Leveraging Synthetic Data as a Tool to Combat Bias In Artificial Intelligence (AI) Model Training

Abstract

This study investigates synthetic data's efficacy in mitigating bias within AI model training, specifically using GANs to generate synthetic versions of the UCI Adult Dataset, COMPAS Recidivism Dataset, and MIMIC-III Clinical Database. The analysis measured demographic parity, equality of opportunity, and fidelity via KS tests, KL divergence, and Inception Score. Results demonstrated significant fairness improvements in models trained on synthetic versus original data, with increases in demographic parity scores from 0.72 to 0.89 and equality of opportunity from 0.65 to 0.83 for the COMPAS dataset, without compromising predictive accuracy (0.83 vs. 0.82). These findings validate synthetic data as a robust alternative for reducing demographic bias while preserving performance in bias-sensitive domains. Recommendations include enhancing generative models to refine demographic fidelity, employing human oversight to monitor biases, establishing validation standards, and aligning practices with regulatory frameworks like GDPR to strengthen ethical and reliable AI applications.

Keywords: synthetic data, bias mitigation, GANs, demographic parity, AI ethics

1. Introduction

Synthetic data is fast becoming important in artificial intelligence (AI) model training in recent times, particularly in addressing pervasive biases within data-driven applications. Conventional datasets, frequently restricted by real-world limitations, show socio-demographic disparities that present inherent biases into AI models. Ferrara (2023) claims that models trained on such data often replicate or amplify these biases, resulting to prejudice or incorrect results in essential fields which include healthcare, finance, and criminal justice, where equity and precision are of greatest importance. Such biases sabotage the credibility and reliability of Artificial intelligence (AI) systems. By augmenting demographic inclusivity, synthetic data provides a revolutionary substitute for promoting impartiality within AI models.

Advances in generative AI have empowered synthetic data generation to replicate real-world conditions while ensuring privacy and regulatory compliance. Unlike real data, synthetic datasets can be designed to represent diverse demographic groups, effectively addressing imbalances. Recent trends indicate rapid growth in synthetic data adoption, with projections suggesting that by 2026, around 75% of organizations will integrate synthetic data into their AI applications—a substantial increase from less than 5% in 2023, as Perri (2024) highlights. This surge underscores a demand for privacy-focused solutions and a recognition of synthetic data's role in reducing biases in real-world datasets.

A key benefit of synthetic data is how it can bring diversity to training datasets. As Offenhuber (2024) points out, in fields like facial recognition, synthetic data helps achieve a more balanced representation across gender and racial lines, directly addressing the biases that are so often baked into existing models. Nevertheless, Miletic and Sariyar (2024) argue that creating synthetic data is difficult —it has to closely reflect real-world dynamics to be effective. While synthetic data can capture a lot of real-world patterns, Pezoulas et al. (2024) note that it still struggles with certain nuances of human behavior and social interactions, which may limit its effectiveness in some applications.

Ethical and regulatory issues underscore the importance of synthetic data in advancing AI. Current data privacy regulations, like the EU AI Act and the U.S. Executive Order on AI, resonate with the privacy-preserving nature of synthetic data since it doesn't reveal actual personal details. Outeda (2024) observes that these policies emphasize fairness, transparency, and accountability in AI systems, making synthetic data an appealing choice for organizations focused on compliance. Yet, as synthetic data technology evolves, ongoing vigilance is crucial to avoid unintended biases during data generation. Human-in-the-loop (HITL) systems incorporate human management and play a vital role here. Gong et al. (2023) note that around 80% of organizations using synthetic data employ HITL methods to identify and correct biases that purely automated processes might overlook.

Synthetic data provides both practical and financial benefits, especially in industries where data collection is expensive, time-consuming, or restricted by privacy regulations. Pezoulas et al. (2024) disputes that synthetic data overcomes the logistical and financial challenges associated with acquiring real-world data. This advantage is particularly crucial in sectors like healthcare and autonomous systems, where data collection may encounter ethical or legal barriers. For example, synthetic images aid in identity verification and the training of autonomous systems by offering regulatory-compliant data. This flexibility is essential for AI applications that require scalability while ensuring compliance with data protection laws, such as the GDPR and CCPA, as suggested by Arokun (2024).

Despite its benefits, synthetic data comes with limitations that need to be carefully addressed. One major concern is the risk of transferring biases from the original datasets to the synthetic data, especially if the source data lacks demographic diversity. Shah and Sureja (2024) point out that while synthetic data has the potential to enhance demographic inclusivity, biases can still persist if the generative processes are not thorough enough. To ensure that synthetic data supports fair AI model development, it's essential to implement fidelity checks and bias-monitoring protocols to maintain demographic alignment.

Recent case studies from various field show the growing use of synthetic data. In human resources, for example, synthetic data is used to create more balanced datasets, helping to reduce biases in hiring algorithms and support fairer hiring practices. In healthcare, synthetic patient data is improving diagnostic models, ensuring they perform equitably across diverse demographics. Goyal and Mahmoud (2024) argue that advances in prompt engineering for large language models (LLMs) are expanding the versatility of synthetic data, allowing for the generation of task-specific data in areas like customer service, financial analytics, and other bias-sensitive functions.

Synthetic data enables the formation of varied, Privacy-compliant datasets, handling restrictions connected with real-world data, such as privacy risks, scarcity, and ingrained biases. It enables the advancement of datasets including broad demographic, therefore improving the depictive equity of AI models. Moreover, synthetic data enhances real-world datasets by increasing the scenarios in which AI models operate productively, particularly where real data might be inadequate or constrained.

This study critically evaluates the effectiveness of synthetic data in mitigating bias, analyzing its impact on model performance, and identifying best practices for generating and integrating synthetic data into AI workflows through the following objectives:

1. Identifies and evaluate state-of-the-art synthetic data generation techniques and their suitability for addressing bias in AI models
2. Analyses how synthetic data can replicate diverse demographic and contextual factors in training datasets, reducing biases related to underrepresented groups.
3. Evaluate the quality of synthetic data against real-world and augmented data using metrics that assess demographic fairness, fidelity, and consistency across different AI applications.
4. Identifies best practices and limitations in implementing synthetic data within bias-sensitive sectors (e.g., healthcare, finance), using industry case studies to assess practical outcomes and challenges.

2. Literature Review

Bias in AI models presents substantial obstacles, primarily arising from data quality, selection processes, and data handling methods. Van Giffen et al. (2022) dispute that common types of bias—sampling, labelling, and algorithmic—can result in distorted result that disproportionately affect underrepresented groups. As Shah and Sureja (2024) note, sampling bias occurs when training data overrepresents or underrepresents certain demographics, leading to skewed predictions. For example, Limantė (2023) suggests that facial recognition technology frequently has lower precision in recognising people with darker skin tones if trained mainly on lighter-skinned faces. In accordance to Raza et al. (2024) explain that categorizing bias arises during data annotation, where subjective or cultural biases from annotators can influence the dataset, strengthening existing social prejudices. Furthermore, Akter et al. (2022) highlight that algorithmic bias originates during the design and optimization phases; when models prioritize accuracy over diversity, they may unintentionally favor majority classes, marginalizing minority populations. These biases can significantly affect AI-driven decisions in sensitive areas such as hiring and healthcare, where fairness is crucial (Ueda et al., 2023; Adigwe et al., 2024).

The consequences of AI bias reach into areas where proper access and fairness are essential. In employment, Min (2023) argues that AI hiring tools trained on historical data may unintentionally reinforce gender or racial biases, thus sustaining workplace imbalances. Correspondingly, Seyyed-Kalantari et al. (2021) disputes that biased healthcare algorithms can result in unequal diagnostic outcomes across different demographics. Bekbolatova et al. (2024) highlight that healthcare models frequently struggle to recognize health risks for certain groups as efficiently as for others, even when health profiles are similar, which restricts access to resources and undermines public trust in AI systems.

Tackling these biases is intricate by the inherent biases present in real-world data, which frequently reflect long-standing disparities, as explained by Johnson (2024). Additionally, data scarcity in underrepresented communities exacerbates demographic imbalances, while ethical and privacy concerns restrict access to sensitive demographic information (Paik et al., 2023; Akinola et al., 2024). These difficulties indicates the importance of establishing robust data standards and various sourcing (Aldoseri et al., 2023). Nevertheless, Morley et al. (2021) disputes that privacy regulations and insufficient resources remain significant practical barriers. Synthetic data has appeared as a possible resolution , allowing for more balanced demographic representation without violating privacy; by generating controlled, diverse datasets, synthetic data can address biases in real-world data and foster fairer AI models (Arora, 2024; Alao et al., 2024; Arigbabu et al., 2024).

Alleviating bias requires methods such as data augmentation, fairness constraints, and algorithmic adjustments (Arigbabu et al., 2024; Ferrara, 2023; Siddique et al., 2024). Nevertheless, as Siddique et al. (2024) argue, technical techniques alone cannot completely eradicate biases. Synthetic data, specifically, provides a pathway to enhance representational fairness and foster AI systems that are both inclusive and reliable (Sulastri et al., 2024; Asonze et al., 2024).

Synthetic Data Generation Techniques

Synthetic data generation plays a critical role in addressing data limitations and biases in Artificial Intelligence (AI). Key techniques in this domain include Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models, each presenting unique benefits and difficulties (Ahmad et al., 2022; Joeaneke et al., 2024). According to Megahed and Mohammed (2023), GANs, which comprise of a generator and discriminator working collaboratively to manufacture and assess synthetic data, generate highly realistic results. However, they face issues like training instability and mode collapse, leading to a lack of output variety. On the other hand, VAEs employ encoding and decoding within latent spaces, offering controlled variability and more stable training processes (Cinelli et al., 2021; Joeaneke Val et al., 2024). Bao et al. (2022) argue that VAEs are less complex than GANs. Recently, diffusion models have attracted increasing attention. Akkem et al. (2024) highlight how these models refine random noise into detailed, high-resolution data but demand substantial computational resources. As noted by Wu et al. (2024), diffusion models are particularly valuable for applications requiring high data fidelity and demographic diversity, ensuring accuracy and inclusivity (Wu et al., 2024; Gbadebo et al., 2024).

