16 Why Might LLMs be Wrong?

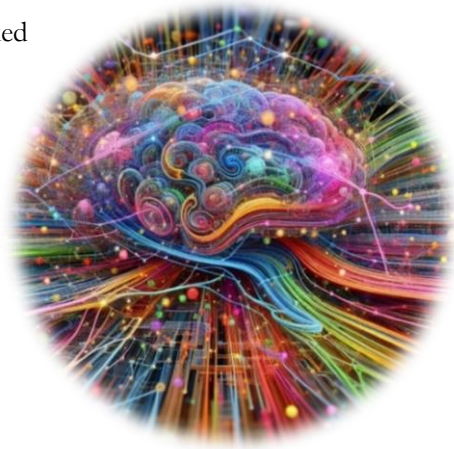
There are several reasons why LLMs are neither 100% consistent nor 100% correct. Here are but a few.

Blame the Trainers

It was humans, after all, who created all the content on which LLMs are trained. We're not always correct, and we don't always agree on things. Two people (or governments) who wrote opposing opinions about the same topic may both have their input into the training model. The complex history of humanity makes it difficult for an LLM to make binary decisions.

Blame Biology

Think of the last two times someone asked you the same question. Did you answer, verbatim, both times? It's likely that your brain will piece together each response differently. Your thoughts connect to different neurons across different synapses in your brain each time, then instruct your mouth to move.



In fact, **there are parallels between the architectural principles of large language models (LLMs) and the biology of the human brain**. Here are a few key points:

- LLMs use artificial neural networks, which are loosely inspired by how neural networks in the brain connect and transfer information.
- LLMs learn statistical patterns from their training data, like how the brain's neural circuits and memories are influenced by sensory inputs and experiences during development and life.
- The transformer architecture in modern LLMs handles sequential data (like language) similarly to how neurons and synaptic pathways in the brain send signals.
- Both LLMs and brains can apply existing knowledge to new situations and make connections between different domains based on learned statistical probabilities.

Don't worry, though. Despite the hype about Artificial General Intelligence (AGI), human brains are (currently) far more capable. Biological neurons are far more specialized and complex than an LLM's generic, artificial nodes. Brains process more than language and handle diverse tasks. We form a deeper understanding of our world by using reinforcement, unsupervised, and biologically constrained learning rules. **We are the pilots!**

Do humans understand LLMs (or brains)?

The short answer is no, to both. Knowledge about both phenomena is (and will be) under development.

Unlike our brain, which evolved through biological processes over millions of years, LLMs are quite new. They are now so complex that researchers treat them like natural phenomena. They conduct experiments to explain LLMs' behavior, like how they study unexplained phenomena in physics or biology.

Researchers have discovered many interesting traits, but there are as many questions as answers. One phenomenon is called "grokking". Sometimes, LLMs appear to fail at a task until suddenly, like a lightbulb switching on, they grasp it. This behavior contradicts conventional deep learning expectations.[1]

LLMs consist of billions of parameters. Parameters are like weighted connections between neurons. Each neuron processes input, the synapses determine the strength of connections, and the next word (or token) in the sentence. However, the exact mapping between input, weights, and output is not straightforward. It involves intricate, nonlinear connections. The sheer number of parameters makes it challenging to dissect their inner workings.

What Happened?! Avoiding Hallucinations

LLMs can occasionally generate text that is not grounded in reality, or that is irrelevant to the context. This can happen for several reasons, and can best be avoided with the following suggestions:

Why LLMs Hallucinate	Tips to Minimize Hallucination
Lack of context	Provide more background information
Unclear instructions	Clearly state the desired outcome
Ambiguous language	Use precise language
Insufficient data	Offer more details and **examples**
Overly complex prompts	Narrow down the scope of the prompt into multiple "turns," or chats back/forth

Still, even with a good prompt, there are times when an LLM goes off the rails. Behind the scenes, and beyond the control of the end-users, there are **reasons the models may hallucinate**.

Data quality

LLMs rely on huge amounts of Internet text that may be noisy, erroneous, biased, or inconsistent. Examples include training data from Reddit and other social media content. LLMs may use these sketchy sources as part of their outputs. One concern of researchers is that if models use their own prior knowledge to train their next version(s), that the incorrect content from past systems will be recursively looped into future systems.

Insufficient data

LLMs may face situations where they are asked to generate content for which they have little or no data. This can cause errors when LLMs go beyond their limits and make nonsense or irrelevant text. If you prompt an LLM without referencing any of your work data, it'll happily reply with content it considers "close enough" from the Internet.

Complexity

LLMs have many parameters, which make them hard to understand and fix. It also raises the chance of overfitting, where LLMs learn specific patterns or examples from the data and do not adapt to new situations.

Overparameterization

LLMs have more parameters (weights) than necessary. This allows them to adapt to the data very well. However, this can also result in poor generalization. Larger models are more likely to fall into the "wrong trap" and have difficulty adjusting themselves. They become less adaptable and more attached to the data. These dependencies make them both powerful and unpredictable. Some studies show LLMs lose accuracy over time.[2] See? They **are** like human brains!

Uncertainty

While LLMs' training set is vast, they can't read minds. LLMs do not always show how confident or unsure they are about their outputs. Sometimes, they'll admit they don't know the answer, and sometimes they'll provide a response anyway. The more specific the input prompt, the more specific the response back can be. While extremely well informed, **LLMs innately have zero common sense.**

Why Copilot May Get It Wrong

There are myriad reasons, but this table provides some common issues and suggests some remediations. See the Troubleshooting section in each app chapter back in Part I for additional workarounds.

Issue	Possible Cause	Solution
Copilot shows an error message when trying to use it.	- Internet connection is unstable or slow. - Copilot service is temporarily unavailable.	- Check your internet connection. - Try again later or check the Copilot status page for updates. - Try accessing Copilot via the web apps at **www.microsoft365.com**.
Copilot doesn't provide relevant or accurate suggestions.	- Query is too vague or complex - Document type or language is not supported by Copilot.	- Provide more details or keywords in your query. - Check the list of supported document types and languages.
Copilot fails to provide the desired information.	The information may not be available in the enterprise or web search results.	Try rephrasing the query or providing more specific information, like a "/" reference to a specific file.
Copilot fails to grasp the request.	The request may be phrased in a way that is difficult for Copilot to understand.	Try rephrasing the request using clear and concise language.
Copilot fails to generate content.	Copilot's guardrails may deem the request inappropriate. The request may be too broad or not specific enough.	Try a different prompt with more specific or appropriate information.
The prompt is too long.	Copilot's prompts are (currently) limited in character counts.	Shorten the prompt, breaking it into multiple parts. For instance, after the first prompt to "Create a proposal about customer xyz," next say "Add information about their company earnings."

Interns Aren't Perfect, Either

Ever give your kid or a colleague a task, and have them come back with something that was woefully off from what you meant? When you took more time explaining your desired outcome, did they come back with better results?

While LLMs dazzle us with their feats, we're still unraveling how they work. It's a scientific puzzle that bridges mathematics, computer science, and cognitive science.

There are some things that humans can do to interact better. **Think of Copilot as an eager robotic intern who can handle a finite set of tasks extremely well, but who's not yet capable of interpreting what you mean.** That should give a good grounding in why results will vary.

Another way to ensure accurate answers is to make it easier for Copilot to find the right data in the organization, which is covered in the next two chapters.

[1] **Large language models can do jaw-dropping things. But nobody knows exactly why. | MIT Technology Review**
[2] **https://arxiv.org/pdf/2307.09009**