SRINIVAS UNIVERSITY
INSTITUTE OF ENGINEERING & TECHNOLOGY
Srinivas Campus,Mukka,Surathkal,Mangaluru-574 146

“Machine Learning Ethics & Bias – Case Study on Biased Hiring Systems ”

AN GROUP TASK

Submitted by:

NAME	TOTADA CHANDRASHEKAR
USN	01SU23CS214

Subject:

Fundamentales of AI and ML

Subject code:

24SBT110

Submitted to:

Prof,Mahesh Kumar

Academic Year 2026

1. Introduction

Machine Learning (ML) is widely used in decision-making systems such as hiring, loan approval, healthcare diagnosis, and criminal justice. These systems promise efficiency, scalability, and objectivity. However, ML models learn from historical data. If the data contains bias, the system may reproduce or even amplify discrimination.

This assignment analyzes a real-world case where a machine learning hiring tool caused gender bias and proposes ethical guidelines and technical solutions to improve such systems.

---

2. Case Study Overview: Biased Hiring Tool

A large technology company developed an AI-based resume screening system to automate recruitment. The model was trained on historical hiring data collected over approximately 10 years.

What Happened?

- **Most of the historical hires were men.**
- **The ML model learned patterns from past data.**
- **It began favoring resumes that included male-associated words.**
- **Resumes containing terms like "women's chess club captain" were penalized.**
- **The system downgraded applications from women.**

As a result, the tool showed clear gender bias and was eventually discontinued.

3. Root Causes of Bias

1. Historical Bias

The training data reflected past gender imbalance in hiring. Since more men were hired historically, the system learned that male candidates were preferable.

2. Imbalanced Dataset

The dataset contained significantly more male resumes than female resumes, causing the model to favor dominant patterns.

3. Proxy Variables

Even if gender was not explicitly included, certain words acted as indirect indicators (proxies) for gender.

4. **Lack of Fairness Testing**

The system was not adequately audited for fairness before deployment.

5. **Over-reliance on Automation**

Human oversight was limited, allowing biased decisions to go unnoticed initially.

4. Ethical Issues Identified

1. **Discrimination**

The system violated principles of equal opportunity by favoring one gender.

2. **Lack of Transparency**

Applicants were unaware that AI was evaluating their resumes.

3. **Accountability Issues**

It was unclear who was responsible — developers, data scientists, or management.

4. **Loss of Trust**

Bias in automated systems reduces public trust in AI technologies.

5. **Legal Risks**

Such bias can violate anti-discrimination and employment laws.

5. Ethical Principles Violated

1. **Fairness – Equal treatment was not ensured.**
2. **Accountability – Responsibility was not clearly defined.**
3. **Transparency – Decision-making process was not explainable.**
4. **Non-maleficence – The system caused unintended harm.**
5. **Justice – Opportunities were unequally distributed.**

6. **Proposed Solutions to Improve the Model**

A. Data-Level Solutions

1. **Use balanced and representative datasets.**
2. **Remove sensitive attributes (gender, race, religion).**
3. **Detect and eliminate proxy variables.**
4. **Use synthetic data augmentation if necessary.**
5. **Conduct bias audits before model training.**

B. **Model-Level Solutions 1. Use fairness-aware algorithms.**

2. **Apply techniques such as: o Re-weighting o Adversarial debiasing o Constraint optimization**
3. **Implement Explainable AI (XAI) tools to interpret predictions.**
4. **Compare fairness metrics during validation.**

C. **Evaluation-Level Solutions**

Measure fairness using:

- **Demographic Parity**
- **Equal Opportunity**
- **Equalized Odds**
- **Disparate Impact Ratio**

Regularly monitor model outputs for bias after deployment.

D. Organizational & Governance Solutions

1. **Establish an AI Ethics Committee.**
2. **Implement Responsible AI policies.**
3. **Ensure human-in-the-loop decision systems.**

4. **Provide transparency notices to users.**
5. **Conduct regular third-party audits.**
6. **Maintain documentation (Model Cards, Data Sheets).**

7. Guidelines for Ethical ML Development

To prevent bias, organizations should follow these guidelines:

1. **Bias Awareness Training for developers.**
2. **Diverse Development Teams.**
3. **Continuous Monitoring after deployment.**
4. **User Feedback Mechanisms.**
5. **Regulatory Compliance with data protection and anti-discrimination laws.**
6. **Impact Assessment Reports before releasing AI systems.**

8. Broader Implications

If not addressed, biased ML systems can:

- **Reinforce societal inequalities**
- **Increase unemployment discrimination**
- **Harm marginalized communities**
- **Create reputational damage for companies**
- **Lead to legal consequences**

Responsible AI development is not just a technical requirement but a social responsibility.

9. Conclusion

This case study demonstrates how machine learning systems can unintentionally inherit and amplify societal bias. The biased hiring tool highlights the importance of fairness testing, transparency, accountability, and ethical governance in AI systems.

Bias mitigation must occur at every stage of the ML lifecycle — from data collection and model development to evaluation and deployment. By implementing ethical guidelines,

fairness metrics, and strong governance frameworks, organizations can build more inclusive, trustworthy, and responsible AI systems.

10. References (Optional for Academic Submission)

1. **Research on Algorithmic Bias in Hiring Systems**
2. **Responsible AI Frameworks**
3. **Fairness in Machine Learning Literature**
4. **Data Ethics and Governance Guidelines**