

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

A Research Paper submitted to the Department of Engineering and Society

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

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

**Claire Yoon**

Fall 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

## Introduction

“We can deceive ourselves into thinking they’re not doing harm, or we can fool ourselves into thinking, because it’s based on numbers, that it is somehow neutral. AI is creeping into our lives. And even though the promise is that it’s going to be more efficient; it’s going to be better, if what’s happening is we’re automating inequality through weapons of math destruction and we have algorithms of oppression, this promise is not actually true and certainly not true for everybody” (“Joy Buolamwini,” 2020).

With the idea of “Can Machines Think?,” the concept of artificial intelligence (AI) was introduced in the 1950s, and the research on AI systems in various fields began to grow in the 1980s (Anyoha, 2017). Over the decades, AI research and applications have evolved continuously, resulting in significant breakthroughs and advances in computer vision and machine learning. In the 21st century, AI is now widely used in various areas from everyday uses such as unlocking mobile devices through facial recognition technology to healthcare, finance, transportation, and education. Because of its extensive versatility and convenience, AI has become a part of our daily lives and has been implemented across many industries.

However, despite all the positives, the following problems have been revealed in AI applications: racial and gender discrimination. As algorithmic bias in AI systems has led to discriminative incidents, it has brought attention to the public. Although there have been some cases that affected negatively minority groups, it is not clear what harms are happening and what people are responding to. As such, it is difficult to say that the existence of algorithmic bias in AI systems itself is actually harmful to society. However, if the results of discriminating against marginalized groups are repeatedly and consistently shown in AI programs, this will certainly affect not only minority groups but also our entire society by creating tension while spreading stereotypes and distrust of the systems. For this reason, it is important to understand what cases are connected with algorithmic bias and what information can be drawn from them. In this paper,

I examine what events have contributed significantly to raising public awareness of the algorithmic bias in AI systems and what they suggest. More specifically, through frequency analysis, I seek to find correlations between those events and public attention on the problem to answer what the problem exactly is and what it reflects, so it will allow us to understand the issue better.

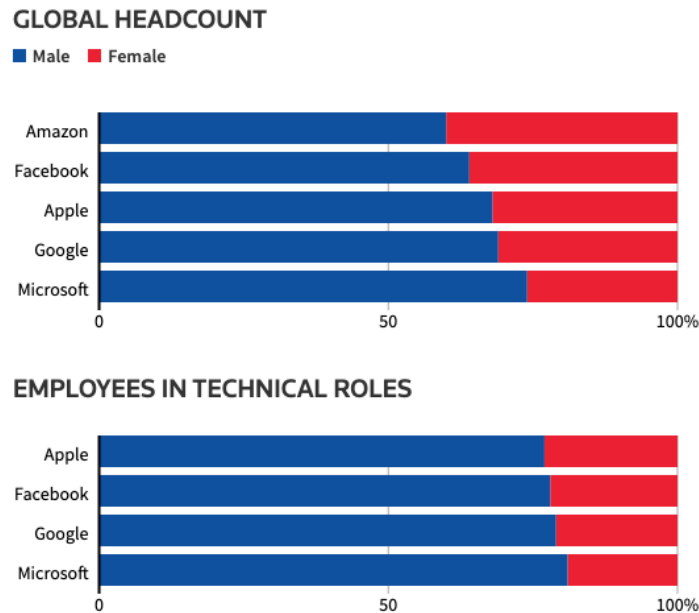
## **Problem Definition**

### *Spreading Algorithmic Bias in AI Systems and the Consequences of Its Impact*

Even in an increasingly developed AI technology and global leading tech companies, it is not easy to mitigate algorithmic bias in AI applications. In the words of Selbst, fixing discrimination in algorithmic systems is not something that can be solved easily and it's a process ongoing, just like discrimination in any other aspect of society (Selbst et al., 2019).

