

## Case Study: AI Incidents and Ethics

### Case Study I Selected:

I have chosen the Amazon's AI Recruiting Tool Bias (2014–2018) based on the case in the AI Incident Case Studies document.

**Summary:** Between the years 2014 and 2018, Amazon developed an AI recruiting tool for streamlining the recruitment process. However at that time the system showed a clear bias against the female applicants particularly for jobs in the technical roles. The model was trained on the past resumes learned historical biases favoring the male applicants data patterns and penalizing resumes with the word female. Despite the internal efforts made to fix the issue, Amazon eventually scrapped the tool recognizing its failure to treat the candidates fairly and objectively. This case became a landmark example of how the biased training data can lead to discriminatory outcomes especially in the hiring process.

### Critical Analysis

#### 1. Bias in the Training Data

- **Why this matters:** I've chose this because biased data can easily go unnoticed until it creates the serious consequences. It's easy to assume that the AI is neutral but this case proved otherwise.
- **My thinking on this:** The model simply used the society's past hiring decisions showing how the machine learning can reinforce inequality if we don't carefully audit the data. What caught my eye was that how subtle the bias was just learning patterns that had already been rewarded.
- **Solution:** Companies should use bias aware data preprocessing techniques and do diversify their training datasets intentionally. Doing the regular audits using fairness metrics can also catch such issues early.

#### 2. Lack of the Transparency

- **Why it matters:** Without the interpretability we can't understand why the system makes certain decisions. In this case even the internal teams struggled to pinpoint why certain resumes were downgraded.
- **My thinking on this:** I've realized how much risk comes from not being able to interpret the model outputs especially in the areas like hiring where this decisions deeply affect the people's lives.
- **Solution:** Incorporating explainable AI methods like SHAP, LIME during the development. Transparency should be a non negotiable part of any model.

### 3. Ethical Oversight and Accountability

- **Why it matters:** The lack of formal oversight allowed this tool to develop and be tested for years before the issues were addressed.
- **My thinking:** I assumed that most companies had tight ethical guardrails but now I understand that organizational priorities don't always align with the responsible AI.
- **Solution:** By forming the dedicated AI ethics review boards that include diverse stakeholders engineers, HR leaders, ethicists who can flag the concerns early and regularly.

**Reflection on the Critical Thinking:** In general I assumed that the biggest risk in AI was just getting the math wrong. But then this case changed my view. It's the human decisions behind the data and designs that determine whether the output is harmful or fair. I used to focus mainly on the technical accuracy but now I understand that ethical context matters just as much. For instance, I hadn't previously considered how historical bias could quietly enter into the training data. In the future, I'll be more careful in asking who is this model serving and who might it unintentionally harm. My approach to the AI development will now include not just validation and performance testing but also the ethical checkpoints throughout the process.

#### References:

1.AI Incident Case Studies Document

2.Reuters article on Amazon's AI hiring bias (October 2018)

<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

3.The Guardian article on Amazon ditched AI recruiting tool that favored men for technical jobs (2018)

<https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>