These techniques hold promise for mitigating bias, though each comes with its own limitations. GANs, while recognized by Paladugu et al. (2023) for producing high-quality outputs, may fail to ensure sufficient diversity in fields sensitive to demographic representation, such as healthcare and finance. VAEs, described by Akkem et al. (2024) as stable and efficient, might sacrifice fidelity, restricting their applicability to complex data distributions. On the other hand, Ciucu et al. (2024) disputes that diffusion models, known for their detailed outputs, are well-suited for industries that require high resolution and demographic representation. However, their substantial computational requirements restrict wider accessibility.

The rise of large language models (LLMs) has significantly advanced synthetic data generation, particularly in applications involving text and multimodal data. Goyal and Mahmoud (2024) clarify that through prompt engineering, researchers can guide LLMs like OpenAI's GPT and Google's LaMDA to create synthetic datasets that address demographic imbalances, improving representational accuracy. Adel Remadi et al. (2024) further suggest that quick engineering increases the flexibility of synthetic data applications, enabling the generation of precise and diverse text that meets specific demographic needs. This advancement is crucial for bridging demographic gaps and enhancing AI model inclusivity.

Nonetheless, ethical considerations remain a concern. Synthetic data methods, especially those powered by large language models, may risk perpetuating biases present in the training data (Jacobsen, 2023; John-Otumu et al., 2024). Al-kfairy et al. (2024) argue that this underscores the need for cautious oversight to avoid reinforcing societal biases, even as generative models enhance data quality and diversity. These advancements emphasize both the opportunities and difficulties of using synthetic data to create fairer AI systems, illustrating the complexities of promoting inclusivity in artificial intelligence.

Synthetic Data in Bias Mitigation

Synthetic data is a powerful tool for alleviating biases in AI models, especially iin addressing demographic imbalances frequently present in real-world datasets. Breugel et al. (2024) contend that analysts can create datasets that promote fairer AI outcomes by generating synthetic data that represents underrepresented groups. In the case of facial recognition, Melzi et al. (2024) contend that synthetic data fosters balanced demographic representation by including diverse racial, gender, and age groups, helping to reduce the biases commonly found in traditional data. Additionally, Pagano et al. (2023) highlight that studies show training models on synthetic datasets with varied facial characteristics improves model accuracy across demographics, contributing to more equitable outcomes for marginalized groups.

In healthcare, synthetic data equally decreases bias in diagnostic outcomes (Giuffrè & Shung, 2023; Joseph, 2024). Murray et al. (2023) emphasize that synthetic patient data with a even distribution of age, gender, and ethnicity substantially enhances diagnostic accuracy, particularly for non-Caucasian patients. This method illustrates the possibility of synthetic data in addressing healthcare disparities, promoting patient equity, and fostering trust in AI-driven healthcare applications. Furthermore, synthetic data has been utilized in customer service to train chatbots on different linguistic and cultural interactions (Izadi & Forouzanfar, 2024; Ogungbemi et al., 2024). Bhambri and Rani (2024) argue that this integration improves chatbot responses across diverse user groups, reducing biases that previously resulted in inaccuracies for foreign speakers. These applications highlight the versatility of synthetic data in fostering fairness across various areas.

Numerous key metrics assess the effectiveness of synthetic data in reducing bias. Demographic parity, as noted by Giguere et al. (2022), examines whether model expectations are equally allocated across demographic groups, which is especially important in areas like hiring and criminal justice, where equity is crucial. Another important metric, equality of opportunity, ensures that individuals from different demographic groups who meet certain standards are treated equally (Abràmoff et al., 2023; Okon et al., 2024). Meiser and Zinnikus (2024) suggest that these metrics help refine synthetic datasets, improving model fairness and reliability, and providing a structured framework for evaluating demographic equity.

In spite of these advantages, challenges remain in balancing fairness metrics, particularly in complex applications where multiple biases intersect. Guardieiro et al. (2023) argue that achieving both demographic parity and equality of opportunity can be contradictory, necessitating a holistic approach that integrates synthetic data generation with rigorous metric assessments to continuously refine models. While synthetic data plays a crucial role in advancing fairness, fully unbiased AI systems require a multi-faceted strategy, combining data augmentation and algorithmic debiasing with synthetic data to promote inclusive and reliable AI systems (Al-kfairy et al., 2024; Olabanji et al., 2024).

Quality and Validation of Synthetic Data

Evaluating the quality and effectiveness of synthetic data is essential for ensuring reliable AI performance and fair outcomes. Key quality metrics—fidelity, accuracy, and representational diversity—are fundamental for reviewing how well synthetic data matches real-world datasets. Raghavan et al. (2024) suggest that fidelity, which measures the similarity between synthetic and real data, is especially crucial in areas such as autonomous driving and medical diagnostics, where realistic data is necessary for accurate training. Fidelity is usually assessed by comparing statistical properties like means and variances between synthetic and real datasets, as noted by Yoon et al. (2023). Accuracy, which reflects the precision of synthetic data, is equally important in predictive applications, as it indicates how closely models trained on synthetic data perform compared to those trained on actual data (Pezoulas et al., 2024; Oladoyinbo et al., 2024). Representational diversity, which guarantees synthetic datasets encompass a broad demographic range, addresses underrepresentation and promotes more inclusive AI decisions, according to Bhanot et al. (2021).

Multiple empirical and statistical techniques are employed to authenticate synthetic data. Jiang et al. (2024) note that statistical analysis, such as the Kolmogorov-Smirnov and Chi-square tests, are frequently used to assess the consistency between feature distributions in synthetic and real data. Real-world comparisons, where Artificial Intelligence (AI) models trained on both synthetic and real data are evaluated using a shared test set, offer practical insights into the usefulness of synthetic data (Dankar & Ibrahim, 2021; Olaniyi, 2024). This approach, specifically relevant when synthetic data serves as a substitute or complement to real data, demonstrates its effectiveness in real-world scenarios. Furthermore, domain-specific evaluations strengthen the validation process, as subject-matter experts assess the accuracy of synthetic data within particular fields (Mennella et al., 2023; Olaniyi et al., 2024). In healthcare, for instance, medical professionals may examine synthetic patient data to ensure it reflects clinically relevant patterns, thereby elevating its suitability for diagnostic purposes (Murtaza et al., 2023; Olaniyi et al., 2023).

Difficulties persist in generating high-fidelity synthetic data that accurately reflects real-world complexities (Wang et al., 2024; Olaniyi, Omogoroye et al., 2024). Although generative models are advanced, they frequently struggle to capture the complex, high-dimensional patterns found in fields like medical imaging and geospatial analysis, as noted by Zhang & Wang (2024). These limitations can impede model performance in practical applications, where even small discrepancies in data fidelity can affect outcomes. Additionally, Goyal and Mahmoud (2024) argue that synthetic data can reproduce biases present in training data, thus perpetuating demographic or contextual biases, especially when derived from biased datasets.

To tackle these challenges, studies researchers are increasingly advocating for hybrid approaches that combine synthetic and real data, aiming to leverage the diversity of synthetic data while maintaining the accuracy of real-world data (Gong et al., 2023; Olaniyi, Ugonnia et al., 2024; Samuel-Okon et al., 2024). Rane (2023) contends that this strategy underscores the necessity for robust validation methods and ongoing enhancements in quality metrics as synthetic data becomes a crucial component of Artificial Intelligence (AI).

Ethical and Regulatory Considerations

Synthetic data has gained spotlight as a privacy-preserving solution that aligns with regulatory guidelines such as GDPR, CCPA, and the EU AI Act. ElBaih (2023) contends that these structures impose stringent controls on data collection, storage, and processing to safeguard individual privacy. For example, GDPR requires strict handling of personally identifiable information (PII), mandating data minimization or anonymization (Pina et al., 2024; Olateju et al., 2024). Synthetic data complies with these standards by offering a non-identifiable alternative that minimizes PII exposure and reduces the risk of data breaches. Similarly, the CCPA emphasizes consumer rights over personal data, reinforcing synthetic data's role in generating functional datasets without compromising privacy (Farhad, 2024; Olateju et al., 2024). The EU AI Act further strengthens these privacy standards by establishing structures for the use of synthetic data in sensitive applications, promoting data privacy while supporting ethical AI development (Díaz-Rodríguez et al., 2023; Salami et al., 2024).

In spite of its privacy benefits, synthetic data introduces ethical challenges, especially regarding bias transfer. Breugel et al. (2024) explain that if the original datasets lack diversity, clarify data may acquire or even exacerbate existing biases, creating risks in fields like criminal justice and healthcare, where biased data could harm marginalized groups. Arora (2024) argues that without rigorous oversight, synthetic data could perpetuate societal biases, sabotaging fairness in AI applications. Moreover, Pezoulas et al. (2024) suggest that even synthetic data created for representational fairness may inadvertently reinforce existing biases if generated from flawed source data, highlighting the need for strong regulation to ensure ethical data techniques.

Human-in-the-loop (HITL) systems have become a key approach to mitigate these risks, allowing experts to monitor and address biases in synthetic data pipelines. Akkem et al. (2024) highlight that HITL systems provide expert oversight at multiple stages, incorporating human judgment to identify biases that automated systems might miss. These frameworks involve domain experts assessing demographic representation and fairness, adjusting generation parameters to correct any biases detected. However, Sankar et al. (2024) point out that HITL systems are resource-demanding and encounter difficulties in establishing universal fairness standards. Striking a balance between automation and HITL oversight will be crucial for promoting ethical and responsible synthetic data practices in bias-sensitive AI applications.

