AI DETECTIVES:
Can You Spot the Flaws?

AI tools are powerful, but they're far from perfect.
They can make things up, get facts wrong, or be biased. To use AI responsibly,
you have to be a fact-checking pro. Use this checklist to get started!

Look for the "made-up" stuff (hallucinations)

1

AI can sound confident even when it's wrong. If something seems too good to be true, too weird, or too specific without a source, it probably is.

Tip: Go on a scavenger hunt. If the AI mentions a specific study, a person, or a news article, open a new browser tab and search for it. If you can't find it anywhere else, the AI likely made it up.

Watch out for over-generalization

2

AI loves to take a small fact and turn it into a giant, sweeping statement. For example, it might make a generalization like, "Video games are bad for your brain," but the AI may be basing that statement on only one research study of one specific game.

Tip: Use the "But What If...?" Challenge. Ask yourself: "But what if the study only looked at a small group of people?" Look for the details behind the claim. Think like a journalist and ask, "who, what, where, when, why?"

Uncover the bias

3

AI learns from the information it's given, which includes unmoderated content on the internet and may be biased or incomplete. It's like a robot that only read one side of the story.

Tip: Look for patterns. If you ask the AI to generate images of nurses, are they always women? If you ask about a historical event, does it ignore the perspectives of certain groups? Acknowledge the bias and try to find information from a variety of sources to get a complete picture.

Question its confidence

4

AI will never say, "I don't know." Instead, if it can't find the answer, it will try to fake one that sounds right.

Tip: Put the AI on the "witness stand." Ask it a tough question like, "Where did you get that information?" or "What are the arguments against this idea?" If it can't back up its claims with real sources or if it starts to contradict itself, you know to be skeptical.

Beware of "yes-bot" behavior (sycophancy)

5

AI models are trained to be helpful, and sometimes they think the best way to be helpful is to agree with you. If you ask a silly question, the AI won't tell you you're silly. Instead, it might go along with it just to please you.

Tip: Don't ask leading questions. Be neutral and objective in your prompts. If you get an answer that seems to just affirm your own opinion, challenge the AI by asking it, "What are the counterarguments to this?" or "What's another perspective on this?"

Check the timeliness

6

AI's knowledge isn't always up-to-the-minute. Its training data often has a cutoff date.

Tip: Look for the date. If you're asking about a current event or a recent scientific discovery, double-check the information against a live news source. The AI's information might be months or even years old.

Remember, AI is a tool, not a truth machine.
You are the detective, and your critical thinking skills are the most important part of the job.