AI DATA

How Do You Stop LLM Hallucinations? An Enterprise Guide to Data-Driven Accuracy

Chatbots make things up because they predict likely words, not true ones. Grounding, guardrails, and human review turn a confident guesser into a system you can rely on.

PREPARED BY

Lifewood Data Technology

INSIGHTS

PUBLISHED

July 2026 · 11 min read

READ ONLINE

lifewood.com/blogs/reducing-llm-hallucinations

In November 2022, Jake Moffatt booked a last-minute flight after his grandmother died. He asked the airline's chatbot about bereavement discounts. It told him he could claim the reduced fare within 90 days of travelling. That was false: the airline had no such policy, refused the refund, and argued it was not responsible for its own chatbot. A tribunal disagreed, and Air Canada was ordered to pay.

The chatbot had not been hacked. It had simply produced a confident, well-written, completely invented answer.

What is an AI hallucination, and why do chatbots make things up?

An LLM, or Large Language Model, is the engine behind ChatGPT and most AI assistants. Picture an extraordinarily well-read autocomplete that has read enormous amounts of text and is very good at predicting the next few words. Chain those predictions together and you get fluent answers, summaries, and emails.

Here is the catch: the model produces the most likely-sounding continuation, not the most truthful one. When it does not know, it does not pause and say "I'm not sure." It fills the gap with something plausible that has never been true.

IBM defines a hallucination as an AI confidently presenting incorrect or invented information as fact. OpenAI's researchers add a memorable reason: it is like a student in an exam that never rewards "I don't know." The smart move is to guess, so the model learns to guess.

Why do hallucinations matter for business?

For a casual user, a wrong answer is an annoyance, but for an enterprise it is exposure. The Air Canada case showed a company can be held legally responsible for what its AI tells a customer. Courts have also sanctioned lawyers who filed documents citing cases an AI invented.

The research is blunt about scale. A 2024 Stanford study of over 800,000 legal questions found general-purpose models hallucinated on 69% to 88% of specific legal queries. On questions about a court's core holding, the rate was at least 75%. Even GPT-4 hallucinated 58% of the time.

Purpose-built tools do better, but not by enough. A follow-up Stanford study in the *Journal of Empirical Legal Studies* found leading legal AI products still hallucinate on 17% to 33% of queries, with accuracy ranging from 65% for the best, to 41% and then 19% for the others. These are tools sold specifically for a high-stakes profession.

- **51%** of organizations using AI report at least one negative consequence (McKinsey, 2025).
- **46%** are willing to trust AI, despite 66% using it regularly (KPMG, 2025).
- **70%** believe AI needs stronger regulation and governance (KPMG, 2025).
- **57%** of organizations say their data is not AI-ready (Gartner, 2025).
- **17–33%** hallucination rate of purpose-built legal AI tools (Stanford, 2025).

In McKinsey's 2025 research, 51% of organizations using AI reported at least one negative consequence, and inaccuracy was the risk most often tied to real harm. Trust is fragile too: in a study of 48,000+ people across 47

countries by KPMG and the University of Melbourne, only 46% said they were willing to trust AI. Yet 66% use it regularly, and 70% want stronger regulation.

Why is reducing hallucinations harder than it looks?

You cannot fully delete the behaviour, because guessing is built into how language models work, so no single setting removes it. Even the providers describe hallucination as a stubborn, ongoing challenge. The realistic goal is to keep it rare, catch it, and never let it reach a customer unchecked.

The model rarely knows your business, because a general AI has read the public internet but never your latest pricing, policy, or contract, so it improvises. IBM, Google, and AWS all note that this gap must be closed with your own trusted data.

AI-ready data is also rarer than people think, and according to Gartner, about 57% of organizations believe their data is not AI-ready. Messy data in means bad answers out, so accuracy is a data problem long before it is a model problem.

How do you stop LLM hallucinations?

The most effective approach is not a secret algorithm, but a discipline the largest tech companies now agree on, called grounding. Never let the model answer from memory alone. Hand it the relevant, trusted facts at the moment of the question, and tell it to answer only from those facts.

The common method is Retrieval-Augmented Generation, or RAG. Before the AI answers, the system retrieves the right documents from your knowledge base, augments the question with them, and only then asks the model to generate a reply. AWS calls RAG a pragmatic way to give an enterprise model accurate context. Google frames it as supplying the facts and grounding the answer on them. NVIDIA notes it reduces the chance of a plausible-but-wrong answer. IBM is candid that RAG lowers the risk of hallucination without making a model error-proof.

A grounded enterprise system has a recognisable shape. A knowledge base holds your trusted documents. An orchestrator fetches facts, then combines question and facts. The language model writes a draft answer. A guardrail checks whether the draft is backed by the facts. A grounded answer is delivered with a source you can verify. A failed check is fixed, flagged, or routed to a human.

