Facial Recognition Software and the Politics of Design in Sociotechnical Systems

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

Grace Ko

Spring 2022

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

Sean M. Ferguson, Department of Engineering and Society

Introduction to bias in facial recognition

Facial recognition software has become a large project that many large technology companies are undertaking to push their names to the forefront of innovative leaders. In 2019, the facial recognition market was estimated to be worth around $3.2 billion and is expected to grow up to $7 billion in revenue by 2024 at a CAGR at 16% (Analytic Insights, 2019). Some top companies include Deep Vision AI, SenseTime, FaceFirst, TrueFace, and Amazon Rekognition, and IBM. However, there are some deep inequities that exist within this developing technology.

Extensive bias in data and data algorithms is a pervasive issue that extends throughout the entire technology field. This bias can affect the quality of daily life for individuals, especially minorities. For example, searching up images of "unprofessional hairstyles" on Google Images will return images of naturally Black hairstyles (Korte, 2020), which displays and reinforces traditional racial stereotypes. Additionally, these inequalities exist regarding gender as well. In video conferencing platforms such as Zoom and Skype, the hardware is coded to recognize lower tones of voices over higher pitched tones. As a result, women in meetings are often silenced if a man has decided to speak at the same time as her (Feldman, 2020).

One specific technology can have particularly harmful effects for minorities, especially Black individuals. Facial recognition technology, when utilized by law enforcement and the government, influence how certain groups of people are treated and viewed in society. Data training bias as well as inherent human bias combine to form a lethal effect on the safety and wellbeing of minorities in world where they are constantly surveilled. The bias that exists in hierarchal company structure, individual thoughts, and in data must be addressed be used to reduce the amount of inequity found in technology for minority individuals. Additionally, large technology companies that play a major role in overall change must commit to addressing this topic.

Background on data racism

Racism in technological institutions mainly exists due to the prominence of bias in the data sets used for creating algorithms. Studies done about the accuracy of facial recognition systems include one by researchers in the MIT media lab, Joy Buolamwini and Timnit Gebru. In their study, they sought to create a bias detector to evaluate the accuracy of current facial recognition systems when used on individuals of different races. Darker-skinned females had the highest error rate of around 34.7% while the maximum error rate for light-skinned males was 0.8%. This discrepancy in error most likely occurred because the datasets used to inform popular facial recognition systems were made up of lighter-skinned subjects, leaving the software unable to correctly recognize darker-skinned individuals (Hardesty, 2018) Another study done by the National Institute of Standards and Technology (NIST) in 2003, found that algorithms had a harder time identifying female subjects than male subjects and that the rate of identification accuracy was also lower for younger subjects. This dataset bias not only exists in facial recognition systems but in other important technology implementations in healthcare settings. For example, Ruha Benjamin also conducted a study in the Science journal discussing how the automated system used by health insurance companies to evaluate the health profiles of individuals is inherently biased. For example, "...if two people have the same risk score that indicates that they do not need to be enrolled in a "high-risk management program", the health of the Black patient is likely much worse than that of their White counterpart" (Benjamin, 2019). This discrepancy occurs because the data set used to train these automated systems is historic, which includes factors such as "segregated hospital facilities, racist medical curricula, and unequal insurance structures" (Benjamin, 2019). Bias in data sets is not limited exclusively to facial recognition software and is instead prominent in many vital societal technologies.

Reasons for data set bias

Data set bias occurs for a multitude of reasons. Firstly, most facial recognition programs are trained using data sets that are not ethnically diverse. There are a larger proportion of white, male faces when compared to the faces of minority women, so the program does not have enough background to accurately identify these individuals correctly.

