AI Ethics Assignment

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

Algorithmic bias refers to systematic and repeatable errors in an AI system that lead to unfair outcomes, often disadvantaging certain groups of people. This bias can arise from biased data, flawed algorithms, or the way the system is used.

Examples:

1. Facial Recognition Bias: Some facial recognition systems have been shown to have higher error rates when identifying people of color, particularly Black individuals, compared to white individuals. This occurs due to underrepresentation of diverse faces in training datasets.
2. Hiring Algorithms: AI systems used for screening job applicants may favor male candidates over female ones if they are trained on historical hiring data that reflects gender bias, such as preferring resumes with male-dominated job titles or past hiring decisions.

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

Transparency refers to how open and accessible information about an AI system is, including its design, data sources, and decision-making processes.

Explainability is the extent to which the internal mechanics or decision-making of an AI system can be understood and interpreted by humans.

Why they are important:

- Transparency builds trust by allowing stakeholders to understand how the system operates and what data it uses.
- Explainability helps users and regulators understand the reasons behind specific decisions, which is critical for accountability, especially in high-stakes areas like healthcare, finance, and law enforcement.

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

GDPR significantly shapes AI development in the EU by:

- Requiring data minimization and purpose limitation, meaning AI systems must only use personal data that is necessary and for specific, lawful purposes.
- Granting individuals rights such as the right to access, the right to be forgotten, and the right to explanation, which influence how AI models are trained and deployed.
- Mandating data protection by design and by default, ensuring AI developers incorporate privacy safeguards from the outset.

These regulations promote responsible AI use but can also limit the use of large-scale personal data for training AI models.

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.