Security and Privacy Concerns of Facial Recognition Technology

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

Ethan Cha

Spring 2024

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

Joshua Earle, Department of Engineering and Society

Introduction

Of the various forms of identification available, biometrics is generally accepted as a reliable identification system. Some common forms of biometrics used in criminal identification are fingerprint analysis, retina scans, and facial recognition technology. Facial recognition technology, in particular, has gained popularity beyond criminal identification. From getting into our phones, bank accounts, to social media, our society has become accustomed to using some sort of facial identification system to authenticate access to our accounts. That being said, this technology is far from perfect. Not only is it one of the most unreliable biometrics, but it also comes with negatives such as privacy concerns and racial bias.

The technology discussed in this paper is facial recognition software, which has sparked concerns regarding privacy. The big privacy questions are: How is facial recognition technology being used in public areas with cameras? Are my personal photos safe and private? Who has access to data involving things such as Apple's FaceID? And many others. Najibi states that, "Police use face recognition to compare suspects' photos to mugshots and driver's license images; it is estimated that almost half of American adults – over 117 million people, as of 2016 – have photos within a facial recognition network used by law enforcement. This participation occurs without consent, or even awareness, and is bolstered by a lack of legislative oversight." (Najibi, 2020, para. 2) This means that government agencies can run facial recognition software on pretty much anyone who is around technology without our consent.

To train image processing software, large sets of data must be fed in, which can inadvertently introduce biases. One prominent example of this is the censorship found in AI chat tools in China. The censorship in China is reflected in these machine-learning tools, finding words such as "surveillance" and "CCP" as positive words and associated words such as

“democracy” with negative words such as “chaos” (Cook, 2023, China AI-generated censorship section). Similarly, this kind of bias can be seen in facial recognition software through the datasets used to train them. In order to train an image processing program to recognize different objects, the developer needs to label the image in order for the machine to understand what it is. This is where politics and bias comes into play. After training a particular image recognition artificial intelligence, “a photograph of a woman smiling in a bikini is labeled a ‘slattern, slut, slovenly woman, trollop.’ A young man drinking beer is categorized as an ‘alcoholic, alky, dipsomaniac, boozer, lush, soaker, souse.’ A child wearing sunglasses is classified as a ‘failure, loser, non-starter, unsuccessful person.’”(Crawford & Paglen, 2021, Introduction section) As a society, we must be mindful of this bias because, “[c]onsciously or not, deliberately or inadvertently, societies choose structures for technologies that influence how people are going to work, communicate, travel, consume, and so forth over a very long time.” (Winner, 1980, page 127)

There are four levels to facial recognition: detection, characterization, verification, and identification in ascending order of complexity. (Leong, 2019, page 110) Detection refers to a software to be able to recognize faces in a camera feed. Characterization is the step where the software is able to make more detailed observations such as gender, age, and emotion characteristics. The verification level is where privacy concerns come into play. This is a one-to-one matching system that compares one face to one template. An example of this is Apple’s FaceID which checks the person in front of the iPhone at a given time to the template saved on the phone. (Hamann & Smith, 2019, How it Works section) Templates are created by taking measurements of the person’s face and translating it into a number. This means that faces cannot be recreated from a template. These templates are saved into a database for verification

processes in the future. Identification is a step more complex than verification. While verification is one-to-one, identification is one-to-many. This is the method that is used in criminal justice. Once a template is created from an image, it is then run against a database of a set of people, for example, criminal offenders. This process is much harder to be precise and it is not uncommon to set a threshold for "matches" and have multiple potential matches which are then verified by a human actor to make a final decision.

I. Methods and Frameworks

The method used in this paper is the reading and synthesization of previous literature related to the research questions. These literatures include various types of articles and books. The pieces of literature were carefully selected to allow for a deep understanding within the specific topics discussed in this paper. I use both explanatory synthesis and argumentative synthesis. I synthesize different articles to answer questions such as: How do the negatives of facial recognition technology affect its use in criminal justice, a field in which this technology is prevalent? To what extent does racial bias play a role in this technology? I also use these resources to form an argument that answers the question: Is it too early to fully utilize this technology due to issues of privacy, consent, and regulation?

