



# AI Hallucinations



## hal·lu·ci·na·tion

*noun*

an experience involving the apparent perception of something not present.

- Occurs when LLMs generate false information.
- Can appear plausible – LLMs are designed to produce fluent, coherent text.
- Often are complete nonsense – i.e. unsubstantiated by the facts.
- Generally caused by bad data in the source content; bias created while training the model; unclear, inconsistent, or contradictory input prompts.

## ■ LLM limitations

### Inaccuracy

LLMs can generate biased, inaccurate, and incomplete data output, and are only as good as training data

### Hallucinations

LLMs are not capable of determining accuracy. Only predict what word ordering has the highest probability of success

### Transparency

LLMs generally cannot tell users why or how they came up with an answer or other response

### Consistency

LLMs will not produce the same output for the same question at the same time because they only work on selecting probabilities

### Security

LLMs utilize user inputs as training data, and have revealed proprietary or confidential data to other users

## ●■ Solving for LLM limitations

### Inaccuracy

- **Trust but verify**

### Hallucinations

- Choose products that include specific citations to matter data

### Transparency

- Evaluate all output, explanations, & citations using legal judgment

### Consistency

Slight variations in responses are not a bug. Try asking the same questions to get a sense of the variation.

### Security

Ensure LLMs are hosted securely and are not training on your confidential client data.