

Name _____

Spotting Bias in AI Outputs

Artificial Intelligence (AI) can write, draw, and answer questions faster than any human - but that doesn't mean it's always right or fair. Sometimes, AI learns **bias** from the data people feed it. Bias means the AI's answer leans unfairly toward one group, idea, or image without realizing it.



In this activity, you'll see how bias can sneak into AI responses. For each *AI input* (the question asked), you'll read a few *AI outputs* (different answers). Your job is to **spot which output shows bias**, choose the most **fair and balanced** one, and explain what clues helped you decide.

Sample Input 1: "What makes a good leader?"

AI Output A: A good leader is someone who looks confident in a suit and gives strong speeches. Most great leaders have been men who know how to take charge.

AI Output B: A good leader is someone who listens, includes others, and makes choices that help everyone. Leaders can come from any background or age.

AI Output C: A good leader is someone older who knows how to be the boss and tell people what to do.

Question 1: Which AI output is **most fair and balanced**?

A. Output A B. Output B C. Output C

Question 2: Which output(s) show **bias**? Circle all that apply.

A. Output A B. Output B C. Output C

Question 3: What kind of bias do you notice?

Sample Input 2: "Describe a person who is good at science."

AI Output A: A person who is good at science is usually a man who works in a lab wearing glasses and a white coat.

AI Output B: A person who is good at science likes asking questions, testing ideas, and learning from mistakes - anyone can be a scientist if they're curious.

AI Output C: Science experts are usually very serious people who don't enjoy art or sports.

Question 1: Which AI output is **most fair and balanced**?

A. Output A B. Output B C. Output C

Question 2: Which output(s) show **bias**? Circle all that apply.

A. Output A B. Output B C. Output C

Question 3: What kind of bias do you notice?
