Part 1: Theoretical Understanding :

1. Short Answer Questions

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

Definition:
Algorithmic bias refers to systematic and unfair discrimination that occurs when an AI system produces results that are prejudiced due to flawed assumptions, biased training data, or underlying societal inequalities embedded in the algorithm.

Examples:

1. **Facial Recognition Bias:** AI systems performing worse on darker-skinned individuals due to underrepresentation in training data.

2. **Hiring Algorithm Bias:** Recruitment algorithms that favor male candidates over female ones because they were trained on historical hiring data from male-dominated industries.

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

- **Transparency** refers to how open and understandable the **design, structure, and operations** of an AI system are to stakeholders. It includes making information about the data, algorithms, and processes available.

- **Explainability** focuses on how easily humans can **interpret and understand** the **outputs or decisions** made by an AI system, especially complex models like deep learning.

Why they are important:

- **Transparency** builds trust and allows for auditing and accountability.

- **Explainability** ensures that users and impacted individuals can understand, contest, or appeal AI decisions—critical in sectors like healthcare, finance, or law.

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

GDPR impacts AI development by enforcing strict requirements around:

- **User consent** before personal data collection or processing.

- **Data minimization**, ensuring only necessary data is used.

- **Right to explanation**, giving individuals the right to know how automated decisions affecting them are made.

- **Accountability**, requiring developers to justify AI decision-making processes and ensure fairness and transparency.

This compels AI developers to adopt **privacy-preserving practices**, ensure **fairness**, and **document model behavior**, especially when personal data is involved.

---

2. Ethical Principles Matching

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.