Thesis Project Portfolio

Algorithmic Bias in Facial Recognition: Exploring Racial and Gender Disparities through CNN Models

(Technical Report)

Gender Shades and George Floyd's Death: Raising Public Awareness of Algorithmic Bias

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Claire Yoon

Fall, 2023

Department of Computer Science

Table of Contents

Sociotechnical Synthesis
(Executive Summary)

Examination of Facial Recognition Technology and Contributing Events to Raising Public Awareness of Algorithmic Bias

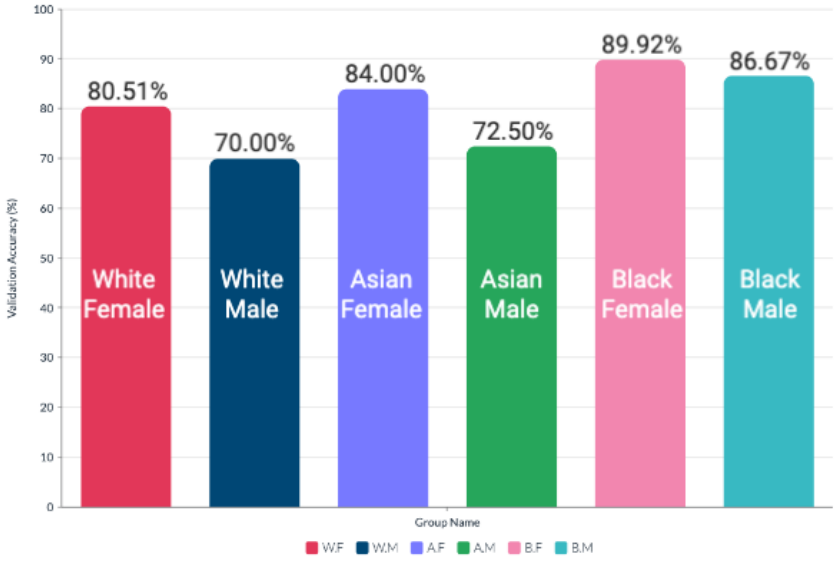
We now have an option for all humanity to "make it" successfully on this planet in this lifetime. ... Whether it is to be Utopia or Oblivion will be a touch-and-go relay race right up to the final moment.
–R. Buckminster Fuller, *Critical Path* (1981)

The introduction of Artificial Intelligence (AI) marked a significant innovation in our society and is used in various areas. Despite its numerous benefits, AI systems have shown discriminatory results. A key study "Gender Shades" by Buolamwini, revealed significant disparities in identifying subjects using Facial Recognition Technology. Specifically, it showed that the accuracy of darker-skinned females was up to 34.4% lower compared to lighter-skinned males. Driven by curiosity, my technical project titled *Celebrity Facial Recognition*, working with two fellow students, focused on verifying whether our machine learning models have similar patterns to prior research showing facial recognition systems misidentified Black faces more than White faces. Expanding the scope from one of AI's technologies, facial recognition technology, to overall AI applications, my STS research investigated correlations between certain events and public awareness of algorithmic bias utilizing frequency analysis.

Inspired by "Gender Shades," my team and I designed machine learning models using a convolutional neural network (CNN), a deep learning architecture commonly used in tasks related to image classification. We aimed to create machine learning models that can identify the correct subject from a group, and to compare whether the machine learning models we developed are consistent from prior research. We divided six different groups based on their race

and gender: White female, White male, Asian female, Asian male, Black female, and Black male. In addition, we chose celebrities as the representatives of each group with similarities in their appearances and age. With the collected photographs of six different groups, we trained and evaluated the models. Overall, the models performed well in identifying the correct celebrity within a group, achieving approximately 80% validation accuracy. This metric measures the model's effectiveness on testing data that was not used during training. However, the technical project produced different results than prior research conducted, which was that facial recognition technology falsely identified Black and Asian faces, particularly those of women, 10 to 100 times more frequently than White male faces. In our project, the lowest validation accuracy was observed in the White male group, while the second lowest accuracy was found in the Asian male group.

Figure 1. Comparison of Validation Accuracy between Six Groups (Created by Author)



Note. The graph shows 6 different racial and gender groups. The lowest validation accuracy is found in the White male group (70.00%), followed closely by the Asian male group (72.50%). This outcome contrasts with prior research, which often found that Facial Recognition Technology often misidentified Black female subjects.

In my STS research project, I investigated correlations between particular events and public awareness of algorithmic bias. As the first step in the research, I looked for when public interest in algorithmic bias increased by year. Through Google Trends, I found that there was a consistent increase in searching "algorithmic bias" starting from 2015. In addition to this, there were substantial spikes in 2018 and 2020. To reveal the factors behind these peaks, I researched what events related to algorithmic bias occurred in those particular years. As a result, I discovered "Gender Shades" and George Floyd's death, both of which significantly contributed to raising public awareness of algorithmic bias in AI systems. "Gender Shades" demonstrated that all the AI programs used in the study showed discriminatory performance towards darker-skinned females, and there was a peak in searching "algorithmic bias" after the publication of the study. Another notable point was Floyd's death on May 25, 2020. There was a remarkable surge in searches for "algorithmic bias" in June 2020 which is immediately after his death. Although the event did not seem directly related to algorithmic bias, his death substantially influenced our society, igniting worldwide movements against racism. Algorithmic bias and his death were connected with racial injustice since showed discriminatory results.

Overall, my projects emphasize the need for a comprehensive approach incorporating technical and social aspects. Although my projects uncover different aspects – one shows a difference from prior studies and the other analyzes the contributing factors to raising public interest in algorithmic bias – they have the same theme and purpose. Both projects helped increase my understanding of the topic and emphasized the need of a deeper understanding of the problem. More specifically, my STS research showed that algorithmic bias is not simply solely based on technical factors such as algorithms and AI; there are also various, complex, and subtle factors behind algorithmic bias that need to be considered.