

Part 2: Case Study Analysis (40%)

Case 1: Biased Hiring Tool

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

Source of Bias: - Historical hiring data was skewed toward male candidates, causing the model to learn gendered patterns. - Feature selection may have inadvertently included proxies for gender. - Model design did not incorporate fairness constraints.

Proposed Fixes: 1. Remove gender and any proxy features from the training dataset. 2. Rebalance the training data to include an equal representation of genders. 3. Implement fairness-aware algorithms or constraints that penalize biased outcomes.

Metrics to Evaluate Fairness: - Statistical parity difference (measure difference in selection rates between genders). - Equal opportunity (check true positive rates across genders). - Disparate impact ratio.

Case 2: Facial Recognition in Policing

Ethical Risks: - Wrongful arrests due to misidentification, disproportionately affecting minorities. - Violation of privacy and potential misuse of surveillance data. - Erosion of public trust in law enforcement.

Policies for Responsible Deployment: 1. Require human oversight in decision-making; AI outputs should support, not replace, human judgment. 2. Regularly audit system performance across demographic groups and report findings publicly. 3. Establish strict data governance, ensuring consent, privacy, and compliance with civil rights laws.