

EXAMINING RACIAL BIAS IN FACIAL RECOGNITION TECHNOLOGY

by

Miela Efraim

Term Paper

**Engineering Management 52 - Technical and Managerial Communication
Tufts University
Medford, MA**

1 May 2022

EXECUTIVE SUMMARY

I wrote this paper as a spring 2022 term paper for Engineering Management 52, Technical and Managerial Communication, taught by Professor Amy Hirschfeld at Tufts University in Medford, MA.

The purpose of this paper is to raise awareness about racial bias in facial recognition technology and to educate readers about the consequences and ramifications of these technologies, which many Americans are unaware of. Racial bias in technology is a widespread issue, and is particularly harmful in facial recognition algorithms. Many Americans unknowingly provide their biometric data to companies everyday. Through education on the existence of racial bias in facial recognition technology, Americans can spread public awareness and take action to combat racial bias in facial recognition technology and encourage governments to introduce more regulations to protect citizens.

I discuss (1) examples of bias in design, (2) facial recognition algorithms, (3) public awareness about racial bias in facial recognition technology, (4) government regulations on the use of facial recognition technology and (5) recent developments regarding new research into racial bias in facial recognition technology in the United States.

Although facial recognition technology has been highly regarded as a revolutionary advancement in artificial intelligence, concerns about privacy and security have become more prominent in recent years. In 2016, a researcher at the MIT Media Lab in Cambridge, MA named Joy Buolamwini released a study titled *Gender Shades* that determined that IBM and Microsoft facial recognition technology was less accurate at analyzing dark-skinned faces compared to light-skinned faces. Ever since then, more attention has been given to the severity of this issue and more research has been conducted to examine the extent to which facial recognition technology causes harm.

Despite the perceived benefits of facial recognition technology, it is incredibly invasive and has shown to have disproportionate consequences for people of color. Racial bias in facial recognition technology can make certain products ineffective for people of color, lead to people of color being mistakenly identified as criminals, etc. The risks associated with facial recognition technology are numerous and unregulated in many places across the country.

Some cities in the United States have explicitly banned the use of facial recognition technology. Many advocates and researchers have called for greater regulations governing the use of facial recognition technology and the sale of biometric data from companies.

I recommend that Americans call for greater public awareness and stricter federal regulations for governing the use of facial recognition technology in the United States in order to protect their privacy and security while also ensuring that people of color are not unfairly discriminated against. Americans need to understand how their biometric data is being used and how this affects their everyday lives.

TABLE OF CONTENTS

	<u>Page</u>
Executive Summary	ii
Table of Contents	iii
Definitions	iv
Abbreviations, Acronyms, and Initialisms	v
1.0 Introduction	6
1.1 Purpose	6
1.2 Scope	6
2.0 Background	7
2.1 Definition of Bias	7
2.2 Artificial Intelligence Concepts	7
2.3 Machine Learning Concepts	8
2.4 Overview of Algorithmic Bias	8
2.5 Importance of Inclusive Design	9
3.0 Facial Recognition Algorithms	10
3.1 Biometric Data	10
3.2 Current Usage of Facial Recognition Algorithms	11
3.3 Privacy and Security Concerns	11
4.0 Public Awareness of Racial Bias in Facial Recognition Technology	12
4.1 Gender Shades Research Study	13
4.1.1 <i>Coded Bias</i> Documentary	13
4.2 Algorithmic Justice League	13
4.2.1 Efforts to Spread Public Awareness	14
5.0 Government Legislation Regulating the Use of Facial Recognition Technology	15
5.1 Cities Banning Use of Facial Recognition Technology	15
5.1.1 Somerville, MA	15
5.1.2 San Francisco, CA	15
5.2 Facial Recognition and Biometric Technology Moratorium Act of 2021	15
6.0 Conclusion	16
List of References	17

LIST OF DEFINITIONS

Algorithm	A set of instructions given to a computer program to complete a task
Artificial Intelligence	Ability of systems to perform tasks that require human intelligence
Artificial Neural Network	Model the relationships between inputs and outputs to identify patterns
Bias	Prejudice against something or someone in a way that is unfair
Biometric Data	Body measurements and calculations of human features and characteristics
Correlation	A relationship between two variables
Deep Learning	Relies on artificial neural networks to extract high level features from datasets
Implicit Bias	Unconsciously having stereotypes and attitudes towards people
Inclusive Design	Design process that aims to make products and systems usable for all users
Machine Learning	Uses algorithm and models to develop systems that are capable of learning without explicit instructions
Training Data	Data that is used to train an algorithm to complete the desired task
Unsupervised Learning	Allows machines to make their own predictions with limited to no human input
Racial Bias	Being biased against someone due to their race
Supervised Learning	Requires human input and does not allow machines to function independently