I conducted a race studies analysis as my framework because it makes sense in the context of facial recognition technology in criminal justice. A race studies analysis refers to when race and racism are the central objects under examination (Hunter, 2023, para. 3). In this paper, I use racial analysis to examine how race plays a factor in the unfairness of facial recognition technology in criminal justice. I explore how race plays a role in the foundations of these AI and also how race plays a role in the practice of using facial recognition technology in criminal justice.

There are multiple cases where the use of facial recognition technology as sole evidence in the conviction of a black person has led to wrongful arrests such as the example of Nijeer Parks in 2019 (Sarlin, 2021). This is because facial recognition software is "more prone to error when trying to match the faces of darker skinned people" (Sarlin, 2021, para. 9), especially women who had error rates 34% higher than white males (Najibi, 2020, para. 3). This is just one example of racial injustice found in the context of facial recognition technology in criminal justice and the rest of this paper will dive deeper into this case and an analysis on why the disparity exists in the first place and what we should do moving forward.

II. Paper Roadmap

In the rest of the paper, I will first discuss what I believe to be key controversies regarding facial recognition technology. This includes the issues of privacy, consent, and the lack of regulation in this field. Then I will go into more detail of how facial recognition technology is used in criminal justice and also the racial bias that comes with this technology. Lastly, I will make my own comments and analysis in the discussion section answering the initial questions asked in the introduction and finally wrap the paper up in a conclusion.

Controversies

As we get familiar with the use cases of facial recognition technology and the criteria required for its implementation, people are, and should be, worried about how this technology affects their rights. To this end, this section will cover issues with privacy, a fatal problem of consent, and the lack of regulation.

I. Privacy

As a technology that utilizes cameras in public and stores images in databases, it is only natural to be concerned with how its use in public and private places affects people's privacy.

The two main areas of concern are governmental and commercial use. (Leong, 2019, page 113) In government use, the misidentification of a person might unfairly place people on watch lists for various crimes and these misidentifications can potentially be targeted both intentionally and unintentionally towards certain people groups. (Leong, 2019, page 113)

Similarly, in the commercial realm, companies may use ambiguous guidelines for placing people on their own watch lists to deny people of service. People may be unaware that they are on watchlists, and these lists can also be shared among companies. Companies are even less regulated than the government and these practices can happen in the dark. This makes it even easier to discriminate against certain groups of people. (Leong, 2019, page 113) These ideas can be seen in China where the surveillance of people is practiced. China is not afraid of using these technologies "for everything from identifying jaywalkers to dispensing toilet paper." (Leong, 2019, page 113) The Chinese government's ability to surveil its citizens has historically enabled them to discriminate against certain groups and individuals. (Leong, 2019, page 113)

II. Consent

With privacy concerns set aside, there is also a consequential problem with the consent of using such technology. Selinger and Hartzog argue that, "The greater the risk to autonomy, the more … a person is entitled to understand." (Selinger & Hartzog, 2020, page 108) This belief is especially relevant in the use of facial recognition technology for surveillance purposes because the loss of anonymity is a big threat to autonomy for a variety of reasons. Not being under constant surveillance allows people to speak and act more freely without fear and affects how relationships between people are formed and maintained in that "it allows us to selectively disclose information and share different aspects of our identity in different contexts" (Selinger & Hartzog, 2020, page 115) Not being under constant surveillance also allows people to be

comfortable with putting themselves in riskier situations that if they fail, they won't be tied to those mistakes forever. Just imagine a situation where many people are watching you, like on a televised game show or talk show. It is so easy to back out of words and actions you would have said or taken just by the mere thought of people perceiving and judging you based on them. This idea is better said by Foucault in an idea called "governmentality" where, "[h]e who is subjected to a field of visibility, and who knows it, assumes responsibility for the constraints of power; he makes them play spontaneously upon himself; he inscribes in himself the power relation in which he simultaneously plays both roles; he become the principle of this own subjection." (Foucault, 1979, page 202)

