

Part 1: Theoretical Understanding: Designing Responsible and Fair AI Systems

Define algorithmic bias and provide two examples of how it manifests in AI systems.	2
Explain the difference between transparency and explainability in AI. Why are both important?	2
How does GDPR (General Data Protection Regulation) impact AI development in the EU?.....	2
Ethical Principles Matching	3

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

Algorithmic bias occurs when AI systems produce unfair or prejudiced outcomes due to biased data, design, or deployment, disproportionately affecting specific groups.

Example 1: Predictive policing tools trained on historical arrest data may over-predict crime in minority neighborhoods, leading to disproportionate surveillance.

Example 2: AI resume screening systems trained on male-dominated hiring data may penalize female candidates by undervaluing their qualifications.

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

Transparency: Openness about an AI system's data sources, algorithms, and decision-making processes (e.g., sharing dataset details).

Explainability: The ability to describe how an AI reaches a decision in human-understandable terms (e.g., explaining why a loan application was rejected).

Importance: Transparency enables external scrutiny and trust, ensuring stakeholders can verify ethical practices.

Explainability empowers users to understand and contest decisions, promoting accountability and fairness.

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

GDPR shapes AI development by:

Requiring explicit user consent for data processing, ensuring ethical data collection.

Enforcing data minimization, limiting AI systems to use only necessary data.

Granting users rights to access, rectify, or delete their data, necessitating transparent and user-centric AI designs.

Imposing penalties for non-compliance, incentivizing bias audits and explainable models to align with fairness and accountability.

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.