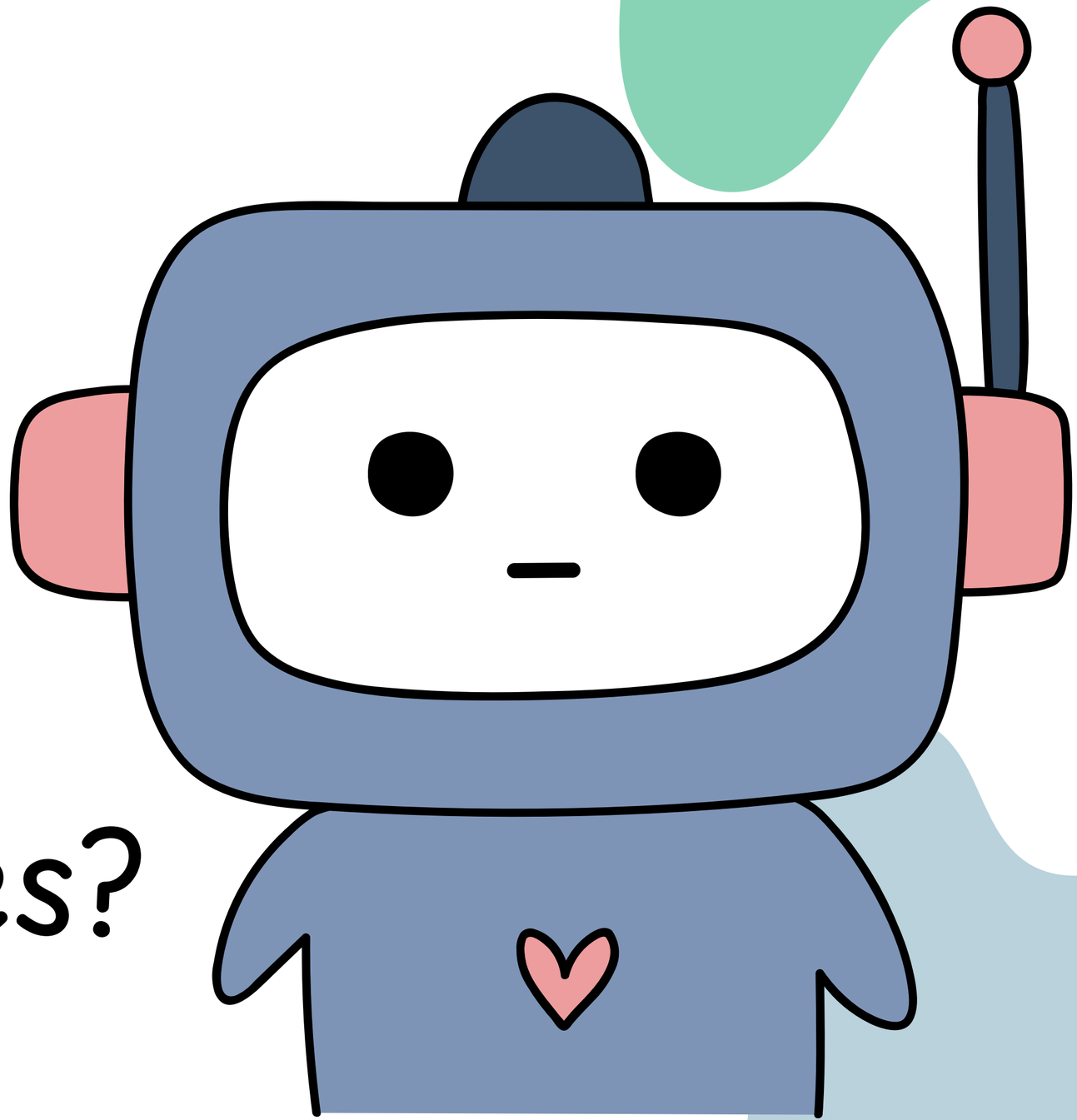
Can AI Make Mistakes?



