Responsible AI Inspector Report

Case 1: The Job Interviewer Who Never Sleeps (But Might Be a Bit Sexist)

What's Happening?

Imagine applying for your dream job and getting ghosted - not by a person, but by a hiring bot. This AI tool automatically screens resumes and filters out candidates before any human even sees them. Well, here's the twist: it's been rejecting a lot of female applicants, especially those with career gaps - like for maternity leave. The AI thinks they're 'less committed.'

What's Problematic?

- Unfairness: Discriminates against women.

- Lack of transparency: Candidates don't know why they were rejected.

- No accountability: Who's responsible?

One Improvement Idea:

Audit and retrain the model using fairness toolkits like IBM's AI Fairness 360. Also, add transparency features so applicants get clear rejection reasons.

Case 2: The Overzealous Cheating Detector

What's Happening?

Online exams use AI proctoring to detect cheating - watching through webcams for 'suspicious' behavior. But students with neurodivergent traits (like ADHD or autism) often get flagged just for behaving differently.

What's Problematic?

- Discrimination: Neurodivergent students unfairly targeted.

- Stress and stigma: False accusations can happen.

- Lack of inclusivity: AI not trained on diverse behavior data.

One Improvement Idea:

Introduce human-in-the-loop reviews. Also, train the system with behaviorally diverse datasets.

Conclusion

Final Thoughts

AI can be smart - but it needs human wisdom to be fair. We can't let machines make life-changing decisions without checks, balances, and empathy.

Next time you meet an AI, remember: it's not magic. It's just math - and math needs a conscience.