In 2014, a team of engineers at Amazon started working on a project for an automated recruiting engine to review job applicants' résumés, but the system learned to favor male candidates, as a result, negatively impacted résumés that included words related to women such as "women's chess club captain." A year later, the company disbanded the team because the recruiting tool did not work in a gender-neutral way, especially discriminating against women applying for technical positions. This is because Amazon's computer models were trained to validate applicants by observing patterns on résumés submitted to the company over a decade. "Most came from men, a reflection of male dominance across the tech industry" (Dastin, 2018). The gender gap with men far outnumbering women in hiring in top U.S. tech companies can be seen in Figure 1.

**Figure 1. Gender Gap in Tech Companies**



*Note.* From “Dominated by Men” (Huang, n.d.). These graphs show that gender breakdown of global headcount and technical workforce in Amazon, Facebook, Apple, Google, and Microsoft. The gender gap is more noticeable in technical positions.



















Amazon’s experimental recruiting tool followed the similar pattern, where it learned to downrank résumés containing the word “women’s” until the company discovered the problem. The sexist recruiting engine brought public attention to algorithmic bias because Amazon could not resolve the problem to making of its algorithm gender-neutral even though the company is at the forefront of AI technology (Lavanchy, 2018).

### *Gender Shades: Unveiling Discrimination in Facial Recognition Technology*

In 2018, Buolamwini, a computer scientist and founder of the Algorithmic Justice League, published her research “Gender Shades” which contributed to increasing public awareness of algorithmic bias significantly. The study examined the performance of facial recognition systems developed by Microsoft, Face++, and IBM. This project aimed to

demonstrate the importance of increasing transparency in the performance of AI products that focused on human subjects. The study focused on how the AI-powered gender classification products perform differently based on the gender and skin type of individuals and evaluated the accuracy of the systems. While these three companies' products appeared to have relatively high accuracy overall, the accuracy for darker-skinned female groups was remarkably lower compared to other racial and gender groups, as illustrated in Figure 2. A significant disparity of 34.4% was observed between the lighter-skinned male group and the darker-skinned female group in IBM's program (Buolamwini and Gebru, 2018).

**Figure 2.** Accuracy and Largest Gap in Four Different Groups

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
 Microsoft	94.0% 	79.2% 	100% 	98.3% 	20.8% 
 FACE++	99.3% 	65.5% 	99.2% 	94.0% 	33.8% 
 IBM	88.0% 	65.3% 	99.7% 	92.9% 	34.4% 

*Note.* From “Gender Shades” (Gender Shades, 2018). This figure shows the accuracy and the largest gap between different groups in facial recognition systems of Microsoft, Face++, and IBM. The largest gap is 34.4% in the IBM software. This gap highlights that there is significant algorithmic bias in the system.

Within a day after receiving the performance results, IBM, one of the companies selected for the project, took steps to address the issues and fairness in AI in the system. With an official statement, IBM acknowledged the presence of bias in its facial recognition technology and recognized the importance of addressing this issue promptly. The company stated as following:

IBM is deeply committed to delivering services that are unbiased, explainable, value aligned, and transparent. We do not view AI ethics simply from the perspective of easily quantifiable distributive fairness results such as accuracy disparities across gender and skin tone. We are actively pursuing a research agenda that includes explainability, computational morality, value alignment, and other topics that will also be translated into product and service offerings (IBM, 2018, pp. 2-3).

Besides IBM's official statement, "Gender Shades" has had significant influences on the fields of AI, technology, and public interest about the algorithmic bias of artificial intelligence because this project directly showed that discriminatory algorithms exist in AI systems.

#### *Murder of George Floyd: Increasing Public Interest on Racial Injustice*

On May 25, 2020, George Floyd, a 46-year-old African American man, was killed by Derek Chauvin, a white police officer, due to brutal suppression during the arrest (Hill et al., 2020). This incident ignited a worldwide movement with the slogan "Black Lives Matter." The protests campaigned against violence, racial inequality, and police brutality towards Black people. Floyd's death impacted the world significantly to highlight the deep systemic issues underlying Black lives. Although this was not entirely new information, it renewed call for action against discrimination toward Black community. As Floyd's death influenced all areas of our society including politics, business, media, culture, and education, public interest in racial inequality was rapidly increasing. Every social media feed, news, and article was dominated by this tragedy, and "park benches became settings for conversation and confrontation" (Garcia, 2021).

