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 is the systematic, repeatable errors in AI systems that create unfair outcomes by privileging certain groups over others due to biased data, design, or human prejudice. Examples:

- Facial recognition systems performing poorly on darker-skinned females due to lack of diverse training data.
- Hiring algorithms favoring one gender or racial group unfairly in recruitment decisions.

Q2. Explain the difference between *transparency* and *explainability* in AI. Why are both important?

- *Transparency* means AI systems reveal how they function and what data they use.
- *Explainability* is the ability to clearly explain the reasons behind AI's decisions or outputs.

Both are important to build trust, ensure accountability, and detect or mitigate bias or errors in AI systems.

Q3. How does GDPR (General Data Protection Regulation) impact AI development in the EU?

GDPR imposes strict data protection rules, requiring AI developers to ensure data privacy, obtain user consent, and avoid discriminatory practices. It mandates transparency, enables data subject rights (like access and correction), and enforces substantial fines for non-compliance, thereby shaping responsible AI development in the EU.

2. Ethical Principles Matching

Match the following principles to their definitions:

- **A) Justice**
- **B) Non-maleficence**
- **C) Autonomy**
- **D) Sustainability**
    1. ***Ensuring AI does not harm individuals or society.***
    2. ***Respecting users' right to control their data and decisions.***
    3. ***Designing AI to be environmentally friendly.***
    4. ***Fair distribution of AI benefits and risks.***

Principle	Definition
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).
    2. Propose three fixes to make the tool fairer.
    3. Suggest metrics to evaluate fairness post-correction.

Case 1: Biased Hiring Tool (Amazon)

Source of bias: The training data was predominantly male resumes from a decade, reflecting historical gender imbalance, causing the model to favor male candidates and penalize terms related to women.

Three fixes to make the tool fairer:

1. Use balanced and diverse training data that represents all genders equally.
2. Remove or neutralize gender-indicative terms and proxies during feature engineering.
3. Implement fairness-aware algorithms that explicitly reduce bias, such as adjusting model weights or using adversarial debiasing.

Metrics to evaluate fairness post-correction:

- Demographic parity (equal selection rates across genders).
- Equal opportunity (similar true positive rates for all groups).
- Disparate impact ratio (ratio of favorable outcomes across 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).
    2. Recommend policies for responsible deployment.

Case 2: Facial Recognition in Policing

Ethical risks:

- Higher false identification rates among minorities can lead to wrongful arrests and convictions.
- Violations of privacy and civil rights due to disproportionate surveillance.
- Erosion of public trust and potential systemic discrimination.

Recommended policies for responsible deployment:

1. Mandate rigorous bias testing before deployment, especially across demographic groups.
2. Limit use cases to clearly defined, high-accuracy scenarios with human oversight.
3. Require transparency, auditability, and regular independent impact assessments.
4. Implement strict regulations to protect privacy and prevent misuse or overreach.
