

Review Form 3

Journal Name:	Journal of Engineering Research and Reports
Manuscript Number:	Ms_JERR_127156
Title of the Manuscript:	Leveraging Synthetic Data as a Tool to Combat Bias In Artificial Intelligence (AI) Model Training
Type of the Article	

General guidelines for the Peer Review process:

This journal's peer review policy states that **NO** manuscript should be rejected only on the basis of '**lack of Novelty**', provided the manuscript is scientifically robust and technically sound. To know the complete guidelines for the Peer Review process, reviewers are requested to visit this link:

<https://r1.reviewerhub.org/general-editorial-policy/>

Important Policies Regarding Peer Review

Peer review Comments Approval Policy: <https://r1.reviewerhub.org/peer-review-comments-approval-policy/> Benefits for Reviewers: <https://r1.reviewerhub.org/benefits-for-reviewers>

Review Form 3

PART 1: Review Comments

Compulsory REVISION comments	Reviewer's comment	Author's Feedback <i>(Please correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)</i>
Please write a few sentences regarding the importance of this manuscript for the scientific community. Why do you like (or dislike) this manuscript? A minimum of 3-4 sentences may be required for this part.	The paper touches on tough questions of data quality and characteristics of synthetic data. The authors evaluate the capacity of synthetic datasets to enhance the fairness of ML classifiers, which is an extremely important topic for applied ML researchers. The authors obtain nontrivial results, showing that the synthetic data can make a difference in fairness metrics associated with a classifier.	
Is the title of the article suitable? (If not please suggest an alternative title)	The empirical analysis proposed by the paper is only applicable to multivariate datasets with mostly integer/categorical features usually stored in relational databases. The paper largely ignores the primary data modalities of AI systems, such as text, sound, image, video, and robotic systems data. The name of the paper would be a lot more concise if it was talking about classifiers instead of AI as a whole.	
Is the abstract of the article comprehensive? Do you suggest the addition (or deletion) of some points in this section? Please write your suggestions here.	The abstract reflects well on the work. The main question I had after reading the abstract was: "I wonder what models the authors trained on synthetic and natural datasets?" Adding that the model was a logistic regression would add clarity to the abstract.	
Are subsections and structure of the manuscript appropriate?	The general structure of the manuscript is appropriate, although the Methodology section should be significantly extended and split into subsections.	
Please write a few sentences regarding the scientific correctness of this manuscript. Why do you think that this manuscript is scientifically robust and technically sound? A minimum of 3-4 sentences may be required for this part.	No factual errors were detected in the manuscript. The only part that caught my attention was the definition of the Inception Score, but it could be a typo or misrendering of my Word application. The manuscript is scientifically robust because the authors introduced the benchmark datasets and metrics they were using throughout their study and provided results in a way that could hypothetically be reproduced in the community.	
Are the references sufficient and recent? If you have suggestions of additional references, please mention them in the review form.	The literature review is extensive and recent.	
Minor REVISION comments Is the language/English quality of the article suitable for scholarly communications?	There is a multitude of typos throughout the text that could be fixed using an AI grammar check.	
Optional/General comments	<ol style="list-style-type: none"> 1. The manuscript is missing a lot of details critical for reproducibility: <ol style="list-style-type: none"> a. The details about training the GAN and the logistic regression are missing. b. How was the synthetic dataset for COMPAS Recidivism Dataset generated? Was GAN used as well? I could not find the details anywhere. c. Generation of synthetic dataset for MIMIC-III Clinical Database is not described. d. How much synthetic data were generated? e. which parts were used for training/testing of the models (synthesizer GAN and classifier Logistic Regression) f. open-sourcing the code would considerably increase impact of the publication 2. Expanding on the datasets under study would improve the paper. It is not clear what the composition of the dataset is, what the features are, what kind of information they contain and what motivations were behind choosing exactly these datasets among dozens of others. 3. Several questions about the interpretation of the results that the authors should reflect upon in the paper: <ol style="list-style-type: none"> a. What are the characteristics of the synthetic dataset that led to improved Demographic Parity? b. The results of Assessment of Quality and Fairness of Synthetic Data essentially duplicating the result reported in Evaluation of Synthetic Data Generation Techniques that the synthetic and natural datasets are similar. What exactly was the hypothesis that motivated this experiment, given the prior results? c. Table 1 shows us numbers for KLD and KST. What are they compared to? This numerical evaluation lacks baselines. It's clear that the distributions are similar, but what is surprising about 	

Review Form 3

	<p>it? What generalizable message can be taken from this fact? Similarly, the Inception Score is hard to interpret by itself without any point of reference. The claim of the authors that it is high does not look too convincing without this interpretation.</p> <p>4. General criticism: The fact that GANs can generate data of a similar distribution is barely novel or surprising in 2024. I do not believe a result like that should be emphasized.</p> <p>5. Comments about representation of the results:</p> <ol style="list-style-type: none"> a. Figures 1 and 2 do not provide any additional information besides that given in Table 1 and only distract attention. b. Figures 3-5 do not help understanding the results. Using bold fonts for higher values in Table 2 might. c. Figures 6 and 7 repeat the results reported in Table 3 without additional information or clarification. They are redundant. 	
--	--	--

PART 2:

	Reviewer's comment	Author's comment <i>(if agreed with reviewer, correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)</i>
Are there ethical issues in this manuscript?	<i>(If yes, Kindly please write down the ethical issues here in details)</i>	

Reviewer Details:

Name:	Igor Molybog
Department, University & Country	University of Hawai'i at Manoa, United States of America