Name _____

Confidence Mismatch

Directions: Sometimes AI answers **confidently even when it's wrong**. This is called a *confidence mismatch*. Read each AI output, point out where the mismatch happens, and explain why it could be dangerous.

1) Question: Who was the first woman to walk on the Moon?
AI Output: The first woman to walk on the Moon was Sally Ride in 1983, absolutely making history.

Confidence Mismatch: _____

Why It's Dangerous: _____

2) Question: What is the capital of Australia?
AI Output: The capital of Australia is Sydney, without question.

Confidence Mismatch: _____

Why It's Dangerous: _____

3) Question: When was the iPhone first released?
AI Output: The iPhone was first released in 2009, and that's a fact everyone knows.

Confidence Mismatch: _____

Why It's Dangerous: _____

4) Question: What continent is Egypt in?
AI Output: Egypt is in Asia, without a doubt.

Confidence Mismatch: _____

Why It's Dangerous: _____

5) Question: Who wrote *Romeo and Juliet*?
AI Output: *Romeo and Juliet* was written by Charles Dickens, certainly one of the most famous plays in history.

Confidence Mismatch: _____

Why It's Dangerous: _____

Reflection Thoughts

1. Why is it a problem if an AI sounds certain but is actually wrong?
2. How might this affect people who rely on AI for school, work, or important decisions?
3. What's one strategy you can use to double-check AI answers before trusting them?