Ethics in AI: Case Studies and Bias Analysis

Group Members:

- Member 1 – Theoretical Understanding

- Member 2 – Case Study: Biased Hiring Tool

- Member 3 – Case Study: Facial Recognition in Policing

- Member 4 – Dataset Bias Audit (COMPAS)

- Member 5 – Ethical Reflection

Table of Contents

1. Part 1 – Theoretical Understanding (Member 1)

Algorithmic Bias
Algorithmic bias refers to systematic and repeatable errors in a computer system that create unfair outcomes, such as favouring one group over others. These biases can come from biased training data, flawed assumptions in the model, or imbalanced design choices.

Real-World Examples
Example 1: Facial Recognition Bias – Some facial recognition systems perform poorly on people with darker skin tones. A 2018 MIT study found that error rates for identifying Black women were as high as 34%, while white men had an error rate of less than 1%.
Example 2: Credit Scoring Systems – Credit scoring algorithms have shown bias against minority communities, offering worse loan terms or denying credit even with similar profiles.

Transparency vs Explainability
Transparency is being open about how the system was built - including the data, model, and logic.
Explainability refers to how well humans can understand why an AI made a particular decision.

Why Both Matter:
- Transparency builds trust.
- Explainability ensures accountability.
- Together, they make AI more ethical and fair.

GDPR and AI in the EU
GDPR enforces strict data rules in the EU:
- Limits data collection to user consent.
- Requires explanation for automated decisions.
- Promotes privacy by design and data minimization.

Ethical Principles Matching

Principle	**Description**	**Example**
Justice	Fair treatment of all individuals	Facial recognition fairness
Non-maleficence	Avoiding harm through safe AI design	Testing AI in healthcare
Autonomy	Respecting user control and informed consent	Data control by users
Sustainability	Supporting long-term social/environmental well-being	Energy-efficient AI systems

2. Part 2 – Case Study: Biased Hiring Tool (Member 2)

Case Overview: Amazon's AI Recruiting System
Amazon's AI recruiting tool showed bias against women because it was trained on resumes submitted over a 10-year period, most of which came from men.

Proposed Fixes
- Balance training data to include more female applicants.
- Remove gender-indicative features like names or school clubs.
- Use fairness-aware algorithms during training.

Fairness Metrics Post-Correction
Disparate Impact Ratio
Equal Opportunity Difference
Demographic Parity

3. Part 2 – Case Study: Facial Recognition in Policing (Member 3)

Ethical Risks

- Wrongful arrests due to misidentification.
- Discrimination against minority groups.
- Loss of public trust in law enforcement.

Responsible Use Policies

- Require human oversight before acting on AI decisions.
- Set high accuracy thresholds before deployment.
- Maintain detailed audit logs and conduct regular bias audits.

4. Part 3 – Dataset Bias Audit: COMPAS (Member 4)

Executive Summary

Analysis of the COMPAS recidivism algorithm revealed racial disparities in false positive and false negative rates.

Key Findings

- African-American defendants: 45% false positives vs. 23% for Caucasians.
- False negatives for Caucasians: 48% vs. 28% for African-Americans.
- Statistical Parity Difference: -0.174
- Disparate Impact: 0.637 (below the 0.8 threshold)

Remediation Recommendations

- Regular bias audits.
- Demographic-specific threshold adjustments.
- In-processing techniques like adversarial debiasing.

5. Part 4 – Ethical Reflection (Member 5)

Ethical Reflection on AI Practices
Working on the COMPAS audit and related case studies made me reflect on ethical AI development.

Data Fairness: COMPAS disproportionately labeled African-American defendants as high-risk. I will prioritize fairness checks and balanced datasets.

Transparency: Many systems are black boxes. I'll advocate for documentation and explainability tools.

Responsible Use: AI should never make high-stakes decisions without human oversight. I'll support audit trails and stakeholder reviews.

User Privacy: GDPR taught me the importance of consent and data minimization. I'll use privacy-first design in future work.

Ethics is not a final step — it's a mindset from day one.

6. Appendices

Appendix A: COMPAS Analysis Code Summary
Appendix B: Extended Visualizations
Appendix C: Full Metric Tables

Attached File: ..\Desktop\Week-7-Assignment\AuditedForBias.ipynb