

AI hallucinations—where models confidently generate false or fabricated information—have become a serious concern in enterprise decision-making. As companies like **Garrett Motion**, **PMI**, and **Pandora** experiment with Large Language Models (LLMs) and retrieval-augmented generation (RAG), the question isn't just about accuracy. It's about trust, explainability, and aligning AI outputs with strict business logic.

Trust: The Real Bottleneck for Enterprise AI Adoption

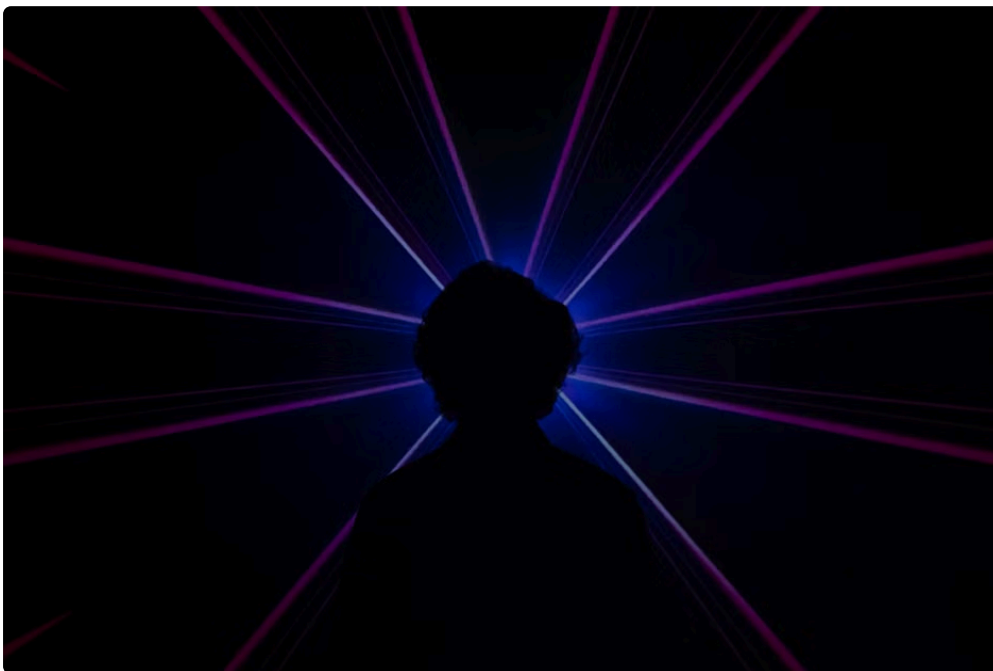
Many enterprises invest heavily in AI to boost efficiency and improve decision quality. But despite achieving higher accuracy in controlled tests, adoption often stalls. Why? Because trust is not automatically granted when a model performs well on metrics alone.

- **Lack of explainability:** Decision-makers don't just want answers; they want to understand why the AI chose a particular course of action.
- **Inconsistent outputs:** Random output variance from identical prompts shakes user confidence.
- **Unaligned business logic:** Answers that contradict established company rules create friction.

For companies like Garrett Motion, which operates in complex engineering domains, blind trust in AI is not an option. Similarly, PMI and Pandora require consistent, transparent AI decisions before scaling these technologies.

Understanding AI Hallucinations

AI hallucinations occur when an AI model, often a Large Language Model (LLM), produces answers that are incorrect but presented with unwarranted confidence. These fabricated responses are especially risky in enterprises where flawed decisions can lead to financial loss, regulatory breaches, or reputational damage.



What Causes Hallucinations?

- **Training Data Limitations:** Gaps or biases in datasets can cause models to “fill in” missing info.
- **Model Architecture:** LLMs are designed to predict probable next words, not verify facts.
- **Prompt Ambiguity:** Vague questions can cause models to hallucinate answers.

- **Outdated Knowledge:** Static training data may not reflect current enterprise realities.

