Nontando Myoli

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

Answer:

Algorithmic bias is when an AI system makes unfair decisions because it was trained on biased or unbalanced data. This can lead to discrimination or unequal treatment of certain groups.

Examples:

1. A hiring AI that Favors male applicants because it learned from past hiring data that mostly included men.
2. Facial recognition systems that misidentify people of colour more often due to lack of diversity in the training images.

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

Answer:

- **Transparency** is about how open and clear an AI system is — like knowing how it was built and what data it uses.
- **Explainability** is about understanding *why* the AI made a specific decision.

Why both matter:
Transparency builds trust and allows for accountability. Explainability helps people understand AI decisions — especially in sensitive areas like healthcare or justice.

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

Answer:

GDPR protects personal data in the EU. It affects AI by:

- Requiring clear consent before using personal data
- Giving people the right to know if decisions were made by AI
- Limiting how automated decisions can affect individuals
- Pushing developers to build more ethical and privacy-focused systems

---

Nontando Myoli

Ethical Principles Matching

Match the principle to the correct definition:

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

Part 2: Case Study Analysis

Source of Bias:

- Biased Training Data:
  The AI learned from past resumes submitted over 10 years — most of which were from men, reflecting gender bias in the tech industry.
- Design Flaw:
  The model picked up gendered patterns (like penalizing resumes that included "women's" clubs or schools).

Three Fixes to Make It Fairer:

1. Use balanced and inclusive training data
   Include resumes from diverse gender, race, and background groups to reduce historical bias.
2. Remove sensitive attributes
   Strip out direct and indirect gender indicators (like names, pronouns, clubs).
3. Add bias monitoring tools
   Regularly test the model for gender bias and retrain when patterns are detected.

Fairness Evaluation Metrics**:**

- **Demographic Parity:** Are all groups getting equal opportunity?
- **Equal Opportunity:** Are qualified individuals from all groups treated fairly?
- **False Positive/Negative Rate Balance:** Are errors spread evenly across demographics?

Nontando Myoli

Case 2: Facial Recognition in Policing

Ethical Risks:

- Wrongful Arrests: Misidentification can lead to innocent people being detained.
- Privacy Invasion: Constant surveillance affects people's sense of freedom.
- Discrimination: Reinforces racial injustice if used without bias controls.

Recommended Policies**:**

1. **Strict usage guidelines:**
   Only use facial recognition with clear, legal justification and human oversight.
2. **Bias audits and transparency:**
   Test systems for accuracy across races and make results public.
3. **Limited deployment in sensitive areas:**
   Avoid using facial recognition in schools, protests, or public spaces where people can't opt out.