AI Ethics Assignment - Part 1 & 2

Part 1: Theoretical Understanding

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

Algorithmic bias occurs when an AI system produces systematically unfair outcomes for certain groups, often due to biased training data or flawed assumptions.

Examples:

1. Facial recognition systems often misidentify Black individuals more frequently than White individuals.

2. Credit scoring models may give lower scores to individuals from low-income zip codes due to historical economic disparities.

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

- Transparency refers to the openness of the AI system's design and data sources.
- Explainability refers to the ability to understand how and why an AI system made a decision.
- Importance: Transparency fosters accountability, while explainability builds trust and helps in debugging or ethical evaluations.

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

- It mandates data minimization, informed consent, and the right to explanation for automated decisions.
- AI systems must respect user privacy, provide justification for decisions, and be auditable for fairness and accountability.

Q4: 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.

Part 2: Case Study Analysis

Case 1: Biased Hiring Tool (Amazon)

- Source of Bias: The training data was based on historical resumes, mostly from male candidates, reinforcing gender stereotypes.

AI Ethics Assignment - Part 1 & 2

- Fixes:
  1. Rebalance training data to include resumes from all genders.
  2. Remove gender-indicative features from input.
  3. Use fairness-aware algorithms during model training.
- Fairness Metrics:
  - Disparate impact ratio
  - Equal opportunity difference
  - Precision and recall by gender

Case 2: Facial Recognition in Policing

- Ethical Risks:
  - Higher misidentification rates for minority groups leading to wrongful arrests.
  - Violation of privacy through mass surveillance.
  - Racial profiling and erosion of public trust.
- Responsible Deployment Policies:
  - Independent audits to ensure accuracy across demographics.
  - Mandatory disclosure and public consent.
  - Use only in high-stakes scenarios with human oversight.