

AI Detective Vibes: Spotting Flaws & Fixing Bias

Case 1: The Biased Hiring Bot

What's Happening:

A tech company deploys an AI-powered hiring assistant to screen job applicants. It scans resumes, ranks candidates, and rejects some automatically. However, the AI keeps flagging and rejecting female applicants who took career breaks - often due to caregiving.

What's Problematic:

This is a classic case of bias baked into the data. The AI likely learned from historical data where women with gaps were under-hired - and now it's repeating that injustice, perpetuating a cycle of discrimination. It fails the fairness test and violates the principle of equal opportunity.

One Way to Fix It:

Inject fairness constraints into the model - and train it on balanced, diverse datasets. Better yet, use counterfactual testing: ask, "Would this applicant have been accepted if the gap weren't there?" This helps the model avoid relying on irrelevant factors like career breaks.

Case 2: The Overzealous Proctoring AI

What's Happening:

A school uses an AI system to monitor students during online exams. It tracks eye movements, background noise, and facial expressions to catch cheaters. But it often flags neurodivergent students or those with ADHD, tics, or different gaze patterns.

AI Detective Vibes: Spotting Flaws & Fixing Bias

What's Problematic:

This AI confuses difference with dishonesty. Neurodivergent behaviors aren't cheating - they're natural variations. This leads to false accusations, stress, and even academic penalties for students who did nothing wrong. It's an accountability and inclusivity failure.

One Way to Fix It:

Redesign the system with human-in-the-loop review - don't let the AI make final decisions. Also, allow students to self-identify accommodations so the system doesn't treat everyone the same when it shouldn't.

Final Thoughts - The AI Vibe Check

Just because AI can see, rank, or decide doesn't mean it should - especially without context or compassion. Whether it's hiring or school exams, transparency, fairness, and inclusivity aren't optional - they're essential.

When in doubt, ask:

Would you trust this AI to judge your mom, your cousin, or your favorite teacher?

If the answer is no - go back, tweak that model, and do right by the people.