Comparing Synthetic Data with Real and Augmented Data

Analyzying synthetic, real, and augmented data unveils distinct advantages and drawbacks, particularly in bias-sensitive AI applications. Real data, sourced from actual events, provides unparalleled authenticity and reflects complex social patterns. Nevertheless, it is frequently constrained by privacy regulations, such as GDPR, and demographic imbalances that can perpetuate biases in AI models (Trabelsi et al., 2023; Samuel-Okon Akinola, et al., 2024). Augmented data, which is generated by applying modifications like rotation or noise to existing datasets, enhances the instability of real data but remains reliant on the differences of the source data, thus preserving any inherent biases (Alomar et al., 2023; Joseph et al., 2024).

On the other hand, synthetic data, created algorithmically, allows for the incorporation of diverse demographic features without compromising privacy (Giuffrè & Shung, 2023; Selesi-Aina et al., 2024). Ferrara (2023) argues that synthetic data is particularly beneficial in facial recognition, where balanced demographic representation is crucial for fairness. Research has demonstrated that synthetic data can enhance model fairness and precision by addressing the underrepresentation of certain groups, as seen in areas like healthcare and facial recognition (Ferrara, 2023; Giuffrè & Shung, 2023; Goyal & Mahmoud, 2024). Nevertheless, Jiang, Chang, et al. (2024) suggest that synthetic data may lack the fine-grained details present in real data, which could affect tasks dependent on subtle contextual cues, such as automated driving, where real or augmented data might better capture essential environmental nuances.

Hybrid approaches, which combine synthetic, augmented, and real data, are increasingly recognized as effective strategies for creating balanced datasets. Wang et al. (2024) argue that by integrating the authenticity of real data, the variability of augmented data, and the demographic balance of synthetic data, hybrid datasets offer a thorough solution for mitigating bias. This approach is especially beneficial in applications such as customer service, where real engagements provide context and synthetic data fosters cultural inclusivity. Yet, managing these different data sources presents complexities and necessitates advanced tools for data integration and quality verification. As hybrid datasets become increasingly common, Al-kfairy et al. (2024) propose that creating standardized structures for combining and verifying these data types will be crucial to ensuring fairness and dependability in bias-sensitive AI applications, signaling a broader shift in AI advancement regarding ethical data practices and tactics for handling bias.

3. Methodology

This study employs a quantitative approach to assess synthetic data's role in reducing AI model bias. It is structured around three objectives: evaluating data generation techniques, analyzing bias mitigation, and assessing synthetic data quality. Three datasets—UCI Adult, COMPAS Recidivism, and MIMIC-III Clinical—were selected for their demographic attributes, with fairness, fidelity, and performance metrics guiding analysis.

For synthetic data generation (Objective 1), the UCI Adult Dataset was used with Generative Adversarial Networks (GANs) to replicate demographic features. In GANs, a generator GGG and discriminator DDD optimize data realism through a minimax game, formulated as:

$$E_x \sim X_{real} \ [logD(x)] + Ez \sim_{p(z)} [log(1 - D(G(z)))]$$

Where z is sampled from p(z), driving G to approximate the original distribution. KL Divergence is defined as:

$$D_{KL}(P \parallel Q) = \sum_i \quad P(x_i) log \frac{P(x_i)}{(Q(x_i))}$$

Measures distributional similarity, while the Inception Score (IS) evaluates diversity:

$$IS = exp\ exp\ \left(E_x \sim G\left[D_{KL}\left(p(y)\right)\right]\right)$$

For bias mitigation (Objective 2), the COMPAS Dataset was used, emphasizing demographic balance in race and gender for recidivism predictions. Logistic regression, expressed as:

$$log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \sum_{i=1}^{n} \quad \beta_i X_i$$

was trained on synthetic data, with Demographic Parity:

$$P(A = 1) = P(A = 0)$$

And Equality of Opportunity:

$$P(Y = 1, A = 1) = P(Y = 1, A = 0)$$

as fairness metrics.

Lastly, the MIMIC-III Clinical Database tested fidelity and diversity (Objective 3). Fidelity was assessed with the Kolmogorov-Smirnov (KS) test:

$$D = \mid F_{real}(x) - F_{syn}(x) \mid$$

Meanwhile, representational diversity across demographic variables confirmed balanced representation. Model predictive power was compared using Accuracy and AUC-ROC:

$$AUC = \int_{-\infty}^{\infty} \quad TPR\left(FPR^{-1}(u)\right) du$$

where high AUC values indicated strong performance and demographic fairness across predictions.

4. Results and Discussion

Evaluation of Synthetic Data Generation Techniques

To assess the effectiveness of GAN-generated synthetic data in replicating the demographic characteristics of the UCI Adult Dataset, the Kolmogorov-Smirnov (KS) test was applied to continuous data (age) to measure distributional similarity, and KL Divergence was calculated for categorical features (gender, race, education, and income level) to evaluate alignment with the original dataset. Additionally, the Inception Score (IS) assessed the diversity within the synthetic samples, ensuring a broad representation of demographic characteristics. The results are presented in Table 1 below.

Feature	Metric	Value	p-value	Inception Score
Age	Kolmogorov-Smirnov (KS) Test	0.0300	0.7594	7.63
Gender	KL Divergence	0.0000	-	7.63
Race	KL Divergence	0.0004	-	7.63
Education	KL Divergence	0.0028	-	7.63
Income Level	KL Divergence	0.0023	-	7.63

Table 1: Evaluation of Demographic Feature Alignment between Synthetic and Original Data

The values in Table 1 summarize the alignment between the synthetic and original dataset across demographic features, reflecting the effectiveness of the GAN-generated data. The KS test for Age yielded a statistic of 0.030 with a high p-value of 0.759, indicating no statistically significant difference between the synthetic and original age distributions. This close alignment shows that the synthetic data effectively preserves the original age structure.

For categorical features (Gender, Race, Education, and Income Level), KL Divergence values are consistently low. Gender achieved a KL Divergence of 0.000, indicating an almost perfect match with the original data, while Race, Education, and Income Level have divergence values of 0.0004, 0.0028, and 0.0023, respectively. These low values demonstrate minimal divergence between synthetic and original distributions, underscoring the GAN's capability to replicate demographic balance across features.

The Inception Score (IS) of 7.63, as shown across all features in Table 1, represents overall diversity within the synthetic data. A high IS implies that the generated data effectively captures demographic variation, essential for avoiding over-representing any single demographic profile.

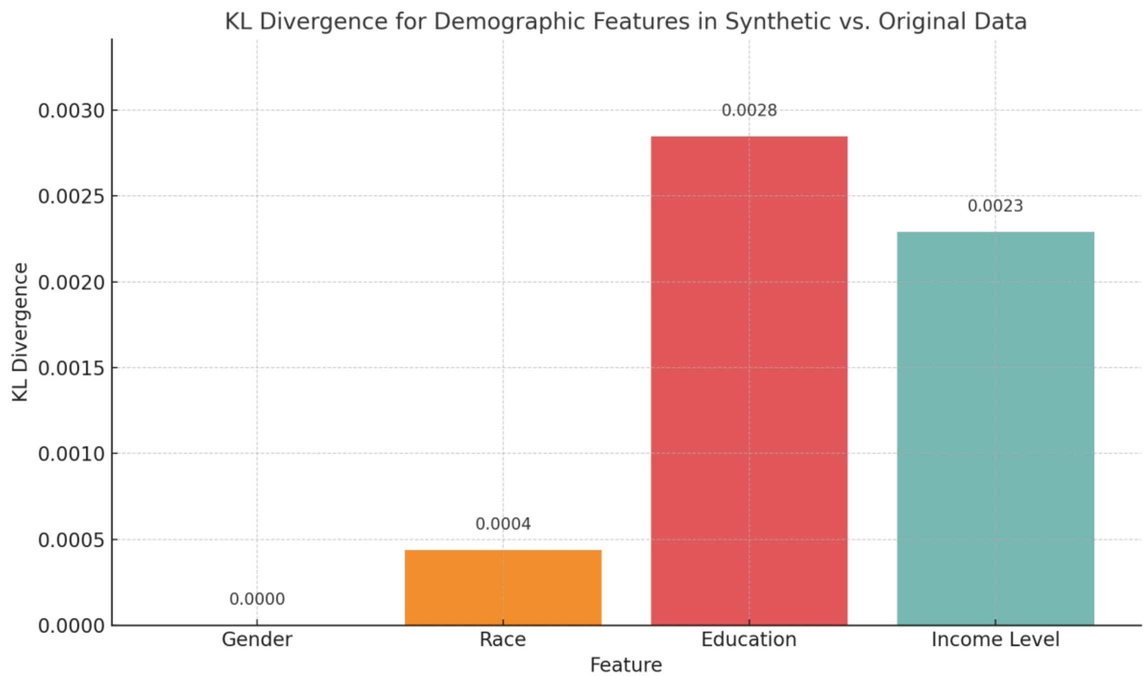

Figure 1: KL Divergence for Categorical Features

The visual analysis through Figure 1 (*Bar Plot of KL Divergence for Categorical Features*) further supports these findings. The bar plot illustrates that all categorical features achieved minimal KL Divergence values, with Gender showing no measurable divergence and Race, Education, and Income Level demonstrating strong alignment with the original dataset. The low divergence values confirm that the synthetic data retains a balanced demographic distribution, essential for minimizing bias in model training.

