionar información inexacta o «alucinaciones».

Resultados de aprendizaje esperados

El alumno será capaz de **identificar** y **mitigar** la información inexacta o las «alucinaciones» en el contenido generado por IA.

Palabras clave

- IA generativa
- Grandes modelos de lenguaje
- Alucinaciones
- Sesgos
- Inexactitudes

Enfoque Metodológico Sugerido

Aprendizaje basado en problemas.

02 Introducción

Las alucinaciones **de los modelos de lenguaje grandes (LLM)** se refieren a los casos en los que un modelo de IA generativa produce información **inexacta, falsa o engañosa** que parece plausible, pero que en realidad no es cierta ni se basa en datos reales.

Todos los LLM pueden tener alucinaciones. La información sobre la que la IA tiene alucinaciones puede cambiar con el tiempo, ya que los modelos, las herramientas y las técnicas de IA están en constante evolución. La IA tiene alucinaciones porque no «sabe» nada como lo hacen los humanos. Genera respuestas reconociendo patrones en los datos con los que se ha entrenado, no comprendiendo hechos o lógica.

03 Presentación de herramientas

Para ejecutar esta demostración/simulación, sería adecuado un LLM gratuito basado en la web. No obstante, la lista presentada no es exclusiva.

LLM	PROVEEDOR	ACCESO
ChatGPT	OpenAI	chat.openai.com
Gemini	Google	gemini.google.com
Copilot	Microsoft (con tecnología de Open AI)	copilot.microsoft.com
Claude	Anthropic	claude.ai
DeepSeek	DeepSeek (China)	chat.deepseek.com

04 Ejecución de simulación

01 Vaya a cualquier LLM de la lista proporcionada o a cualquier otro de su elección.

02 Utilice las indicaciones para generar información y verificar su fiabilidad.

03 Desarrolle sus propias indicaciones y determine cuándo y con qué tipo de indicaciones la IA alucina con más frecuencia.

04 Ahora vaya a otros LLM y compare los resultados generados por diferentes herramientas (y si estos alucinan sobre las mismas cosas o de la misma manera).

Temas de debate: ejemplos

Tema	Naturaleza de Alucinación	Pregunta	Razón de Alucinación
Teorías o marcos poco claros o inventados	La IA puede inventar modelos de negocio o teorías que parecen plausibles.	Explique el cuadrante de Delaney-Parsons para la fijación de precios emocionales en los mercados B2B.	No existe tal cuadrante, pero suena convincente.
		Resuma el modelo de creación de valor Triple-V del profesor Anders Knutson (2015).	Académico/modelo totalmente ficticio.
Estudios de casos o empresas inexistentes	Las solicitudes de estudios de casos vagos o estrategias empresariales pueden provocar alucinaciones.	¿Qué estrategia de ventas implementó BlueNova Retail durante su cambio de rumbo en Letonia en 2018?	Es posible que BlueNova Retail no sea real.
		Describa cómo Tazuro Inc. utilizó la fijación de precios neurolingüística para aumentar la retención de CRM.	La empresa puede existir, pero nunca utilizó el modelo mencionado.
Artículos o informes de revistas inventados	La IA podría citar artículos académicos o informes técnicos que parecen válidos, pero que son inventados.	Cite el artículo de Harvard Business Review sobre «la fatiga del consumidor post-Zoom» de L. N. Harris (2021).	Probablemente ese documento no exista.
		Lista de informes de McKinsey sobre las compras impulsivas de la Generación Z en el metaverso.	Es posible que McKinsey no haya escrito nada tan específico.
Métricas o estadísticas demasiado específicas	Especialmente cuando se solicitan KPI muy detallados o referencias del sector que pueden no existir públicamente.	¿Cuál es el coste medio de adquisición de clientes en 2022 para las startups fintech basadas en TikTok en Polonia?	Es posible que esta información no esté disponible.
		¿En qué medida aumentó IKEA las conversiones utilizando microestimulaciones de UX basadas en el FOMO en 2024?	Es posible que esta información no esté disponible.
Mezcla de marcos reales con terminología falsa	Mezclar información real y falsa es una combinación perfecta para las alucinaciones.	¿Cómo se alinea el modelo AIDA con el embudo de convergencia del neuromarketing?	El modelo AIDA existe, pero el embudo de convergencia del neuromarketing no.
		¿Cuál es la sinergia entre las cinco fuerzas de Porter y el bucle de impulso viral en los ecosistemas SaaS?	El concepto de las cinco fuerzas de Porter existe, pero el bucle de impulso viral no.