Another reason that technology bias is so hard to eradicate is because of the decades long history of racist surveilling. In a study done by Alvaro M. Bedoya, the Director of the Center on Privacy & Technology at Georgetown University discusses the idea that "the burdens of government surveillance have fallen overwhelmingly on the shoulders of immigrants, heretics, people of color, the poor, and anyone else considered "other" (Bedoya, 2020). One of the earliest examples Bedoya discusses is the reign of Elizabeth I in 1558. She set up an entire network of spies and informers that targeted Catholics and Puritan Separatists for their religious beliefs. Suspected religious defectors were monitored by law enforcement and placed under arrest if any suspicious activity occurred. Soon, race and social unrest began to play a huge factor in surveillance decisions. Some examples include the wiretapping and bugging of Martin Luther King Jr. by the FBI under President Hoover. The agents following King created a recording of his extramarital affairs and mailed it to his office with a note that threatened blackmail. In 2017, Immigrations and Customs Enforcement (ICE) proposed a system where they would automatically scan immigrants' social media platforms and flag a minimum of 10,000 individuals a year for deportation investigation (Bedoya, 2020). This long history has continuously displayed an innate bias towards minorities that has persisted throughout history.

Current affairs

The detrimental consequences of this data set bias has been evident in recent years for minority individuals. Robert Julian-Borchak Williams was one of these individuals. Williams was working at his automotive supply company office when he got a call from the Detroit Police Department demanding that he come to the station because he had been arrested. Williams faced emotional distress when police arrested him in front of his wife and kids, telling him that he faced charges of “larceny” and that he faced a “federal warrant” (Hill, 2020). Once he reached the police station, the officers accused him of stealing watches from Shinola, an expensive, trendy boutique that Williams had only visited once in 2014 when it had opened. The officers then proceeded to show Williams a screenshot of the suspect, who was a black man with a similar build as Williams. However, it was obviously not him.

This error was not borne from a purely human mistake, instead, law enforcement was led astray by a misidentification by facial recognition software from the company DataWorks Plus. Past studies have shown how unreliable and inaccurate this system is, with the algorithm falsely identifying African-American and Asian faces 10 times to 100 times more than Caucasian faces. The software had identified Williams as a possible match, and when his picture was included in a lineup and shown to a Shinola employee, Williams was incorrectly identified. The detectives quickly realized the mistake but Williams was forced to stay in jail for 30 hours and had to be released on a $1,000 bond.

Williams' case is not an isolated incident. In 2019, Amara K. Majeed, a Brown University student, was mistakenly identified as the Sri Lankan bombing suspect, Fathima Qadiya. According to Majeed, she woke up to "…35 missed calls, all frantically informing me that I had been falsely identified as one of the terrorists involved in the recent Easter attacks in my beloved motherland, Sri Lanka" (Fox, 2019). However, her life was forever changed when a group of investigators used facial recognition software that matched Majeed's picture with a picture of the suspect. Majeed began receiving death threats and was worried for the safety of her family back home. Majeed was wrongly identified through a biased facial recognition and as a result her life was unfairly risked with no consequence to the manufacturers of the facial recognition software in question.

In a more recent incident, facial recognition technology has been utilized to identify and collect data on protestors at the Black Lives Matter protests in 2020. After the wrongful killing of George Floyd by police officers in May 2020, Black Lives Matter protests began to rise all over the United States, continuing on for several months. In August 2020, NYPD police officers tried to arrest Derrick Ingram, a co-founder of the social justice organization Warriors in the Garden, who was accused of "allegedly assaulting a police officer by shouting loudly into a megaphone at a June protest" (Amnesty International, 2021). It was found that the police had used a facial recognition system that had used web scraping to match a photo of Derrick to one found on his private Instagram. The police then proceeded to violate his rights by "misinforming him about his [rights], threatened to break down his door, attempted to interrogate him without a lawyer, used at least one police helicopter and drones, and stationed dozens of officers in his home" (Amnesty International, 2021). This is just one of many cases where facial recognition technology was used in a way that was detrimental to minorities.

IBM case study

Some large tech giants are now realizing the detriment of the facial recognition systems that they have created. In June 2020, IBM CEO Arvind Krishna released a letter to Congress stating that “IBM firmly opposes and will not condone uses of any [facial recognition] technology, including facial recognition technology offered by other vendors, for mass surveillance, racial profiling, violations of basic human rights and freedoms, or any purpose which is not consistent with our values and Principles of Trust and Transparency” (Peters, 2020). Additionally, Krishna made the statement that “…[We] believe now is the time to begin a national dialogue on whether and how facial recognition technology should be employed by domestic law enforcement agencies” (Peters, 2020). IBM chose to support the use and production of technology that advocates for transparency, such as body cameras for police officers and an emphasis on data analytics.

