AI Ethics Assignment - Theory Part

Part 1: Theoretical Understanding (30%)

Q1: Define algorithmic bias and provide two examples.

Algorithmic bias occurs when an AI system produces systematically unfair outcomes that favor certain groups over others due to biased data, design, or human assumptions.

Examples: 1. Hiring tools that penalize female applicants because historical data reflected male-dominated hiring patterns. 2. Facial recognition systems that misidentify people with darker skin tones more often due to imbalanced training datasets.

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

- Transparency: Openness about how an AI system operates, including data sources, algorithms used, and decision-making logic.
- Explainability: Making AI decisions understandable to humans, explaining why a specific output was produced.

Importance: Both are essential for trust, accountability, and ethical use. Transparency ensures oversight and compliance, while explainability helps detect unfair or incorrect outcomes.

Q3: How does GDPR impact AI development in the EU?

GDPR enforces strict data protection rules, requiring: - User consent for data collection and processing - Right to explanation for automated decisions - Data minimization and privacy by design

This pushes developers to design privacy-conscious, transparent, and accountable AI systems.

Ethical Principles Matching:

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.