Assuming there is sufficient understanding of the risk to autonomy, the next issue is giving valid consent. Selinger and Hartzog argue that, "In order for consent to data and surveillance practices to be knowing and voluntary, at least three pre-conditions should exist: (1) such a request should be infrequent, (2) the harms to be weighed must be vivid, and (3) there should be incentives to take each request for consent seriously." (Selinger & Hartzog, 2020, page 116) Even if, technically, a person signs a form saying they consent to being under surveillance, if they do not meet these three conditions, this "consent" would be considered defective under Selinger and Hartzog's three pre-condition argument. For the first condition, if requests are too incessant, people will begin to disregard subsequent requests. Take terms and conditions for example. Many apps nowadays have users sign user agreements before being able to even create an account. But, how many people actually read the whole terms and conditions? We see these agreements way too often for people to even care about them. Relating back to face surveillance, if government agencies and commercial companies constantly ask for this consent before being able to use their facilities, people are way less likely to take each of these instances seriously and

will become desensitized. For the second condition, if the potential harms that come from granting consent are too abstract, it is hard for people to imagine what this means for them practically speaking. What does potential loss of privacy mean to someone? Or, what does inhibition of autonomy mean? And so on. Then, "people's cost/benefit calculus may be corrupted by an inability to take adequate stock of the risks." (Selinger & Hartzog, 2020, page 116) It is hard for people to weigh abstract risks to concrete benefits without first analyzing these abstract risks and coming up with practical, concrete ways in which these risks manifest. The final condition refers to the fact that there should be proper incentive to take each consent request seriously. If the amount of harm is spread across multiple decisions, there would not be enough incentive because each decision does not seem like a significant threat on its own, but over time and over many decisions "society is exposed to death by a thousand cuts". (Selinger & Hartzog, 2020, page 116)

III. Regulation

As facial recognition technology becomes more popular and its use cases become more widespread, it is important to regulate this technology, especially because of how powerful it can be. Facial recognition technology has the potential to introduce new avenues of abuse in both the government and the commercial spheres if left unregulated. According to a report by the National Academies Press, "when [facial recognition technology is] applied broadly and without safeguards, the technology can allow repressive regimes to create detailed records of people's movements and activities and block targeted individuals from participation in public life." ("Advances in Facial", 2024, Areas of Concern section) This tracking of people's movements alone opens up a plethora of different ways of harming people. There are other ways that unregulated facial recognition technology can open up pathways for abuse including

discriminatory practices and security breaches that can leak crucial biometric data just to name a few.

As of January 2024, “the U.S. does not currently have authoritative guidance, regulations, or laws to adequately address issues related to facial recognition technology use” (“Advances in Facial”, 2024, para. 3) apart from a few exceptions. In the report by the National Academies Press, they suggest the consideration of many different types of legislation such as limits on the storing of facial data, requiring certification of use especially in fields where errors can result in serious harm to people such as in law enforcement, and general privacy regulation in commercial practices. (“Advances in Facial”, 2024, Ensuring Responsible Use section) The same report also calls for a standard for image quality and regulation of acceptable accuracy rates for different phenotypes which likely aims to offset some of the inherent racial bias in this technology that will be discussed in more detail later.

