Part 2: Case Study Analysis (40%)

Case 1: Biased Hiring Tool (Amazon)

Scenario:
Amazon used an AI system to help with hiring. But the system was **penalizing women's resumes** because it was trained using **10 years of past data,** mostly from **male candidates**, since tech is male-dominated.

Task 1: Identify the Source of Bias

- The **bias came from the training data**.
- The resumes used were **mostly from men**, so the AI **learned that men are more desirable**, even if that's not true.
- The **model copied past hiring patterns**, which were already biased.

Task 2: Propose 3 Fixes to Make the Tool Fairer

1. **Use a balanced dataset**: Train the model on an equal number of resumes from both **men and women**.
2. **Remove gendered keywords**: Delete things like "women's chess club" or names that clearly show gender.
3. **Review model decisions**: Regularly **audit** the AI's hiring suggestions to make sure it's not favoring one gender.

Task 3: Suggest Fairness Metrics

- **Disparate impact**: Measure if one group is **harmed more** than others by the tool.
- **Selection rate**: Check if **both genders are selected equally**.
- **Equal opportunity**: Make sure **qualified people** from all groups have the **same chance** of being selected.

Case 2: Facial Recognition in Policing

Scenario:
A facial recognition tool used by police **makes more mistakes** when identifying **Black and minority individuals**. This can lead to **wrong arrests** or the surveillance of innocent people.

Task 1: Discuss Ethical Risks

1. **Wrongful arrests** – Innocent people might go to jail because of a **false match**.
2. **Privacy violation** – People’s faces are being scanned and stored **without permission**.
3. **Discrimination** – The system could target certain **racial or ethnic groups** more than others.

Task 2: Recommend Responsible Policies

1. **Independent testing** – Before police use the system, it should be **tested by experts** to ensure it works fairly for all skin tones.
2. **Public transparency** – People should know when and where facial recognition is being used.
3. **Clear rules and limits** – Police should only use the tech for **serious crimes**, not just general surveillance.
4. **Opt-out options** – Allow people to **refuse** to be part of facial recognition databases, where possible.