

Part 2: Case Study Analysis

Case 1: Biased Hiring Tool

Source of Bias:

- **Training data bias:** The model learned from historical hiring data favoring male candidates.

Proposed Fixes:

1. **Debias training data** by oversampling female applicants or removing gender-indicating features.
2. **Use fairness-aware algorithms** (e.g., adversarial debiasing).
3. **Regular audits** with diverse test cases to detect bias.

Fairness Metrics:

- **Disparate impact ratio** (selection rates across genders).
 - **False positive/negative parity** (equal error rates).
-

Case 2: Facial Recognition in Policing

Ethical Risks:

- **Wrongful arrests** due to misidentification.
- **Privacy violations** from mass surveillance.
- **Reinforcement of racial profiling.**

Policy Recommendations:

1. **Ban use in high-stakes decisions** (e.g., arrests without human review).
2. **Mandate accuracy thresholds** for different demographics.
3. **Public oversight** to prevent misuse.