Part 1: Theoretical Understanding

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

Definition: Algorithmic bias is when an AI system produces results that are systematically prejudiced due to biases in data, assumptions, or model design.

Examples:

1. **Hiring Tools:** AI recruiting systems trained on biased historical data might favor male candidates over female candidates.
2. **Facial Recognition:** Systems trained on datasets lacking diversity misidentify people with darker skin tones more frequently.

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

- **Transparency:** The openness in revealing how an AI system is built, including data sources, models, and algorithms.
- **Explainability:** The ability of the AI system to provide understandable reasons for its decisions.

Importance: Both ensure trust, accountability, and allow users and regulators to understand, challenge, or improve AI outputs.

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

- **Data Minimization:** AI models can only use necessary data.
- **Consent:** Explicit permission is needed from users.
- **Right to Explanation:** Users can request explanations for automated decisions.
- **Penalties:** Non-compliance leads to heavy fines, enforcing ethical standards.

Ethical Principles Matching

Principle	Definition

Justice	Fair distribution of AI benefits and risks.
Non-maleficence	Ensuring AI does not harm individuals/society.
Autonomy	Respecting users' right to control data/decisions.
Sustainability	Designing AI to be environmentally friendly.