These causes combine to produce AI systems that sound authoritative but generate inaccurate output—exposing enterprises to *hallucination statistics* that remain frustratingly high.

Hallucination Statistics: What Do The Numbers Say?

Study / Source Model Type Hallucination Rate Context Wang et al., 2023 LLM (GPT-family) 15-25% Factual question answering Enterprise Pilot by PMI RAG-enhanced LLM Under 10% Domain-specific documentation retrieval Internal Garrett Motion Study Custom LLM + RAG ~12% Engineering decision support Analysis by Pandora AI Team Baseline LLM ~20% Customer service chatbot

These numbers are not trivial. Even a 10% hallucination rate can cause significant business risk if the hallucinations trigger crucial decisions without human oversight.

Retrieval-Augmented Generation (RAG) to the Rescue?

RAG integrates external data retrieval systems with LLMs, attempting to ground responses in factual context. PMI's pilot project using RAG-enhanced LLMs reduced hallucinations significantly by retrieving verified documents before generating answers.

However, while RAG improves **accuracy**, [o9solutions](#) it does not automatically fix issues of trust or explainability.

- **Transparency:** The retrieval process must be auditable—what documents were used and why?
- **Consistency:** Does the system repeat the same answers upon repeated queries?
- **Business-Logic Alignment:** Do retrieved facts comply with enterprise rules?

Garrett Motion's engineering teams found that even when retrieval was sound, decision support recommendations sometimes conflicted with internal policies, causing hesitation.

Why Accuracy Gains Do Not Automatically Create Trust

Improving accuracy numbers is necessary but insufficient for enterprise adoption:

1. **Explainability:** Executives want clear reasoning steps, not just a score or confidence metric.
2. **Repeatability:** Random variance in output example after example breeds doubt.
3. **Alignment with Business Logic:** AI outputs must respect company-specific rules and regulations. Pandora's customer service chatbot learned that even subtle contradictions with company policy can lead to losses in trust.

Trust develops when AI decisions are predictable, interpretable, and consistently aligned with business goals.

Hallucinations and the Risk of Fabricated Certainty

Far worse than providing wrong answers is providing wrong answers with undue confidence—fabricated certainty. This "black box excuse" of "the model decided" masks the risk and makes audit and oversight nearly impossible.



Here are the major risks:

- **Regulatory Compliance:** Enterprises like PMI face strict audits. Stochastic outputs that cannot be traced back to verifiable data pose compliance risks.
- **Reputational Damage:** Garrett Motion cannot afford to have AI-driven decisions in product safety be questioned publicly.
- **Financial Impact:** Erroneous decisions in Pandora’s billing or licensing models could cause revenue loss.

Without transparent decision pathways—linking each output back to trusted data and business rules—the hallucination risk remains a showstopper.

What Would I Show in an Audit?

Auditability is crucial. The enterprise AI systems we trust must be able to demonstrate:

- **Data Lineage:** Which documents, rules, or records influenced this decision?
- **Decision Traceability:** The step-by-step logic flow and intermediate outputs.
- **Consistency Logs:** Historical records of when the AI provided different answers to the same query.

If you see black-box explanations like “the model decided,” add that to the running list of black box excuses—and demand real evidence.

Conclusion: Hallucinations Are Bad, But Trust Is Worse

AI hallucinations present a non-trivial risk to enterprises adopting LLM-based decision tools. While retrieval-augmented generation (RAG) can improve accuracy numbers, this does not guarantee trust. Companies like Garrett Motion, PMI, and Pandora are pioneering methods to embed explainability, consistency, and business logic alignment into their AI deployments.

Fixing hallucinations means more than just fewer errors. It requires transparent AI decisions that are auditable, repeatable, and aligned with enterprise risk tolerance.

Enterprises must evaluate AI not just on accuracy or hallucination statistics, but on their ability to build—and prove—trust.