# Part 1: Theoretical Understanding

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

**Answer:**
Algorithmic bias is when an AI system produces unfair, prejudiced, or unequal results because of biased data or design choices.

**Example 1:** A facial recognition system that performs poorly on darker-skinned people because it was mostly trained on lighter-skinned faces.

**Example 2:** A hiring algorithm that favors male candidates over equally qualified female candidates because it was trained on past hiring data that reflected gender bias.

---

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

**Answer:**
**Transparency** means being open about how an AI system is designed, trained, and operates. It includes sharing details like data sources, algorithms used, and decision-making processes.

**Explainability** refers to making the AI's individual decisions understandable—so people can see *why* a particular result was given.

**Importance:**
Both are important because they help build trust, allow users to understand and challenge decisions, and support accountability. Without transparency and explainability, AI systems can act as "black boxes," making it hard to detect bias or errors.

---

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

**Answer:**
GDPR requires that personal data is collected and processed lawfully, fairly, and transparently. For AI development in the EU, this means:

- Getting **user consent** before collecting personal data.
- Allowing users to access, correct, or delete their data.
- Ensuring **data minimization** (only necessary data is collected).
- Providing **explanations** for automated decisions that affect people.
  These rules ensure AI systems respect privacy and protect users' rights.

---

# ✅ 2. Ethical Principles Matching

✅ **Match the principles to their definitions:**

- **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.