



MITIGATING ALGORITHMIC BIAS IN MACHINE LEARNING THROUGH SYNTHETIC TABULAR DATA GENERATION

Harer Savita Laxman

Research Scholar, Department of Computer Science Engineering, Glocal University, Saharanpur, U.P.

Dr. Shashank Swami

Research Supervisor, Department of Computer Science Engineering, Glocal University, Saharanpur, U.P.

ARTICLE DETAILS

Research Paper

Received: **23/08/2025**

Accepted: **24/09/2025**

Published: **30/09/2025**

Keywords: Algorithmic Bias, Machine Learning, Synthetic Data, Tabular Data, Fairness

ABSTRACT

The study titled "Mitigating Algorithmic Bias in Machine Learning through Synthetic Tabular Data Generation" probes the ways in which synthetic data methods might improve the accuracy and fairness of machine learning models. The study incorporates survey and experimental studies to assess the practical effect of synthetic data and perceptions of bias, using a quantitative research approach. Data scientists, artificial intelligence researchers, and machine learning practitioners were among the 135 people polled in a structured survey on their views on the use of synthetic data and their familiarity with the causes of bias. Model training on both natural and artificially enhanced datasets utilizing techniques like SMOTE, GANs, and Variational Autoencoders (VAE) was also part of the experimental assessment. To investigate causal links between data augmentation methods and model fairness results, statisticians used tools including frequency analysis, t-tests, and analysis of variance (ANOVA). It was shown that synthetic data production significantly improves fairness and lowers data imbalance without sacrificing accuracy. According to the results, synthetic tabular data provides a happy medium between model performance, data privacy, and fairness, making it a promising method for ethical AI research. Adding to the continuing body of work in ethical AI, this study provides evidence that synthetic data may help reduce algorithmic bias in ML applications.

I. INTRODUCTION

In the rapidly evolving landscape of artificial intelligence (AI) and machine learning (ML), data has emerged as the cornerstone upon which intelligent systems are built. However, the quality, diversity, and representativeness of data significantly influence how these systems perform and, more importantly, how fairly they make decisions. A growing body of research has exposed the prevalence of algorithmic bias—systematic and unfair discrimination embedded within machine learning models due to imbalanced or biased data. Such bias can lead to inequitable outcomes in crucial sectors like healthcare, hiring, finance, and criminal justice, where automated decisions directly impact human lives. As the awareness of this problem intensifies, researchers and practitioners are increasingly turning to synthetic data generation, particularly for tabular data, as a promising solution to mitigate bias, ensure fairness, and enhance privacy in machine learning applications. Algorithmic bias originates from multiple sources throughout the machine learning pipeline. Often, it stems from the historical and structural inequalities present in real-world data. For example, if a dataset used for credit scoring reflects societal disparities—such as unequal access to education or employment—then the resulting model may replicate or even amplify those inequalities. Additionally, bias may arise from data imbalance, where certain demographic groups are underrepresented or overrepresented, leading to skewed model predictions. Feature selection, label quality, and even the algorithms themselves can introduce or exacerbate bias. Traditional approaches to bias mitigation, such as re-weighting samples, modifying algorithms, or post-processing model outputs, have shown limited success when the underlying data itself is fundamentally flawed or insufficiently diverse.

In this context, synthetic tabular data generation offers a novel pathway to address these challenges at their root—the data level. Synthetic data refers to artificially generated data that mimics the statistical properties and relationships of real datasets without directly copying individual records. For tabular data, which is structured into rows and columns (as in spreadsheets or databases), synthetic data generation involves creating new samples that preserve the distributions, correlations, and dependencies among features. Techniques such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), copula-based models, and Bayesian networks are increasingly being used to produce high-quality synthetic tabular data that can supplement or replace real datasets. The use of synthetic data can significantly reduce algorithmic bias in several ways. First, it enables data balancing by artificially increasing the representation of minority or under-sampled groups. For instance, if a dataset used to train an employment screening model contains disproportionately fewer samples from women or marginalized communities, synthetic data can be generated to ensure a more equitable distribution of demographic features. Second, synthetic data can remove sensitive correlations that perpetuate bias. By controlling the data generation process, practitioners can explicitly decorrelate protected attributes such as gender, race, or age from outcome variables, preventing models from learning discriminatory associations. Third, synthetic data generation supports privacy preservation by allowing organizations to share or use data for model training without exposing sensitive personal information, which is especially vital under data protection regulations such as GDPR and HIPAA.