A study found that there was the surge of searching for the term "racism" across 101 countries and 32 languages after Floyd's death (Barrie, 2020). The research used "pytrends," a Python library, to translate the data from Google Trends. Through analyzing the data, the

research found that public interest in “racism” rose for about four weeks after Floyd’s death, peaking between June 1 and June 4. This increase indicated a significant impact on global awareness and interest in issues of racial injustice. The study highlighted that “episodes of mass unrest may lead to enhanced interest in issues of injustice globally” (Barrie, 2020, p. 3). Furthermore, the study raised a question about the long-term effects of such events on societal awareness and participation.

### **Research Approach**

To answer the research question of what events have contributed significantly to raising public awareness of the algorithmic bias in AI systems, frequency analysis is used through Google Trends.

Frequency analysis is a method to examine how often particular values of a variable phenomenon are likely to occur and evaluate the patterns in that phenomenon (“What is Frequency Analysis?,” n.d.). This method deals with the number of occurrences and investigates measures of tendency. The frequency analysis used in this research paper allows reading trends of how public attention and recognition towards algorithmic bias in AI systems changed when notable events occurred since the flourishing era of AI. To do so, measuring the frequency of the term “algorithmic bias” must be a prerequisite. This helps identify correlations between algorithmic bias and the specific events. Focusing on the peaks of searching keywords and the dates that the events occurred will show whether they are related or not. To find out whether the events mentioned above and the subsequent increase in public interest on the algorithmic bias in AI systems have had relationships, I firstly look for the trends of algorithmic bias since the introduction of AI. Next, I focus on the period that the specific events mentioned above happend.

If there was a substantial increase in searches for “algorithmic bias” immediately following the events, it can be inferred that there are correlations between them.

**Results**

*Understanding the Trends in Public Attention towards AI Bias Using Frequency Analysis*

To see whether there are connections between the events mentioned above with algorithmic bias, examining the tendency of public interest and awareness about algorithmic bias must be the initial step. The following graph in Figure 3 shows the frequency of searches for “algorithmic bias” through Google Trends from 2004 to the present.

**Figure 3.** Search Interest for “Algorithmic Bias” between 2004 and the Present

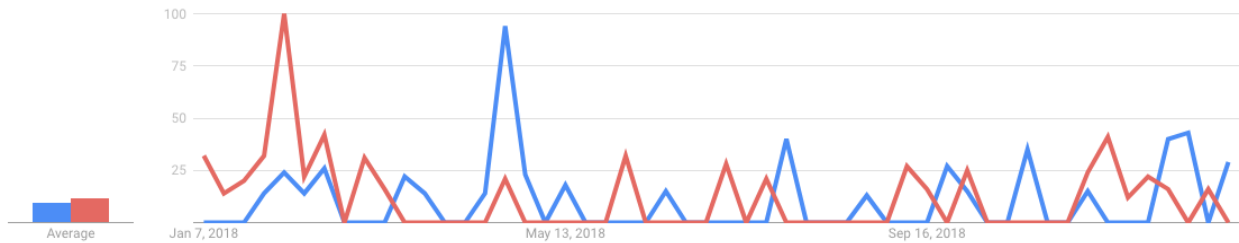


*Note.* From “Google Trends.” This graph helps understand the public interest in algorithmic bias after AI technology was introduced.

We can see that the graph hits the spike in the early 2000s, and there is a consistent increase since 2015. The spike around 2004 is aligned with the period when significant advancements in machine learning and artificial intelligence started. In addition, the steady increase since 2015 can be the result of the influence of Amazon’s sexist recruiting engine, which was disbanded in 2015. It can be seen that the frequency of “algorithmic bias” searches has continuously increased since 2018.



**Figure 4.** Comparison of Search Engine Trends: Google Count Containing “Algorithmic Bias” and “Gender Shades” in 2018



*Note.* From “Google Trends.” This graph compares the term “algorithmic bias” with the term “Gender Shades.” The blue line refers to algorithmic bias and the red line refers to Gender Shades. They have the similar patterns in general and there is a big peak of blue line after Gender Shades was published.

