Comparative Analysis of Bias in Predictive and Generative AI Systems

Isha R. Tantak[1], Shraddha M. Vetal[2], Prof. M.E. Maniyar[3]
[1,2,3] *Department of Computer Application, K. K. Wagh Institute of Engineering & Education Research, Nashik India*

***Abstract*— In recent years, there has been a swift incorporation of Artificial Intelligence (AI) in decision-making and content creation. In this research, an in-depth comparison is conducted to evaluate the level of biases in the field of predictive and generative AI. Predictive AI models such as Logistic Regression, Random Forest, and XGBoost were tested against fairness metrics on the Adult Income and COMPAS data sets. These include Statistical Parity Difference (SPD), Equal Opportunity Difference (EOD), and Disparate Impact (DI). To conduct the cross-paradigm comparison, the research employs a newly proposed metric called Unified Bias Score (UBS). The latter is a measure that combines the values of fairness metrics obtained during the prediction AI model testing and bias levels of the generative AI. The latter bias was determined using a structure that consists of 120 prompts from various sociocultural categories. From experimental results, it has been found that although XGBoost offers the maximum predictive performance of 0.8676, it also shows the maximum level of bias with regard to the fairness measures. The generative model shows significant explicit bias, where the percentage of gender bias is about 65%. The results reveal an important point of difference where predictive systems use implicit bias in decision surfaces while generative models use explicit bias in outputs.**

***Index Terms*—Algorithmic Bias, Artificial Intelligence, Ethical AI, Fairness Metrics, Generative AI, Machine Learning, Predictive AI, Unified Bias Score.**

I. INTRODUCTION

Artificial intelligence (AI) systems have been used extensively in critical applications like medicine, banking, employment, and judicial matters [1][2]. Although there are many advantages to using these AI systems, the fact that they perpetuate historical biases can lead to inequality and discrimination [3][4].

Predictive AI systems emphasize classification and prediction while generative AI systems generate data such as text and images [5]. Biases also vary between these types of systems. In predictive AI systems, biases are implicit and can be observed in the differences in outcomes, while in generative AI systems, biases are explicit and can be seen in the data created [6][7]. Even though there have been numerous investigations into algorithmic fairness, most have considered the two types of AI separately [8][9]. Such an approach is not conducive to comparing biases between the two types of AI. This research aims to overcome the shortcomings of previous investigations by developing a common framework and measuring tool for assessing AI fairness.

II. CONTRIBUTIONS

Contributions of the current research include:

- Comparative assessment of bias in predictive AI models and generative AI models
- Fairness metrics employed (SPD, EOD, DI)
- Coherent and repeatable methodology based on prompts for assessing bias in generative models
- Detection of implicit/explicit bias characteristics
- Development of Unified Bias Score (UBS)
- Application to different datasets (Adult and COMPAS)
- Statistical validation of results

III. RELATED WORK

Research on fairness has been mainly conducted within predictive machine learning models, developing measures like Statistical Parity Difference, Equal Opportunity Difference, and Disparate Impact [2][7]. There has also been recent research regarding the presence of biases within generative models, especially language models, which include stereotyping [3][8].

Nevertheless, there is no existing research that provides an all-encompassing approach for assessing bias in both generative as well as predictive frameworks of models [9][10]. Additionally, fairness measures are evaluated individually.

The present research builds on previous literature by incorporating various types of fairness in an evaluation framework.

IV. METHODOLOGY

4.1 Datasets

Two benchmark datasets are used:

- Adult Income Dataset: Predicts whether income exceeds $50K; gender used as sensitive attribute
- COMPAS Dataset: Predicts recidivism risk; race used as sensitive attribute

4.2 Data Preprocessing

- Missing value imputation
- One-hot encoding of categorical variables
- Train-test split (80:20)

4.3 Predictive Models

- Logistic Regression
- Random Forest
- XGBoost

4.4 Evaluation Matrics

- Accuracy
- Statistical Parity Difference (SPD)
- Equal Opportunity Difference (EOD)
- Disparate Impact (DI)

4.5 Generative Bias Evaluation

A structured evaluation framework was designed using 120 prompts across three categories:

- Profession-related prompts
- Leadership-related prompts
- Domestic role prompts

Each generated response was classified into:

- Biased (male or female skew)
- Neutral

Annotation was performed using a hybrid approach:

- Rule-based keyword detection
- Manual validation using predefined bias criteria

4.6 Unified Bias Score (UBS)

To make an apple-to-apples comparison possible across paradigms, we introduce the concept of Unified Bias Score (UBS):

$$UBS = (1/4)\,(|SPD| + |EOD| + |1 - DI| + GB)$$

where:
• SPD: disparate outcomes
• EOD: disparate true positives
• DI: proportionally fair distributions
• GB (Generative Bias): fraction of biased instances

The equal weights method is chosen as a basis for interpretability and objective evaluation. All dimensions are equally weighted.