Moreover, synthetic tabular data generation enhances transparency and accountability in AI systems. Since synthetic data can be engineered to test specific hypotheses or scenarios, it can be used for fairness auditing and bias stress-testing of models before deployment. This proactive approach allows developers to identify unfair behaviors in algorithms under diverse

demographic or contextual conditions that may not be adequately represented in the original dataset. Furthermore, synthetic data facilitates reproducibility and open research by enabling public sharing of realistic datasets that do not compromise privacy or intellectual property, fostering broader collaboration in bias mitigation research. Despite its potential, the generation and application of synthetic data are not without challenges. One major concern is fidelity versus fairness: ensuring that the synthetic data is both statistically accurate and ethically fair. If the generative model learns biased distributions from real data, it may reproduce those biases synthetically. Therefore, careful design of the data generation pipeline is essential—incorporating fairness-aware objectives, controlled sampling, and rigorous evaluation metrics. Another issue lies in model over fitting or mode collapse in generative models, which can result in unrealistic or redundant data samples. Evaluating the quality and fairness of synthetic data remains an active area of research, requiring robust metrics for measuring not just statistical similarity but also fairness and privacy guarantees.

To address these challenges, emerging hybrid approaches combine synthetic data generation with fairness-aware machine learning techniques. For example, fairness-constrained GANs and differential privacy-enhanced data generators are being developed to create datasets that are simultaneously realistic, unbiased, and privacy-preserving. Additionally, regulatory and ethical frameworks are being proposed to govern the responsible use of synthetic data, ensuring that synthetic data practices align with principles of transparency, accountability, and inclusiveness. The convergence of these technological and ethical advancements marks a significant step toward realizing responsible AI—systems that are not only intelligent but also just and trustworthy. Ultimately, synthetic tabular data generation represents more than a technical solution; it embodies a paradigm shift in how we conceptualize data in the era of AI. Instead of being passive consumers of historical data, researchers and practitioners can actively engineer fairness into datasets, shaping models that better reflect the diversity and complexity of human society. As machine learning continues to permeate every facet of modern life, the adoption of synthetic data methodologies offers a powerful means to mitigate bias at its source and foster equitable outcomes across domains.

II. REVIEW OF LITERATURE

Duong, Manh & Conrad, Stefan. (2024) we provide and explain solutions for reducing bias in tabular datasets, inspired by recital (67) of the current EU corrigendum to the AI Act. Our primary interest is in datasets that include many protected characteristics, including gender, age, and nationality. Since most current approaches are only suitable for one protected property, this makes assessing and reducing bias more difficult. Two things are added to this paper: It all starts with the introduction of new anti-discrimination policies. To help academics and practitioners choose the most appropriate metric to evaluate the dataset's fairness, our framework classifies these metrics alongside current ones. Additionally, a fresh take on an established bias reduction technique, FairDo, is shown. By modifying the dataset, we demonstrate that this approach may reduce the impact of any kind of prejudice, including intersectional discrimination. The feasibility of de-biasing datasets with several protected features is shown by experimental results on real-world datasets (Adult, Bank, Compas). Also, as compared to the original datasets, the converted fair datasets do not noticeably lower the performance of any of the machine learning models that were evaluated. In our experiments, we found that discrimination might be eliminated by as much as 83%. A minimum of 7% and a maximum of 27% reduction in the gap between protected groups was seen in the majority of tests. Overall, the results demonstrate the efficacy of the mitigation

method, and this research adds to the continuing debate about the application of the AI Act in the European Union.

Panagiotou, Emmanouil et al., (2024) Machine Learning (ML) models are prone to data-inherited bias since they are data-driven. This is particularly true in classification tasks where group and class imbalances are common. Inequity in the categorization goal or protected factors, such as race or gender, might compromise the usefulness and fairness of ML. There aren't many solutions for when class and group imbalances coexist in real-world tabular information. While most approaches rely on oversampling techniques, such as interpolation, to reduce imbalances, new developments in synthetic tabular data production show potential but have not been thoroughly investigated in this regard. In order to tackle class and group imbalances, this research uses state-of-the-art methods to generate synthetic tabular data and employs several sampling procedures in a comparative comparison. Findings from experiments conducted on four datasets show that generative models may effectively reduce bias, opening the door to further research in this area.