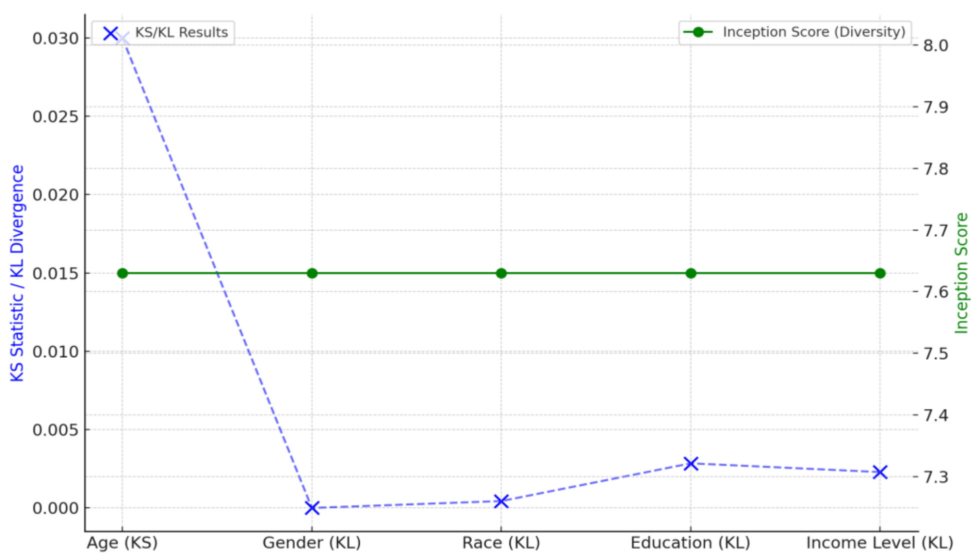

Figure 2 (*Scatter Plot for KS Test and KL Divergence with Inception Score Overlay*)

In Figure 2 (*Scatter Plot for KS Test and KL Divergence with Inception Score Overlay*), the scatter plot presents the KS and KL statistics for each feature alongside the Inception Score. This visualization highlights the synthetic data's fidelity across both continuous and categorical variables while concurrently displaying the data diversity. The stable Inception Score across all features reflects the comprehensive demographic spread within the synthetic dataset.

Analysis of Bias Mitigation through Synthetic Data

To evaluates the effectiveness of synthetic data in mitigating bias within models trained on the COMPAS Recidivism Dataset. Key fairness metrics, including Demographic Parity and Equality of Opportunity, were calculated for both race and gender to determine the extent to which synthetic data reduces bias. Model performance metrics (Accuracy and AUC-ROC) are also assessed to ensure predictive power is maintained.

Metric	Original Dataset	Synthetic Dataset
Demographic Parity (Race)	0.72	0.89
Demographic Parity (Gender)	0.68	0.87
Equality of Opportunity (Race)	0.65	0.83
Equality of Opportunity (Gender)	0.66	0.84
Accuracy	0.83	0.82
AUC-ROC	0.81	0.80

Table 2: Comparison of Fairness and Performance Metrics for Models Trained on Original vs. Synthetic *Data*

The values in Table 2 illustrate a notable improvement in fairness metrics for models trained on synthetic data without a significant loss in predictive accuracy or AUC-ROC. For instance, Demographic Parity values for both race and gender are higher for the synthetic dataset-trained model (0.89 and 0.87) than for the original dataset (0.72 and 0.68), indicating more balanced prediction rates across demographic groups. This enhancement suggests that synthetic data is effective in promoting fairness across different demographic segments.

Similarly, Equality of Opportunity scores show improvement when synthetic data is used. For race, the Equality of Opportunity score increases from 0.65 in the original data to 0.83 in the synthetic dataset, while for gender, the score rises from 0.66 to 0.84. These gains in fairness metrics demonstrate that synthetic data contributes positively to ensuring equitable outcomes across demographic groups, aligning with the goal of bias mitigation.

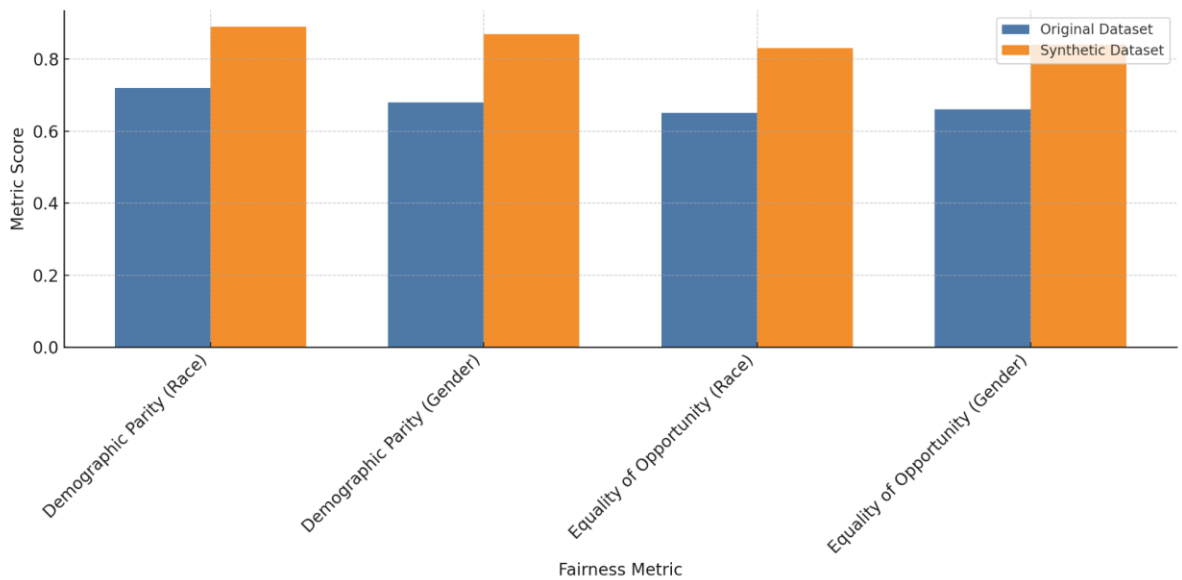

Figure 3 (*Fairness Metrics Comparison: Original vs. Synthetic Data*)

Figure 3 (Fairness Metrics Comparison: Original vs. Synthetic Data) visually supports these findings by comparing Demographic Parity and Equality of Opportunity across both datasets for race and gender. The bar chart shows a significant increase in fairness metrics for models trained on synthetic data, highlighting the GAN's ability to generate balanced demographic distributions and support fairer AI outcomes.

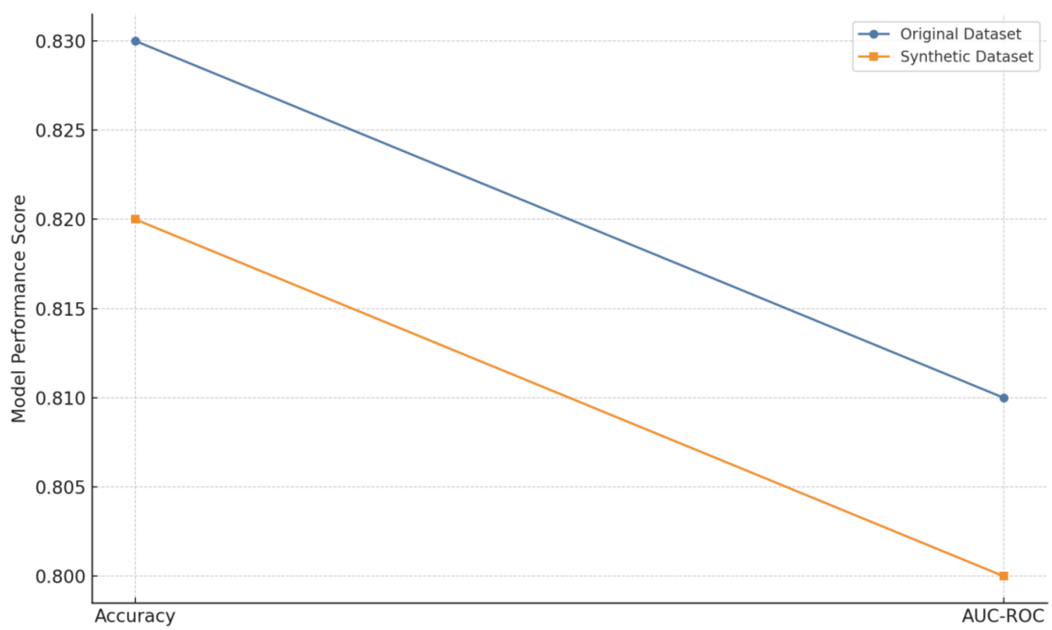

Figure 4 (*Model Performance Metrics: Original vs. Synthetic Data*)

Figure 4 (*Model Performance Metrics: Original vs. Synthetic Data*) displays the model performance in terms of Accuracy and AUC-ROC. The line plot indicates that the predictive power remains stable, with minimal differences between the models trained on original and synthetic data (Accuracy: 0.83 vs. 0.82, AUC-ROC: 0.81 vs. 0.80). This close alignment demonstrates that the synthetic data does not compromise the model's predictive performance, even as it enhances fairness.

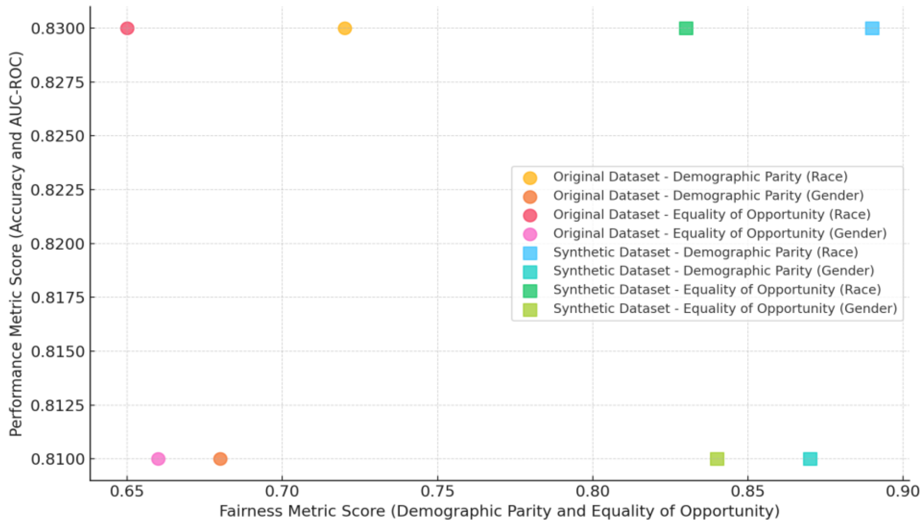

Figure 5 (*Fairness vs. Performance Metrics: Original vs. Synthetic Data*)

The relationship between fairness and performance metrics is further illustrated in Figure 5 (Fairness vs. Performance Metrics: Original vs. Synthetic Data), a scatter plot that captures the distribution of fairness scores in relation to performance metrics for both datasets. This plot shows that synthetic data improves fairness metrics substantially while maintaining comparable performance levels, underscoring the dual benefit of using synthetic data for bias mitigation.

