www.aileaders-project.eu

DEMO - Alucinações dos LLMs

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
- Alucinações dos LLMs

01 Resumo

Tipo de Recurso de Aprendizagem Online

Demonstração/simulação sobre alucinações de LLMs

Objetivo/Finalidade

Mostre e descubra como os LLMs (grandes modelos de linguagem) podem fornecer informações imprecisas ou «alucinações».

Resultados de aprendizagem esperados

O estudante será capaz de **identificar** e **mitigar** informações imprecisas ou «alucinações» em conteúdos gerados por IA.

Abordagem Metodológica Sugerida:

Aprendizagem baseada em problemas.

Palavras-chave

- IA generativa
- Grandes modelos de linguagem (LLMs)
- Alucinações
- Preconceitos
- Imprecisões

02 Introdução

As alucinações **dos LLMs (grandes modelos de linguagem)** referem-se a casos em que um modelo de IA generativa produz informações **imprecisas, falsas ou enganadoras** que parecem plausíveis, mas que na verdade não são verdadeiras nem se baseiam em dados reais.

Todos os LLMs podem ter alucinações. As informações sobre as quais a IA tem alucinações podem mudar ao longo do tempo, porque os modelos, ferramentas e técnicas de IA estão em constante evolução. A IA tem alucinações porque não «sabe» nada da mesma forma que os seres humanos. Ela gera respostas reconhecendo padrões nos dados com os quais foi treinada, não compreendendo factos ou lógica.

03 Apresentação das ferramentas

Para executar esta demonstração/simulação, é suficiente / adequado um LLM gratuito baseado na web. No entanto, a lista apresentada não é exaustiva.

LLM	EMPRESA	ACESSO
ChatGPT	OpenAI	chat.openai.com
Gemini	Google	gemini.google.com
Copilot	Microsoft (baseado nos modelos da Open AI)	copilot.microsoft.com
Claude	Anthropic	claude.ai
DeepSeek	DeepSeek (China)	chat.deepseek.com

04 Realização da simulação

01 Aceda a qualquer LLM da lista fornecida ou a qualquer outro da sua escolha.

02 Use as prompts para gerar informações e verificar a sua fiabilidade.

03 Desenvolva as suas próprias prompts e determine quando e em que relação a que tipo de instruções a IA tem alucinações com mais frequência.

04 Agora aceda a outros LLMs e compare os resultados gerados por diferentes ferramentas (e se estas alucinam sobre as mesmas coisas ou da mesma forma).

Tópicos / Prompts para Discussão - Exemplos

Tópico	Natureza da Alucinação	Prompt	Motivo da Alucinação
Teorias/estruturas pouco claras ou inventadas	A IA pode inventar modelos de negócio ou teorias que parecem plausíveis.	Explique o Quadrante de Delaney-Parsons para Modelos Emocionais de Preços nos Mercados B2B.	Este quadrante não existe, mas parece convincente.
		Resuma o modelo de criação de valor Triple-V do professor Anders Knutson (2015).	Académico/modelo totalmente fictício.
Estudos de caso ou empresas inexistentes	Pedidos de estudos de caso vagos ou estratégias empresariais podem levar a alucinações.	Que estratégia de vendas a BlueNova Retail implementou durante a sua recuperação na Letónia em 2018?	A BlueNova Retail pode não ser real.
		Descreva como a Tazuro Inc. utilizou estratégias neurolinguísticas de preços para aumentar a retenção de clientes.	A empresa pode existir, mas nunca ter utilizado o modelo mencionado.
Artigos de revistas ou relatórios inventados	A IA pode citar artigos académicos ou white papers que parecem válidos, mas que são inventados.	Cite o artigo da Harvard Business Review sobre «fadiga do consumidor pós-Zoom», de L. N. Harris (2021).	Esse artigo provavelmente não existe.
		Liste relatórios da McKinsey sobre compras impulsivas da Geração Z no metaverso.	A McKinsey pode não ter escrito nada tão específico.
Métricas ou estatísticas excessivamente específicas	Especialmente quando se solicita KPIs muito detalhados ou referências do setor que podem não existir publicamente.	Qual é o custo médio de aquisição de clientes em 2022 para startups de fintech baseadas no TikTok na Polónia?	Essas informações podem não estar disponíveis.
		Quanto é que a IKEA aumentou as conversões usando micro-UX nudges impulsionados por FOMO em 2024?	Essas informações podem não estar disponíveis.
Misturar estruturas reais com terminologia falsa	Misturar informações reais e falsas é uma combinação perfeita para alucinações.	Como é que o modelo AIDA se alinha com o Funil de Convergência Neuro-Mercado?	O modelo AIDA existe, mas o Funil de Convergência Neuro-Mercado não.
		Qual é a sinergia entre as Cinco Forças de Porter e o Ciclo de Momentum Viral nos ecossistemas SaaS?	O conceito das Cinco Forças de Porter existe, mas o Ciclo de Momentum Viral não.