The potential harms that come from the lack of legislation have caused state and local governments to take action. Although bans in US governance are quite rare, some cities have started to ban the use of facial recognition technology. (Selinger & Hartzog, 2020, page 119) Sarlin states that, "Virginia recently became the fifth state to curtail the use of the facial recognition by police, while Portland, San Francisco, Oakland, and Boston are among the cities outlawing it." (Sarlin, 2021, para. 9)

Criminal Justice

One of the most well known applications of facial recognition technology is its use in the criminal justice sector. Given a world where facial recognition technology is perfect and that there is full time surveillance of people around the world, law enforcers would simply be able to pull up footage of the crime in question and immediately identify the perpetrator and locate them

immediately. This may sound good in theory, however, in its current practice, aside from the ethical questions of full time surveillance discussed earlier, facial recognition technology just is not ready to be used as sole evidence in criminal justice.

Many conversations about the immediate problems that may arise from facial recognition technology stem from the potential harm that may come from persistent production of inaccurate results. (Selinger & Hartzog, 2020, page 110) These problems include innocent citizens being "put on government watchlists, deprived of due process in court" (Selinger & Hartzog, 2020, page 110), or in Nijeer Park's case, being wrongfully arrested. His charges were very serious which included, "aggravated assault, unlawful possession of weapons, using a fake ID, possession of marijuana, shoplifting, leaving the scene of a crime, resisting arrest. On top of that, Parks was accused of nearly hitting a police officer with a car." (Sarlin, 2021, para. 5) To summarize the investigation, two eyewitnesses claimed that a photo on a fake ID matched the actual criminal they had seen. The police then sent this photo ID to be scanned using facial recognition technology and got a match with Nijeer Parks. "With seemingly no other evidence", the local law enforcement signed off on a warrant for Parks' arrest. (Sarlin, 2021, A fleeing suspect section) Although in the end Parks was deemed innocent due to sufficient evidence, it is frightening how this technology could potentially ruin a law-abiding citizen's life.

Parks' case is not unique. Similar arrests occurred such as Robert Williams and Michael Oliver who were arrested due to incorrect facial recognition results. (Sarlin, 2021, Police use in question section) Although cases where facial recognition results are used as sole evidence are rare, according to a report by the National Association of Criminal Defense Lawyers, police not mentioning to the defense that facial recognition was used at all is more common. (Sarlin, 2021, Police use in question section) Even if facial recognition is used in criminal justice, "the

restrictions on facial recognition being presented as evidence are lax, and there is little transparency about how those matches are being used in criminal cases." (Sarlin, 2021, para. 10)

Racial Bias

All technology reflects human bias. Facial recognition artificial intelligence is no exception. In fact, artificial intelligence has "an unusually great capacity to amplify [human bias]." (Williams, 2023, para. 5) Artificial intelligence is designed to give statistically likely answers that they were trained to give. If the foundations on which the AI was trained was biased, then the results too will be biased. The biases shown in facial recognition technology is especially dangerous because these systems dictate who is more likely to be prosecuted by the police and who is more likely to be wrongfully arrested due to shortcomings in the artificial intelligence.

We see this bias in facial recognition technology in particular towards darker skinned individuals. Among four categories, darker-skinned females, darker-skinned males, lighter-skinned females, and lighter-skinned males, error rates for dark-skinned females were up to 34% higher than with lighter-skinned males. (Najibi, 2022, Inequity in face recognition algorithms section) This is a result of many factors, one of which being that the default camera settings are not optimal for the capturing of darker toned individuals which in turn results in a lower-quality database. (Najibi, 2022, Building a more equitable face recognition landscape section) Starting in the early twentieth century, much of the "skin-color" balance was done by referencing a pale-skin Caucasian lady in order to measure and calibrate skin-tone in printed photos. (Benjamin, 2019, page 275) These light-skinned Caucasian women were referred to as "Shirley" and were the "ideal standard for skin color in most North American and international analog photo labs". (Benjamin, 2019, page 275) Because of this bias, analog color film stock and

digital camera design had a lot of trouble distinguishing darker-skin facial features. (Benjamin, 2019, page 275) It would make sense that this idea of using "Shirley" as the sole reference for testing camera technology would have become less of a problem, but this is not the case. There are a number of cases where lack of diverse test subjects were so obviously apparent. One example is the HP webcam that was launched in 2009, which was supposed to be able to detect faces and track them, however, after a bit of consumer testing, it was revealed that it was only able to detect light-skinned individuals and could not do the same for darker-skinned individuals. (Benjamin, 2019, page 296) When it comes to facial recognition technology, "if the very basis of photographic technology doesn't see dark skin very well, how can you possibly be surprised at the biased results?" (Williams, 2023, Birth of the Transformer section)

