Ethical AI Assignment – Structured Breakdown

■ Part 1: Short Answer Questions

Q1: Define Algorithmic Bias + Examples
Definition: Algorithmic bias occurs when AI systems produce unfair or prejudiced outcomes due to biased data, flawed model design, or systemic inequalities embedded in training inputs.
Examples:
- Hiring Bias: Amazon's AI recruiting tool penalized resumes with terms like "women's" or graduates from women's colleges.
- Healthcare Disparities: A US healthcare algorithm underestimated Black patients' needs by using cost as a proxy for medical need.

Q2: Transparency vs Explainability
Transparency: Openness about how an AI system is built (data sources, algorithms, decision-making processes).
Explainability: Making individual AI decisions understandable to users.
Why Both Matter:
- Transparency builds trust and accountability.
- Explainability allows users to challenge or verify decisions in high-stakes domains.

Q3: GDPR's Impact on AI in the EU
- Data Protection: Compliance with data minimization, purpose limitation, lawful processing.
- Rights of Individuals: Access, correction, and erasure of personal data.
- Risk Assessments: Conduct DPIAs for high-risk AI systems.

■■ Part 2: 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

■ Part 2: Case Study Analysis (40%)

Case 1: Amazon's Biased Hiring Tool
Source of Bias: Biased training data, penalized female terms, reinforcement of historical patterns.
Three Fixes:
- Debias Training Data
- Fairness-Aware Algorithms
- Human Oversight
Fairness Metrics: Disparate Impact Ratio, Equal Opportunity Difference, False Positive Rate by gender.

Case 2: Facial Recognition in Policing
Ethical Risks: Wrongful arrests, privacy violations, disproportionate targeting of minorities.
Recommended Policies:
- Mandatory Bias Audits
- Human-in-the-loop Verification
- Transparency Reports
- Community Oversight Boards

■ Part 3: Practical Audit (25%)

COMPAS Bias Audit with AI Fairness 360
Steps: Load dataset, encode race, compute FPR disparity, Disparate Impact, Mean Difference.
Visualization Ideas: Bar chart of FPR by race, heatmap of risk scores.
Report Summary (300 words): Audit revealed racial bias in predictions. Black defendants were nearly twice as likely to be mislabeled high-risk. Disparate Impact = 0.76, Mean Difference = -0.19. After applying reweighing and fairness classifiers, metrics improved (DI=0.92). Recommendations: ongoing audits, diverse datasets, transparent deployment.

■ Part 4: Ethical Reflection (5%)

In my upcoming AgriDrone project, I'll ensure transparency by documenting how drone data is collected and processed. To maintain fairness, I'll test across diverse farming regions. I'll prioritize autonomy by giving farmers control of their data and will conduct regular bias audits with stakeholder feedback.

■ Bonus Task: Ethical AI in Healthcare (1-Page Guideline)

Title: Ethical AI Use in Healthcare – Guiding Principles
1. Patient Consent Protocols: Informed consent, clear AI role explanation, opt-out options.
2. Bias Mitigation Strategies: Diverse datasets, fairness-aware algorithms, interdisciplinary validation.
3. Transparency Requirements: Disclose system limits, provide explainable outputs, publish performance metrics.
Closing Note: Ethical AI must prioritize patient safety, equity, and trust.