**Assignment: “AI in the Real World — Judge the Bot"**

Description

**Assignment Brief (student-facing)**

You’re now a **Responsible AI Inspector** Your job is to investigate how AI is used in a scenario, spot anything suspicious (bias? lack of transparency?), and help fix it. You'll be given 2 short cases.

Your mission:

1. Describe what the AI is doing.
2. Spot what could go wrong (hint: check fairness, privacy, accountability...)
3. Suggest 1 way to improve it responsibly.

Bonus points if you write your answer like a short **blog post** explaining the issue in a fun and clear way.

Think of it as being a detective — but make it **vibe**.

**Example Scenario Prompts**

1. **Hiring Bot**: A company uses an AI to screen job applicants. It tends to reject more female applicants with career gaps.
2. **School Proctoring AI**: A system flags students as "cheating" based on eye movement — but it often flags neurodivergent students.

**Deliverables**

* Short write-up for 2 cases:
  + What’s happening
  + What’s problematic
  + One improvement idea
* Bonus format: blog

## **Case 1: The Hiring Bot Mystery**

**What’s happening:**  
A company has deployed an AI hiring assistant to filter job applicants. It looks at résumés and decides who moves forward. But… there’s a catch: it keeps rejecting more women who have career gaps (like time off for childcare).

**What’s problematic:**  
Our bot detective senses **bias** here. The AI is unfairly penalizing a specific group — women with career breaks. That means qualified candidates are being unfairly filtered out just because the system learned from biased historical hiring data. This reduces fairness, diversity, and equal opportunity.

**How to fix it:**  
Solution clue: Train the AI on **more balanced and inclusive data**. Specifically, adjust the model to recognize that career gaps don’t equal poor performance. Also, add a human review step for flagged applications. In detective terms: don’t let the bot be the final judge!

## **Case 2: The School Proctoring Bot**

**What’s happening:**  
In online exams, a school uses an AI proctoring tool that flags students for “suspicious behavior.” The trigger? Too much eye movement away from the screen.

**What’s problematic:**  
Our AI inspector finds this shady. Students who are **neurodivergent, anxious, or visually impaired** may naturally move their eyes more. The system mislabels them as cheaters, which is unfair, stressful, and even discriminatory. This creates a **false accusation problem** and damages trust in the exam system.

**How to fix it:**  
Solution clue: Use AI as a **support tool, not judge and jury**. Eye movement should be one signal among many, not the only factor. Combine with other evidence (e.g., unusual network activity, unexplained audio). Plus, allow human invigilators to review alerts before accusing students.

**Closing note from the AI Detective:**  
Both cases show the same lesson — AI isn’t magic. It reflects the data and rules we feed it. To keep it responsible, we need fairness checks, privacy safeguards, and human oversight. Otherwise, the bot becomes more of a suspect than a helper.

**REFERENCE**

Silberg, J., & Manyika, J. (2019). Notes from the AI frontier: Tackling bias in AI (and in humans). McKinsey Global Institute, 1(6), 1-31.