Part 1: Theoretical Understanding

Q1: Define algorithmic bias and provide two examples of how it manifests in AI systems.
Algorithmic bias refers to systematic and unfair discrimination in AI outcomes caused by flawed data, design, or decision-making processes. It occurs when AI models produce results that disadvantage certain individuals or groups.

- **Example 1:** A hiring algorithm trained on past employee data favors male candidates over female candidates, replicating historical gender imbalances.
- **Example 2:** A facial recognition system misidentifies people of colour more frequently than white individuals, leading to disproportionate risks of wrongful arrests.

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

- **Transparency** means making the inner workings of an AI system visible and understandable to stakeholders (e.g., what data is used, what model is applied).
- **Explainability** refers to the ability to provide clear, interpretable reasons for specific AI outputs or decisions.

Importance:

- Transparency builds trust and accountability by revealing how decisions are made.
- Explainability ensures users and regulators can understand, challenge, or contest decisions—especially in high-stakes contexts like healthcare, finance, and justice.

Q3: How does GDPR (General Data Protection Regulation) impact AI development in the EU?
GDPR affects AI development by enforcing **data protection, fairness, and accountability**. Key impacts include:

- **Right to explanation:** Users can request explanations of AI-driven decisions.
- **Data minimization & consent:** AI developers must collect only necessary data with explicit consent.
- **Accountability:** Organizations must demonstrate lawful processing and bias mitigation.
This ensures AI systems respect privacy, fairness, and human rights.

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.