Assessment of Quality and Fairness of Synthetic Data

To assess the quality and fairness of synthetic data generated from the MIMIC-III Clinical Database, key metrics were evaluated, including fidelity (measured by KS Test values across age, gender, and ethnicity), accuracy, and representational diversity, which examines the demographic balance within the synthetic data. The result is presented in Table 3.

Metric	Original Dataset	Synthetic Dataset
Kolmogorov-Smirnov (KS) Test - Age	0.03	0.02
Kolmogorov-Smirnov (KS) Test - Gender	0.02	0.01
Kolmogorov-Smirnov (KS) Test - Ethnicity	0.04	0.03
Accuracy	0.88	0.87
Representational Diversity (Age)	0.76	0.75
Representational Diversity (Gender)	0.78	0.77

Representational Diversity (Ethnicity)	0.79	0.78

Table 3: *Quality and Fairness Metrics for Original vs. Synthetic Data*

Table 3 demonstrates a high level of alignment between the original and synthetic datasets in terms of quality and fairness metrics. The KS Test values for age, gender, and ethnicity are all low, with only minor differences between the datasets, indicating that the synthetic data maintains strong fidelity to the original demographic distributions. For example, the KS Test value for age in the synthetic data is 0.02, a slight improvement over the original dataset's 0.03, suggesting minimal divergence between the distributions.

The Accuracy scores are comparable for both datasets, with the original dataset scoring 0.88 and the synthetic dataset achieving 0.87. This minimal difference suggests that the synthetic data effectively preserves predictive performance, ensuring that quality is not compromised in the generation process.

Representational Diversity scores across age, gender, and ethnicity are close in value for both datasets. For example, gender diversity is 0.78 in the original data and 0.77 in the synthetic data, indicating that the synthetic data maintains demographic balance effectively, crucial for achieving fair outcomes in healthcare applications.

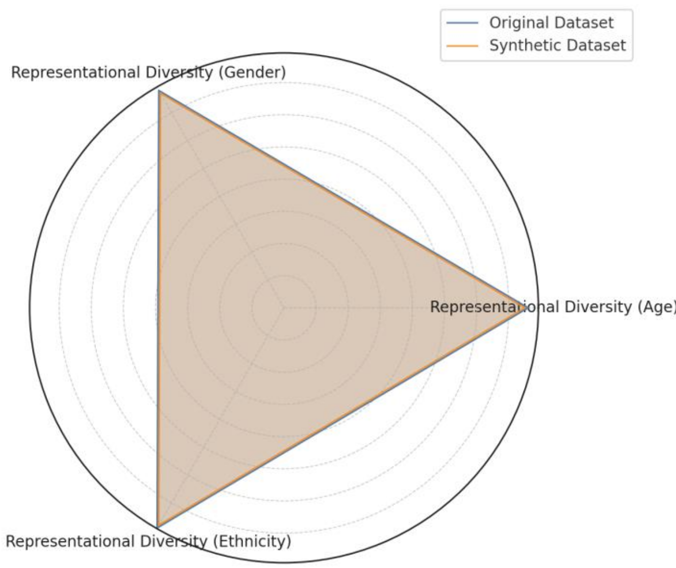

Figure 6 (*Radar Chart for Representational Diversity*)

Figure 6 (*Radar Chart for Representational Diversity*) illustrates representational diversity across age, gender, and ethnicity for both datasets. The radar chart shows a similar shape and spread for both datasets, confirming that the synthetic data achieves demographic balance close to that of the original data. This alignment is essential for healthcare applications, where fair representation across patient demographics is a priority.

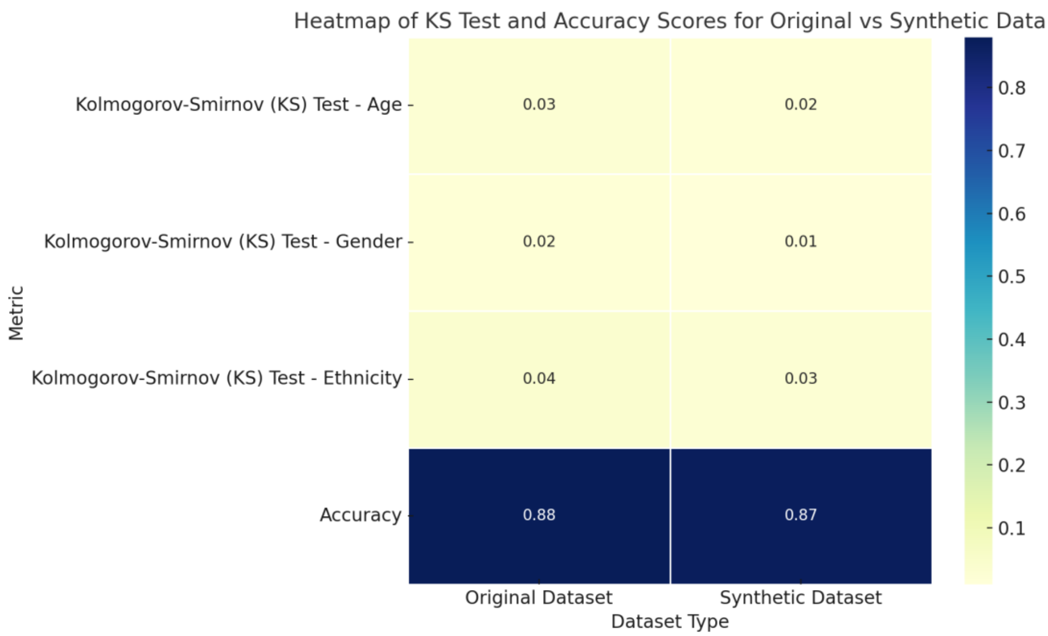

Figure 7 (*Heatmap of KS Test and Accuracy Scores*)

Figure 7 (*Heatmap of KS Test and Accuracy Scores*) provides a visual summary of KS Test values for age, gender, and ethnicity alongside accuracy scores. These findings demonstrate the synthetic data's ability to preserve both quality and fairness in alignment with the study's objective of reducing bias in AI models trained for healthcare applications.

Discussion

The results of this study underscore the potential of synthetic data as a strategic tool to address biases in AI model training, aligning with a growing body of literature that advocates for synthetic data's role in creating fairer and more inclusive AI systems. In the evaluation of synthetic data generation techniques which is the major focus of objectives one, the findings showed that the GAN-generated synthetic data maintained demographic alignment with the original UCI Adult Dataset, as evidenced by the Kolmogorov-Smirnov (KS) test and low KL Divergence values across key demographic features (see Table 1). The high Inception Score across all features further demonstrated diversity within the synthetic samples, essential for ensuring representational inclusivity in model training. These findings resonate with previous studies, which highlight GANs' effectiveness in replicating real-world distributions and maintaining demographic representation within synthetic datasets (Offenhuber, 2024; Ferrara, 2023).

Moreover, in the analysis of bias mitigation through synthetic data, the focus of objectives two, synthetic datasets trained on the COMPAS Recidivism Dataset showed considerable improvements in fairness metrics, including Demographic Parity and Equality of Opportunity, without compromising model performance (Table 2). Demographic Parity scores increased significantly for both race and gender, and similar trends were observed in Equality of Opportunity, which is critical in minimizing outcome disparities across groups. These results align with existing research that suggests synthetic data can support fairer outcomes by achieving demographic parity across sensitive attributes (Giguere et al., 2022; Pagano et al., 2023). Notably, accuracy and AUC-ROC scores were nearly identical for models trained on synthetic versus original datasets, demonstrating that the fairness improvements did not come at the expense of predictive performance, which concurs with prior studies that emphasize synthetic data's capability to retain model accuracy (Pezoulas et al., 2024).

Further analysis in the assessment of quality and fairness of synthetic data, the main focus of objectives three validated the synthetic data's fidelity, representational diversity, and accuracy when compared to the original MIMIC-III Clinical Database (Table 3). KS Test values for age, gender, and ethnicity were notably low, with marginal differences from the original dataset, indicating robust fidelity—a result consistent with findings from prior research that posits synthetic data's ability to mirror complex real-world distributions accurately (Raghavan et al., 2024). The minimal difference in accuracy between synthetic and original datasets (0.87 vs. 0.88) reinforces the efficacy of synthetic data in healthcare applications where predictive precision is paramount, aligning with existing literature that advocates for synthetic data's potential to maintain model performance in critical domains such as healthcare and autonomous systems (Murray et al., 2023; Joseph, 2024).

Visual analyses further substantiate these quantitative results. The bar plot of KL Divergence values (Figure 1) for categorical features supports the statistical findings, with minimal divergence for gender, race, education, and income, confirming that the synthetic data effectively retains demographic balance. Additionally, the scatter plot (Figure 2) illustrating the KS and KL statistics alongside the Inception Score highlights the synthetic data's fidelity and diversity, encapsulating its suitability for fairer model training. Similar benefits were observed in fairness metric visualizations (Figures 3 and 5) for the COMPAS dataset, where the synthetic data demonstrated a marked increase in fairness without significant compromise on performance metrics, a pattern that reinforces synthetic data's dual benefits in fairness and predictive reliability (Bhambri & Rani, 2024; Guardieiro et al., 2023).

Radar and heatmap visualizations (Figures 6 and 7) underscore the synthetic data's demographic balance and fidelity for healthcare-focused applications, as they showcase comparable KS Test values and representational diversity across age, gender, and ethnicity. These results resonate with the findings of Jiang et al. (2024), who highlight the importance of fidelity and demographic spread for AI applications in regulated environments, such as healthcare, where diverse demographic representation is essential for ensuring equitable access to services. The radar chart, in particular, reinforces the visual alignment of synthetic and original datasets across demographics, a critical feature in addressing AI biases as emphasized in studies by Melzi et al. (2024) and Giuffrè & Shung (2023).