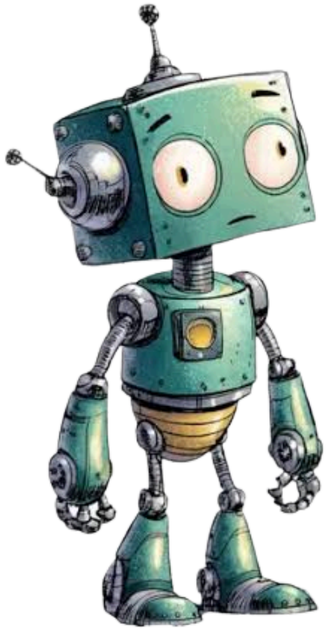
WARM-UP QUESTION

Are computers always right?

AI CAN MAKES MISTAKES

How?

- Relies on old data they've used before
- Relies on guesses
- It doesn't think, it predicts patterns



TYPES OF MISTAKES AI CAN MAKE

- Factual errors: missing important information
- Bias: unfair patterns learned from data
- Hallucinations: making information up
- Misunderstanding context

REAL LIFE EXAMPLES

- Factual errors: Told a customer they could get a refund under a real policy, but got the policy details wrong. Airline was held responsible for what its AI told the customer.
- Bias: Amazon built an AI to help sort job applications. But it had learned from old hiring data that was mostly men, so it started downgrading resumes that had the word "women's" in them (like "women's chess club"). Amazon found out and shut the tool down before it was fully used.
- Hallucinations: A lawyer used ChatGPT for legal research. It invented 6 fake court cases with fake quotes. When he asked the AI to double check, it said they were real. He got fined because he never checked if the cases were real before submitting them to court

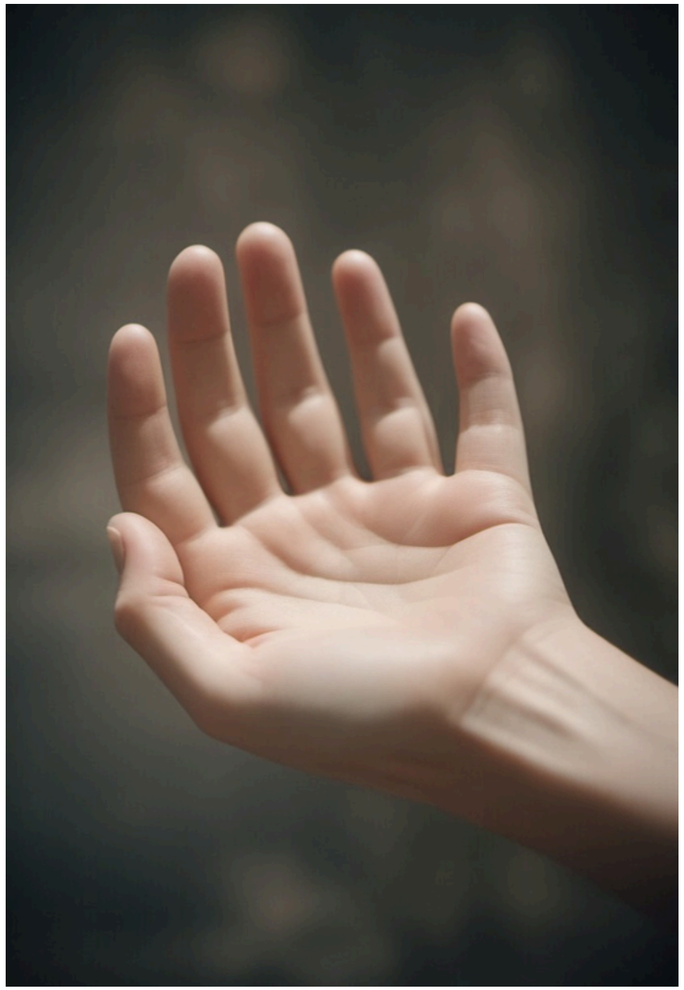
HOW TO SPOT AN AI PICTURE

Look closely at:

- Hands & fingers: too many, too few, or bent wrong
- Text & symbols: letters that look like scribbles or gibberish
- Patterns: things that repeat weirdly (extra chair legs, doubled objects)
- Light & shadows: shadows going the wrong way, or missing
- Edges: blurry or melted-looking where two things meet

EXAMPLES





EXAMPLES





EXAMPLES





HOW HUMANS HELP AI

- If we notice a mistake from the AI, we write it down in the prompt
- The AI will try its best to notice the mistake and start over.



REFLECTION QUESTION

Why should we check AI answers?