Analysis and Discussion

Given what I discussed throughout this paper, let us revisit the questions mentioned earlier: How do the negatives of facial recognition technology affect its use in criminal justice, a field in which this technology is prevalent? To what extent does racial bias play a role in this technology? Is it too early to fully utilize this technology due to issues of privacy, consent, and regulation?

I. Negative Effects on Criminal Justice

Lack of regulation, racial bias, and the context of criminal justice make an appalling combination and currently, facial recognition technology has all of them. These issues work together to create cases like the one with Nijeer Parks. The racial bias is responsible for inaccurate results, criminal justice amplifies the gravity of the results from facial recognition technology, and lack of regulation allows all of it to happen.

The presence of racial bias in facial recognition technology lessens its reliability for targeted racial minorities. The incorrect results disproportionately affect those targeted racial minorities. In the context of criminal justice, when these incorrect results are treated as conclusive evidence, it leads to wrongful convictions. This not only erodes trust in the criminal justice system, but it also perpetuates systemic injustices. The lack of regulation exacerbates these issues, allowing these issues to be practiced without proper guidelines for responsible use.

II. Racial Bias Analysis

It is clear that this technology has a flaw regarding racial bias, most notably the decrease in accuracy when tested on darker-skinned people. The bias traces back to the early adoption of photographic technology. Because the global standard of skin tone was “Shirley”, camera settings and color calibration were tailored a lot more toward lighter skin resulting in lower quality pictures of darker-skinned individuals where facial features are much harder to distinguish. However, to what extent is this an issue? The prevalence of racial bias in facial recognition technology is enough to significantly undermine its credibility as a form of identification system. This is especially true in criminal justice where it is sometimes, whether explicitly disclosed or not, it is used as evidence for an investigation. The fact that this technology is more prone to error for a certain group of people should be enough to deter law enforcement from using it in their practices but it does not.

Beyond issues with the technology itself, the presence of racial bias and continued use through its flaws reflects a larger societal issue when it comes to racial discrimination. This shows a tolerance of racial bias, almost to a point of acceptance, in our society. Continuing to use technology that disproportionately harms racial minorities is detrimental to our society.

III. Is It Too Early?

In short, yes. I believe this technology is far too unrefined to be used by the government and commercial companies especially because it affects people's livelihoods. A term coined by Meridith Broussard in her book *Artificial Unintelligence* that I believe is relevant to facial recognition technology is "technochauvinism" which refers to the flawed assumption that technology is always the solution to every problem. Broussard argues that, "[b]lind optimism about technology and an abundant lack of caution about how new technologies will be used are the hallmarks of technochauvinism", (Broussard, 2019, page 93) which I think is especially relevant to the issues I discussed earlier about the misuse in criminal justice and inherent racial bias coupled with lack of regulation. I think the emergence of new technology excites many people and they immediately think of the potential applications but ignore the potential harms until it is too late. I think this is especially true with the rise of artificial intelligence where everyone is trying to integrate AI faster than legislation is able to protect people from vectors of abuse.

