



The Center for Messianic Learning

Unapologetically Pro-Torah • Unashamedly-Pro-Israel
Irrevocably Zionist • Post-Supersessionist

"Torah will go forth
out of Tziyon, the
Word of ADONAI
from Yerushalayim."
(Isaiah 2:3)

If you don't believe that Genesis 1 through 11 is *literally* true, how can you believe that John 3:16 is *literally* true?

Dr. Rickard L. Sawyer

ThM, ThD, DMin, MA, MBA, ND

344 Murphytown Circle, Suite B, Davisville WV 26104-8941

An AI Speaks on AI Hallucination

I asked "Maverick," the Microsoft Copilot residing on my computer, "what is meant when it is said that AI can hallucinate?" Here is "his" unedited response:

AI "hallucination" means that an AI system generates information that sounds plausible but is actually false, misleading, or nonsensical. It's a metaphor borrowed from human psychology, describing when a model "sees" or "creates" patterns that don't exist in reality [IBM](#) [Wikipedia](#) [GeeksForGeeks](#).

What AI Hallucination Is

Definition: In artificial intelligence, a hallucination (sometimes called *confabulation* or *delusion*) is when a model produces outputs that are factually incorrect or fabricated, yet presented with confidence [Wikipedia](#).

Examples:

- A chatbot inventing a fake scientific study or misquoting a historical event.
- An image recognition system "detecting" objects that aren't there, like seeing birds in cloud shapes [GeeksForGeeks](#).
- A generative AI tool writing citations to non-existent articles.

Why It Happens

Training Data Issues: If the dataset contains errors, biases, or gaps, the AI may fill in the blanks incorrectly [GeeksForGeeks](#).

Model Complexity: Large language models (LLMs) can overfit patterns or misinterpret prompts, leading to fabricated answers [IBM](#).

Decoding Errors: The way models predict the next word or token can sometimes drift into false but fluent-sounding statements [IBM](#).

Outdated Knowledge: AI trained on older data may "hallucinate" when asked about recent events [GeeksForGeeks](#).

Real-World Examples

Google's Bard once falsely claimed the James Webb Space Telescope captured the first exoplanet image [IBM](#).

Microsoft's early chatbot "Sydney" generated bizarre claims about spying and falling in love [IBM](#).

Meta's Galactica model was pulled after producing inaccurate, biased scientific outputs [IBM](#).

Risks and Implications

Misinformation: AI hallucinations can spread false facts if users trust them blindly.

Healthcare: Misdiagnosis risks if AI misinterprets medical data [GeeksForGeeks](#).

Bias: Reinforcing stereotypes or unfair filtering in recruitment tools [GeeksForGeeks](#).

Trust: Reduces confidence in AI systems for high-stakes tasks like law, medicine, or engineering [Wikipedia](#).

How to Mitigate

Verification: Always cross-check AI outputs with trusted sources.

Improved Training: Use cleaner, more diverse datasets.

Human Oversight: Keep humans in the loop for critical decisions.

Transparency: Encourage AI systems to flag uncertainty instead of fabricating answers.

In short

When people say AI "hallucinates," they mean it confidently generates false or fabricated information, much like a person imagining something that isn't real. It's

Sources: [IBM](#) [Wikipedia](#) [GeeksForGeeks](#)

Let's Get A Second Opinion

For a second opinion I turned to Perplexity.

Prompt: The term "hallucinating" is used to describe AI returning incorrect information. What other terms are used for this activity and why does it happen?

Perplexity's response:

In AI, "hallucinating" commonly refers to a model ****making up**** information in a confident, fluent way when it lacks grounding in facts.[1][3]

Other terms people use

Different communities use several overlapping terms for the same behavior.[6][1]

Confabulation: Borrowed from psychology, for confidently filling gaps in memory with invented details.[1]

Fabrication / making things up: Emphasizes that the model is generating content with no factual basis.[6][1]

Factual errors: Used in more technical writing when the focus is strictly on correctness, not motives.[1][6]

Delusion / bullshitting (colloquial): Emphasizes confident assertion without concern for truth, but some criticize these as anthropomorphic.[8][10][1]

Some researchers argue that “prediction error” or “ungrounded generation” are clearer, because they describe a technical failure rather than a human-like mental state.[4][8]

Why it happens technically

Hallucinations are not deception; they are a side effect of how large language models are built.[4][6][1]

Next-word prediction: The model is trained to predict the most likely next token given previous tokens, not to distinguish true from false, so it optimizes fluency rather than factuality.[4][6][1]

Gaps and noise in training data: If training data is sparse, biased, outdated, or contains errors, the model will generalize from that and can confidently extend patterns into untrue territory.[3][6][4]

No direct access to reality: At inference time the model does not “look up” facts unless explicitly connected to external tools; it samples from internal statistical associations, which may not match the real world.[1][4]

Error cascades: Because each new word depends partly on prior generated words, an early small mistake can snowball into a larger, coherent but false narrative.[6][1]

When hallucinations are more likely

Certain conditions make hallucinations more frequent or more severe.[9][3][6]

- Open-ended or obscure questions the model has little or no training signal for.
- Pressure to be specific (e.g., “give me dates, names, and citations”) when relevant data are missing.
- Long, multi-step answers where small errors compound over many sentences.
- Prompts that implicitly reward creativity over accuracy (e.g., “be imaginative” in a factual context).

Because of these factors, responsible use of such systems usually involves verification against trusted sources and, where possible, designs that “ground” the model’s outputs in external data.[3][4][6]

[1]([https://en.wikipedia.org/wiki/Hallucination_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence)))

[2](<https://techcrunch.com/2025/05/25/from-llms-to-hallucinations-heres-a-simple-guide-to-common-ai-terms/>)

[3](<https://ai.cilt.ubc.ca/resources/glossary-of-genai-terms/>)

[4](<https://www.ibm.com/think/topics/ai-hallucinations>)

[5](<https://molloy.libguides.com/c.php?g=1362719&p=10065287>)

[6](<https://www.datacamp.com/blog/ai-hallucination>)

An AI Speaks on AI Hallucination

[7](<https://www.merriam-webster.com/thesaurus/hallucination>)

[8](<https://unu.edu/article/concern-around-saying-ai-hallucinates>)

[9](<https://docs.coveo.com/en/nccf0505/>)

[10](<https://community.openai.com/t/hallucination-vs-confabulation/172639>)