Hoitsma, Fabian et al., (2024) When AI systems are trained on biased data, they will make biased choices that discriminate against certain groups or people. The two main types of bias are explicit bias, in which a protected characteristic or features directly impact decision-making, and implicit bias, in which a protected feature or features indirectly impact decision-making. Biased patterns are notoriously hard to identify and eliminate. This work explores the possibility of concurrently mitigating explicit and implicit biases against one or more protected characteristics in structured classification data sets while maintaining the discriminatory power of the data. Specifically, this paper's key contribution is a training instance reweighting optimization-based bias reduction technique. This technique may concurrently minimize implicit and explicit bias against many protected characteristics and works with nominal and numerical data. By adjusting the parameters of the goal function, one may manage the trade-off between reducing bias and sacrificing accuracy. Numerical simulations conducted on real-world data sets demonstrate that explicit bias against protected features may be eliminated entirely and implicit bias reduced by up to 77% without compromising the performance of a wrapper classifier trained on the same data. When it comes to the chosen datasets, the suggested strategy generally performs better than the state-of-the-art bias mitigation techniques.

Fonseca, Joao & Bação, Fernando. (2023) Anonymization, regularization, oversampling, semi-supervised learning, self-supervised learning, and many more applications are among the many that may benefit from synthetic data production. Because of its enormous potential, new algorithms were developed to produce data in a way that is specific to various data kinds and machine learning applications. Despite Tableau data's prevalence in industrial applications, it is often overlooked. Modern methods for generating synthetic data are not well-defined, and there is a dearth of literature reviews on the subject. Additionally, these methods are scattered across several machine learning tasks and domains. In this article, we examine tabular and latent space approaches to data synthesis. We take a look at 70 generation algorithms for six different ML problems, sort them into six categories, explain what each one does, suggest a new taxonomy that builds on earlier ones, talk about metrics to measure the quality of synthetic data, and suggest some areas for further study. Academics and practitioners, we hope, will walk away from this research with a clearer picture of the gaps in the literature and a better grasp of how to use synthetic data to develop more effective strategies.

Micheletti, Nicolo et al., (2023) the presence of representation bias in health data makes it difficult to draw broad conclusions from studies and increases the likelihood of skewed findings. Consequently, female-or ethnically-underrepresented subpopulations do not equally benefit from therapeutic advances. Traditional resampling methods like SMOTE and more recent ones based on generative adversarial networks (GAN) are among the many strategies that have been suggested to mitigate the effects of representation bias. But, creating synthetic high-dimensional time-series health data remains a challenging task. We came up with CA-GAN, a novel architecture that can synthesize high-dimensional time series data, as a solution. Without the mode collapse, a major issue with GANs, CA-GAN outperforms state-of-the-art methods in quantitative and qualitative assessments. We evaluate 7535 patients with hypotension and sepsis from two separate clinical datasets derived from the actual world. Independent tests on Black and female subpopulations confirm that our CA-GAN's synthetic data improves model fairness. Also, by accurately representing minority classes in data while carefully maintaining their original distribution, CA-GAN boosts performance in a downstream prediction task. Summary of the key points More and more, healthcare clinicians are turning to artificial intelligence (AI) to aid with patient diagnosis, treatment prescription, and risk assessment. In particular, these AI systems may not be gender, racial/ethnic, or socioeconomically representative since they learn from current health information. This has the potential to worsen current health disparities and make AI less effective for certain communities. To address this issue, we developed AI software specifically for this purpose, which can create fake patient data. Protecting our patients' privacy is our first priority; therefore we've developed software that can mimic their data without really reproducing it. Our synthetic data, when added to current datasets, will help AI systems learn to be more fair and inclusive for all patients.

