

Part 1: Theoretical Understanding

Q1: Define algorithmic bias and provide two examples.

Definition:

Algorithmic bias refers to systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one group over another.

Examples:

- **Hiring Algorithms:** An AI trained on historical hiring data might favor male candidates if the data reflects past gender discrimination.
 - **Credit Scoring Models:** A model might give lower credit scores to minorities due to biased training data reflecting historical economic disparities.
-

Q2: Difference between transparency and explainability. Why are both important?

- **Transparency:** Refers to how open and accessible the design, data, and decision-making processes of an AI system are to stakeholders.
- **Explainability:** Refers to the ability to understand how and why an AI system reached a particular decision or prediction.

Importance:

- Transparency builds trust and accountability.
 - Explainability helps users and regulators understand decisions, especially in high-stakes areas like healthcare or justice.
-

Q3: How does GDPR impact AI development in the EU?

- **Right to Explanation:** Users can demand explanations for automated decisions.
- **Consent Requirements:** AI systems must have explicit user consent for personal data processing.
- **Data Minimization & Fairness:** Developers must ensure data is relevant and bias-free.

Impact: GDPR enforces transparency, accountability, and user rights, shaping ethical AI development in the EU.

Ethical Principles Matching:

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

Part 2: Case Study Analysis

Case 2: Facial Recognition in Policing

1. Ethical Risks:

- Wrongful arrests due to misidentification (esp. minorities).
- Erosion of privacy and consent.
- Disproportionate surveillance and systemic discrimination.

2. Policy Recommendations:

- Mandatory **third-party bias audits** before deployment.
- **Consent-based surveillance** in public spaces.
- Transparent **accountability framework** and **error disclosures**.
- Ban in high-risk settings unless accuracy across demographics is proven.