IV. Next Steps

If scrapping this technology as a whole is not an option, how can we improve this technology as it stands? I think, currently, the biggest problem lies in the regulation. We should write and enforce legislation that protects against the issues discussed in this paper. I believe that we need to place limitations on where face images and templates are stored, who gets to store them, and for what purpose they are stored. This will help protect people's privacy when they are out in the public, say at a commercial business. As mentioned in the "Regulation" section, I think we need to establish standards for acceptable image quality and diversity in the data set when it comes to training these AI models. This should help reduce the racial bias that these systems

have learned and increase accuracy rates across all tones of skin. There should be strict guidelines on how this technology is used in criminal justice, especially regarding the transparency when using this technology in a case. The investigation should be clear about if they use facial recognition technology and how it is being used. I do not think that results from facial recognition software should be sufficient evidence for a warrant for somebody's arrest as sole evidence. I think we could benefit from a federal government body that is responsible for regular audits and enforcing ethical and responsible practices.

Conclusion

As a society, we are quick to innovate but slow to regulate. I do not think there is something necessarily bad about regulation lagging behind new technological advances but I do think that people should take more caution and think about the potential downsides rather than just being focused on the excitement of new things. In the case of facial recognition technology, there are new applications in biometric authentication, entertainment like virtual reality systems, and law enforcement but we should stop to think about how it affects people's privacy and ability to give consent or how inherent biases can affect different demographic groups.

The high profile cases where facial recognition was used incorrectly in criminal justice serve as a reminder of the prevalence of racial bias, algorithm inaccuracy, and potential for corruption and abuse of power in law enforcement. Moreover, it is an indication that this technology is not ready for use as it stands.

Moving forward, I think we need a proper plan for regulatory frameworks surrounding this technology to mitigate racial biases and avenues of abuse in criminal justice as well as protect individuals' privacy. With effective legislation to back it up, I believe facial recognition technology can be harnessed into a powerful tool in many applications.

Works Cited

Advances in Facial Recognition Technology Have Outpaced Laws, Regulations; New Report Recommends Federal Government Take Action on Privacy, Equity, and Civil Liberties Concerns. National Academies. (2024, January 17). https://www.nationalacademies.org/news/2024/01/advances-in-facial-recognition-technology-have-outpaced-laws-regulations-new-report-recommends-federal-government-take-action-on-privacy-equity-and-civil-liberties-concerns

Benjamin, R. (2019). *Captivating technology: Race, carceral technoscience, and liberatory imagination in everyday life*. Duke University Press.

Broussard, M. (2019). *Artificial unintelligence: How computers misunderstand the world*. The MIT Press.

Cook, S. (2023, February 27). China's Censors Could Shape the Future of AI-Generated Content. The Japan Times. https://www.japantimes.co.jp/opinion/2023/02/27/commentary/world-commentary/china-artificial-intelligence/

Crawford, K., & Paglen, T. (2021). Excavating AI: The Politics of Images in Machine Learning Training Sets. Ai & Society. https://link.springer.com/article/10.1007/s00146-021-01162-8

Foucault, M. (1979). Panopticism. Penguin.

Hamann, K., & Smith, R. (2019). Facial recognition technology: where will it take us. Criminal Justice, 34(1), 9-13.

Hunter, B. (2023, November 20). *How we can engage with race and racism in research: Developing a racial analysis*. William T. Grant Foundation.

https://wtgrantfoundation.org/how-we-can-engage-with-race-and-racism-in-research-developing-a-racial-analysis

Najibi, A. (2020, October 26). Racial Discrimination in Face Recognition Technology. Science in the News. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

Leong, B. (2019). Facial recognition and the future of privacy: I always feel like ... somebody's watching me. Bulletin of the Atomic Scientists, 75(3), 109-115. DOI: 10.1080/00963402.2019.1604886

Sarlin, J. (2021, April 29). A false facial recognition match sent this innocent black man to jail | CNN business. CNN. https://www.cnn.com/2021/04/29/tech/nijeer-parks-facial-recognition-police-arrest/index.html

Selinger, E., & Hartzog, W. (2020). The inconsentability of facial surveillance. Loyola Law Review, 66(1), 33-54.

Williams, D. P. (2023, July 1). Bias optimizers. *American Scientist*, *111*(4), 204–207.

Winner, L. (1980). Do Artifacts Have Politics? Daedalus, 109(1), 121–136.