However, this decision has faced some critique from those who believe that IBM is the main company behind expanding the spread of surveillance technologies in major cities. Eva Blum-Dumontet from Privacy International states that IBM “pushed [for] a model for urbanisation which relied on CCTV cameras and sensors processed by police forces, thanks to the smart policing programs that IBM was selling them” (BBC, 2020). By advocating for urban surveillance, IBM was creating additional profit for themselves by selling their technology for governmental use. Additionally, she points out that IBM’s statement didn’t completely reject the idea of facial recognition technology all together, implying that they will be continuing working with this idea in the future.

Despite these criticisms, IBM has maintained that they are committed to creating change within the technology industry. In a ThinkPolicy blog post written in September 2020, IBM reiterated their many achievements within the AI field which included partnering with the University of Notre Dame to establish an innovative research lab that worked to create best practices in technology ethics and being one of two signatories for the Vatican's Rome Call for AI Ethics, a partnership that called for "[the] regulation of intrusive technologies such as a facial recognition and to promote the ethical development of artificial intelligence" (Pullella, 2020). Additionally, they have been consulting with the U.S. Department of Commerce to work on limiting the export of facial recognition systems and focusing on improving "1-to-many" facial recognition software specifically, which is the code most commonly used in mass surveillance systems. "1-to-many" facial recognition software entails taking a singular photo of an individual and comparing it with a large database of facial data to find a match. Law enforcement databases utilize mass surveillance systems, which lead to the high rate of racial profiling cases similar to the case studies mentioned previously.

However, their commitment to recognizing and fixing the bias in their systems has not come without some setbacks. Previously in 2019, IBM had released a data set named Diversity in Faces that had over 1 million diverse face images to help train their facial recognition technology. However, the company came under fire after it was discovered that the images making up the data set had been scraped without consent from the website Flickr (Hao, 2020). This controversy brought into another question of how deeply bias was ingrained in the idea of facial recognition.

Clearview AI case study

While IBM appears to be taking positive steps towards decreasing the overall level of bias in facial recognition systems, there are some companies that have been under fire for contributing to dangerous bias levels in law enforcement and violating individual privacy laws. Clearview AI is a facial recognition software company originating in America that gathers image data from social media platforms and sells the database and the facial recognition technology to a multitude of customers such as law enforcement, universities, and even certain individuals. In 2021, the Office of the Australian Information Commissioner (OAIC) discovered that Clearview AI had violated Australian privacy rules. This process is called image scraping, which is where data from the web is extracted. Clearview AI was doing this without the consent of any Australian citizens, and selling their faces to law enforcement agencies. However, Clearview AI is maintaining their innocence, stating that the images that they collected were publicly available so that no breach of privacy occurred. Additionally, they are arguing that the use of the photos all occurred in the United States, so no Australian laws were broken.

This case study calls into question the ideas of privacy and what it really means. Clearview AI was able to take photos off of the internet and sell them to various law enforcement agencies, who's intentions with the pictures are unknown. Social media platforms such as Instagram and TikTok have become places to document lives and whether or not that data should be available to companies is a heated debate.

Amazon case study

One of the largest tech companies in the world, Amazon, has a poor track record with advocating for a decrease of bias in their facial recognition system Rekognition. In 2018, the American Civil Liberties Union (ACLU) delivered former CEO Jeff Bezos with a petition that had over 150,000 signatures demanding that Amazon stop providing the government with facial recognition technology to combat the misuse of these systems for the deportation of immigrants. This petition was accompanied by letters from almost 70 different civil rights and research organizations, concerns from Amazon shareholders, and an internal memo from Amazon employees. Even after widespread concern about the usage of Amazon's technology refused to respond, instead deciding to continue providing government agencies with their technology and even using their subsidiary Ring to allow law enforcement agencies to use footage from home security cameras (Hao, 2020).

