GenAI for Synthetic Data to Mitigate Bias

John Staunton | Technological University Dublin | X00229681@mytudublin.ie

In collaboration with CeADAR, Ireland's Centre for AI

Abstract: This project investigates using GenAI to create Synthetic Data (SD) to mitigate bias in AI models in the healthcare sector. We will undertake a review of the types of biases that can arise and how these can be addressed using GenAI and SD. We will work with a real-world medical dataset to firstly assess it for bias and to then generate new SD to help address this. We will assess the quality of the SD and use it to re-train the AI model. The desired outcome is that the AI model is more accurate and robust, with much reduced levels of bias, leading to improved health outcomes for patients in the future.

Sources of Bias

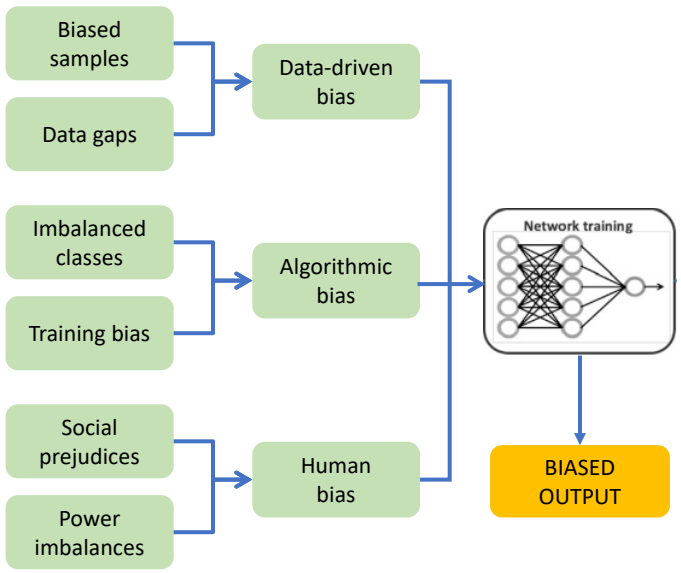


Fig. 1 Sources of bias in training ML algorithms [4]

Example of Algorithmic Bias

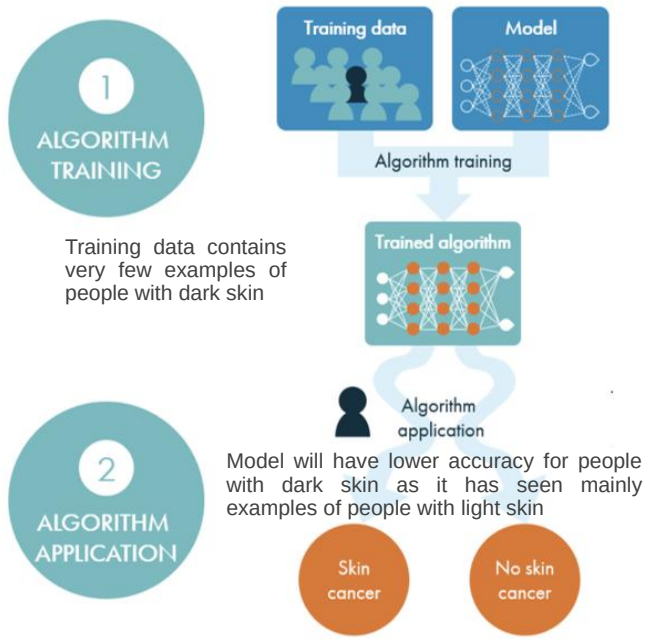
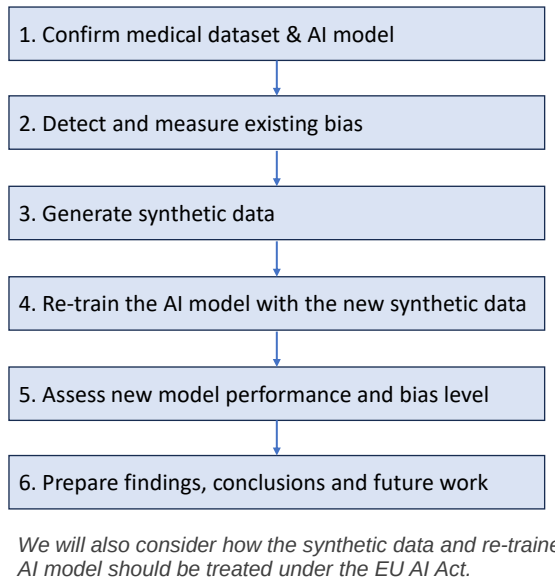


Fig. 2 Example of Bias in Skin Cancer Detection [5]

Research Methodology



We will also consider how the synthetic data and re-trained AI model should be treated under the EU AI Act.

