What is an LLM hallucination, and how do you reduce it?

1 November 2026 · 5 min read · blogs.iictp.org/what-is-llm-hallucination

An LLM hallucination is an answer that sounds confident and fluent but is wrong or invented, such as a citation that does not exist. It happens because a language model generates likely text, not verified facts, so it can produce plausible statements it has no basis for.

TL;DR The short version

- A hallucination is a confident, fluent answer that is wrong or invented, because a model generates likely text, not verified facts.
- Grounding the model in retrieved source text with citations, and allowing 'I don't know', reduces it.
- It cannot be eliminated today, so verify claims against sources, run generated code and tests, and sample answers for human review.

Read online, listen to it, or change the text size: https://blogs.iictp.org/what-is-llm-hallucination

Why do language models hallucinate?

A model predicts the next piece of text from patterns in its training data. When it lacks the facts, it can still produce a fluent answer, and nothing inside it checks that answer against a source. Gaps in its knowledge and ambiguous questions make this more likely.

What do hallucinations look like?

Typical cases are invented references, made-up functions or settings in code, wrong numbers presented precisely, and details of a document that was never provided. They are hard to spot because the tone is the same as for correct answers.

How does retrieval help?

Giving the model relevant source text and asking it to answer from that text, with citations, grounds the answer in something checkable. It reduces hallucination but does not remove it, because the model can still misread or ignore the source. See RAG explained.

What prompt and design choices help?

Ask for answers only from the supplied material, allow “I don’t know” as a valid reply, and request quotations that can be matched against the source. Lower sampling randomness for factual tasks, and break large tasks into smaller checked steps.

How should output be checked?

Verify claims against the source, run generated code and its tests, and sample answers for human review. For high-stakes uses, such as legal, medical or financial content, a person should confirm the result before anyone relies on it.

Can hallucination be eliminated?

Not today. It can be reduced and made detectable, so the aim is to design the system so that a wrong answer is caught before it causes harm. Measure the rate on your own questions, and keep measuring after each change.

> Related credential: AI Practitioner. Explore the exam facts and syllabus.

Sources and further reading

1. Ji et al. Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 2023.
2. NIST. AI Risk Management Framework. nist.gov/itl/ai-risk-management-framework.

Published by IICTP, an independent technical credential body, on 1 November 2026. The current version of this post is at https://blogs.iictp.org/what-is-llm-hallucination. More guides: blogs.iictp.org.