Theoretical Understanding

1. Short Answer Questions

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

Algorithmic bias refers to systematic and unfair discrimination in the outcomes of an AI system, often caused by biased training data, flawed model design, or lack of diversity in development processes.

Examples:

1. **Hiring Algorithms**: An AI recruiting tool trained on past hiring data may favor male candidates if historical data reflected gender bias in hiring decisions.
2. **Facial Recognition**: Systems trained primarily on lighter-skinned faces tend to misidentify darker-skinned individuals, leading to higher false positives for minority groups.

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

- **Transparency** is the degree to which the internal workings, data sources, and decision processes of an AI system are open and understandable to stakeholders (e.g., developers, regulators).
- **Explainability** is the extent to which users (e.g., patients, consumers) can understand why an AI made a specific decision or prediction.

Importance**:**

- Transparency builds trust and allows for accountability, especially for high-impact systems (e.g., in healthcare or finance).
- Explainability helps users challenge or accept AI decisions, debug errors**,** and ensure fairness**.**

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

The **GDPR** shapes AI development by enforcing user rights and data protections:

- Users have a right to explanation for automated decisions.
- AI developers must implement privacy by design and obtain explicit consent for data use.
- It restricts the use of personal data for profiling unless necessary and lawful.
- Non-compliance can result in heavy fines**,** pushing companies toward more ethical, transparent AI systems.

2. Ethical Principles Matching

Match the following principles to their definitions:

- **A) Justice**
- **B) Non-maleficence**
- **C) Autonomy**
- **D) Sustainability**
  1. *Ensuring AI does not harm individuals or society.*
  2. *Respecting users' right to control their data and decisions.*
  3. *Designing AI to be environmentally friendly.*
  4. *Fair distribution of AI benefits and risks.*

Principle	Definition
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.
A) Justice	Fair distribution of AI benefits and risks.