Case 1: Biased Hiring Tool (Amazon Recruiting System)

1. Source of Bias

The bias originated from **the training data**.
Amazon's AI hiring tool was trained on **historical resumes** submitted to the company over a 10-year period. Since the tech industry had been **male-dominated**, the data reflected this imbalance. The model "learned" to favor male candidates and penalize resumes containing words like "women's," as in "women's chess club captain."
Additionally, **model design** contributed to bias because it **relied too heavily on past hiring patterns**, rather than objective skills or competencies.

2. Three Fixes to Make the Tool Fairer

1. **Balanced and Representative Training Data**
    - Rebuild the dataset to include an **equal representation of male and female applicants**, as well as diversity in roles, experience, and education.
    - Use **data augmentation** or reweighting techniques to correct for historical imbalances.
2. **Bias Detection and Mitigation Techniques**
    - Apply **fairness-aware algorithms**, such as adversarial debiasing or re-sampling methods, to reduce the model's reliance on gender-related features.
    - Regularly audit the model for biased patterns during development and deployment.
3. **Human Oversight and Transparency**
    - Involve **human HR experts** in the final decision-making process, ensuring that the AI tool supports rather than replaces human judgment.
    - Provide **explainable AI outputs** so recruiters can understand and verify why the model ranks a candidate in a certain way.

3. Metrics to Evaluate Fairness Post-Correction

To ensure the tool remains unbiased, use **quantitative fairness metrics**, including:

Metric	Purpose	Example Use

Demographic Parity	Measures whether selection rates are equal across groups.	Check if male and female candidates are shortlisted at similar rates.
Equal Opportunity	Ensures qualified candidates have equal chances regardless of gender.	Compare true positive rates between male and female applicants.
Disparate Impact Ratio	Compares favorable outcomes between protected and unprotected groups.	Ensure ratio ≥ 0.8 (Four-Fifths Rule standard).
Bias Audit Reports	Regularly assess outcomes across gender and other demographics.	Ongoing monitoring to prevent new biases from emerging.

Summary:
The Amazon hiring bias stemmed from **biased historical data** and **flawed model design**. Fixes should focus on **data balance**, **algorithmic fairness**, and **human oversight**, while fairness metrics like **demographic parity** and **equal opportunity** ensure continuous improvement.

Case 2: Facial Recognition in Policing

1. Ethical Risks

1. **Wrongful Arrests and Discrimination**
    - Facial recognition systems have been shown to **misidentify people of color, women, and young individuals** more frequently.
    - This can lead to **false accusations or wrongful arrests**, damaging reputations and trust in law enforcement.
2. **Privacy Violations**
    - Constant surveillance using facial recognition can **infringe on individuals' right to privacy**, as it tracks movements and identities without consent.
    - Data collected can be misused, leaked, or shared with third parties without transparency.
3. **Erosion of Public Trust**

- Over-reliance on automated systems may cause communities—especially minorities—to **lose confidence in the fairness of policing**.
- Lack of accountability can make it difficult for affected individuals to challenge AI-based decisions.

4. **Bias Amplification**
    - If the training data used to build the system underrepresents certain groups, **the algorithm reinforces social and racial inequalities**, leading to systemic bias in law enforcement decisions.

2. Policies for Responsible Deployment

1. **Independent Auditing and Bias Testing**
    - Require **regular third-party audits** of facial recognition systems to detect and correct bias.
    - Mandate **accuracy benchmarks** across all demographic groups before any operational use.
2. **Transparency and Accountability Measures**
    - Law enforcement agencies should **publicly disclose** how and when facial recognition is used.
    - Implement **appeal and oversight mechanisms** so individuals can contest AI-based identifications.
3. **Strict Data Protection and Privacy Controls**
    - Limit data collection to **specific, legally justified investigations**.
    - Prohibit the use of facial recognition for mass surveillance or monitoring peaceful protests.
    - Ensure data is **encrypted, stored securely**, and deleted after investigations conclude.
4. **Human-in-the-Loop Decision Making**
    - Ensure that **facial recognition results are only advisory**.
    - Final identification or arrest decisions must be made by trained human officers, not automated systems.

5. **Legal and Ethical Frameworks**

    - Develop national or regional **AI ethics laws** that define acceptable use cases and enforce penalties for misuse.

    - Require **informed consent and community consultation** before deploying such technologies.

Summary:
Facial recognition in policing presents serious ethical challenges, including **wrongful arrests**, **privacy violations**, and **loss of public trust**. Responsible deployment requires **bias audits, transparency, data protection, and human oversight** supported by strong **legal frameworks**.