Integrating Automation Scripts in Production Environments:
Increase Innovation and Streamline Ideas

Reframing Algorithmic Bias with Cultural Clothes in Public Facial Recognition
Technology: Improving Public Goods through a Private Sector Lens

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Joshua Matthew De Vera

05/09/2023

ADVISORS

Kathryn A. Neeley, Department of Engineering and Society

Alice Fox, Department of Engineering and Society

Analysis of Bias in Facial Recognition and Evolution of Bias Reduction

Overview:

      This project analyzes the utilization of biometric security in the modern age as well as controversial effects due to its development. The rise of biometric security through facial recognition is due to a need for stronger security measures without decreasing ease of utilization. As cyber exploits evolve, the use of biometric security, a feature specific to only one individual with limited external reproducibility, has increased due to the uniqueness of each person's static and dynamic physical attributes. This research aims to analyze the bias in facial detection software development and modern methods utilized to reduce bias.

Positionality:

      My research goals to reduce the technological gap between different societal classes stem from my diverse socio-economic background. As a Filipino man raised in the middle class of the United States and whose parents were in the upper and lower economic classes in the Philippines, I have experienced the disparities and opportunities different socio-economic groups are given in developed and developing countries. By living in the US, I have access to an abundance of technology in schools, funded research opportunities, and programming competitions that allow me to explore my curiosity in engineering and foster software development skills. The drastic differences between the socio-economic classes between the US and Philippines within each country motivate me to create sustainable and accessible technology that can be used to bridge these gaps across the world. I acknowledge that I have not personally experienced these challenges but have furthered my understanding through talking with family and community members who have experienced it. Likewise, I found it important to diversify my perspective to understand the many factors that lead to these disparities. Alongside my pursuit of computer science in college, I took multiple courses including nutrition, public health in the US, and sustainable energy systems to further my knowledge in the different factors that can significantly aid these marginalized groups. The information and perspectives taken away from these courses allow me to translate my software development and research projects into realistic, actionable products that will benefit these communities. By acknowledging that the choices groups make are limited by the choices they are given, my research aims to lower barriers to higher education and improve quality of life.

Problematization:

      The use of biometric security has become a principle technique for security on all levels. Most phones have facial recognition software to unlock the device, but it is also being implemented in high-security government buildings that require federal clearances. Whether it's used by the common person or important personnel, facial recognition has become very common in security for all devices. In recent years, there has been growing controversy over how these facial detection software has been created due to its inability to track people with darker complexion and facial accessories (hijabs, Bindi, kippah, etc.). Bias in facial detection software development has led to failures in the private and public sector including loss of access to hardware and software and the criminalization of these marginalized groups.

Guiding Question or Main Argument:

What internal and external factors have influenced the development of facial recognition technology and bias reduction techniques and what are its effects on marginalized communities?

Projected Outcomes:

By analyzing the development of facial recognition software, I aim to recognize bias within these systems and analyze pathways to reduce this bias. By addressing bias in the algorithms, security devices with this software will become more secure and accessible for marginalized communities that aren't properly represented and allow for more accurate results. Likewise, this will benefit facial recognition users and decrease the financial and social cost of failures.

Technical Project Description:

At the University of Virginia, I have taken a machine learning course that explores the datasets, data processing, and training algorithms of image recognition software as well as bias that different models can produce. The problem with reducing bias through an increased training dataset is the cost (energy, computing resources, and time). Instead, I analyzed a very naïve approach to reducing bias by utilizing external bias adjusting algorithms. This is limited by the knowledge of bias known to the developer that induces bias in the adjustor algorithm. Likewise, I have had an internship that utilized image recognition software to aid nurses in nursing homes. Through these experiences, I've noticed the bias in facial recognition and how it can negatively impact underrepresented populations as failure can increase costs of repeated entry and avoidable injuries. After studying static image recognition through the college course and dynamic image recognition through my internship, I understand the software processes that go into creating image recognition software and how bias can be introduced, implicitly and explicitly. With this technical knowledge of the domain, I intend to do research on how bias is introduced in different steps of development and modern methodologies that provide cheap, adaptive approaches to bias reduction.