LIST OF ABBREVIATIONS, ACRONYMS, AND INITIALISMS

AI	Artificial Intelligence
AJL	Algorithmic Justice League

1.0 INTRODUCTION

1.1 Purpose

I wrote this paper as a spring 2022 term paper for Engineering Management 52, Technical and Managerial Communication, taught by Professor Amy Hirschfeld at Tufts University in Medford, MA. The purpose of this paper is to raise awareness about the implications of racial bias in facial recognition technology, which Americans are less familiar with, and to educate readers about how to protect their privacy and security when interacting with technology. As Americans learn more about the harsh realities of how facial recognition technology lacks legal structure and violates human rights, they can encourage the government to enact more regulations and take action to protect themselves and their right to privacy.

1. 2 Scope

I focus on educating readers about the severe consequences of racial bias in facial recognition technology. In the second section, I go over relevant concepts in artificial intelligence and machine learning, provide an overview of algorithmic bias, and highlight the importance of inclusive design. In the third section, I provide an overview of facial recognition algorithms, including the collection of biometric data, current usage of facial recognition algorithms, and resulting privacy and security concerns. In the fourth section, I examine how public awareness of racial bias in facial recognition technology has developed ever since the release of the *Gender Shades* research study, and how the founding of the Algorithmic Justice League has made efforts to educate the public and encourage the government to enact strict regulations. In the fifth section, I examine current government legislation regarding the use of facial recognition technology and provide examples of some cities that have banned the use of this technology. In

the final section, I summarize recent research findings from the NCBI and NIST about racial bias in facial recognition technology.

This paper does not include an in-depth discussion of research into other forms of bias in facial recognition technology, because the purpose of this paper is to alert readers to the severe consequences people of color are facing due to the existence of racial bias in facial recognition technology. I do not address how other marginalized user groups, such as women and people with disabilities, are negatively impacted by facial recognition technology, because it is outside the scope of this paper.

2.0 BACKGROUND

2.1 Definition of Bias

Bias is prejudice against something or someone in a way that is unfair. Racial bias is prejudice against someone on the basis of their race. Racial bias is a form of implicit bias, which describes how people have preconceived notions and attitudes towards others that are unconscious. In technology, racial bias occurs when products and systems unfairly discriminate against people of color.

2.2 Artificial Intelligence Concepts

Artificial intelligence is the ability of systems to perform tasks that require human intelligence. By searching through an array of possible solutions, many problems in artificial intelligence can be identified and solved. In artificial intelligence, an algorithm is a set of rules that are used to produce a certain output. Algorithms make predictions and provide insights about data. A cluster is a group of data points that share a common characteristic. Artificial intelligence sorts through

data in order to identify patterns and connections that humans may not be able to identify. Image recognition is the process of programming computer systems to understand the contents of an image. Machine learning is a subset of artificial intelligence defined as the ability of machines to imitate human intelligence.

2.3 Machine Learning Concepts

Machine learning uses algorithms and statistical models to develop systems and machines that are capable of learning without following explicit instructions or guidelines. Machine learning algorithms comb through large quantities of sample data to make predictions and decisions. Supervised learning describes the forms of machine learning that require human input and do not function independently. Unsupervised learning allows machines to make their own predictions and decisions from the data with limited or no human input. Machines learn from the sample data in order to carry out certain tasks. In some cases, machines develop their own algorithms instead of programmers specifying each step required to solve the specific problem. Deep learning is a type of machine learning that relies on artificial neural networks to extract high level and complex features from datasets. Artificial neural networks, similar to neurons in the human brain, model the relationship between inputs and outputs in datasets to identify patterns. In deep learning, there are multiple layers in the artificial neural network where each level transforms the input data into more composite representations.

2.4 Overview of Algorithmic Bias

Artificial intelligence programs learn from real-world data. Since real-world data is often biased, artificial intelligence programs can inadvertently become biased based on the training data that is

used. Correlations are another example of algorithmic bias. A correlation is a relationship between two variables. Artificial intelligence classifies individual data points into groups and then uses these classifications to make predictions on the assumption that each data point resembles other data points within the group. However, these assumptions may be false and do not accurately represent the diversity and variability of the individual data points within the dataset.