Pagano, Tiago et al., (2023) the elimination of prejudice and undue advantage from AI models is an ongoing task. Researchers examine datasets, metrics, approaches, and tools to find and fix algorithmic unfairness and prejudice. This research explores the current level of understanding on bias and unfairness in machine learning algorithms. The PRISMA-compliant systematic review is available on the OSF platform. From 128 publications published between 2017 and 2022, 45 were chosen using search string optimization and inclusion/exclusion criteria. Scopus, IEEE Xplore, Web of Science, and Google Scholar were the databases searched in early 2022 from 2021. The majority of the publications that were retrieved focus on strategies for recognizing and diminishing injustice and prejudice. Commonly used in bias studies are these approaches, which include tools, statistical procedures, important measures, and datasets. Data, algorithms, and user interaction were the primary focuses of the preprocessing, in-processing, and postprocessing mitigation measures aimed at reducing bias. In order to reduce bias, sensitive qualities are crucial, as shown by the basic fairness metrics of Equalized Odds, Opportunity Equality, and Demographic Parity. Health, education, criminal justice, economics, and criminal justice system image enhancement are just a few of the many diverse themes covered by the 25 datasets, the majority of which include sensitive information.

Barbierato, Enrico et al., (2022) when developing algorithms to augment or even replace human judgment using machine learning approaches, it is crucial to keep data bias and fairness in mind. While there are several techniques to detect and evaluate bias in scientific literature, the approaches that produce intentionally skewed datasets have received less attention. This is despite the fact that data scientists may use these datasets to develop and test decision-making algorithms that are fair and unbiased. One novel approach to building a

synthetic dataset capable of representing bias is shown in this study, which makes use of probabilistic networks that use structural equation modeling. To demonstrate how changing parameters affects bias and fairness, the suggested method was tested on two datasets: one more realistic, based on loan approval status data, and one simpler. When compared to other procedures, this one requires a minimal number of parameters to generate datasets with a controlled amount of bias and fairness.

Mandhala, Venkata et al., (2022) the proposed method for identifying and mitigating bias in datasets was investigated in this study. It is feasible to increase a model's performance rate while decreasing its bias, according to the findings of this research. Kolmogorov-Smirnov (KS), class imbalance, KL divergence, and sample disparity are the pre-training measures used in the research. By finding the maximum weightage for each metric, the traits may be discovered. The model shows a significant improvement in the system's performance after being trained using objective data. The ROC curve, together with the false positive and false negative rates, is used to accomplish the bias trade-off. Results from comparing FPR and FNR with and without bias reduction show a significant improvement.

III. RESEARCH METHODOLOGY

- **Research Design**

This study employed a quantitative research design to investigate the impact of synthetic tabular data on mitigating algorithmic bias in machine learning models. The research focused on analyzing participants' perceptions of bias sources, adoption of synthetic data techniques, and subsequent improvements in model fairness and accuracy. A descriptive and inferential approach was used, combining frequency distributions, percentage analysis, t-tests, and ANOVA to generate insights and test the significance of observed differences across synthetic data techniques.

- **Population and Sample**

The study targeted data scientists, machine learning practitioners, and AI researchers involved in tabular data modeling and fairness evaluation. Using purposive sampling, a total of 135 participants were selected based on their experience with machine learning and familiarity with bias mitigation practices. This sample size allowed for sufficient statistical power for both descriptive and inferential analyses while maintaining practical feasibility.

- **Data Collection**

Data were collected using a structured survey and experimental model evaluation. The survey captured participants' perceptions of algorithmic bias, awareness of synthetic data techniques, and perceived improvements in fairness. For experimental evaluation, machine learning models were trained on original and synthetic-augmented datasets generated via SMOTE, GANs, and Variational Auto encoders (VAE). Model performance, including accuracy and fairness metrics, was recorded to compare the impact of synthetic data on mitigating bias.

IV. DATA ANALYSIS AND INTERPRETATION

Table 1: Sources of Perceived Algorithmic Bias

Source of Bias	Frequency (n)	Percentage (%)
Data Imbalance	50	37.0
Historical Bias	40	29.6
Measurement Errors	25	18.5
Model Design/Algorithm	20	14.8
Total	135	100%