The question never goes straight to the model. First the system pulls the right facts from your documents, then hands them to the model with the question, and finally checks the draft against those facts before anyone sees it. Grounding is the foundation, and the strongest systems add two more layers. Microsoft's Azure groundedness detection flags any part of an answer not supported by the sources, and can rewrite it before the user sees it. NVIDIA's open-source NeMo Guardrails adds fact-checking rails to RAG systems. The second layer is the most reliable safeguard of all: a human in the loop for high-stakes answers.

What belongs in an enterprise accuracy playbook?

- **Ground every answer.** Connect the AI to trusted, current data with RAG so it answers from facts, not memory.

- **Show the source and add guardrails.** Make the AI cite where each answer came from, and use automated checks to flag or rewrite unsupported claims.
- **Keep a human in the loop.** For money, legal, medical, or customer-facing decisions, a person reviews before it ships.
- **Start with the data.** Clean, well-labeled, AI-ready data is the input that makes everything above work.

How does Lifewood turn data-driven accuracy into reality?

Trustworthy AI is built on trustworthy data, and that is where Lifewood works. Founded in 2004 and refocused as an AI-data specialist, the company now operates across 30+ countries and 40+ delivery centers. It combines a worldwide human workforce with an industrialized methodology and its proprietary LiFT platform.

AIGC means AI-Generated Content: the text, images, audio, and video that generative models produce. Lifewood supports AIGC at both ends, preparing the high-quality data that makes generated content accurate, and applying full-time human-in-the-loop quality control so the output holds up.

- **Data collection.** Multilingual, multi-modal gathering across text, audio, image, and video in 50+ languages.
- **Annotation and labeling.** Labeling, tagging, transcription, and sentiment analysis: the structured truth a model learns from.
- **LLM training data.** Supervised fine-tuning sets, human-preference (RLHF) data, and model-evaluation datasets.
- **AIGC and QA.** Enterprise AI-generated content, including video at scale, backed by full-time human review and validation.

Hallucinations are usually a data-quality problem, not only a model problem. Lifewood's specialists attack the problem at its source: structuring and validating the documents a RAG system retrieves from, which addresses the 57% of organizations whose data is not AI-ready; building industry- and language-specific datasets plus evaluation sets that teach models to ground answers and admit uncertainty; and providing a global workforce as the verification layer that catches errors automated checks miss, in 50+ languages.

The pipeline runs from collection, to annotation, to training and evaluation, to human QA. Those stages feed a grounded knowledge base, which supports an assistant answering from verified facts. When the AI does slip, errors flow back into the data process, so the next version is more accurate.

What comes next for enterprise AI accuracy?

Guardian agents: AI that checks AI. A fast-emerging idea is AI that supervises other AI. Guardian agents review, monitor, and can block another model's risky outputs.

AI-ready data becomes the priority. Attention is shifting from flashy models to the unglamorous foundation: clean, well-governed data. Given that 57% of organizations say their data is not AI-ready, the organizations that invest there will quietly pull ahead.

What is the bottom line?

Hallucinations are not a sign that AI is broken, but a predictable and manageable feature of how language models work. The strategy is broadly agreed by McKinsey, Gartner, OpenAI, Google, AWS, IBM, NVIDIA, and Microsoft alike. Ground the model in trusted data. Show the source. Add guardrails. Keep a human in the loop. Govern the whole thing.

Underneath every step sits the same requirement: high-quality, well-prepared, AI-ready data. With 57% of organizations short of it, and 51% already reporting a negative AI consequence, that is where the work begins. The breakthrough is not a cleverer model, but better data, handled with care. You cannot prompt your way out of a data problem — accuracy is built upstream, in the data, long before the model ever speaks.

Sources

- Forbes / Civil Resolution Tribunal of B.C. *Moffatt v. Air Canada*, 2024 BCCRT 149.
- IBM. *What is Retrieval-Augmented Generation (RAG)?*
- OpenAI. *Why Language Models Hallucinate.* 2025.
- McKinsey & Company. *The State of AI*, 2025 global survey.
- KPMG & University of Melbourne. *Trust, Attitudes and Use of AI: A Global Study 2025.*
- Harvard Business Review. *What Are Your Company's AI Nightmares?* 2026.
- Dahl, M. et al. *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models.* Journal of Legal Analysis 16, 64 (2024).
- Magesh, V. et al. *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools.* Journal of Empirical Legal Studies 22, 216–242 (2025).
- Google Cloud, AWS, NVIDIA, Microsoft Azure AI Content Safety, and GitHub / NVIDIA NeMo Guardrails product documentation.
- Gartner. *Hype Cycle for Artificial Intelligence*, 2025.

Editorial note: this article explains how widely used techniques reduce AI errors. No method removes them entirely, and results depend on each organization's data and use case. Statistics are drawn from the publicly available sources listed above and attributed in the text.