2.5 Importance of Inclusive Design

Inclusive design is a design process with the specific goal of making products and systems usable for as many users as possible, with a particular focus on traditionally marginalized user groups. Some of these user groups include women, people of color, and people with disabilities. Instead of trying to make the product or system usable for as many users as possible, designers aim to fulfill as many user needs as possible. Inclusive design considers all aspects that affect a user's ability to use the product or system, such as culture, ability, language, gender, age, etc. The Inclusive Design Research Center has identified the three key dimensions of inclusive design methodology as: recognizing, respecting, and designing with human uniqueness and variability in mind, using inclusive, open, and transparent processes while co-designing with people from a diverse range of perspectives, and understanding that designing within a complex adaptive system requires adapting to changes that influence larger systems. Inclusive design is incredibly important, and has gained more attention in corporate spaces in recent years, because it makes products and systems more inclusive, equitable, and accessible for all users.

Design and technology perpetuate bias and discrimination. Technology reflects those who design and create it. Since able-bodied white men have been at the forefront of the digital age,

technology has primarily been designed for these types of users. When a diverse range of perspectives are not considered or reflected, products and systems inadvertently discriminate against marginalized user groups. There are countless examples of bias in everyday life. Many products and systems have inadvertently been designed for white users. This is mostly due to products and systems not being designed by a diverse set of individuals, in addition to biased research and failing to conduct testing with a diverse range of target users. These failures result in products and systems failing to accommodate, serve, and protect marginalized user groups such as people of color. One especially problematic example is Google's computer vision system labeling African-Americans as gorillas. Another example is most nude products, such as band-aids and bras, only being available in white skin tones. Many automatic soap dispensers are unable to detect dark skin tones. When Apple Emojis were first introduced, the default skin tone, and the only skin tone available, was white. Although many companies are making greater strides to adapt their products and systems for people of color, systemic bias needs to be addressed at the start by stopping the practice of designing with white users as the default. Racial bias in facial recognition technology is the direct result of the sample data not including enough people of color. Therefore, the algorithms are better trained to identify Caucasian faces, and less accurate at correctly identifying faces of people of color.

3.0 FACIAL RECOGNITION ALGORITHMS

3.1 Biometric Data

Biometrics are body measurements and calculations of human features and characteristics.

Biometric authentication is used as a form of identification. Examples of biometric data include

face recognition, fingerprints, and hand geometry. In many cases, biometric data is used for reasons and in contexts that the user has not consented to.

3.2 Usage of Facial Recognition Algorithms

Facial recognition is a method of recognizing human faces using an algorithm that uses biometric data to map out facial features and measurements from a picture or video. These algorithms are trained to recognize faces based on the sample data they are fed through machine learning. Facial recognition technology reads the geometry of faces, such as the distance between eyes, the distance from forehead to chin, the distance between ears, the distance between nose and lips, etc. By identifying key facial landmarks, facial recognition algorithms compare this information with the database of known faces in order to find a match and recognize the face of the person.

Facial recognition technology has been around since the 1960s, where the framework for modern facial recognition software was established. In the 2010s, facial recognition became a more standard feature and the accuracy and capabilities of facial recognition technology greatly advanced and improved. In 2011, facial recognition technology was used to identify Osama bin Laden. In 2015, Android introduced the Trusted Faces feature that allowed users to log into their phones by aiming their faces at the camera. In 2017, Apple introduced face ID for users.

Currently, facial recognition technology is used to unlock iPhones, to analyze pictures in Google Photos, to assist criminal investigations, and in many other applications. In recent years, facial recognition technology has drawn widespread attention for being invasive, due to many companies selling facial recognition technology and biometric data to law enforcement agencies.

3.3 Privacy and Security Concerns

There are numerous privacy and security concerns arising from the use of facial recognition technology. These concerns include data security, privacy rights, reduced anonymity, tracking, lack of transparency, and misuse. Although other forms of data can be encrypted, faces cannot be. Since faces cannot be encrypted, data breaches involving facial recognition make people more prone to identity theft or stalking since people cannot change their faces the way that they can change their passwords or account numbers. Furthermore, facial recognition technology has become so advanced that the algorithms are able to recognize people based on partial features, such as when people are wearing a mask or sunglasses and concealing part of their face. Since facial recognition is used so widely nowadays, from social media to security cameras, facial recognition reduces the anonymity of people and makes their identity more known and recognizable. Additionally, facial recognition has emerged as a robust tracking method since it is difficult for people to avoid facial recognition technology or prevent unwanted tracking of their face. A main privacy and security concern is lack of transparency, because many people are unaware of how their biometric data and facial scans are being used and in what contexts. Facial scans are discreet, easy to capture, and can be taken from far away without the knowledge of the people whose faces are being scanned. Many people may not be aware of their faces being included in large databases of social media profiles, drivers license pictures, etc.