In Figure 4, the red line refers to “Gender Shades” and the blue line refers to “algorithmic bias.”

Buolamwini’s “Gender Shades” is one of the most contributing events to raising social awareness of biased results in AI applications because the project revealed the algorithmic bias by comparing four groups of different genders and races. Since both keywords were searched frequently in 2018 and the frequency of searching “algorithmic bias” peaked after Buolamwini’s project was published, it can be inferred there is a strong correlation between them.

Filtering the trends of the last five years displays in Figure 5 that there is a peak between April 26, 2020 and October 24, 2021.

**Figure 5.** Search Engine Trends in Interest of Algorithmic Bias: Google Count Containing “Algorithmic Bias” between October, 2018 and the Present (Past 5 years)



*Note.* From “Google Trends.” This graph shows the search interest in algorithmic bias in the past 5 years. There is consistent interest in algorithmic bias.

Changing the timeline to 2020 tells us that the spike is around June, as shown in Figure 6. We can infer that searching the keyword “algorithmic bias” is related to the murder of George Floyd since the event happened on May 25, 2020 and the “Black Lives Matter” movement went viral through social media influencing the world significantly and it led to widespread protests and nationwide condemn against police brutality and racial injustice.

**Figure 6.** Search Engine Trends in Interest of Algorithmic Bias: Google Count Containing “Algorithmic Bias” in 2020



*Note.* From “Google Trends.” This graph shows the search interest in algorithmic bias in 2020. Zooming in to 2020 shows that there was a spike around June.

Although his death and algorithmic bias seem to have nothing to do with it, they can be linked to each other on a common theme of racial injustice. In other words, the problems of both

algorithmic bias and Floyd's death were derived from discrimination because both negatively affected minority groups especially Black people. As Barrie showed there was a significant increase in searching the term "racism" after Floyd's death in June 2020, it is aligned with the increase of public awareness due to the connection between algorithmic bias and racism even though algorithmic bias is related to technology. In addition to this, the false arrest of Robert Julian-Borchak Williams, an African-American man, due to a flawed facial recognition algorithm in 2020 would have contributed to increasing public interest in algorithmic bias (Hill, 2020). Due to the fact that the facial recognition program used by the U.S. police for 20 years worked well for recognizing white men but did not identify other races well, public anger against police brutality and abuse of power that led to racial injustice, coupled with George Floyd's death, sparked public awareness and attention.

Figure 1 to 6 graphs through Google Trends serve as evidence for the analysis of the cause-and-effect relationship. As a result of examining the correlation between events using frequency analysis, it can be answered that Buolamwini's Gender Shades conducted in 2018 and a series of events such as a wrongfully arrested case due to a flawed facial recognition system directly linked to algorithmic bias in 2020 greatly contributed to raising public awareness of the algorithmic bias of artificial intelligence. However, the death of George Floyd caused by police brutality did not appear to fall under any of the criteria of "technology" and "artificial intelligence." Although the event was not directly related to algorithmic bias, there are bridge terms such as racism, racial injustice, and discrimination that can connect to algorithmic bias. In addition, his death has had a great impact on our society, reflecting public perceptions of racial

injustice that is actual harm to spread general anxiety, which aligned with the consequences of the other events. From the analysis, algorithmic bias is not enough to say it is only related to technology or AI. Rather than this, algorithmic bias is a more complicated problem that we need to consider other cases that are somewhat irrelevant to it. Therefore, understanding it with different events is important because there are a lot of factors that can cause algorithmic bias and many events can have the direct potential to bring public attention to the problem, such as research on algorithmic bias, as well as somewhat indirect factors can be related to raising public interest toward AI bias like George Floyd incident. Analyzing the correlation of the above graphs, their relationship is not simply clear to define as cause and effect, but it is important to understand the relationship between all of the different events and the growing social interest collectively as they occur. The implications of the above graphs answered my research question of “what cases are connected with algorithmic bias and what information can be drawn from them.” More specifically, how the frequency analysis of those events and investigating the correlations between them contribute to understanding the problem better and what they suggest. Both “Gender Shades” and Floyd’s death had a substantial impact to raise public awareness, but two events were different: while one is directly related to algorithmic bias itself, the other one is not. Based on the observation in frequency analysis, it can be answered that algorithmic bias is complicated so considering many factors is important to understand it better.

