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

Algorithmic bias refers to systematic and unfair discrimination produced by an AI system due to flawed data, model design, or assumptions. It results in outcomes that favor or disadvantage certain groups.

Examples:

1. **Recruitment AI rejecting women more often:**
   If a hiring algorithm is trained on historical data dominated by male applicants, it may learn to rank men higher and automatically downgrade female candidates.

2. **Facial recognition performing poorly on darker skin tones:**
   Many facial recognition systems have been shown to misidentify Black individuals at much higher rates because the training datasets contain mostly lighter-skinned faces.

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

- **Transparency** refers to openness about how an AI system is built—its data sources, algorithms, training process, and decision-making workflow. It answers the question: *"What's inside the AI?"*

- **Explainability** refers to the ability to clearly interpret and understand *why* an AI made a specific decision. It answers: *"Why did the AI give this result?"*

Why both matter:

- They build **trust** in AI systems.

- They help identify **errors, biases, and unfair outcomes**.

- They are essential for **accountability**, especially in high-risk areas like healthcare, finance, and criminal justice.

- They support compliance with laws and policies requiring justification for automated decisions.

Q3: How does GDPR impact AI development in the EU?

GDPR affects AI development in several major ways:

1. **Requires user consent for data collection and processing.**
   AI systems must justify why they need personal data and obtain clear permission.

2. **Limits the use of personal data.**
   Data must be necessary, relevant, and minimized—no collecting data “just in case.”

3. **Provides the “right to explanation.”**
   Individuals can request an explanation for automated decisions affecting them, pushing developers to build explainable AI.

4. **Mandates data protection and privacy-by-design.**
   AI developers must integrate privacy and security measures from the beginning of the project.

5. **Enforces strict penalties for misuse of personal data.**
   Non-compliance can lead to heavy fines, encouraging responsible and ethical AI practices.

Ethical Risks of Misidentification in Facial Recognition Systems

When a facial recognition system misidentifies minorities at higher rates, several serious ethical risks arise:

1. Wrongful Arrests and Legal Injustice

Biased recognition can lead to innocent individuals—often from minority communities—being mistakenly flagged as suspects. This increases the risk of wrongful arrests, unjust detentions, and long-term harm such as criminal records or trauma.

2. Discrimination and Inequality

If errors disproportionately affect specific racial or ethnic groups, the system reinforces existing social inequalities. This can further marginalize communities already facing discrimination.

3. Privacy Violations

Facial recognition often operates without explicit consent, raising concerns about mass surveillance. Minorities living in heavily monitored areas may experience disproportionate intrusion into their personal privacy.

4. Loss of Public Trust

Repeated misidentification can undermine trust in technology, law enforcement, and public institutions. Communities may feel targeted and less willing to cooperate with authorities.

5. Lack of Accountability

Opaque algorithms make it difficult to determine who is responsible for harms—developers, vendors, or law enforcement—creating loopholes in accountability.

Recommended Policies for Responsible Deployment

1. Mandatory Bias Testing and Audits

Governments should require regular third-party audits to check performance across demographic groups. Systems failing fairness benchmarks should not be deployed.

2. Transparency Requirements

Agencies must disclose how the technology works, what data was used for training, and known error rates. Public oversight helps prevent abuse.

3. Human-in-the-Loop Decision-Making

Facial recognition results should **never** be used as the sole basis for arrest or identification. A trained human must verify and validate all matches.

4. Strict Data Protection and Consent Policies

Adopt privacy-by-design principles:

- Limit data collection
- Require consent where possible
- Prohibit storage of facial images without legal justification

5. Restrict High-Risk Use Cases

Ban or strictly regulate facial recognition in:

- Public mass surveillance
- Schools
- Protest environments
  Deployment should only occur in low-risk, controlled settings.

6. Accountability and Redress Mechanisms

Create clear procedures for:

- Reporting and investigating harms
- Compensating individuals misidentified
- Holding developers and institutions liable for misuse

7. Diverse and Representative Training Data

To reduce bias, training datasets must include balanced demographic representation, particularly of minority groups.

If you'd like, I can convert this into a shorter exam answer, a presentation slide, or a paragraph-style summary.