4.0 PUBLIC AWARENESS OF RACIAL BIAS IN FACIAL RECOGNITION TECHNOLOGY

4.1 *Gender Shades* Research Study

In 2018, Joy Buolamwini collaborated with Timnit Gebru on a research study titled *Gender Shades* that uncovered racial and gender bias in facial recognition technology from Microsoft

and IBM. Joy Buolamwini is a computer scientist and digital activist who works at the MIT Media Lab in Cambridge, MA and Timnit Gebru is an AI ethicist from Ethiopia who has worked at Apple and Google. Their research found that the facial recognition technology used by Microsoft and IBM was much less accurate at analyzing dark-skinned faces compared to light-skinned faces. This landmark research study launched a movement in spreading awareness about the existence of racial bias in facial recognition technology and the severe consequences of this technology on people's lives.

4.1.1 *Coded Bias* Documentary

Coded Bias is a documentary by Shalini Kantayya that was released in 2020 at the SunDance Film Festival. The documentary examines the implicit bias embedded into artificial intelligence, and opens with a video of Joy Buolamwini discovering that her face is unrecognizable to a facial recognition program until she puts a white mask over her face. The rest of the documentary provides examples of how racial bias in facial recognition technology is impacting people of color all over the world, in areas such as housing, jobs, healthcare, credit, education, and in the legal system. The release of the documentary led to widespread attention to the issue of racial bias in facial recognition technology and has educated many Americans and people all over the world to the concerning nature of facial recognition algorithms.

4.2 Algorithmic Justice League

In 2016, Joy Buolamwini founded the Algorithmic Justice League (AJL) a digital advocacy organization headquartered in Cambridge, MA that aims to raise public awareness about the implications of implicit racial bias in facial recognition technology and artificial intelligence.

AJL works to expose the existence of racial bias in facial recognition technology and how it threatens basic human rights.

4.2.1 Efforts to Spread Public Awareness

The AJL has been at the forefront of spreading public awareness on this issue through several campaigns and projects. In 2018, the AJL and the Center on Technology and Privacy at Georgetown Law School drafted the Safe Face Pledge to encourage organizations to publicly commit to reducing the harm caused by facial recognition technology by prohibiting lawless police use and requiring transparent government use. The Safe Face Pledge asks organizations to show value for human life, dignity, and rights in addition to addressing harmful biases, facilitating transparency, and embedding these commitments into their business practices. In 2019, Joy Buolamwini testified at a congressional hearing for the US House of Representatives Committee on Science, Space and Technology about racial bias in facial recognition technology and the ethical implications of artificial intelligence. In 2021, American skincare brand Olay collaborated with the AJL on a “Decode the Bias” campaign to determine if the Skin Advisor System used by Olay demonstrated any bias against women of color. The results of the campaign found that the skin age estimate of the system was more accurate for women with lighter skin tones, and was also more accurate for younger women. This campaign was particularly successful in raising public awareness due to the size and prominence of Olay, and has encouraged other big name brands to examine the existence of racial bias in their own systems and technology.

5.0 GOVERNMENT LEGISLATION REGULATING USE OF FACIAL RECOGNITION TECHNOLOGY

5.1 Cities Banning Use of Facial Recognition Technology

5.1.1 San Francisco, CA

In 2019, the city of San Francisco in California unanimously approved the Stop Secret Surveillance ordinance, which requires public notice, clear use policies, and approval from the Board of Supervisors before city departments can be allowed to use surveillance technology and prevents city departments from using facial recognition technology for surveillance reasons.

5.1.2 Somerville, MA

In 2019, the city of Somerville in Massachusetts became the first place on the East Coast to explicitly ban government use of facial recognition technology for surveillance purposes. This marked Somerville as part of the growing nationwide trend to bring facial recognition technology under democratic control and to protect the privacy and security of citizens.

5.2 Facial Recognition and Biometric Technology Moratorium Act of 2021

In 2021, the Facial Recognition and Biometric Technology Moratorium Act, also known as H.R. 3907, was introduced before the House of Representatives “to prohibit biometric surveillance by the Federal Government without explicit statutory authorization and to withhold certain Federal public safety grants from State and local governments that engage in biometric surveillance”.¹

1

[https://www.congress.gov/bill/117th-congress/house-bill/3907/text#:~:text=Introduced%20in%20House%20\(06%2F15%2F2021\)&text=To%20prohibit%20biometric%20surveillance%20by,that%20engage%20in%20biometric%20surveillance.](https://www.congress.gov/bill/117th-congress/house-bill/3907/text#:~:text=Introduced%20in%20House%20(06%2F15%2F2021)&text=To%20prohibit%20biometric%20surveillance%20by,that%20engage%20in%20biometric%20surveillance.)

