**Q1: Define algorithmic bias and give two examples**

**Definition:**
Algorithmic bias is a systematic error in an AI system that produces unfair outcomes for certain individuals or groups, often defined by race, gender, age, or other sensitive attributes. Bias can occur due to historical data, model design, or deployment context, resulting in advantages or disadvantages for specific populations.

**Examples:**

1. **Hiring Tools:** An AI resume scanner trained on historical male-dominated hiring data may rank female candidates lower, even if equally qualified.

2. **Facial Recognition:** A system trained mostly on light-skinned faces may misidentify darker-skinned faces, leading to unequal treatment or wrongful arrests.

**Sources of Bias (Optional Notes):**

- Training-data bias
- Measurement bias
- Sampling bias
- Modeling choices
- Deployment & feedback loops

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

**Transparency:**
Transparency means that the AI system's inner workings, data sources, and decision-making process are open and accessible to stakeholders. A transparent AI allows auditors, regulators, and users to see what data and logic the system uses.

**Explainability:**
Explainability is the ability of the AI system to provide understandable reasons for its specific outputs or decisions. It translates complex model computations into human-understandable terms.

**Why both are important:**

- **Accountability:** Helps organizations take responsibility for AI outcomes.
- **Trust:** Users are more likely to accept AI decisions if they can understand and verify them.
- **Bias Detection:** Transparency and explainability allow identification of errors or unfair treatment before harm occurs.
- **Regulatory Compliance:** Many laws (like GDPR) require both transparency and explainability for automated decisions.
-
-

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

**Answer:**
The GDPR regulates how personal data is collected, stored, processed, and shared. For AI systems, this has several impacts:

1. **Data Minimization:** AI models must only use data that is necessary for their purpose, avoiding excessive collection.
2. **Consent Requirements:** Users must provide informed consent for their data to be used in AI training or analysis.
3. **Right to Explanation:** Users have the right to understand automated decisions that affect them, requiring explainable AI models.
4. **Data Privacy and Protection:** Sensitive data must be protected using encryption, anonymization, or pseudonymization.
5. **Accountability:** Organizations deploying AI are legally accountable for how their models use personal data and the fairness of outcomes.

**Impact on AI Development:**

- Encourages ethical data handling.

- Drives development of privacy-preserving AI techniques.
- Forces organizations to implement explainable and auditable AI systems.

**Ethical Principles Matching**

| **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. |