Caution About AI Hallucinations

Are you Trippin?

Artificial Intelligence can hallucinate!

What are hallucinations?

Just like we can see patterns in clouds or images where they do not exist, AI will fill in gaps with sometimes plausible information and sometimes complete nonsense ("What are AI Hallucination?"). You need to double-check everything AI generates for misleading or inaccurate statements.

Some Examples

- AI created completely false legal briefs - up to 69% of the time with GPT3.5 (Shapero).
- "Google's Bard chatbot incorrectly claiming that the James Webb Space Telescope had captured the world's first images of a planet outside our solar system.
- Microsoft's chat AI, Sydney, admitting to falling in love with users and spying on Bing employees.
- Meta pulling its Galactica LLM demo in 2022, after it provided users inaccurate information, sometimes rooted in prejudice." ("What are AI Hallucinations).

More general issues...

- Incorrect predictions: AI will predict something that is unlikely to happen
- False positives: Like saying AI wrote a paper you personally wrote
- False Negatives: Failing to identify some deadly medical issue
- Provide incorrect dates
- Use fictious sources
- Reference "famous" resources that were not used in the answer (e.g. books no normal person owns)
- Incorrectly profiling individuals as potential criminals
- Express racist, sexist or otherwise prejudiced responses
- Again, AI can see patterns that do not exist and assert things completely made up

Your Responsibility

Double check the data - AI makes up things and can be worse on topics for which there are no precise answers like ethics, politics, philosophy, and religion (see Shevlin).

Double check all listed resources - You cannot trust AI to provide accurate information or list credible resources used to create the response.

Do your own research in credible resources and do your own writing!

References

Shapero, Julia, "AI models frequently 'hallucinate' on legal queries, study finds." The Hill 1 January 2024 https://thehill.com/policy/technology/4403776-ai-models-frequently-hallucinate-onlegal-queries-study-finds/

Shevlin, Henry "What are 'hallucinations" and what more can we expect from AI?" Cambridge Dictionary 14 November 2023 https://www.youtube.com/watch?v=bOfCgWc-JKQ