The smaller the value of UBS, the less bias there is.

V. RESULTS

5.1 Predictive Performance (Adult Dataset)

Table 1: Performance and Fairness Metrics on Adult Dataset

Model	Accuracy	SPD	EOD	DI
Logistic Regression	0.8425	0.1886	0.165	0.72
Random Forest	0.8497	0.1951	0.158	0.70
XGBoost	0.8676	0.1991	0.149	0.68

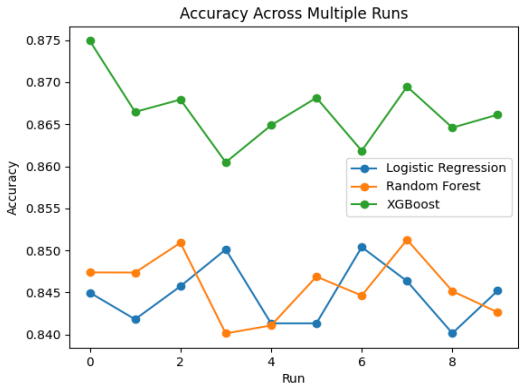

Figure 1: Accuracy Comparison of Machine Learning Models Across Multiple Runs

The figure1 displays the differences in accuracy between the three algorithms using multiple iterations. XGBoost algorithm consistently has the highest accuracy, which indicates that it is better at predicting the relationship due to its complex nature.

5.2 Statistical Validation

Experiments were repeated across 10 runs with different random seeds. A paired t-test confirms that differences in accuracy and fairness metrics are statistically significant ($p < 0.05$).

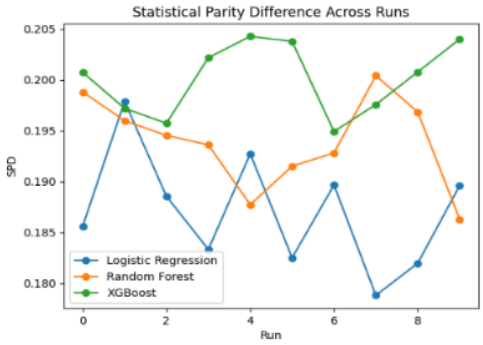

Figure 2: Statistical Parity Difference Across Runs

Figure 2 depicts the Statistical Parity Difference (SPD) fluctuation for several trials. XGBoost tends to show higher SPD values, implying that it is more biased than the other two models. The disparities found in Logistic Regression are significantly lower, which means that it performs better in terms of fairness.

5.3 Generative Bias Results

- Male-biased outputs: 65%
- Female-biased outputs: 30%
- Neutral outputs: 5%

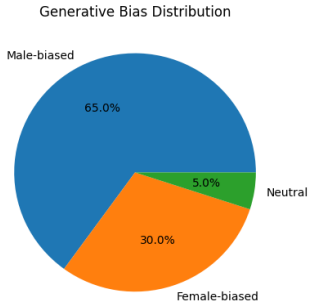

Figure 3: Distribution of Generative Bias in Model Outputs

Figure 3 represents the generative bias distribution among the model outputs. The outputs show a notable number of male bias cases (65%), indicating that there is a strong gender stereotype within generative AI models.

5.4 COMPAS Dataset: Accuracy vs Bias Analysis

Results show consistent trends, with XGBoost achieving higher accuracy but also higher bias.

