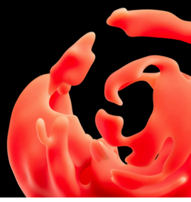
AI and Ethics

Short Answer Questions and Principles



Short Answer Questions

(a) Define algorithmic bias and provide two examples of how it manifests in AI systems.

Algorithmic bias refers to systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others. This bias can arise from several sources, including biased training data, flawed algorithms, or biased human input.

Examples:

1. **Facial Recognition Software:** Many facial recognition systems have been shown to exhibit bias based on race and gender, with higher error rates for individuals with darker skin tones and for women. This can lead to misidentification and unfair treatment in law enforcement and security contexts.
2. **Recruitment Algorithms:** AI-powered recruitment tools may perpetuate existing societal biases if they are trained on historical hiring data that reflects past discrimination. For example, an algorithm trained primarily on male resumes might unfairly penalize female applicants, even if they are equally qualified.

(b) Explain the difference between transparency and explainability in AI, and why both are important.

- **Transparency** in AI refers to the degree to which one can understand the inner workings of an AI system, including its architecture, data sources, and algorithms. A transparent AI system is like a glass box, where the process from input to output is visible.
- **Explainability**, on the other hand, refers to the ability to provide clear and understandable reasons for specific AI decisions or predictions. An explainable AI system not only makes a decision but also offers insights into *why* that decision was made.

Both transparency and explainability are crucial for fostering trust, accountability, and ethical AI development. Transparency helps to identify potential biases and vulnerabilities in the system, while explainability allows users to understand and challenge AI decisions, ensuring fairness and preventing unintended consequences.

(c) How does GDPR impact AI development in the EU?

The General Data Protection Regulation (GDPR) significantly impacts AI development in the EU by imposing strict rules on the processing of personal data. Key impacts include:

- **Data Minimization:** GDPR requires AI developers to collect and process only the data necessary for a specific purpose, limiting the potential for misuse of personal information.
- **Right to Explanation:** Although debated, GDPR is often interpreted as providing individuals with a right to an explanation for decisions made by automated systems, particularly those with significant legal or similar effects.
- **Increased Transparency:** GDPR mandates transparency about data processing activities, requiring AI developers to inform individuals about how their data is being used and for what purposes.
- **Data Security:** AI developers must implement appropriate technical and organizational measures to ensure the security of personal data, protecting it from unauthorized access, loss, or destruction.
- **Privacy by Design:** GDPR promotes a 'privacy by design' approach, requiring AI developers to consider privacy implications from the outset of a project and to integrate privacy safeguards into the design of AI systems.

Ethical Principles Matching

Match the ethical principle to its definition:

A) Justice

B) Non-maleficence

C) Autonomy

D) Sustainability

Ensuring equitable distribution of resources, opportunities, and outcomes, preventing discrimination, and promoting fairness.

Avoiding causing harm, whether physical, psychological, or societal, and minimizing potential risks associated with AI systems.

Respecting individuals' rights to make their own decisions, providing them with adequate information, and empowering them to control their interactions with AI systems.

Developing AI systems in a way that considers long-term environmental, social, and economic impacts, ensuring they contribute to a sustainable future.

Answers:

A) Justice - Ensuring equitable distribution of resources, opportunities, and outcomes, preventing discrimination, and promoting fairness.

B) Non-maleficence - Avoiding causing harm, whether physical, psychological, or societal, and minimizing potential risks associated with AI systems.

C) Autonomy - Respecting individuals' rights to make their own decisions, providing them with adequate information, and empowering them to control their interactions with AI systems.

D) Sustainability - Developing AI systems in a way that considers long-term environmental, social, and economic impacts, ensuring they contribute to a sustainable future.

Summary

This document explored key ethical considerations surrounding AI development and deployment. It addressed algorithmic bias, the importance of transparency and explainability, and the impact of GDPR on AI in the EU. Additionally, it matched fundamental ethical principles to their definitions, providing a concise overview of ethical guidelines for AI practitioners.