

Part 2: Case Study Analysis (40%)

Case 1: Biased Hiring Tool

Scenario: Amazon's AI recruiting tool penalized female candidates.

Tasks:

- *Source of Bias:* The model was trained on past hiring data that reflected male-dominated patterns, embedding gender bias into the model.

- *Three Fixes:*
 - Remove gendered keywords and attributes from training data.

 - Rebalance the dataset to reflect diverse candidates equally.

 - Introduce fairness constraints during model training.

- *Fairness Evaluation Metrics:*
 - **Demographic Parity:** Equal selection rate across genders.

 - **Equal Opportunity:** Equal true positive rates for different groups.

 - **Disparate Impact Ratio:** Should be close to 1 for all groups.

Case 2: Facial Recognition in Policing

Scenario: A facial recognition system misidentifies minorities at higher rates.

Tasks:

- *Ethical Risks:*
 1. Wrongful arrests and legal consequences.
 2. Invasion of privacy and civil liberties.
 3. Erosion of trust in law enforcement and justice systems.
- *Policy Recommendations:*
 1. Ban the use of facial recognition in high-risk scenarios until bias is resolved.
 2. Mandatory third-party audits for fairness and accuracy.
 3. Require informed consent and transparency from law enforcement agencies.
 4. Provide human oversight and appeals process for all identifications.