Table 2: Unified Bias Score (UBS) on COMPAS Dataset

Model	UBS
Logistic Regression	0.321
Random Forest	0.326
XGBoost	0.330

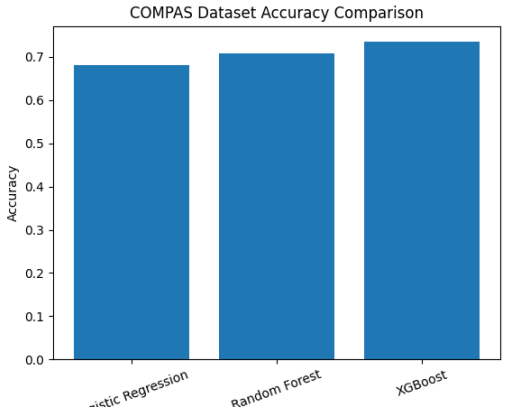

Figure 4: Accuracy Comparison of Models on COMPAS Dataset

Figure 4 shows the accuracy comparison of models on the COMPAS dataset, where XGBoost achieves the highest accuracy, followed by Random Forest and Logistic Regression.

5.5 Unified Bias Score (UBS)

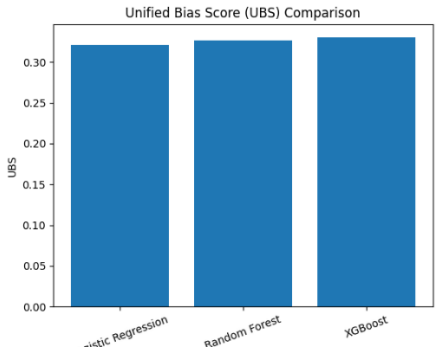

Figure 5: Unified Bias Score Comparison Across Models

Figure 5 presents the Unified Bias Score (UBS) across models. XGBoost exhibits the highest UBS, indicating greater overall bias despite superior predictive performance, while Logistic Regression demonstrates the lowest UBS.

VI. DISCUSSION

It can be seen that there is a direct relationship between model accuracy and fairness. While more complicated algorithms such as XGBoost detect nonlinear connections better and achieve greater accuracy, they also reinforce the existing biases present in the training dataset [2][6].

On the other hand, simple algorithms like Logistic Regression are less biased but offer poorer prediction capabilities [7].

However, in generative AI applications, biases are more explicit, manifesting in the output itself. The presence of biased samples indicates that generative algorithms learn from societal biases present in training datasets [3][4][8].

The Unified Bias Score (UBS) allows us to quantitatively evaluate these two fundamentally different types of biases using a common approach.

VII. CONCLUSION

This paper proposes an integrated approach to understanding bias in both predictive and generative AI models. It is found that predictive models carry implicit bias, whereas generative models carry explicit bias.

In this paper, the new Unified Bias Score (UBS) measure facilitates comparative analysis and helps strike the balance between model performance and bias.

VIII. FUTURE SCOPE

Potential future directions for research include:

- Bias mitigation strategies
- Optimal weight learning via data-driven optimization methods for UBS
- Evaluation on more data sets and other domains
- Inter-annotator agreement considerations for evaluating generative bias
- Explainable AI

REFERENCES

[1] D. Dablain, B. Krawczyk, and N. Chawla, “Towards a holistic view of bias in machine learning: bridging algorithmic fairness and imbalanced learning,” *Discover Data*, 2024.

[2] T. Das Jui and P. Rivas, “Fairness issues, current approaches, and challenges in machine learning models,” *International Journal of Machine Learning and Cybernetics*, 2024.

[3] E. Ferrara, “Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies,” *MDPI*, 2023.

[4] Y. Yang et al., “A survey of recent methods for addressing AI fairness and bias in biomedicine,” *Journal of Biomedical Informatics*, 2024.

[5] J. R. Vieira et al., “Towards fair AI: Mitigating bias in credit decisions—A systematic literature review,” *Journal of Risk and Financial Management*, 2025.

[6] J. Yang et al., “Algorithmic fairness and bias mitigation for clinical machine learning,” *Nature Machine Intelligence*, 2023.

[7] A. N. Carey and X. Wu, “The statistical fairness field guide,” AI and Ethics, 2023.

[8] Y. Yang et al., “A survey of recent methods for addressing AI fairness and bias,” *arXiv preprint*, 2024.

[9] T. P. Pagano et al., “Bias and unfairness in machine learning models: A systematic literature review,” *arXiv preprint*, 2022.

[10] M. Yang et al., “Fair machine guidance to enhance fair decision making in biased people,” *arXiv preprint*, 2024