

MODULE 1 · WEEK 1 · LESSON 6

Hallucinations, Bias and Garbage In: Why AI Gets It Wrong

The failure modes every user must know

WHAT YOU WILL BE ABLE TO DO

- Define hallucination and give two causes
- Explain how training data bias becomes unfair outcomes
- Apply a 4-step verification habit to any AI answer

LESSON CONTENT

Hallucination: confident, fluent, false

Because the model predicts plausible text, it can produce fluent statements that are simply untrue, with complete confidence. Causes include gaps in training data and prompts that assume facts the model fills in to be helpful. Fluency is not evidence of truth.

Bias: the mirror problem

Models learn from human-produced data, which carries human biases. If historical hiring data favoured one group, a model trained on it can reproduce that preference at scale. Bias is not a glitch in the data; it is the data, faithfully learned.

Your verification habit

Step one: check whether the claim is the kind of thing the model could know. Step two: look for a primary source - a document, a statute, an official page. Step three: cross-check with a second independent source or tool. Step four: sanity-check numbers with your own arithmetic. Four steps, twenty seconds, and you stop being a victim of confident nonsense.

KEY TERMS

Hallucination	Fluent, confident output that is factually false.
Bias	Systematic unfairness learned from training data.
Grounding	Tying model output to verified source documents.
Verification habit	Check, source, cross-check, arithmetic.

HANDS-ON TASK

Catch it red-handed · Tools: Any free chatbot + a search engine

1. Ask a chatbot for a reference on a niche local topic (e.g. a small town's by-law).
2. Record any citation or statistic it gives.
3. Verify each one against a real source and log what was fabricated.

KEY TAKEAWAYS

- ✓ Before you go, three takeaways from this lesson.
- ✓ One: Define hallucination and give two causes.

- ✓ Two: Explain how training data bias becomes unfair outcomes.
- ✓ Three: Apply a 4-step verification habit to any AI answer.
- ✓ And your hands-on task: catch it red-handed.
- ✓ Ask a chatbot for a reference on a niche local topic (e.g.
- ✓ a small town's by-law).
- ✓ When it is done, mark the lesson complete on your dashboard, keep the artifact for your portfolio, and step into the next lesson.
- ✓ I will see you there.

Watch the lesson video on your dashboard, complete the task, then mark the lesson complete. Your artifact belongs in your Portfolio of Evidence.