Part 1: Theoretical Understanding (30%)
1. Short Answer Questions

Q1: Define algorithmic bias and provide two examples of how it manifests in AI systems.

> **Algorithmic bias** refers to systematic and repeatable errors in an AI system's outputs that result in **unfair or discriminatory treatment** of certain individuals or groups, often based on protected characteristics like race, gender, or socioeconomic status.

Q2: Difference between transparency and explainability in AI. Why are both important

Transparency	Explainability
Transparency is about the openness and clarity of the AI system's design, data, and processes. It answers the question: "What is in this model and how was it built?"	Transparency is about the openness and clarity of the AI system's design, data, and processes. It answers the question: "What is in this model and how was it built?"

Q3: How GDPR (General Data Protection Regulation) impacts AI development in the EU.

> The **General Data Protection Regulation (GDPR)** significantly impacts AI development in the EU by setting stringent rules for the processing of personal data. Since most AI models, especially machine learning systems, rely on large volumes of data (often including personal data), developers must comply with several core. GDPR principles

2. Ethical Principles Matching

Match the following principles to their definitions:

A) Justice: Fair distribution of AI benefits and risks.

B) Non-maleficence: Ensuring AI does not harm individuals or society.

C)Autonomy: Respecting users’ right to control their data and decisions.

D) Sustainability: Designing AI to be environmentally friendly.

Part 2: Case Study Analysis (40%)

Case 1: Biased Hiring Tool

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

- **Tasks**:

    1. Identify the source of bias (e.g., training data, model design).
        - The bias comes from its training data - because it was trained on resumes submitted over a 10-year period, most of which came from men.

    2. Propose three fixes to make the tool fairer.

        - Train the model with unbiased data
        - Identifying Bias: AI can help identify and mitigate human biases in decision-making processes.
        - Implementing transparency in the AI.**Explainable AI (XAI):** Techniques that provide insights into how AI models make decisions, such as feature importance in a decision tree.

    3. Suggest metrics to evaluate fairness post-correction

        - **Demographic Parity** : .Ensures the tool is not systematically favoring one group over another in its output. This is a key metric for ensuring a diverse candidate pool.
        - **Equality of Opportunity:**This is often fairer than Demographic Parity because it focuses on identifying qualified candidates. It ensures that the tool is equally good at finding qualified candidates from all groups.

Case 2: Facial Recognition in Policing

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

1. Discuss ethical risks (e.g., wrongful arrests, privacy violations).
    - **Wrongful accusation ,arrests and convictions** : This constitutes a fundamental breach of justice and due process. It undermines the principle of "innocent until proven guilty" and can lead to life-altering consequences—loss of liberty, job, reputation, and psychological trauma—for an innocent person.
    - **Erosion of Trust in Institutions and Technology:** This corrodes the social contract. Citizens cannot have faith in a system that uses a demonstrably biased tool, leading to broader societal cynicism and instability.

2. Recommend policies for responsible deployment.
    - Prohibit the use of facial recognition for live, mass surveillance of public spaces by government entities. This is the most effective way to prevent its use for tracking protests, monitoring religious gatherings, or general population surveillance.
    - Before any deployment, law enforcement agencies must engage in a transparent public consultation process and establish a civilian oversight board with the power to review and approve FRT use policies and audit outcomes.