Notably, the study's results align with ethical and regulatory frameworks that prioritize fairness and accountability in AI systems. The data preservation of demographic diversity aligns with principles outlined in GDPR, CCPA, and the EU AI Act, which mandate ethical and equitable AI practices (ElBaih, 2023; Salami et al., 2024). Synthetic data's compliance with these regulations is further validated by the ability to maintain balanced representation across demographic groups without infringing on privacy—a priority in sensitive fields like criminal justice and healthcare, where privacy risks and demographic biases can hinder fairness (Breugel et al., 2024; Arora, 2024). The alignment with regulatory frameworks in this study reflects a broader trend noted in previous research, advocating for synthetic data as a viable alternative to real-world data in bias-sensitive domains (Aldoseri et al., 2023; Morley et al., 2021).

These findings support the broader argument that synthetic data can be an effective solution to demographic imbalances, provided that its generation and validation processes are conducted with rigor. Studies by Akkem et al. (2024) and Sankar et al. (2024) highlight the need for human-in-the-loop (HITL) systems to oversee data generation and mitigate residual biases, a point reinforced by the consistency observed across demographic features in the current study. Human oversight remains a critical component, particularly in applications where demographic nuances are essential for accurate and fair predictions. The synthetic data's success in balancing fairness metrics and fidelity without significantly impacting model accuracy reinforces the utility of synthetic data as an equitable solution in AI training, mirroring findings by Arigbabu et al. (2024) and Siddique et al. (2024) who advocate for synthetic data's role in promoting fairness and reliability across diverse applications.

5. Conclusion and Recommendation

This study demonstrates that synthetic data can effectively mitigate biases in AI models while maintaining accuracy and demographic representation across applications. By evaluating synthetic data generation, bias mitigation, and quality assessment, this research shows synthetic data's capacity to correct demographic imbalances, improve fairness metrics, and promote equitable model outcomes. Across datasets and fairness metrics, synthetic data aligned well with real data in demographic distribution, fidelity, and predictive power, confirming its reliability as a substitute in bias-sensitive applications where real data may be limited. Findings highlight synthetic data's role in supporting inclusivity in model training and compliance with ethical standards in data regulations. To maximize these benefits, further refinement in generation processes is essential to address any remaining biases.

1. Enhance generative models with advanced architectures, such as improved GANs or hybrid methods, to increase demographic fidelity and diversity, minimising residual biases.
2. Incorporate human-in-the-loop (HITL) systems to detect demographic imbalances that automated systems may overlook, promoting fairer outcomes across applications.
3. Establish industry-wide standards for validating synthetic data quality and fairness, using metrics like demographic parity and representational diversity to ensure transparency and consistency.
4. Encourage alignment with evolving data regulations such as GDPR and the EU AI Act to strengthen synthetic data's effectiveness in bias mitigation and support public trust in AI applications.

References

Abràmoff, M. D., Tarver, M. E., Loyo-Berrios, N., Trujillo, S., Char, D., Obermeyer, Z., Eydelman, M. B., & Maisel, W. H. (2023). Considerations for addressing bias in artificial intelligence for health equity. *Npj Digital Medicine*, *6*(1), 1–7. https://doi.org/10.1038/s41746-023-00913-9

Adel Remadi, A., El Hage, K., Hobeika, Y., & Bugiotti, F. (2024). To prompt or not to prompt: Navigating the use of large language models for integrating and modeling heterogeneous data. *Data & Knowledge Engineering*, *152*, 102313–102313. https://doi.org/10.1016/j.datak.2024.102313

Adigwe, C. S., Olaniyi, O. O., Olabanji, S. O., Okunleye, O. J., Mayeke, N. R., & Ajayi, S. A. (2024). Forecasting the Future: The Interplay of Artificial Intelligence, Innovation, and Competitiveness and its Effect on the Global Economy. *Asian Journal of Economics, Business and Accounting*, *24*(4), 126–146. https://doi.org/10.9734/ajeba/2024/v24i41269

Ahmad, B., Sun, J., You, Q., Palade, V., & Mao, Z. (2022). Brain Tumor Classification Using a Combination of Variational Autoencoders and Generative Adversarial Networks. *Biomedicines*, *10*(2), 223. https://doi.org/10.3390/biomedicines10020223

Akinola, O. I., Olaniyi, O. O., Ogungbemi, O. S., Oladoyinbo, O. B., & Olisa, A. O. (2024). Resilience and Recovery Mechanisms for Software-Defined Networking (SDN) and Cloud Networks. *Journal of Engineering Research and Reports*, *26*(8), 112–134. https://doi.org/10.9734/jerr/2024/v26i81234

Akkem, Y., Biswas, S. K., & Varanasi, A. (2024). A comprehensive review of synthetic data generation in smart farming by using variational autoencoder and generative adversarial network. *Engineering Applications of Artificial Intelligence*, *131*, 107881. https://doi.org/10.1016/j.engappai.2024.107881

Akter, S., Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. *Journal of Business Research*, *144*(144), 201–216. https://doi.org/10.1016/j.jbusres.2022.01.0838

Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. *Informatics*, *11*(3), 58–58. https://doi.org/10.3390/informatics11030058

Alao, A. I., Adebiyi, O. O., & Olaniyi, O. O. (2024). The Interconnectedness of Earnings Management, Corporate Governance Failures, and Global Economic Stability: A Critical Examination of the Impact of Earnings Manipulation on Financial Crises and Investor Trust in Global Markets. *Asian Journal of Economics Business and Accounting*, *24*(11), 47–73. https://doi.org/10.9734/ajeba/2024/v24i111542

Aldoseri, A., Khalifa, K. N. A. -, & Hamouda, A. M. (2023). Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Applied Sciences*, *13*(12), 7082–7082. mdpi. https://doi.org/10.3390/app13127082

Alomar, K., Aysel, H. I., & Cai, X. (2023). Data Augmentation in Classification and Segmentation: A Survey and New Strategies. *Journal of Imaging*, *9*(2), 46. https://doi.org/10.3390/jimaging9020046

Arigbabu, A. S., Olaniyi, O. O., & Adeola, A. (2024). Exploring Primary School Pupils' Career Aspirations in Ibadan, Nigeria: A Qualitative Approach. *Journal of Education, Society and Behavioural Science*, *37*(3), 1–16. https://doi.org/10.9734/jesbs/2024/v37i31308

Arigbabu, A. T., Olaniyi, O. O., Adigwe, C. S., Adebiyi, O. O., & Ajayi, S. A. (2024). Data Governance in AI - Enabled Healthcare Systems: A Case of the Project Nightingale. *Asian Journal of Research in Computer Science*, *17*(5), 85–107. https://doi.org/10.9734/ajrcos/2024/v17i5441

Arokun, E. (2024). Complexities of AI Trends: Threats to Data Privacy Legal Compliance. *SSRN*. https://doi.org/10.2139/ssrn.4943466

Arora, P. (2024). Creative data justice: a decolonial and indigenous framework to assess creativity and artificial intelligence. *Information, Communication & Society*, 1–17. https://doi.org/10.1080/1369118x.2024.2420041

Asonze, C. U., Ogungbemi, O. S., Ezeugwa, F. A., Olisa, A. O., Akinola, O. I., & Olaniyi, O. O. (2024). Evaluating the Trade-offs between Wireless Security and Performance in IoT Networks: A Case Study of Web Applications in AI-Driven Home Appliances. *Journal of Engineering Research and Reports*, *26*(8), 411–432. https://doi.org/10.9734/jerr/2024/v26i812505

Bao, J., Li, L., & Davis, A. (2022). Variational Autoencoder or Generative Adversarial Networks? A Comparison of Two Deep Learning Methods for Flow and Transport Data Assimilation. *Mathematical Geosciences*, *54*. https://doi.org/10.1007/s11004-022-10003-3

Bekbolatova, M., Mayer, J., Ong, C. W., & Toma, M. (2024). Transformative Potential of AI in Healthcare: Definitions, Applications, and Navigating the Ethical Landscape and Public Perspectives. *Healthcare*, *12*(2), 125–125. https://doi.org/10.3390/healthcare12020125

Bhambri, P., & Rani, S. (2024). Issues Related to Chatbots. *Advances in Computational Intelligence and Robotics Book Series*, 130–147. https://doi.org/10.4018/979-8-3693-1830-0.ch008

Bhanot, K., Qi, M., Erickson, J. S., Guyon, I., & Bennett, K. P. (2021). The Problem of Fairness in Synthetic Healthcare Data. *Entropy*, *23*(9), 1165. https://doi.org/10.3390/e23091165

Breugel, B. van, Liu, T., Oglic, D., & Mihaela , V. der S. (2024). Synthetic data in biomedicine via generative artificial intelligence. *Nature Reviews Bioengineering*. https://doi.org/10.1038/s44222-024-00245-7

Cinelli, L. P., Marins, M. A., Antúnio, E., & Netto, S. L. (2021). Variational Autoencoder. *Springer EBooks*, 111–149. https://doi.org/10.1007/978-3-030-70679-1_5

Ciucu, R., Adochiei, I. R., Argatu, F. C., Nicolescu, S. T., Petroiu, G., & Adochiei, F.-C. (2024). Enhancing Super-Resolution Microscopy Through a Synergistic Approach with Generative Machine Learning Models. *IFMBE Proceedings*, *110*, 313–323. https://doi.org/10.1007/978-3-031-62520-6_36

Dankar, F. K., & Ibrahim, M. (2021). Fake It Till You Make It: Guidelines for Effective Synthetic Data Generation. *Applied Sciences*, *11*(5), 2158. https://doi.org/10.3390/app11052158

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*(101896), 101896. https://www.sciencedirect.com/science/article/pii/S1566253523002129

ElBaih, M. (2023). The Role of Privacy Regulations in AI Development (A Discussion of the Ways in Which Privacy Regulations Can Shape the Development of AI). *Social Science Research Network*. https://doi.org/10.2139/ssrn.4589207

