Case 1: Biased Hiring Tool – Amazon's AI Recruiting System
Problem Summary:
Amazon's AI recruiting tool penalized female candidates, particularly for technical roles. The model downgraded resumes with terms like “women’s” (e.g., “women’s chess club”) and favored male-dominated language and experience.

1. Source of Bias:
Training Data Bias: The model was trained on historical hiring data, which reflected male-dominated hiring trends at Amazon.

Label Bias: Success was defined using past hiring decisions, which were biased.

Feature Selection Bias: The model learned that male-associated terms correlated with successful candidates.

2. Three Fixes to Improve Fairness:
Debias Training Data

Use a balanced dataset that includes equal representation of genders.

Remove biased features (e.g., gendered terms, names, pronouns).

Introduce Fairness Constraints During Model Training

Use tools like IBM AI Fairness 360 or Fairlearn to reduce disparate impact and enforce group fairness.

Human-in-the-Loop Review

Combine automated screening with human evaluation to catch unfair exclusions before final decisions are made.

3. Metrics to Evaluate Fairness Post-Correction:
Demographic Parity: Hiring rates should be similar across gender groups.

Equal Opportunity: Qualified candidates of all genders should have equal chances of being recommended.

Disparate Impact Ratio: Ensure ratio of positive outcomes between groups is above 80% (the “four-fifths rule”).

Case 2: Facial Recognition in Policing
Problem Summary:
Facial recognition tools have shown significantly higher error rates for minority groups, especially Black and Asian individuals. This has led to wrongful arrests, misidentification, and erosion of public trust.

1. Ethical Risks:
Wrongful Arrests & Discrimination: Misidentification can lead to innocent individuals being detained, prosecuted, or surveilled unfairly.

Privacy Violations: Mass surveillance using facial recognition can occur without consent, infringing on civil liberties.

Bias Reinforcement: If biased data from arrests is used to retrain the system, it reinforces systemic inequality.

2. Recommended Policies for Responsible Deployment:
Bias Auditing Before Deployment

Test accuracy across different demographic groups and publish the results.

Human Oversight Mandate

Ensure that AI-generated matches are always reviewed and validated by trained officers before any action is taken.

Transparency & Accountability

Public disclosure of where and how facial recognition is used.

Establish oversight committees and provide mechanisms for individuals to challenge wrongful matches.

Restrict Use to High-Stakes Situations

Use facial recognition only for serious investigations, not general surveillance.