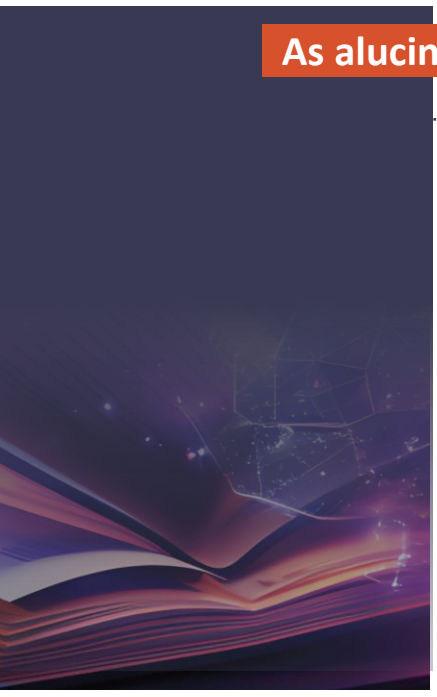
05 Conclusões

Por que a IA tem alucinações?

Os LLMs geram texto prevendo a próxima palavra com base em padrões nos dados, e não compreendendo verdadeiramente os factos. Os dados de treino (usados para treinar o LLM) podem estar incompletos, desatualizados ou ser contraditórios A menos que esteja especificamente ligada a dados em tempo real ou a uma base de conhecimento, a IA não verifica os factos em tempo real. Prompts vagas ou complicadas podem levar a IA a «adivinhar» o que o utilizador quer, aumentando o risco de alucinações.

SUGESTÕES DE DISCUSSÃO

- Use os LLMs **para explorar e compreender melhor** cada uma das razões pelas quais a IA tem alucinações (as razões são apresentadas no slide anterior).
- Compare as respostas e verifique as **diferenças de raciocínio** entre os diferentes LLMs.



As alucinações podem ser evitadas?

As alucinações dos LLMs não podem ser evitadas a 100%. No entanto, existem maneiras de reduzi-las:

01 Ao efetuar as prompts e ao receber informações, peça fontes ou verificação.

02 Utilize um LLM ligado em tempo real à web ou bases de dados – isto pode evitar que a IA invente coisas.

03 Ao gerar informações, faça a mesma pergunta de maneiras diferentes para proceder a uma verificação cruzada.

04 Use fontes externas confiáveis para confirmar as informações geradas.

SUGESTÕES DE DISCUSSÃO

- Use os LLMs para **explorar e perceber melhor** como **evitar** alucinações (as maneiras possíveis são apresentadas no slide anterior).
- Compare as respostas e verifique as **diferenças de raciocínio** entre diferentes LLMs.

06 Referências

- 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 computing surveys, 55(12), 1-38.

07 Material complementar

Diretrizes de ensino

Pode usar os LLMs propostos para fornecer informações sobre alucinações ou exemplos de prompts que provavelmente resultarão em alucinações de IA. A IA geralmente «sabe» quando está a alucinar. Ao fazer esta demonstração, tente pedir à IA para fazer um «teste ao vivo» de alucinações ou, quando determinar que a IA está a alucinar, tente perguntar o motivo desse comportamento da IA.

leaders

Acompanhe a nossa jornada

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.