This legislation is in direct response to reports of unregulated use of facial recognition technology by law enforcement agencies and increasing amounts of research determining the existence of racial bias in facial recognition technology. However, several groups have expressed severe opposition to this bill, citing the benefits provided by facial recognition technology such as identifying participants in the January 6th Capitol Riot, contributing to counterterrorism efforts, reuniting human trafficking victims, etc.

6.0 CONCLUSION

Facial recognition technology is widely viewed as a groundbreaking technological advancement with transformative applications in many areas. However, many Americans are unaware of the racial bias embedded into this technology and how the lack of regulations is threatening the privacy and security of biometric data. Researchers have found evidence of facial recognition algorithms discriminating against people of color, which has led to detrimental consequences in areas such as housing, healthcare, law enforcement, etc. Not only is the implementation of facial recognition technology widely unregulated, but also many Americans are unaware of how their biometric data is collected and shared with their knowledge. I recommend that Americans encourage developers to train facial recognition algorithms with more diverse datasets in order to eliminate racial bias and push for the government to enact more stringent regulations for the implementation of this technology in order to protect the privacy and security of citizens.

LIST OF REFERENCES

1. Najibi, A. (2020). *Racial discrimination in face recognition technology*. Science in the News.
<https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/>
2. Kantayya, S. (2020). *Coded Bias*. 7th Empire Media. <https://www.codedbias.com/>
3. J. G. Cavazos, P. J. Phillips, C. D. Castillo and A. J. O'Toole. (2021). *Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias?* IEEE Transactions on Biometrics, Behavior, and Identity Science.
<https://ieeexplore.ieee.org/abstract/document/9209125>
4. Rauenzahn, B., Chung, J., & Kaufman, A. (2021). *Facing Bias in Facial Recognition Technology*. The Regulatory Review.
<https://ieeexplore.ieee.org/abstract/document/9209125>
5. Lunter J. (2020). Beating the bias in facial recognition technology. *Biometric Technology Today*. [https://doi.org/10.1016/S0969-4765\(20\)30122-3](https://doi.org/10.1016/S0969-4765(20)30122-3)
6. Harwell, D. (2019). *Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use*. The Washington Post. Retrieved February 20, 2022, from
<https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/>
7. Learned-Miller, E., Ordóñez, V., Morgenstern, J., & Buolamwini, J. (2020). *Facial Recognition Technologies in the Wild: A Call for a Federal Office*. Algorithmic Justice League.
https://assets.website-files.com/5e027ca188c99e3515b404b7/5ed1145952bc185203f3d009_FRTsFederalOfficeMay2020.pdf
8. Perkowitz, S. (2021). *The Bias in the Machine: Facial Recognition Technology and Racial Disparities*. MIT Case Studies in Social and Ethical Responsibilities of Computing. <https://doi.org/10.21428/2c646de5.62272586>
9. Hardesty, L. (2018). *Study finds gender and skin-type bias in commercial artificial-intelligence systems*. Massachusetts Institute of Technology.
<https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>
10. Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face recognition Vendor Test (FRVT) Part 3: Demographic Effects*. National Institute of Standards and Technology.
<https://doi.org/10.6028/nist.ir.8280>
11. Zhou, Y., Tianyu, G., Zhang, T., Li, W., Taoyu, W., Han, X., & Shihui, H. (2020). *Neural dynamics of racial categorization predicts racial bias in face recognition and altruism*. Nature Human Behaviour. <http://dx.doi.org/10.1038/s41562-019-0743-y>

12. Andrejevic, M. & Selwyn, N. (2020) *Facial recognition technology in schools: critical questions and concerns*. Journal of Learning, Media, and Technology.
[10.1080/17439884.2020.1686014](https://doi.org/10.1080/17439884.2020.1686014)
13. Castelvechi, D. (2020). *Is facial recognition too biased to be let loose?* Nature News.
<https://www.nature.com/articles/d41586-020-03186-4>
14. Castelvechi, D. (2020). *BEATING BIOMETRIC BIAS*. Nature Publishing Group.
<http://dx.doi.org/10.1038/d41586-020-03186-4>
15. Libby, C., & Ehrenfeld, J. (2021). *Facial Recognition Technology in 2021: Masks, bias, and the future of Healthcare*. Journal of Medical Systems.
<https://doi.org/10.1007/s10916-021-01723-w>