Preliminary Literature Review & Findings:

Bias in facial recognition can be introduced in many steps throughout development. Biased datasets that underrepresented populations poorly assess faces it has not experienced and therefore not trained to recognize. In early development of facial recognition software, sensitive feature parameters in data processing is a common technique used to mitigate outliers and increase accuracy, but important for facial recognition for South Asian and American populations. Algorithms that utilize neural networks have integrated bias based on the neural network architecture of each program. Notably, ResNet50 and VGG-16 are architectures whose neural network is able to reduce racial bias the most despite not being implemented in commercial facial recognition software due to costs needed to implement it. Finally, the introduction of dynamic images with suboptimal facial recognition angle causes a lot of failures when used outside of production. Scientists determined that the best facial recognition software has its accuracy reduced by approximately 20% when in these common, suboptimal positions

when used in practice. Despite the growing advancements in technology used to reduce facial recognition bias has improved in theory, the results from in-field use are still below standard.

The social implications of bias in facial recognition have been studied by a few scholars in ethical engineering. The effect of failure in facial recognition to determine race results in these misclassified populations feeling less human and not worthy of being accounted for in technology. With facial recognitions' wide use and integration in technology and society, the continuation of biased software acknowledges that it is acceptable to overlook these groups and sets the precedent for future biometric technology. My research acknowledges the internal and external factors of bias reduction while emphasizing the social effects it may have as it is integrated. By investigating techniques to reduce failure and increase representation, I present the possible sociological effects of integrating more unbiased technology in society.

STS Project Proposal:

STS defines the relationship between technology and society and how this relationship shapes each other. By understanding this relationship in STS, we can develop technology that better serves society and a society that better develops technology. This research utilizes STS principles to analyze facial recognition technologies and strategies utilized to reduce image recognition bias and better serve underrepresented groups.

Currently analysis and research on the bias of facial recognition demonstrates that people of darker complexion as well as people who aren't cis-white or of Anglo-European descent have lower image recognition accuracy scores (Leslie 2020 & Salvador et al. 2021). The effect of this bias is increased cost of failure especially in police enforcement (Gwyn 2022) and the dehumanization of these underrepresented groups on the individual and social levels (Waelen 2023). I will approach this problem with a technological and ethics lens. From a technological focus, this research aims to analyze models used to reduce bias within facial recognition while acknowledging the costs associated with each method. Likewise, it is important to approach this research from a race and ethics focus to determine how software reflects the bias of its creators, the social effects of mass-utilization of racially-biased software, and the ethical use and creation of new development methodologies to reduce bias. Primary authors in this research are Dr. David Leslie who provides insight on the development of facial recognition software and acknowledges the racial-disparities that affected its development from a technological standpoint (Leslie 2020). Another important author to reference is Dr. Rosalie Waelen who has done research on ethical technology and analysis on the sociological effect of racially-biased systems on underrepresented groups. These two authors' work build the foundation for the purpose of my research, and it allows for more analysis on techniques created to reduce the bias as well as its social implications.

This research will be conducted with Friedman's version of Value Sensitive Design methodology that emphasizes the important values utilized when developing facial recognition technology.

From the origins of facial recognition technology that was trained for white, cis-males populations, the technology has evolved based on the values of the inventors, so it is important to assess how these values determined design choices (Leslie 2020). This methodology will contribute to the analysis of modern facial recognition architecture development and emphasize the tradeoffs in implementations with respect to database size, database diversity, neural network architecture, and societal norms after each iteration. By conducting research with this methodology in mind, I will be able to produce research that emphasizes how bias is introduced into these technological systems through its creators and analyze how biased facial recognition software influences societal norms and racial representation. Friedmans's Value Sensitive Design will center this research on the values of this technology's creators, the implicit values of society and the perpetuation of their values in future technology.

The method I will be using is value source analysis in order to determine the values of the different stakeholders in the development of facial recognition software. I plan to do this through literature review and discourse analysis that will allow me to understand the values of different groups and how each was able to influence the different decisions made while developing different iterations of the software. The literature review will provide insight on the final design decisions and

Barriers & Boons

The energy, time, and movement needed to conduct this project limit this research. Due to the expensive computational power needed for facial recognition software, time and computational power constraints limit the creation of created software with negligible bias. Therefore, this research will focus on the theoretical creation of the software as well as describe the process and algorithms that can be implemented to achieve this goal. Utilizing the literature in this prospectus as a baseline can serve as a platform for finding more plausible techniques to reduce bias in facial recognition technology for commercial use in order to supplement my research.

