



Real-World AI Alert: AI Inspector in Action!



Hello, tech-curious folks! I'm the AI Inspector, here to unravel the mysteries and ethical dilemmas that arise when artificial intelligence meets reality. Today, we'll put on our detective hats and investigate two intriguing cases!

Case 1: The Forgetful Hiring Bot

What's Going On?

A company is using an AI bot to perform the initial screening of job candidates, seeking to streamline the selection process.

What could go wrong?

This hiring bot has demonstrated a problematic bias, disproportionately rejecting more female candidates, especially those with gaps in their career history. This raises serious **fairness** issues, as the system is not evaluating individuals equitably, perpetuating existing biases in the training data and resulting in the unfair exclusion of qualified talent, in addition to negatively impacting workforce diversity.

How can we improve this responsibly?

- ☐ It is crucial to **audit and retrain the algorithm** using a more diverse and representative dataset, including successful examples of women with and without career interruptions.
 - ☐ Additionally, a **human review** system should be implemented for candidates initially discarded by the AI, especially those in groups at risk of bias, ensuring that no talent is unfairly overlooked.
-

Case 2: The School's Suspicious Eagle Eye

What's Going On?

An AI monitoring system is being employed in educational settings to detect alleged "cheating" during online assessments by analyzing students' eye movements.

What could go wrong?

This system presents critical **fairness** and **privacy** issues. Neurodivergent students are frequently mistakenly flagged as "cheating" due to eye movement patterns or fidgeting that are inherent to their condition, but that the AI misinterprets. This not only causes stress and unfair academic consequences for these students, but the constant surveillance and collection of eye movement data also raises significant

concerns about student **privacy** and the lack of transparency in the AI's decision-making criteria.

How can we improve this responsibly?

- ☐ The system needs to be redesigned and retrained to be more inclusive, considering the diversity of neurotypical and neurodivergent behaviors.
- ☐ Crucially, the final decision about "cheating" should never be solely algorithmic; it is imperative to have human oversight that reviews the AI's alerts, analyzes the student's individual context, and makes the final decision.
- ☐ **Transparency** about how the algorithm works and the students' right to challenge the flags raised by the AI are essential elements of **responsible** use.

I hope this detailed analysis helps you better understand the ethical challenges of AI and how we can work to make it more fair and **accountable**! Stay tuned for more investigations from the AI Inspector!