Fig. 3 Key Steps in Research Methodology

INTRODUCTION

In the rapidly changing world of AI in healthcare, data forms the bedrock upon which accurate and robust AI models are built [1]. However, acquiring sufficient high-quality data to train the new AI models can be challenging, time consuming and may inadvertently introduce bias that impacts the performance of the AI models. GenAI can be used to create additional synthetic data to retrain the AI models, with the goal of reducing the level of bias and improving the health outcomes. As the synthetic data is new, we must carefully consider how to measure its quality and effectiveness. It is also important to understand how this synthetic data will be treated under the EU AI Act and what regulatory, legal and ethical responsibilities apply to it.

LITERATURE REVIEW

The origins of synthetic data are often attributed to Rubin in 1993 who devised a method to create "synthetic" versions of US census data so it could be made public without compromising individual privacy [2] [3]. Synthetic data is now also being used to augment real data when training AI models and is particularly useful in addressing data-driven and algorithmic biases [4]. A real-world example of bias from having a class imbalance in the training data is shown above [5]. Techniques for generating synthetic data can be grouped into Statistical Models (e.g. Interpolation, Monte Carlo simulation and Gaussian Kernels) and Deep Learning Based Generation (e.g. GANs, Diffusion Models, LLMs) [6]. A comprehensive survey of 417 synthetic data generation models over the last decade [7] will help us in selecting the most suitable techniques to use, depending on the data formats and subject matter involved. Synthetic data, however, is not the only way to mitigate bias in ML models e.g. a recent neuro-imaging study found that re-weighing the sample weights in a CNN classification model was a successful bias mitigation approach [8].

HUMAN CENTERED APPROACH

Fairer and more equitable AI models | Responsible AI aligned with EU AI Act | UN Sustainable Dev Goals

PRELIMINARY CONSIDERATIONS

A number of options for the medical dataset are available from CeADAR, TUD & Messidor / Nathean. These will be further assessed in the coming weeks.

CONCLUSIONS

This project will tell us which synthetic data generation techniques are most suitable to address bias in healthcare AI models. It will apply these to a real-world dataset and provide practical insights and results. These can be used to build more accurate and robust AI models in healthcare and deliver improved health outcomes for all.

References.

[1] Z. Wang et al., "High-Fidelity Synthetic Data Applications for Data Augmentation," Jan. 2024, doi: 10.5772/INTECHOPEN.113884.
[2] J. Drechsler and A. C. Haensch, "30 Years of Synthetic Data," Statistical Science, vol. 39, no. 2, pp. 221–242, 2024, doi: 10.1214/24-STS927.
[3] Rubin, "RUBIN 1993 - discussion-statistical-disclosure-limitation2," Journal of Official Statistics, Vol. 9 No.2 1993, pp 461 - 468, 1993.
[4] N. Norori, Q. Hu, F. M. Aellen, F. D. Faraci, and A. Tzovara, "Addressing bias in big data and AI for health care: A call for open science," Patterns, vol. 2, no. 10, Oct. 2021, doi: 10.1016/J.PATTER.2021.100347/ASSET/ED1F8E34-A75E-4232-8EF4-39A1F7D56420/MAIN.ASSETS/GR2.JPG.
[5] R. De Boer, "Understanding the role of AI bias in healthcare," Quantib.com Blog. Accessed: Jan. 29, 2025. [Online]. Available: https://www.quantib.com/blog/understanding-the-role-of-ai-bias-in-healthcare
[6] S. Hao et al., "Synthetic Data in AI: Challenges, Applications, and Ethical Implications," Jan. 2024, [Online]. Available: http://arxiv.org/abs/2401.01629
[7] A. Bauer et al., "Comprehensive Exploration of Synthetic Data Generation: A Survey," Jan. 2024, [Online]. Available: http://arxiv.org/abs/2401.02524
[8] E. A. M. Stanley et al., "Towards objective and systematic evaluation of bias in artificial intelligence for medical imaging," Journal of the American Medical Informatics Association, vol. 31, no. 11, pp. 2613–2621, Nov. 2024, doi: 10.1093/JAMIA/OCAE165.

Acknowledgements. The HCAIM (the Human-Centered AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This poster was created as part of the Blended Intensive Programme organized under the Erasmus + Programme of the European Union