MediMind Group

Members:
Siyabonga Mahlangu
Koketso Bambo
Lindiwe Mkuzangwe
Georgia Legodi
Konke Maphisa
Banele Magubane

Group Project: " Bias Audit Report"

500-word ethics statement connecting findings to broader AI ethics principles .

Artificial Intelligence (AI) systems are increasingly deployed in high-stakes decision-making domains such as hiring, lending, social welfare, and healthcare. While these systems promise efficiency, scalability, and consistency, they also carry significant risks of reinforcing or amplifying societal inequities. In this Bias Audit Report, our analysis of an income prediction model revealed substantial disparities across gender and racial groups, highlighting the critical ethical need to detect, quantify, and mitigate bias before real-world deployment.

Our exploratory data analysis (EDA) and model evaluation revealed clear gender and racial biases. Females were systematically underpredicted for high-income outcomes (>50K), with model accuracy ranging between 0.71 and 0.88 across racial groups. Minority racial groups, including Black, Amer-Indian-Eskimo, and Other, also experienced lower prediction accuracy (0.71–0.82) compared to White and Asian-Pac-Islander individuals (0.88–0.94). The demographic parity difference and equalized odds difference further confirmed systematic disparities in predictions, indicating that historical inequalities and imbalanced class distributions in the dataset influenced algorithmic outcomes. High-income individuals from underrepresented groups were disproportionately misclassified, demonstrating the potential for AI to compound existing social inequities if left unmitigated.

To address these ethical concerns, we applied two mitigation strategies. First, Reweighing adjusted the training data to assign higher weights to underrepresented groups, including females, minority races, and high-income individuals. This approach ensures the model gives proportional importance to all groups during learning. Second, Threshold Optimization, a post-processing fairness technique, customized decision thresholds for sensitive groups to equalize true positive and false positive rates. Post-mitigation results showed significant improvements: gender demographic parity difference decreased from 0.134 to 0.044, racial parity difference dropped from 0.300 to 0.043, and equalized odds differences approached near-zero values. Crucially, overall model accuracy (~81%) remained stable, demonstrating that fairness improvements can be achieved without compromising predictive performance.

From an ethical perspective, these findings emphasize the principles of justice, fairness, accountability, transparency, and non-discrimination in AI. Unchecked biases in predictive

models risk perpetuating structural inequalities, disproportionately affecting marginalized populations. Mitigating bias is therefore not only a technical challenge but also a moral responsibility, ensuring AI tools operate equitably and responsibly.

Moreover, this analysis highlights the importance of continuous monitoring and iterative auditing. Bias can emerge or shift over time as societal conditions or input data evolve. Organizations must proactively monitor fairness metrics, update models with new data, and retrain predictive algorithms regularly. Providing actionable fairness insights to stakeholders enables informed, ethical decision-making and strengthens public trust in AI systems.

In conclusion, our bias audit demonstrates that systematic disparities can be identified, quantified, and mitigated using evidence-based strategies. By implementing reweighing and threshold optimization, we achieved measurable improvements in fairness while preserving model performance. These interventions align with broader AI ethics principles, emphasizing equity, accountability, and responsible AI deployment. The findings and recommendations presented in this report provide a roadmap for organizations seeking to deploy predictive models in a socially responsible, equitable, and transparent manner, ensuring AI benefits all groups fairly and ethically.