**✅ Part 2: Case Study Analysis**

**Case 1: Biased Hiring Tool**

**Scenario:** Amazon’s AI recruiting tool penalized female candidates.

**Source of Bias:**

* Biased training data: Historical hiring data reflected male-dominated hiring patterns.
* The model learned from patterns that favored male candidates, reproducing existing discrimination.
* Lack of bias checks during model design and deployment.

**Proposed Fixes:**

* Use balanced and diverse training datasets that represent all genders equally.
* Include bias detection and mitigation techniques during model training (e.g., reweighting or adversarial debiasing).
* Implement regular human audits and override mechanisms to review automated decisions.

**Metrics to Evaluate Fairness:**

* Disparate impact ratio to measure selection fairness across genders.
* False positive and false negative rates for different demographic groups.
* Regular fairness audits comparing outcomes before and after corrections (based on general fairness metrics from AI ethics best practices).

**Case 2: Facial Recognition in Policing**

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

**Ethical Risks:**

* Risk of wrongful arrests due to false positives for minority groups.
* Violation of individual privacy and potential misuse of surveillance data.
* Erosion of public trust in law enforcement and technology providers.

**Recommended Policies:**

* Establish strict guidelines for when and how facial recognition can be used (e.g., only with court orders).
* Mandate independent audits and accuracy testing for different demographic groups.
* Require transparency, public accountability, and clear channels for individuals to contest wrongful identification.

**General Reference:** Based on well-known cases reported by *The New York Times*, *MIT Media Lab research*, and AI ethics frameworks.