SLIDE 1

Chicago Sun-Times, May 2025

A syndicated summer reading supplement recommended 15 books. Ten of them did not exist.

Fake titles. Fake plots. AI-generated recommendations. The journalist did not check. Nobody did.

Source: Axios, AP, May 2025

AI with Kyle

newspaper mockup uses confirmed reported examples

SLIDE 2

The Scale

92%

of people never verify AI answers

45%

of AI responses had significant problems

BBC/EBU, 3,000 responses tested

34%

more confident when wrong than when right

MIT, Jan 2025

486+

court cases globally now involve AI-fabricated citations

Sources: Exploding Topics; BBC/EBU Oct 2025; MIT Jan 2025; Charlotin Database

SLIDE 3

Hallucination

AI presents made-up information with complete confidence.

Type 1

It invents something that does not exist.

- Fake book
- Fake court case
- Fake statistic

Type 2

It contradicts or drifts from a document you actually gave it.

- Wrong summary
- Misquoted passage
- Changed numbers

Same result either way. Fiction delivered with the tone of fact.

Source: OpenAI, Why Language Models Hallucinate, Sept 2025

SLIDE 4

The Mechanism

Search engine

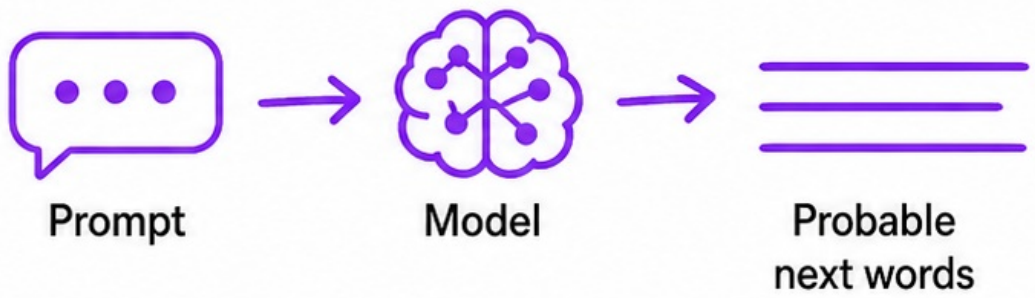
Find the thing. Show you the thing.

Search retrieves real pages and
points you to real sources.

Large language model

Predict the next token.

AI was trained on billions of words.
It learned what confident answers,
legal citations, and statistics look like.
Whether the fact is real is a separate question.



Source: OpenAI, *Why Language Models Hallucinate*, Sept 2025

SLIDE 5

The AI Has No Concept of Truth

A liar knows the truth and hides it.
AI has no concept of truth at all.

User: Are these cases real?

ChatGPT: Yes.

ChatGPT: They can be found on Westlaw and LexisNexis.

In 2023, a lawyer filed a legal brief citing six fake court cases generated by ChatGPT. The lawyer was fined $5,000.

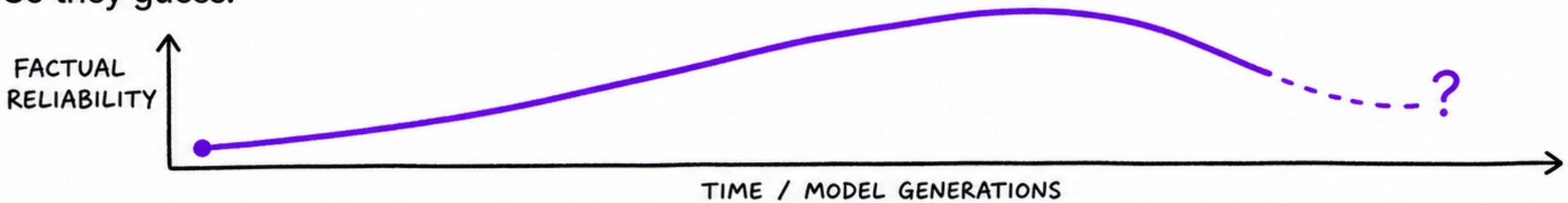
Source: Mata v. Avianca, 678 F. Supp. 3d 443 (S.D.N.Y. 2023)

	Liar	AI
Knows the truth?	Yes	No
Hiding it?	Yes	Does not apply

SLIDE 6

Will This Get Fixed?

OpenAI said in September 2025 that hallucinations are mathematically **inevitable** with current training methods. Models are trained to answer. Abstaining gets penalised. So they guess.



Getting better on...

Some factual questions with web search. GPT-5 with web search makes around **45% fewer factual errors** than GPT-4o on certain question types.

Still getting worse on...

Other tasks. Some newer models hallucinate more, not less. There is no clean upward line.

This is the baseline to build habits around, not something to wait out.

Sources: *OpenAI, Why Language Models Hallucinate, Sept 2025; GPT-5 System Card, Aug 2025*

SLIDE 7

When It’s Most Likely

Four situations where hallucination risk spikes.

Specific facts

Citations, statistics, names, dates, quotes. The more precise it sounds, the more worth checking.

Niche topics

The rarer the subject, the less training data. The model fills gaps.

Recent events

The model often does not know where its knowledge ends, and it may not say so.

Wrong assumptions

AI tends to agree with your premise. Ask a leading question, get a led answer.

Sources: Stanford RegLab, Large Legal Fictions, 2024; Nature npj Digital Medicine, 2025

SLIDE 8

What Actually Helps

Turn on web search

ChatGPT, Claude, and Gemini all have a toggle. GPT-5 makes around 45% fewer factual errors with it on. Single biggest change, zero effort.

Use stronger models for high-stakes tasks

Claude Opus, GPT-5 with thinking on, Gemini 2.5 Pro.
Slower, but worth it when it matters.

Use deep research or extended thinking

Helpful for complex factual questions and reasoning tasks.
It can be worse on summarisation, so use it selectively.

NotebookLM

Use NotebookLM with your own sources

Paste in your documents. Ask questions from them.
Citation hallucination drops dramatically when the model is grounded.
This is the plug-and-play version of RAG.

Sources: GPT-5 System Card, Aug 2025; Vectara HHEM, 2026

SLIDE 9

How To Prompt Better

RISKY

What's the stat on X?

→

IMPROVED

Here is the report. Pull the key stat and cite the exact line.

RISKY

What does the law say about Y?

→

IMPROVED

Summarise this statute in plain English.

RISKY

Is this a good idea?

→

IMPROVED

Here are two options. Which is stronger and why?

Add:
"If you're not certain, say so."

Giving the model permission to abstain helps.

Sources: Anthropic Claude docs on *reducing hallucinations*; OpenAI, *Why Language Models Hallucinate*, Sept 2025

AI with Kyle

SLIDE 10

The Reframe

For 25 years, the default way to get information online was retrieval.
Find the real thing. Show it to you.

That mental model does not apply here.

AI is a writing tool. A drafting tool. A thinking tool.

It is very good at working with information you give it,
and unreliable at sourcing information from its own memory.

The people getting the most out of AI are not treating it as a smarter search engine.
They are treating it as a brilliant drafter with an imperfect memory,
and they read the output before it goes anywhere.

AI with Kyle

Use AI to work with information. Do not trust it to source information from memory.