AI Ethics Assignment – Person A

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 repeatable errors in AI systems that lead to unfair outcomes, such as privileging one group over another. It often stems from biased training data, model assumptions, or deployment practices.

Example 1: A hiring algorithm that favors male candidates because it was trained on historical data dominated by male hires.
Example 2: A facial recognition system that performs poorly on darker-skinned individuals because the training dataset lacked diversity.

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

Transparency in AI refers to the openness of the system's design, algorithms, and data sources. It allows stakeholders to understand how decisions are made. Explainability, on the other hand, is the ability of the system to provide understandable and interpretable reasons for its outputs or decisions.

Both are important because transparency builds trust with users and stakeholders, while explainability ensures accountability and enables human oversight.

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

The GDPR affects AI development by enforcing strict guidelines on personal data use. Key impacts include:
- Requiring informed user consent for data collection.
- Granting users the right to access, correct, and delete their data.
- Mandating data protection by design and default.
- Enabling the right to explanation for automated decisions.

This compels AI developers to prioritize privacy, fairness, and transparency in their systems.

Part 2: Ethical Principles Matching

Match the following principles to their definitions:

A) Justice

B) Non-maleficence

C) Autonomy

D) Sustainability

Matching:

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.