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 unfair discrimination that occurs when AI systems produce results that are prejudiced against certain groups or individuals due to flawed assumptions, incomplete data, or biased training processes.

Two examples:

1. **Hiring algorithms**: AI recruitment systems have shown bias against women and minorities by learning from historical hiring data that reflected past discriminatory practices, leading to the systematic rejection of qualified candidates from underrepresented groups.
2. **Criminal justice risk assessment**: Predictive policing and recidivism algorithms have demonstrated racial bias, disproportionately flagging Black defendants as high-risk for reoffending compared to white defendants with similar profiles, perpetuating systemic inequalities in the justice system.

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

Transparency refers to the openness and accessibility of information about an AI system's development, data sources, decision-making processes, and limitations. It involves making the system's workings visible and understandable to stakeholders.

Explainability refers to the ability to provide clear, understandable explanations for specific AI decisions or predictions, allowing users to comprehend why a particular outcome was reached.

Why both are important:

- **Transparency** builds trust by allowing stakeholders to understand how the system was built and operates generally

- **Explainability** enables accountability by providing specific reasoning for individual decisions
- Together, they enable users to make informed decisions about AI system usage, identify potential biases, and ensure responsible deployment in critical applications like healthcare and criminal justice

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

GDPR significantly impacts AI development through several key requirements:

Data Processing Principles: AI systems must comply with lawfulness, fairness, and purpose limitation, requiring explicit consent for data processing and restricting use to specified purposes.

Right to Explanation: Individuals have the right to understand automated decision-making processes that significantly affect them, requiring AI systems to provide meaningful explanations for their decisions.

Data Minimization: AI developers must collect and process only the minimum data necessary for their specific purpose, limiting extensive data collection practices.

Privacy by Design: AI systems must incorporate privacy considerations from the design phase, including data protection impact assessments for high-risk processing activities.

Data Subject Rights: Individuals have rights to access, rectify, erase, and port their data, requiring AI systems to be designed with these capabilities in mind.

2. Ethical Principles Matching

A) Justice → 4. Fair distribution of AI benefits and risks *Justice ensures equitable access to AI benefits and fair distribution of potential risks across different groups and populations.*

B) Non-maleficence → 1. Ensuring AI does not harm individuals or society *Non-maleficence follows the principle of "do no harm," requiring AI systems to be designed and deployed without causing harm to individuals or society.*

C) Autonomy → 2. Respecting users' right to control their data and decisions *Autonomy respects individual agency and the right to make informed decisions about one's own data and how AI systems affect them.*

D) Sustainability → 3. Designing AI to be environmentally friendly *Sustainability focuses on creating AI systems that minimize environmental impact through efficient resource use and energy consumption.*