Another limitation is assessing the effectiveness of this research since I do not have the knowledge or resources to conduct research on the effectiveness of reducing bias on reversing the effects of current social issues. However, the utilization of more literature from ethical engineering and sociologists to better understand the effects of misrepresentation in society and assess the potential effects of reducing bias in facial recognition technology. Since this research won't be replicable on a testing-scale within one semester, it will be important to find research that can translate the potential effects of these improvements.

References

1. Leslie, D. (2020). Understanding bias in facial recognition technologies. *arXiv preprint arXiv:2010.07023*.

2. Perkowitz, S. (2021). The Bias in the Machine: Facial Recognition Technology and Racial Disparities. *MIT Case Studies in Social and Ethical Responsibilities of Computing, Winter 2021.* https://doi.org/10.21428/2c646de5.62272586

3. Waelen, R.A. The struggle for recognition in the age of facial recognition technology. *AI Ethics* 3, 215–222 (2023). https://doi.org/10.1007/s43681-022-00146-8

4. Gwyn, T., & Roy, K. (2022). Examining Gender Bias of Convolutional Neural Networks via Facial Recognition. *Future Internet*, *14*(12), 375. MDPI AG. Retrieved from http://dx.doi.org/10.3390/fi14120375

5. Kalinga, A. T. N. (2021). *Towards An Ethnic-Bias Free Approach For Facial Recognition* (Doctoral dissertation).

6. Singh, R., Majumdar, P., Mittal, S., & Vatsa, M. (2022). Anatomizing Bias in Facial Analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(11), 12351-12358. https://doi.org/10.1609/aaai.v36i11.21500

7. A. Khalil, S. G. Ahmed, A. M. Khattak and N. Al-Qirim, "Investigating Bias in Facial Analysis Systems: A Systematic Review," in IEEE Access, vol. 8, pp. 130751-130761, 2020, doi: 10.1109/ACCESS.2020.3006051.

8. Xu, T., White, J., Kalkan, S., Gunes, H. (2020). Investigating Bias and Fairness in Facial Expression Recognition. In: Bartoli, A., Fusiello, A. (eds) Computer Vision – ECCV 2020 Workshops. ECCV 2020. Lecture Notes in Computer Science(), vol 12540. Springer, Cham. https://doi.org/10.1007/978-3-030-65414-6_35

9. Wang, A., Liu, A., Zhang, R. *et al.* REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets. *Int J Comput Vis* 130, 1790–1810 (2022). https://doi.org/10.1007/s11263-022-01625-5

10. J. G. Cavazos, P. J. Phillips, C. D. Castillo and A. J. O'Toole, "Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias?," in *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 1, pp. 101-111, Jan. 2021, doi: 10.1109/TBIOM.2020.3027269.

11. Singh, R., Agarwal, A., Singh, M., Nagpal, S., & Vatsa, M. (2020). On the Robustness of Face Recognition Algorithms Against Attacks and Bias. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(09), 13583-13589. https://doi.org/10.1609/aaai.v34i09.7085

12. Wehrli, S., Hertweck, C., Amirian, M. *et al.* Bias, awareness, and ignorance in deep-learning-based face recognition. *AI Ethics* 2, 509–522 (2022). https://doi.org/10.1007/s43681-021-00108-6

13. Salvador, T., Cairns, S., Voleti, V., Marshall, N., & Oberman, A. (2021). Bias Mitigation of Face Recognition Models Through Calibration. *arXiv preprint arXiv:2106.03761*.

14. Vangara, K., King, M. C., Albiero, V., & Bowyer, K. (2019). Characterizing the variability in face recognition accuracy relative to race. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 0-0).

15. Kolla, M., & Savadamuthu, A. (2023). The Impact of Racial Distribution in Training Data on Face Recognition Bias: A Closer Look. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 313-322).