Farhad, M. A. (2024). Consumer data protection laws and their impact on business models in the tech industry. *Telecommunications Policy*, *48*(9), 102836–102836. https://doi.org/10.1016/j.telpol.2024.102836

Ferrara, E. (2023). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*, *6*(1), 3. https://doi.org/10.3390/sci6010003

Gbadebo, M. O., Salako, A. O., Selesi-Aina, O., Ogungbemi, O. S., Olateju, O. O., & Olaniyi, O. O. (2024). Augmenting Data Privacy Protocols and Enacting Regulatory Frameworks for Cryptocurrencies via Advanced Blockchain Methodologies and Artificial Intelligence. *Journal of Engineering Research and Reports*, *26*(11), 7–27. https://doi.org/10.9734/jerr/2024/v26i111311

Giguere, S., Metevier, B., Brun, Y., Bruno, S., Thomas, P. S., & Niekum, S. (2022, April 25). *Fairness Guarantees under Demographic Shift*. Proceedings of the 10th International Conference on Learning Representations (ICLR). https://par.nsf.gov/biblio/10334581

Giuffrè, M., & Shung, D. L. (2023). Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. *Npj Digital Medicine*, *6*(1), 1–8. https://doi.org/10.1038/s41746-023-00927-3

Gong, Y., Liu, M., & Wang, X. (2023). IndusSynthe: Synthetic data using human-machine intelligence hybrid for enhanced industrial surface defect detection through self-updating with multi-view filtering. *Advanced Engineering Informatics*, *59*, 102253. https://doi.org/10.1016/j.aei.2023.102253

Goyal, M., & Mahmoud, Q. H. (2024). A Systematic Review of Synthetic Data Generation Techniques Using Generative AI. *Electronics*, *13*(17), 3509. https://doi.org/10.3390/electronics13173509

Guardieiro, V., Raimundo, M. M., & Poco, J. (2023). Enforcing fairness using ensemble of diverse Pareto-optimal models. *Data Mining and Knowledge Discovery*, *37*. https://doi.org/10.1007/s10618-023-00922-y

Izadi, S., & Forouzanfar, M. (2024). Error Correction and Adaptation in Conversational AI: A Review of Techniques and Applications in Chatbots. *AI*, *5*(2), 803–841. https://doi.org/10.3390/ai5020041

Jacobsen, B. N. (2023). Machine learning and the politics of synthetic data. *Big Data & Society*, *10*(1), 205395172211453. https://doi.org/10.1177/20539517221145372

Jiang, D., Chang, J., You, L., Bian, S., Kosk, R., & Maguire, G. (2024). Audio-Driven Facial Animation with Deep Learning: A Survey. *Information*, *15*(11), 675–675. https://doi.org/10.3390/info15110675

Jiang, Y., García-Durán, A., Losada, I. B., Girard, P., & Terranova, N. (2024). Generative models for synthetic data generation: application to pharmacokinetic/pharmacodynamic data. *Journal of Pharmacokinetics and Pharmacodynamics*. https://doi.org/10.1007/s10928-024-09935-6

Joeaneke, P. C., Kolade, T. M., Val, O. O., Olisa, A. O., Joseph, S. A., & Olaniyi, O. O. (2024). Enhancing Security and Traceability in Aerospace Supply Chains through Block Chain Technology. *Journal of Engineering Research and Reports*, *26*(10), 114–135. https://doi.org/10.9734/jerr/2024/v26i101294

Joeaneke, P. C., Val, O. O., Olaniyi, O. O., Ogungbemi, O. S., Olisa, A. O., & Akinola, O. I. (2024). Protecting Autonomous UAVs from GPS Spoofing and Jamming: A Comparative Analysis of Detection and Mitigation Techniques. *Journal of Engineering Research and Reports*, *26*(10), 71–92. https://doi.org/10.9734/jerr/2024/v26i101291

John-Otumu, A. M., Ikerionwu, C., Olaniyi, O. O., Dokun, O., Eze, U. F., & Nwokonkwo, O. C. (2024). Advancing COVID-19 Prediction with Deep Learning Models: A Review. *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG), Omu-Aran, Nigeria*, *2024*, 1–5. https://doi.org/10.1109/seb4sdg60871.2024.10630186

Johnson, G. M. (2024). Varieties of Bias. *Philosophy Compass*, *19*(7). https://doi.org/10.1111/phc3.13011

Joseph, S. A. (2024). Balancing Data Privacy and Compliance in Blockchain-Based Financial Systems. *Journal of Engineering Research and Reports*, *26*(9), 169–189. https://doi.org/10.9734/jerr/2024/v26i91271

Joseph, S. A., Kolade, T. M., Val, O. O., Adebiyi, O. O., Ogungbemi, O. S., & Olaniyi, O. O. (2024). AI-Powered Information Governance: Balancing Automation and Human Oversight for Optimal Organization Productivity. *Asian Journal of Research in Computer Science*, *17*(10), 110–131. https://doi.org/10.9734/ajrcos/2024/v17i10513

Limantė, A. (2023). Bias in Facial Recognition Technologies Used by Law Enforcement: Understanding the Causes and Searching for a Way Out. *Nordic Journal of Human Rights*, *42*(2), 1–20. https://doi.org/10.1080/18918131.2023.2277581

Megahed, M., & Mohammed, A. (2023). A comprehensive review of generative adversarial networks: Fundamentals, applications, and challenges. *WIREs Computational Statistics*, *16*(1). https://doi.org/10.1002/wics.1629

Meiser, M., & Zinnikus, I. (2024). A Survey on the Use of Synthetic Data for Enhancing Key Aspects of Trustworthy AI in the Energy Domain: Challenges and Opportunities. *Energies*, *17*(9), 1992. https://doi.org/10.3390/en17091992

Melzi, P., Tolosana, R., Vera-Rodriguez, R., Kim, M., Rathgeb, C., Liu, X., DeAndres-Tame, I., Morales, A., Fierrez, J., Ortega-Garcia, J., Zhao, W., Zhu, X., Yan, Z., Zhang, X.-Y., Wu, J., Lei, Z., Tripathi, S., Kothari, M., Zama, M. H., & Deb, D. (2024). FRCSyn-onGoing: Benchmarking and comprehensive evaluation of real and synthetic data to improve face recognition systems. *Information Fusion*, *107*, 102322–102322. https://doi.org/10.1016/j.inffus.2024.102322

Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2023). Generating a novel synthetic dataset for rehabilitation exercises using pose-guided conditioned diffusion models: A quantitative and qualitative evaluation. *Computers in Biology and Medicine*, *167*, 107665–107665. https://doi.org/10.1016/j.compbiomed.2023.107665

Miletic, M., & Sariyar, M. (2024). Challenges of Using Synthetic Data Generation Methods for Tabular Microdata. *Applied Sciences*, *14*(14), 5975. https://doi.org/10.3390/app14145975

Min, A. (2023). Artifical Intelligence and Bias: Challenges, Implications, and Remedies. *Journal of Social Research*, *2*(11), 3808–3817. https://doi.org/10.55324/josr.v2i11.1477

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2021). Operationalising AI ethics: barriers, Enablers and next Steps. *AI & Society*, *38*. https://doi.org/10.1007/s00146-021-01308-8

Murray, A., Francks, L., Hassanein, Z. M., Lee, R., & Wilson, E. (2023). Breast cancer surgical decision-making. Experiences of Non-Caucasian women globally. A qualitative systematic review. *European Journal of Surgical Oncology*, *49*(12), 107109–107109. https://doi.org/10.1016/j.ejso.2023.107109

Murtaza, H., Ahmed, M., Khan, N. F., Murtaza, G., Zafar, S., & Bano, A. (2023). Synthetic data generation: State of the art in health care domain. *Computer Science Review*, *48*, 100546. https://doi.org/10.1016/j.cosrev.2023.100546

Offenhuber, D. (2024). Shapes and frictions of synthetic data. *Big Data & Society*, *11*(2). https://doi.org/10.1177/20539517241249390

Ogungbemi, O. S., Ezeugwa, F. A., Olaniyi, O. O., Akinola, O. I., & Oladoyinbo, O. B. (2024). Overcoming Remote Workforce Cyber Threats: A Comprehensive Ransomware and Bot Net Defense Strategy Utilizing VPN Networks. *Journal of Engineering Research and Reports*, *26*(8), 161–184. https://doi.org/10.9734/jerr/2024/v26i81237

Okon, S. U., Olateju, O. O., Ogungbemi, O. S., Joseph, S. A., Olisa, A. O., & Olaniyi, O. O. (2024). Incorporating Privacy by Design Principles in the Modification of AI Systems in Preventing Breaches across Multiple Environments, Including Public Cloud, Private Cloud, and On-prem. *Journal of Engineering Research and Reports*, *26*(9), 136–158. https://doi.org/10.9734/jerr/2024/v26i91269

Olabanji, S. O., Marquis, Y. A., Adigwe, C. S., Abidemi, A. S., Oladoyinbo, T. O., & Olaniyi, O. O. (2024). AI-Driven Cloud Security: Examining the Impact of User Behavior Analysis on Threat Detection. *Asian Journal of Research in Computer Science*, *17*(3), 57–74. https://doi.org/10.9734/ajrcos/2024/v17i3424

Oladoyinbo, T. O., Olabanji, S. O., Olaniyi, O. O., Adebiyi, O. O., Okunleye, O. J., & Alao, A. I. (2024). Exploring the Challenges of Artificial Intelligence in Data Integrity and its Influence on Social Dynamics. *Asian Journal of Advanced Research and Reports*, *18*(2), 1–23. https://doi.org/10.9734/ajarr/2024/v18i2601

Olaniyi, O. O. (2024). Ballots and Padlocks: Building Digital Trust and Security in Democracy through Information Governance Strategies and Blockchain Technologies. *Asian Journal of Research in Computer Science*, *17*(5), 172–189. https://doi.org/10.9734/ajrcos/2024/v17i5447