However, after IBM announced that they would no longer be providing their facial recognition technology to the government and would be ceasing development on it entirely, Amazon decided to follow suit. Their decision was that there would be a one-year moratorium on police use of Rekognition. This demonstrates the power of large companies when it comes to starting social change by influencing each other. Although Amazon made the correct social decision regarding taking away the use of Rekognition from the police, there was originally pushback from Amazon on the critique from various social organization and studies. When the study done by MIT researchers Joy Buolawnmi and Timnit Gebru that revealed the lack of accuracy in many facial recognition systems came out, including Rekognition, Amazon decided to deny the accusations. On the AWS Machine Learning blog, Michael Punke, the Vice President of Global Public Policy at Amazon, created a post that advocated for the use of Rekognition by law enforcement agencies and attacked the studies that were done by researchers, insisting that they had not used the system correctly.

While other companies had taken the research done seriously and made concrete efforts to address bias existing in their companies, Amazon pushed back and decided to try to discredit Black researchers.

Microsoft case study

While some companies seem to be resisting positive change regarding bias in their data, other tech giants such as Microsoft are embracing this new revolution. In 2018, Microsoft released a statement committing to updating their facial recognition technology, specifically surrounding the system's abilities to recognize different genders with different skin tones. This focus is directly linked to studies mentioned earlier in this paper (NIST, Buolamwini and Gebru) that discovered the large discrepancy in accuracy when facial recognition systems were given white males versus females of color. Through their changes, Microsoft has been able to reduce the error rates for men and women with darker skin by up to 20 times. For all women, the error rate was reduced by at least nine times (Roach, 2018). To make these improvements, Microsoft also factored in aspects such as eyewear, jewelry, and hairstyle to more accurately identify individuals.

In conjunction with focusing on the issue of gender and race bias in their database, Microsoft is taking their commitment to representation a step further by creating the Face API by Azure Cognitive Services team that "worked with experts in bias and fairness across Microsoft called the gender classifier, focusing specifically on getting better results for all skin tones" (Roach, 2018). According to Hanna Wallach, a senior researcher at Microsoft's AI Lab, the three major changes made to improve the accuracy of Microsoft's facial recognition technology included: expanding and revising training for datasets, creating new data collection efforts to improve training data by focusing on skin tone, gender, and age, and improving the gender classifier to have more precise results. While these technical changes have successfully mitigated the dataset concerns of bias, there still exists the issue of innate human bias. Wallach states that “If we are training machine learning systems to mimic decisions made in a biased society, using data generated by that society, then those systems will necessarily reproduce its biases”. This brings forth the question of whether or not deconstructing bias in facial recognition technology is possible due to the deep-rooted bias that is ingrained in human society after decades of racial profiling and surveillance. However, Microsoft is working to create the fairest system possible by also focusing on idea creation and data collection as well.

Discussion

Through various case studies discussed in this paper, it is shown how multiple different technology giants are adapting their views and practices to consider how bias can affect underrepresented groups of individuals. Some, like Microsoft and IBM, are committed to doing the internal work to alleviate the effect of bias in their facial recognition systems, which have been notoriously used for morally questionable ends. However, other companies like Amazon and Clearview still have work to do when it comes to acknowledging the flaws in their systems.

From the wrongful arrest of Robert Williams to the incorrect identification of Amara Majeed, facial recognition software has proved time and time again that it holds an inherent bias against people of color and women. However, many of the large companies that develop the algorithms for this software are unwilling to work to destroy the bias in their data. This is a hard lesson that must be learned as it can only cause further prejudice for minorities in an evolving society. As demonstrated through the multiple case studies and research studies that have been done, politics of designs in our society must be considered to create technology that lacks bias and caters towards our diverse, growing society.

References

BBC. (2020, June 9). *IBM abandons 'biased' facial recognition tech*. BBC News. Retrieved February 26, 2022, from https://www.bbc.com/news/technology-52978191

Bedoya, Alvaro, Privacy as Civil Right (May 12, 2020). New Mexico Law Review, Vol. 50, No. 3, 2020, Available at SSRN: https://ssrn.com/abstract=3599201

Benjamin, R., & Ruha Benjamin, P. U. (2019, October 25). *Assessing risk, automating racism*. Science. Retrieved September 28, 2021, from http://www.science.org/doi/abs/10.1126/science.aaz3873.

Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in ...* Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Retrieved October 17, 2021, from http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

Daub, A. (2021, April 27). *How sexism is coded into the tech industry*. The Nation. Retrieved September 28, 2021, from https://www.thenation.com/article/society/gender-silicon-valley/.

Feldman, F. (n.d.). *The silencing of Zoom*. Kin + Carta. Retrieved November 1, 2021, from https://www.kinandcarta.com/en-us/insights/2020/05/the-silencing-of-zoom/.

Hao, K. (2020, April 2). *Amazon is the invisible backbone of Ice's immigration crackdown*. MIT Technology Review. Retrieved February 26, 2022, from https://www.technologyreview.com/2018/10/22/139639/amazon-is-the-invisible-backbone-behind-ices-immigration-crackdown/

Hao, K. (2020, December 10). *The two-year fight to stop Amazon from selling face recognition to the police*. MIT Technology Review. Retrieved February 26, 2022, from https://www.technologyreview.com/2020/06/12/1003482/amazon-stopped-selling-police-face-recognition-fight/

Hardesty, L. (n.d.). *Study finds gender and skin-type bias in commercial artificial-intelligence systems*. MIT News | Massachusetts Institute of Technology. Retrieved April 3, 2022, from https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212

Lohr, S. (2021, December 8). *Group backed by top companies moves to combat A.I. Bias in hiring*. The New York Times. Retrieved February 26, 2022, from https://www.nytimes.com/2021/12/08/technology/data-trust-alliance-ai-hiring-bias.html

Myers, V. (n.d.). *Inclusion takes root at netflix: Our first report*. About Netflix. Retrieved November 1, 2021, from https://about.netflix.com/en/news/netflix-inclusion-report-2021.

Najibi, A. (2020, October 26). *Racial discrimination in face recognition technology*. Science in the News. Retrieved February 26, 2022, from https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

Padilla, C. A. (2020, September 11). *A precision regulation approach to controlling facial recognition technology exports*. THINKPolicy Blog. Retrieved February 26, 2022, from https://www.ibm.com/blogs/policy/facial-recognition-export-controls/

Phillips , J., Grother, P., Michaels, R. J., Blackburn, D. M., Tabassi, E., & Bone, M. (n.d.). *Face recognition vendor test 2002: Evaluation report - NIST*. Face Recognition Vendor Test. Retrieved October 17, 2021, from https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir6965.pdf.

Pullella, P., & Dastin, J. (2020, February 28). *Vatican joins IBM, Microsoft to call for Facial Recognition Regulation*. Reuters. Retrieved February 26, 2022, from https://www.reuters.com/article/us-vatican-artificial-intelligence/vatican-joins-ibm-microsoft-to-call-for-facial-recognition-regulation-idUSKCN20M0Z1

Punke, M. (2008). *Some Thoughts on Facial Recognition Legislation*. Amazon. Retrieved February 26, 2022, from https://aws.amazon.com/blogs/machine-learning/some-thoughts-on-facial-recognition-legislation/

Roach, J. (2020, December 11). *Microsoft improves facial recognition to perform well across all skin tones*. The AI Blog. Retrieved February 26, 2022, from https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/#:~:text=Microsoft%20announced%20Tuesday%20that%20it,recognize%20gender%20across%20skin%20tones.&text=With%20the%20new%20improvements%2C%20Microsoft,by%20up%20to%2020%20times.

Technology's built-in Machine bias reflects racism, Scholar Says. American Association for the Advancement of Science. (2020, September 3). Retrieved September 28, 2021, from https://www.aaas.org/news/technologys-built-machine-bias-reflects-racism-scholar-says.

Vincent, J. (2021, November 3). *Clearview ai ordered to delete all facial recognition data belonging to Australians*. The Verge. Retrieved February 26, 2022, from https://www.theverge.com/2021/11/3/22761001/clearview-ai-facial-recognition-australia-breach-data-delete#:~:text=Clearview%2C%20which%20claims%20to%20have,October%202019%20and%20March%202020.