ERICK WAMBUGU
POWER LEARN PROJECT AFRICA
AI FOR SOFTWARE ENGINEERING
WEEK 7 – ETHICAL AI

Part 1: Theoretical Understanding (30%)

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

- Algorithmic bias refers to systematic and unfair discrimination produced by an AI system due to biased data, flawed model design, or inequitable deployment practices. It leads to outcomes that disproportionately disadvantage certain groups.

  Examples:
  - Facial Recognition Bias: Systems misidentify darker-skinned individuals at a much higher rate because training datasets contain mostly lighter-skinned faces.
  - Biased Loan Approval Models: Credit scoring algorithms deny loans to certain ethnic or socioeconomic groups because historical financial data reflects existing societal inequalities.

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

- Transparency refers to how openly an AI system reveals its internal workings—such as data sources, model architecture, training processes, and decision rules.
- Explain-ability is the ability to clearly justify or interpret an AI system's outputs in a human-understandable way.

  Importance:
  - Transparency builds trust, enables accountability, and allows stakeholders to audit the system.
  - Explain-ability ensures users, developers, and regulators can understand why a model made a specific decision—critical for fairness, error correction, and ethical compliance.

Together, they ensure AI systems are trustworthy, audit-able, and aligned with human values.

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

- GDPR influences AI development by enforcing strict rules on how personal data is collected, stored, and used. Key impacts include:
  - ✔ Data Minimization: Developers must collect only data that is necessary for the AI task.
  - ✔ Right to Explanation: Users can challenge automated decisions and request meaningful information about how the decision was made.
  - ✔ Consent & Lawful Basis: AI models must use data obtained through explicit, informed user consent.
  - ✔ Accountability: Organizations must document model processes, safeguards, and risk assessments.
  - ✔ Bias and Fairness Requirements: GDPR prohibits discriminatory automated decision-making.

ERICK WAMBUGU
POWER LEARN PROJECT AFRICA
AI FOR SOFTWARE ENGINEERING
WEEK 7 – ETHICAL AI

Match the principles to their definitions:

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.