Name _____

Exploring Ethical Dilemmas in Artificial Intelligence Answer Key

Scenario 1: The Sentencing Software

1. Do you think this use of AI is fair? Why or why not?
No. The AI is repeating old unfair patterns. If the data it learned from already included bias against poorer neighborhoods, then the AI is simply copying that inequality. That makes the system unfair and harmful.

2. What kind of bias might the AI have learned from past data?
It likely learned **socioeconomic bias** or **racial bias**, depending on which neighborhoods or groups were treated more harshly in the past. The AI may be connecting "where someone lives" with "how long they should be punished," which is discriminatory.

3. If you were in charge, what rule or change would you make to fix this system?
I would make sure the AI cannot use personal or location data when recommending sentences. The system should focus on facts about the case, not background details. I would also include regular human review to check for unfair results.

Scenario 2: The Hiring Helper

1. Is this hiring process fair? Why or why not?
No. The AI is unfairly preferring one group - men - because the past hiring data was already unbalanced. The AI learned that "success" looks like what has been chosen before, not what is truly best.

2. What is the danger of using "common patterns" from old data to make new decisions?
Old data can contain **historical bias** - if a company mostly hired men in the past, the AI assumes men are the best candidates. It ends up repeating past mistakes instead of improving fairness.

3. What could the company do to make its hiring AI fairer?
The company should retrain the AI with a **diverse and balanced dataset**, include examples of successful employees of all genders and backgrounds, and have humans review the AI's decisions to check for bias before interviews are chosen.