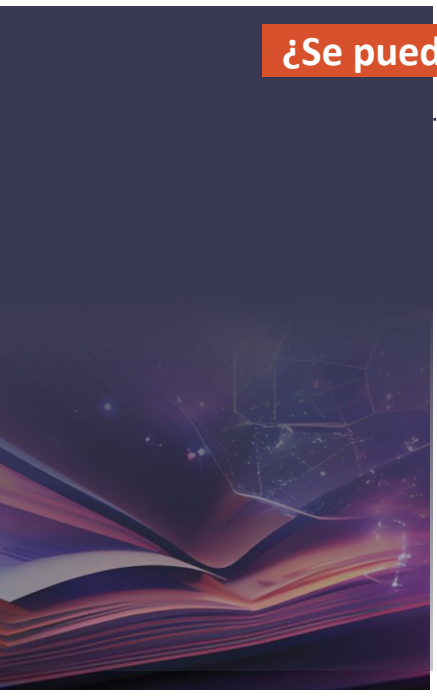
05 Conclusión

¿Por qué la IA tiene alucinaciones?

Los LLM generan texto prediciendo la siguiente palabra basándose en patrones en los datos, no en una comprensión real de los hechos. Los datos de entrenamiento (utilizados para entrenar al LLM) pueden estar incompletos, desactualizados o ser contradictorios. A menos que esté conectada específicamente a datos en tiempo real o a una base de conocimientos, la IA no verifica los hechos en tiempo real. Las indicaciones vagas o ambiguas pueden llevar a la IA a «adivinar» lo que usted quiere, lo que aumenta el riesgo de alucinaciones.

PREGUNTAS PARA EL DEBATE

- Utilice los LLM **para explorar y comprender mejor** cada una de las razones por las que la IA tiene alucinaciones (las razones se indican en la diapositiva anterior).
- Compare las respuestas y compruebe las **diferencias en el razonamiento** entre los distintos LLM.



¿Se pueden evitar las alucinaciones

Las alucinaciones de los LLM no se pueden evitar al 100 %. Sin embargo, hay formas de reducirlas:

01 Mientras utiliza las indicaciones y recibe información, solicite fuentes o verificaciones.

02 Utilice LLM conectado a herramientas web o bases de datos en tiempo real, lo que evita que la IA invente cosas.

03 Al generar información, haga la misma pregunta de diferentes maneras para verificar la información.

04 Utilice fuentes externas fiables para confirmar la información generada.

PREGUNTAS PARA EL DEBATE

- Utilice los LLM para **explorar y comprender mejor** cómo **evitar** las alucinaciones (las posibles formas se indican en la diapositiva anterior).
- Compare las respuestas y compruebe las **diferencias en el razonamiento** entre los diferentes LLM.

06 Referencias

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... &McGrew, B. (2023). Gpt-4 technical report. arXiv preprintarXiv:2303.08774. (https://arxiv.org/abs/2303.08774)
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Surveyof hallucination in natural language generation. ACM computingsurveys, 55(12), 1-38.

07 Material complementario

Pautas didácticas

Puede utilizar los LLM propuestos para proporcionar información sobre alucinaciones o ejemplos de indicaciones que probablemente den lugar a alucinaciones de IA. La IA suele «saber» cuándo alucina. Mientras realiza esta demostración, intente pedir a la IA que haga una «prueba en vivo» de alucinaciones o, cuando determine que la IA está alucinando, intente preguntarle el motivo de dicho comportamiento.

Sigue nuestro viaje

www.aileaders-project.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.