Table 1 presents the distribution of participants' perceptions regarding the sources of algorithmic bias in machine learning systems. The majority of respondents, 37.0%, identified data imbalance as the primary source, highlighting the widespread concern that unequal representation of data categories can lead to skewed model outputs and unfair predictions. Historical bias was recognized by 29.6% of participants, emphasizing that pre-existing social and systemic inequalities encoded in training data can propagate discriminatory patterns in algorithmic decisions. Measurement errors, accounting for 18.5%, indicate that inconsistencies in data collection and labeling processes are also perceived as contributing factors to bias, although to a lesser extent. Finally, 14.8% of respondents attributed bias to model design or algorithmic choices, suggesting awareness that algorithmic architecture, hyper parameter selection, or feature engineering can inadvertently introduce unfairness. Collectively, these findings underscore that while all listed sources contribute to algorithmic bias, data imbalance and historical bias remain the most prominent challenges. The total frequency of 135 participants confirms comprehensive coverage, and the percentages summing to 100% validate the integrity of the survey. This distribution reflects the participants' nuanced understanding that algorithmic bias is multifactorial and that mitigation strategies must address both data-centric and model-centric factors.

Table 2: Adoption of Synthetic Data Techniques

Synthetic Data Technique	Frequency (n)	Percentage (%)
SMOTE	35	25.9
GAN-Based Generation	50	37.0
Variational Autoencoders (VAE)	30	22.2
No Synthetic Data Used	20	14.8
Total	135	100%

Table 2 summarizes participants' adoption of different synthetic data generation techniques aimed at mitigating algorithmic bias. The GAN-Based Generation approach emerged as the most frequently used method, with 37.0% of respondents implementing it, demonstrating its popularity due to its capacity to create realistic and diverse tabular data while preserving complex relationships. SMOTE (Synthetic Minority Oversampling Technique), used by

25.9%, reflects participants' focus on balancing class distributions in datasets, particularly for minority categories prone to underrepresentation. Variational Autoencoders (VAE) were applied by 22.2% of participants, suggesting moderate uptake of deep generative models that efficiently capture underlying data distributions. Interestingly, 14.8% of participants reported not using any synthetic data techniques, indicating either a reliance on original datasets or potential barriers to adoption such as lack of technical expertise, computational resources, or concerns regarding data privacy. The total frequency of 135 ensures full representation, and the percentages totaling 100% allow a clear comparative assessment. This table highlights that synthetic data techniques are actively leveraged by a majority of practitioners, reflecting a growing recognition of their role in reducing bias and improving fairness in machine learning pipelines, though some gaps in awareness or implementation remain.

Table 3: Perceived Improvement in Model Fairness

Improvement Level	Frequency (n)	Percentage (%)
Low	25	18.5%
Moderate	80	59.3%
High	30	22.2%
Total	135	100%

Table 3 illustrates participants' perceptions of the improvement in model fairness after the application of synthetic data. A clear majority, 59.3%, reported a moderate improvement, indicating that while synthetic data helps enhance fairness, its impact may not always be dramatic or universally consistent across models and datasets. High improvement was observed by 22.2% of respondents, suggesting that in specific contexts, synthetic data can substantially reduce bias and improve equitable treatment of all subgroups. Conversely, 18.5% perceived low improvement, highlighting scenarios where synthetic augmentation may have limited effectiveness, possibly due to inherent dataset limitations, poor synthetic data quality, or insufficient alignment with model objectives. These results emphasize that while synthetic data is a valuable tool for fairness enhancement, it is not a standalone solution; complementary strategies such as bias-aware model design, post-processing adjustments, and careful evaluation are necessary. The total frequency of 135 participants validates the comprehensiveness of the responses, and the percentages sum to 100%, confirming the reliability of the survey distribution.

Table 4: Model Accuracy Before vs After Synthetic Data

Accuracy Level	Before Synthetic Data	After Synthetic Data
<70%	40 (29.6%)	15 (11.1%)
70–80%	50 (37.0%)	40 (29.6%)
>80%	45 (33.3%)	80 (59.3%)

Table 4 provides a comparison of model accuracy levels before and after incorporating synthetic data techniques. Prior to synthetic data augmentation, 29.6% of models exhibited accuracy below 70%, reflecting underperforming models likely affected by bias or class imbalance. A larger proportion, 37.0%, achieved moderate accuracy (70–80%), while 33.3% exceeded 80%, indicating a relatively strong baseline performance. After applying synthetic data, there is a notable improvement: only 11.1% remained below 70%, 29.6% fell in the 70–80% range, and a significant 59.3% achieved above 80% accuracy. This shift demonstrates that synthetic data not only addresses bias but can also enhance predictive performance by improving training data quality and balance. The table underscores the dual benefit of synthetic data generation: it supports both fairness and overall model robustness. The pre- and post-comparison clearly illustrates the positive impact, suggesting that model developers can expect tangible gains in performance metrics alongside ethical improvements when using appropriate synthetic data techniques.

