



The ghost in the machine: Hallucinations

Article 3 of 6

BY ENRIQUE FARFAN, PhD, PE



In the ever-evolving landscape of the geoprofessional industry, artificial intelligence continues to drive innovation, efficiency, and curiosity. This article, focusing on hallucinations in AI, is the third in a six-part series written to educate Geoprofessional Business Association (GBA) members on the implications of AI on the geoprofessions. The article was prepared by GBA Business Technology Committee member, Enrique Farfan, PhD, PE, ENV SP, of HDR.

In his short stories, Isaac Asimov explores the ethical system between humans and robots. For example, in his short history "Liar" (1941), Asimov explores scientists' confoundments with the advanced capabilities of a robot that they have created. The robot, named Herbie, develops the ability to lie. Herbie is motivated to lie to protect the humans who interact with him. As explained at the end of the story, Herbie lies and misleads because of the "First Law of Robotics," (a robot may not injure a human being or, through inaction, allow a human being to come to harm). Herbie did not want to cause any of the scientists a "mental injury." Although Asimov wrote science fiction, his ideas are influential and practical today.

With AI, *hallucinations* is the term that refers to instances in which the Large Language Model (LLM) generates content that seems real but consists of false statements or is outside the context of the text. Concisely, hallucinations are lies.

A few types of hallucinations in AI include:

- **Factual hallucinations** – incorrect information.
- **Contextual hallucinations** – out of context or inappropriate information.
- **Interpretative hallucinations** – detailed responses with fictional, unsupported explanations.
 - **Creative hallucinations** – creative but not factual information.

Hallucinations raise the following questions:

- How does the LLM develop the capacity to lie?
- How can we avoid falling for inaccurate information?

LLM hallucinations are a product of a variety of factors, including:

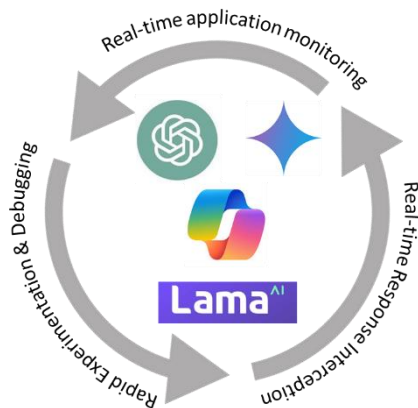
- **Temperature:** LLMs can explore responses that could include less likely or potentially less accurate or relevant responses. Temperature controls how broad an LLM's responses can be and balances creativity and coherence.
- **Outdated Data:** An LLM can be limited to the training data; this limitation could create problems for applications that require up-to-date information, such as the latest engineering design manual or design code, if the latest information is not included in the training dataset.
- **Overfitting:** An LLM can fixate on limited data and be unable to generalize new data.
- **Training Bias:** An LLM can be bound by training data. If training data lacks diversity, an LLM responses will represent those biases, thereby limiting the model from providing generalized data.
- **Compression:** An LLM can use a compressed mathematical representation of input data, which might cause the original information to be reconstructed incorrectly.

AI hallucinations are problematic because they might mislead decision-making processes. Geotechnical engineering relies on factual information to evaluate and design safe and effective structures.

Safeguards and strategies used to limit hallucinations in AI models have been developed. The following techniques can be easily implemented to reduce hallucinations:

1. **Advanced prompting methods** – Ask questions or give instructions so that the AI more clearly “understands” what kind of information or response is needed.
2. **Retrieval-augmented generation (RAG)** - Combine the AI's ability to generate answers with a system that looks up factual information from reliable sources, thereby helping the AI provide more accurate responses.
3. **Few-shot and zero-shot learning** – Provide a few examples (few-shot) or no examples (zero-shot) to help the AI recognize tasks. Providing examples helps the AI handle new types of questions more effectively.
4. **Fine-tuning LLMs** - Adjust the AI model after its initial training with more specific data or examples. This adjustment improves the AI's accuracy in a particular area or for certain types of questions.

Companies that work with LLM and generative models have been developing tools that reduce the frequency of hallucinations. Several third-party companies have stepped up to offer programs that can be implemented as an additional layer to “catch and kill” hallucinations.



A strategy to limit hallucinations was adapted from the Galileo application [www.rungalileo.io] (ChatGPT, Copilot, Gemini, and Lama logo).

The following are examples of hallucinations in an LLM (ChatGPT4):

In geotechnical engineering, an important parameter to consider when designing sheet piles is the maximum deflection of the wall. Designers typically assume tolerable deflections ranging from 1.0 inch to 2.5 inches. However, this range is not standardized in the design manuals.

When asked about the allowable deflection for a sheet pile, ChatGPT gave the following response:

Prompt

> Imagine you're the ultimate geotechnical engineer with a vast knowledge of biographies and design manuals at your fingertips. You can answer any question with absolute accuracy, and you're not afraid to admit if there's something you don't know. Use an informal tone in your responses.

> What is the recommended maximum deflection for a cantilever sheet pile?

Response:

The maximum deflection for a cantilever sheet pile isn't universally standardized, but generally, designers aim to limit it to a range of around 1 to 2.5 inches. This range keeps the structure stable without compromising performance or causing excessive movement that could damage adjacent structures. Keep in mind that these values depend on the project, soil conditions, and design specifications, so it's crucial to tailor them to the specific situation.

Prompt

> Provide two references that support the deflection range that you mentioned. Use a ASCE style.