Olaniyi, O. O., Ezeugwa, F. A., Okatta, C. G., Arigbabu, A. S., & Joeaneke, P. C. (2024). Dynamics of the Digital Workforce: Assessing the Interplay and Impact of AI, Automation, and Employment Policies. *Archives of Current Research International*, *24*(5), 124–139. https://doi.org/10.9734/acri/2024/v24i5690

Olaniyi, O. O., Olaoye, O. O., & Okunleye, O. J. (2023). Effects of Information Governance (IG) on Profitability in the Nigerian Banking Sector. *Asian Journal of Economics, Business and Accounting*, *23*(18), 22–35. https://doi.org/10.9734/ajeba/2023/v23i181055

Olaniyi, O. O., Omogoroye, O. O., Olaniyi, F. G., Alao, A. I., & Oladoyinbo, T. O. (2024). CyberFusion Protocols: Strategic Integration of Enterprise Risk Management, ISO 27001, and Mobile Forensics for Advanced Digital Security in the Modern Business Ecosystem. *Journal of Engineering Research and Reports*, *26*(6), 32. https://doi.org/10.9734/JERR/2024/v26i61160

Olaniyi, O. O., Ugonnia, J. C., Olaniyi, F. G., Arigbabu, A. T., & Adigwe, C. S. (2024). Digital Collaborative Tools, Strategic Communication, and Social Capital: Unveiling the Impact of Digital Transformation on Organizational Dynamics. *Asian Journal of Research in Computer Science*, *17*(5), 140–156. https://doi.org/10.9734/ajrcos/2024/v17i5444

Olateju, O. O., Okon, S. U., Igwenagu, U. T. I., Salami, A. A., Oladoyinbo, T. O., & Olaniyi, O. O. (2024). Combating the Challenges of False Positives in AI-Driven Anomaly Detection Systems and Enhancing Data Security in the Cloud. *Asian Journal of Research in Computer Science*, *17*(6), 264–292. https://doi.org/10.9734/ajrcos/2024/v17i6472

Olateju, O. O., Okon, S. U., Olaniyi, O. O., Samuel-Okon, A. D., & Asonze, C. U. (2024). Exploring the Concept of Explainable AI and Developing Information Governance Standards for Enhancing Trust and Transparency in Handling Customer Data. *Journal of Engineering Research and Reports*, *26*(7), 244–268. https://doi.org/10.9734/jerr/2024/v26i71206

Outeda, C. C. (2024). The EU's AI act: A framework for collaborative governance. *Internet of Things*, *27*, 101291–101291. https://doi.org/10.1016/j.iot.2024.101291

Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (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. https://doi.org/10.3390/bdcc7010015

Paik, K. E., Hicklen, R. S., Kaggwa, F., Puyat, C. V., Nakayama, L. F., Ong, B. A., Shropshire, J. N., & Villanueva, C. (2023). Digital Determinants of Health: Health data amplifies existing health disparities—A scoping review. *PLOS Digital Health*, *2*(10), e0000313–e0000313. https://doi.org/10.1371/journal.pdig.0000313

Paladugu, P., Ong, J., Nelson, N. G., Kamran, S. A., Waisberg, E., Zaman, N., Kumar, R., Dias, R. D., Lee, A. G., & Tavakkoli, A. (2023). Generative Adversarial Networks in Medicine: Important Considerations for this Emerging Innovation in Artificial Intelligence. *Annals of Biomedical Engineering*, *51*. https://doi.org/10.1007/s10439-023-03304-z

Perri, L. (2024, April 12). *Gartner*. Https://Www.gartner.com/En/Articles/3-Bold-And-Actionable-Predictions-For-The-Future-of-Genai#:~:Text=By%202026%2C%2075%25%20of%20businesses%20will%20use%20generative,Data%2C%20up%20from%20less%20than%205%25%20in%202023. https://www.bing.com/ck/a?

Pezoulas, V. C., Zaridis, D. I., Mylona, E., Androutsos, C., Apostolidis, K., Tachos, N. S., & Fotiadis, D. I. (2024). Synthetic data generation methods in healthcare: A review on open-source tools and methods. *Computational and Structural Biotechnology Journal, 23*, 2892–2910. https://doi.org/10.1016/j.csbj.2024.07.005

Pina, E., Ramos, J., Jorge, H., Váz, P., Silva, J., Wanzeller, C., Abbasi, M., & Martins, P. (2024). Data Privacy and Ethical Considerations in Database Management. *Journal of Cybersecurity and Privacy, 4*(3), 494–517. https://doi.org/10.3390/jcp4030024

Raghavan, K., Balasubramanian, S., & Veezhinathan, K. (2024). Explainable artificial intelligence for medical imaging: Review and experiments with infrared breast images. *Computational Intelligence*, *40*(3). https://doi.org/10.1111/coin.12660

Rane, N. (2023). Transformers for Medical Image Analysis: Applications, Challenges, and Future Scope. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4622241

Raza, S., Garg, M., Reji, D. J., Bashir, S. R., & Ding, C. (2024). Nbias: A natural language processing framework for BIAS identification in text. *Expert Systems with Applications*, *237*(Part B), 121542. https://doi.org/10.1016/j.eswa.2023.121542

Salami, A. A., Igwenagu, U. T. I., Mesode, C. E., Olaniyi, O. O., & Oladoyinbo, O. B. (2024). Beyond Conventional Threat Defense: Implementing Advanced Threat Modeling Techniques, Risk Modeling Frameworks and Contingency Planning in the Healthcare Sector for Enhanced Data Security. *Journal of Engineering Research and Reports*, *26*(5), 304–323. https://doi.org/10.9734/jerr/2024/v26i51156

Samuel-Okon, A. D., Akinola, O. I., Olaniyi, O. O., Olateju, O. O., & Ajayi, S. A. (2024). Assessing the Effectiveness of Network Security Tools in Mitigating the Impact of Deepfakes AI on Public Trust in Media. *Archives of Current Research International*, *24*(6), 355–375. https://doi.org/10.9734/acri/2024/v24i6794

Samuel-Okon, A. D., Olateju, O. O., Okon, S. U., Olaniyi, O. O., & Igwenagu, U. T. I. (2024). Formulating Global Policies and Strategies for Combating Criminal Use and Abuse of Artificial Intelligence. *Archives of Current Research International*, *24*(5), 612–629. https://doi.org/10.9734/acri/2024/v24i5735

Sankar, B. S., Gilliland, D., Rincon, J., Hermjakob, H., Yan, Y., Adam, I., Lemaster, G., Wang, D., Watson, K., Bui, A., Wang, W., & Ping, P. (2024). Building an Ethical and Trustworthy Biomedical AI Ecosystem for the Translational and Clinical Integration of Foundation Models. *Bioengineering*, *11*(10), 984–984. https://doi.org/10.3390/bioengineering11100984

Selesi-Aina, O., Obot, N. E., Olisa, A. O., Gbadebo, M. O., Olateju, O. O., & Olaniyi, O. O. (2024). The Future of Work: A Human-centric Approach to AI, Robotics, and Cloud

Computing. *Journal of Engineering Research and Reports*, *26*(11), 62–87. https://doi.org/10.9734/jerr/2024/v26i111315

Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*, *27*(12), 2176–2182. https://doi.org/10.1038/s41591-021-01595-0

Shah, M., & Sureja, N. (2024). A Comprehensive Review of Bias in Deep Learning Models: Methods, Impacts, and Future Directions. *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-024-10134-2

Siddique, S., Haque, M. A., George, R., Gupta, K. D., Gupta, D., & Faruk, M. J. H. (2024). Survey on Machine Learning Biases and Mitigation Techniques. *Digital*, *4*(1), 1–68. https://doi.org/10.3390/digital4010001

Sulastri, R., Janssen, M., van de Poel, I., & Ding, A. (2024). Transforming towards inclusion-by-design: Information system design principles shaping data-driven financial inclusiveness. *Government Information Quarterly*, *41*(4), 101979. https://doi.org/10.1016/j.giq.2024.101979

Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach for Student's Behavior Recognition. *Big Data and Cognitive Computing*, *7*(1), 48. https://doi.org/10.3390/bdcc7010048

Ueda, D., Kakinuma, T., Fujita, S., Kamagata, K., Fushimi, Y., Ito, R., Matsui, Y., Nozaki, T., Nakaura, T., Fujima, N., Tatsugami, F., Yanagawa, M., Hirata, K., Yamada, A., Tsuboyama, T., Kawamura, M., Fujioka, T., & Naganawa, S. (2023). Fairness of

Artificial Intelligence in healthcare: Review and Recommendations. *Japanese Journal of Radiology*, *42*(1). https://doi.org/10.1007/s11604-023-01474-3

Van Giffen, B., Herhausen, D., & Fahse, T. (2022). Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *Journal of Business Research*, *144*(1), 93–106. https://doi.org/10.1016/j.jbusres.2022.01.076

Wang, Z., Draghi, B., Rotalinti, Y., Lunn, D., & Myles, P. (2024). High-Fidelity Synthetic Data Applications for Data Augmentation. *Artificial Intelligence (London)*. https://doi.org/10.5772/intechopen.113884

Wu, S., Kurugol, S., & Tsai, A. (2024). Improving the radiographic image analysis of the classic metaphyseal lesion via conditional diffusion models. *Medical Image Analysis*, *97*, 103284. https://doi.org/10.1016/j.media.2024.103284

Yoon, J., Mizrahi, M., Ghalaty, N. F., Jarvinen, T., Ravi, A. S., Brune, P., Kong, F., Anderson, D., Lee, G., Meir, A., Bandukwala, F., Kanal, E., Arık, S. Ö., & Pfister, T. (2023). EHR-Safe: generating high-fidelity and privacy-preserving synthetic electronic health records. *Npj Digital Medicine*, *6*(1), 1–11. https://doi.org/10.1038/s41746-023-00888-7

Zhang, Q., & Wang, T. (2024). Deep Learning for Exploring Landslides with Remote Sensing and Geo-Environmental Data: Frameworks, Progress, Challenges, and Opportunities. *Remote Sensing*, *16*(8), 1344. https://doi.org/10.3390/rs16081344