Part 1: Theoretical Understanding (30%)

1. Short Answer Questions

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

Answer:

Algorithmic bias refers to systematic and unfair discrimination in the outcomes of AI systems, often caused by biased training data, model design, or deployment context. These biases can lead to unjust treatment of certain individuals or groups, reinforcing existing societal inequalities.

Examples:

1. **Facial Recognition Systems:** Studies have shown that some facial recognition systems have significantly higher error rates for people with darker skin tones, leading to misidentification or failed recognition.

2. **Hiring Algorithms:** An AI system trained on historical hiring data may favor male candidates over females if the past data reflects gender-biased hiring practices.

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

Answer:

- **Transparency** refers to the openness and clarity regarding how an AI system works—its design, goals, data sources, and development process.

- **Explainability** is the extent to which the internal mechanics of an AI model can be understood and interpreted by humans, especially regarding why it made a particular decision.

Importance:

- **Transparency** builds **trust** by allowing stakeholders to see how and why an AI system was built.

- **Explainability** is critical for **accountability**, especially in high-stakes domains like healthcare and finance, where users must understand and challenge decisions when necessary.

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

Answer:

GDPR enforces strict regulations on how personal data is collected, processed, and stored, directly impacting AI development. Its main effects include:

- **Data Consent:** AI developers must obtain explicit consent from users before collecting or using their personal data.

- **Right to Explanation:** Users can demand explanations of automated decisions that significantly affect them.

- **Data Minimization:** Developers must limit data collection to what is strictly necessary.

- **Penalties:** Non-compliance can lead to substantial fines, incentivizing responsible AI practices.

These regulations ensure **privacy**, **accountability**, and **fairness**, aligning AI systems with ethical and legal standards.

2. Ethical Principles Matching

Match the following principles to their definitions:

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.