Response:

Here's an example of how to cite two sources in ASCE style that discuss sheet pile deflection ranges:

Reference 1:

- Bowles, J. E. (1996). Foundation Analysis and Design. 5th ed., McGraw-Hill, New York.

Reference 2:

- Lambe, T. W., and Whitman, R. V. (2008). Soil Mechanics. Wiley, New York.

These references provide foundational knowledge on soil mechanics and foundation analysis, which often include guidelines or ranges for sheet pile deflection, depending on soil conditions, structural requirements, and safety factors. For more accurate and specific ranges, check recent design manuals or technical papers specific to geotechnical engineering and sheet pile design.

Someone who is familiar with these publications knows that none of these books directly address the deflection limits. Additionally, note that the publication year for Lambe and Whitman (2008) refers to a publication in India (SI Version) by Wiley India. This example shows the extent of the database in ChatGPT. This response is a hallucination because the information provided is inaccurate but appears authentic.

When queried on the validity of its response, ChatGPT provided the following answer:

Prompt:

> review your references something is wrong.

Response:

Let me refine that for you. Here's a corrected version:

Reference 1:

- Bowles, J. E. (1996). Foundation Analysis and Design. 5th ed., McGraw-Hill, New York.

While this reference covers a range of foundational engineering principles, it doesn't specifically outline the deflection range mentioned above. However, it remains a reputable source for foundation design strategies.

Reference 2:

- Ebeling, R. M., and Morrison, E. E. (1992). The Seawalls, Bulkheads, and Quaywalls. US Army Corps of Engineers, Washington, DC.

This reference includes insights into deflection tolerances of sheet pile structures and is more closely aligned with the deflection guidelines discussed.

These are suggestions for related material where further verification can be sourced.

After the second query, ChatGPT provided some insight into one of the references and gave an additional reference—the second reference is another hallucination. The following citations are the correct references for the ones that ChatGPT has mixed in its response.

Ebeling, R. M., and Morrison, E. E. (1992). *The Seismic Design of Waterfront Retaining Structures*. U.S Army Corps of Engineers, Report ITL-92-11 Washington, DC.

NAVFAC (1981). *The Seawalls, Bulkheads, and Quaywalls*. Naval Facilities Engineering Command. DM-25.4 July 1981. Alexandria.

As previously discussed, several factors contribute to these hallucinations. The AI companies that are developing LLMs and working with generative models have been trying to reduce the hallucinations in their algorithms.

As of when this article was written, OpenAI has launched ChatGPT4o, and one of its prospective enhancements is that the new model has fewer hallucinations.

When given the same prompts as above, ChatGPT4o provided the following responses:

Response:

Alright, let's dive into the nitty-gritty of cantilever sheet piles. The recommended maximum deflection for a cantilever sheet pile typically falls in the range of 1/200 to 1/100 of the exposed height of the pile. So, if you've got a sheet pile that's 10 meters high, you're looking at a maximum deflection of about 50 to 100 millimeters.

Keep in mind, this isn't a one-size-fits-all answer. The actual acceptable deflection can depend on a bunch of factors like the type of soil, the load conditions, and what the sheet pile is supporting. For more precise guidance, it's always best to consult specific design manuals or guidelines like the "Steel Sheet Piling Design Manual" by the U.S. Army Corps of Engineers or the relevant sections in the Eurocode 7.

If you've got any more questions or need a deeper dive into this topic, just let me know!

When prompted for two references that support the above response, GPT4o provided the following:

Response:

Sure thing! Here are two references that support the deflection range I mentioned:

- U.S. Army Corps of Engineers. (1994). *Design of Sheet Pile Walls* (EM 1110-2-2504). Washington, D.C.: U.S. Army Corps of Engineers.
- EAU 2012 Committee. (2013). *Recommendations of the Committee for Waterfront Structures - Harbours and Waterways* (8th ed.). Berlin: Ernst & Sohn.

These sources should give you a solid foundation on the recommended deflection limits for cantilever sheet piles. If you need more detailed info, those manuals are great resources to check out.

Although the responses are correct, the references provided do not support the recommended information on the maximum deflection. However, ChatGPT4o does provide more precise references.

AI hallucinations are perceived as a byproduct of AI's creativity. AI is expected to be free to generate content independently and with human-like ability while also maintaining accuracy and precision. Can we have it both ways?

We will see more tools that will allow users to choose between creativity and conformity.

Certainly, limiting AI hallucinations has become an important topic when developing applications with LLM. As more tools are developed, users will be able to choose between creativity and conformity. However, the ethical aspects of human interaction with AI models should always start with Asimov's First Law of Robotics.

Previous Articles in the Series

[AI Unveiled: The Wizardry Behind Chatbots and Intelligent Systems](#)

[How Large Language Models Work: In a nutshell](#)

Upcoming Articles

Superprompts: How to Talk with the Genie and Keep it in the Bottle.

Role Play Your Way: How to Talk with Terzaghi

Applications of LLM in Geotechnical Engineering: Use it or Lose

ABOUT THE AUTHOR



Enrique Farfan is a Geotechnical and Structural Engineer with a Ph.D. in Engineering and a Master's in Civil Engineering from the University of New Mexico. His expertise spans across various projects in mining, energy, water resources, and transportation, including work on dams, levees, canals, waterfront structures, bridges, foundations, and seismic analysis. Enrique has implemented diverse design solutions using customized computer programs and database implementations. During his graduate years, he developed an interest in fuzzy logic, optimization, inverse problems, neural networks, and computer programming. Now, he enjoys exploring the fascinating universe of AI and its applications.