Table 5: T-Test Comparison of Model Fairness (Original vs Synthetic Data)

Group	Mean Fairness Score	SD	t-value	p-value
Original Data	65.2	8.5	4.32	0.0001
Synthetic Data Augmented	73.8	7.2		

Table 5 presents the results of an independent sample t-test comparing model fairness scores between original data and synthetic data-augmented datasets. The mean fairness score increased from 65.2 in the original data to 73.8 following synthetic data augmentation, with standard deviations of 8.5 and 7.2, respectively. The calculated t-value of 4.32 with a p-value of 0.0001 indicates that this improvement is statistically significant, confirming that synthetic data augmentation has a measurable and meaningful effect on enhancing fairness. This finding suggests that concerns regarding algorithmic bias can be effectively mitigated using synthetic techniques, rather than relying solely on traditional preprocessing or model adjustments. Moreover, the reduced standard deviation in the augmented group points to more consistent fairness across predictions, reflecting the ability of synthetic data to stabilize model behavior for diverse subgroups. Overall, this t-test provides empirical validation of synthetic data's role in improving ethical performance metrics, bridging the gap between theoretical bias mitigation and observable results in real-world applications.

Table 6: ANOVA Test across Different Synthetic Techniques

Technique	Mean Fairness Score	SD	F-value	p-value
SMOTE	70.1	6.5	5.12	0.007
GAN-Based Generation	74.2	7.0		
Variational Autoencoders (VAE)	72.8	6.8		

Table 6 summarizes the results of an ANOVA analysis comparing mean fairness scores across three synthetic data techniques: SMOTE, GAN-Based Generation, and Variational

Autoencoders (VAE). The mean fairness scores range from 70.1 for SMOTE to 74.2 for GAN-based methods, with an intermediate 72.8 for VAEs. The F-value of 5.12 and p-value of 0.007 indicate that the differences among techniques are statistically significant, suggesting that certain methods are more effective in improving fairness than others. Specifically, GAN-based generation appears to provide the highest fairness enhancement, likely due to its ability to capture complex data distributions and generate high-quality synthetic instances. SMOTE, while effective for balancing minority classes, shows a comparatively lower mean improvement, highlighting its limitations in multidimensional or high-complexity datasets. These findings imply that researchers and practitioners should carefully select synthetic data methods according to dataset characteristics and fairness objectives, rather than assuming uniform efficacy across techniques. The ANOVA analysis provides robust statistical evidence supporting targeted selection of synthetic data methods for bias mitigation.

V. CONCLUSION

Mitigating algorithmic bias in machine learning through synthetic tabular data generation presents a transformative approach to creating fair, balanced, and privacy-conscious AI systems. By generating representative and controllable datasets, synthetic data allows practitioners to overcome the limitations of biased, incomplete, or sensitive real-world data. It not only supports the inclusion of underrepresented populations but also enables explicit de-biasing strategies during data creation. However, the success of this approach depends on maintaining a careful balance between realism and fairness, ensuring that synthetic data accurately reflects genuine patterns while avoiding the reinforcement of historical inequities. As research advances, combining synthetic data generation with fairness-aware learning, explainability, and regulatory compliance can lead to truly responsible AI ecosystems. Ultimately, synthetic tabular data generation offers a foundation for equitable and trustworthy machine learning—transforming bias mitigation from a corrective afterthought into an integral component of the data design process.

REFERENCES:

1. Azelpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophical Compass*, 16, e12760.
2. Barbierato, E., Della Vedova, M., Tessera, D., Toti, D., & Vanoli, N. (2022). A methodology for controlling bias and fairness in synthetic data generation. *Applied Sciences*, 12(4), 1–19.
3. Delgado-Rodriguez, M., & Llorca, J. (2004). Bias. *Journal of Epidemiology & Community Health*, 58, 635–641.
4. Duong, M., & Conrad, S. (2024). Measuring and mitigating bias for tabular datasets with multiple protected attributes. *arXiv preprint*.
5. Elhassan, T., & Aljurf, M. (2016). Classification of imbalance data using Tomek link (T-link) combined with random under-sampling (RUS) as a data reduction method. *Global Journal of Technology & Optimization*, 1, 100011.
6. Engel, C., Linhardt, L., & Schubert, M. (2024). Code is law: How COMPAS affects the way the judiciary handles the risk of recidivism. *Artificial Intelligence and Law*, 32, 1–22.
7. Fernandez, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, 61, 863–905.