## **Conclusion**

In this paper, I demonstrate the effectiveness of frequency analysis for analyzing correlations between public awareness towards algorithmic bias and a series of events that are directly or indirectly related to algorithmic bias. This research on algorithmic bias and public

awareness culminated in the key finding. By incorporating the sources discussed in the above analysis, we can find patterns that there are strong correlations between certain events with public awareness about the algorithmic bias in AI systems since the surge of searching “algorithmic bias” in Google Trends was similar to or immediately after the events happened. When an event was directly related to artificial intelligence or algorithmic bias such as “Gender Shades” and Amazon’s sexist recruiting engine, there was public interest in algorithmic bias right after the event happened. One notable point is the murder of George Floyd. This was one of the significant events in raising public awareness of the problem in AI systems even though it did not seem to related to artificial intelligence. This suggests that finding a direct causal relationship alone is not enough as an answer to the growing public interest in algorithmic bias. As previously mentioned, there seems to be no relationship between them when it comes to discussing algorithmic bias and George Floyd’s death, specifically algorithmic bias belongs to technical criteria and the murder of George Floyd belongs to social/political criteria. However, these two categories alone cannot simplify the relationship between them because they can be connected to a common theme of racial injustice.

Though this research is limited to frequency analysis through finding patterns between certain events and their influences on society, it is meaningful to reveal the fact that algorithmic bias is still unclear and complex because there was a somewhat unrelated event that contributed to increasing public perception of AI substantially. Future work can be done that involves an examination of more cases and factors regardless of direct relationships with algorithmic bias while considering the long-term effects of algorithmic bias on society.

## References

- Anyoha, R. (2017, August 28). *The History of Artificial intelligence - Science in the News*. Science in the News. <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>
- Barrie, C. (2020). Searching Racism after George Floyd. *Socius: Sociological Research for a Dynamic World*, 6(6). <https://doi.org/10.1177/2378023120971507>
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 77–91. <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>
- Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *U.S.* <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- Garcia, M. (2021, May 25). *The Monumental Impact of George Floyd's Death on Black America*. NBC News. <https://www.nbcnews.com/news/nbcblk/monumental-impact-george-floyds-death-black-america-rcna1021>
- Hill, E., Tiefenthäler, A., Triebert, C., Jordan, D., Willis, H., & Stein, R. (2020, May 31). How George Floyd was Killed in Police Custody. *The New York Times*. <https://www.nytimes.com/2020/05/31/us/george-floyd-investigation.html>
- Hill, K. (2020, August 3). *Wrongfully Accused by an Algorithm*. The New York Times. <https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>
- Huang, H. (n.d.) *Dominated by men*. Reuters. <https://fingfx.thomsonreuters.com/gfx/rngs/AMAZON.COM-JOBS-AUTOMATION/010080Q91F6/index.html>
- IBM. (2018). *IBM Response to "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification."* <http://gendershades.org/docs/ibm.pdf>

*Joy Buolamwini: How Do Biased Algorithms Damage Marginalized Communities?* (2020).  
NPR.org. <https://www.npr.org/transcripts/929204946>

Lavanchy, M. (2018). *Amazon's sexist hiring algorithm could still be better than a human*. IMD  
Business School for Management and Leadership Courses.  
<https://imd.widen.net/view/pdf/z7itobahi6/tc061-18-print.pdf>

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness  
and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness,  
Accountability, and Transparency - FAT\* '19*. <https://doi.org/10.1145/3287560.3287598>

*What is Frequency Analysis?* | *Research Optimus*. (n.d.). [Www.researchoptimus.com](http://www.researchoptimus.com).  
<https://www.researchoptimus.com/article/frequency-analysis.php>