8. Fonseca, J., & Bação, F. (2023). Tabular and latent space synthetic data generation: A literature review. *Journal of Big Data*, 10(6), 1–9.
9. Ha, T., & Kim, S. (2023). Improving trust in AI with mitigating confirmation bias: Effects of explanation type and debiasing strategy for decision-making with explainable AI. *International Journal of Human–Computer Interaction*, 39, 1–12.
10. Hameed, M., Qureshi, A., & Kaushik, A. (2024). Bias mitigation via synthetic data generation: A review. *Electronics*, 13(19), 3909.
11. Hoitsma, F., Nápoles, G., Guven, C., & Salgueiro, Y. (2024). Mitigating implicit and explicit bias in structured data without sacrificing accuracy in pattern classification. *AI & Society*, 40, 2551–2570.
12. Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, 31, 388–409.
13. Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, 30, 25–36.
14. Lin, Z., Jung, J., Goel, S., & Skeem, J. (2020). The limits of human predictions of recidivism. *Science Advances*, 6, eaaz0652.
15. Mandhala, V., Bhattacharyya, D., Midhunchakkaravarthy, D., & Kim, H.-J. (2022). Detecting and mitigating bias in data using machine learning with pre-training metrics. *Ingénierie des Systèmes d’Information*, 27, 119–125.
16. Meyer, J. G., Urbanowicz, R. J., Martin, P. C. N., O’Connor, K., Li, R., Peng, P.-C., Bright, T. J., Tatonetti, N., Won, K. J., Gonzalez-Hernandez, G., et al. (2023). ChatGPT and large language models in academia: Opportunities and challenges. *BioData Mining*, 16, 1–19.
17. Micheletti, N., Marchesi, R., Kuo, N., Barbieri, S., Jurman, G., & Osmani, V. (2023). Generative AI mitigates representation bias using synthetic health data. *Unpublished manuscript*.
18. Miletic, M., & Sariyar, M. (2024). Challenges of using synthetic data generation methods for tabular microdata. *Applied Sciences*, 14(14), 5975.
19. Pagano, T., Bessa Loureiro, R., Lisboa, F., Peixoto, R., Guimaraes, G., Cruz, G., Araujo, M., Santos, L., Cruz, M., Oliveira, E., Winkler, I., & Sperandio Nascimento, E. G. (2023). Bias and unfairness in machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big Data and Cognitive Computing*, 7(1), 15.
20. Panagiotou, E., Roy, A., & Ntoutsis, E. (2024). Synthetic tabular data generation for class imbalance and fairness: A comparative study. *Unpublished manuscript*. 1(2).
21. Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and futurescope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154.
22. Schelter, S., & Stoyanovich, J. (2020). Taming technical bias in machine learning pipelines. *Bulletin of the Technical Committee on Data Engineering*, 43, 39–50.
23. Tahir, M. A., Kittler, J., & Yan, F. (2012). Inverse random under sampling for class imbalance problem and its application to multi-label classification. *Pattern Recognition*, 45, 3738–3750.
24. Yang, X., Kuang, Q., Zhang, W., & Zhang, G. (2017). AMDO: An over-sampling technique for multi-class imbalanced problems. *IEEE Transactions on Knowledge and Data Engineering*, 30, 1672–1685.
25. Yee, K., Tantipongpipat, U., & Mishra, S. (2021). Image cropping on Twitter: Fairness metrics, their limitations, and the importance of representation, design, and agency. In



- Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–24.
26. Yen, S., & Lee, Y. (2006). Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset. *Lecture Notes in Control and Information Science*, 344, 731.
 27. Yen, S.-J., & Lee, Y.-S. (2009). Cluster-based under-sampling approaches for imbalanced data distributions. *Expert Systems with Applications*, 36, 5718–5727.