Accountability and the Development of Facial Recognition Technology:

How The Road to Oppressive Technology is Paved with Good Intentions

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

By

Daniel McNamara

Spring, 2021

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:

> *"There is a culture in the technology industry which influences people and appears to override their backgrounds and upbringings." -Ruha Benjamin (2019)*

Facial recognition technology (FRT) promises incredible utility in many fields, but the implementation—especially in policing, military, and medicine—often fails to consider ethical implications such as racial bias and personal privacy. While FRT could prove to be a powerful tool, the inherent issue of biased algorithms opens up a debate concerning whether its benefits outweigh its potential for harm. Researchers looking to use FRT to find trends are often misguided and naive about the implications of their work. Since algorithms and code serve as a layer of abstraction between computer scientists and the real world, the ethical implications of their work is not always clearly defined. However, these issues are directly tied to existing systemic issues in society, so it is important for programmers to understand these issues. As Kelly Gates, author of *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*, succinctly put it: "Individual identification is always tied to social classification. It's always there for some specific purpose, and that's usually to determine someone's level of access or privilege. The ethical questions in facial recognition relate to those social hierarchies and how they're established," (Gates, 2011).

FRT is already being used in ethically questionable ways in countries including the United States, China, and India. In the United States, several law enforcement agencies and border protection points use FRT systems to identify individuals (USGAO, 2020, pg. 1). The Chinese government uses FRT to identify and track ethnic minorities in order to further

marginalize them (Stinson, 2020, pg. 27). Similarly, India is creating a massive database of faces in order to spy on and police their citizens despite the technology not being entirely accurate (Sangomla, 2020, pg. 30). This presents a clear and present danger in other countries that are considering FRT for police or military purposes. These systems have the potential to compound on existing societal issues and biases.

Computer scientists must seriously think about the real-world implications of their work. If an algorithm is used in a way that disproportionately harms an underprivileged group, some of the responsibility falls on the people who wrote the algorithm. Even as the technology continues to become more refined and powerful, the main issues surrounding the field are in its real-world applications. The purpose of the research is to explore the culture of FRT and its connections to real world issues. This includes the identifying the intentions of programmers, researchers, and consumers of FRT and how these intentions consider the stakes of its implementation in society.

The State of FRT:

Research and development of facial recognition technology (FRT) is fraught with ethical missteps and misguided intentions. Because of the potential for misuse in controversial areas such as police surveillance and remote military action, computer scientists working in FRT must consider the applications of their algorithms. FRT research has implications beyond the technology itself, so researchers must consider the negative

Prominent examples of questionable FRT projects include algorithms developed by researchers that aim to discern an individual's sexual preference or even proclivity to

criminal behavior (Stinson, 2020). If an algorithm that can discern criminals from law-abiding civilians sounds too good to be true, that's because it is. The data set for training the algorithm relied on mugshots and identification photos. This may sound reasonable in theory, but a flawed data set can unintentionally cause an algorithm to make inaccurate connections. In the case of the criminality algorithm, the researchers failed to account for differences in lighting and angle between mugshots and personal identification pictures. The researchers also defined a criminal as any person with a criminal conviction in court. This assumption introduced an inherent bias to the algorithm because it did not account for societal factors that may lead to higher conviction rates for certain groups of people. When facing criticism regarding the study, the researchers suggested that the results must be examined strictly academically. It is naive to suggest that FRT—or any technology—can be fully understood and applied without considering the implications of its real-world implementation in society.

Understanding how FRT affects society is monumentally important in its development. Even a minuscule false positive rate in a controlled environment can translate to multiple lives being adversely and unfairly affected (Garvie, et al., 2016). This approach requires a deliberate and conscientious effort to examine and analyze social systems and biases that may inform bias in FRT algorithms. For example, Figure 1 (below) graphs the disparity between races in false match rates for two algorithms.

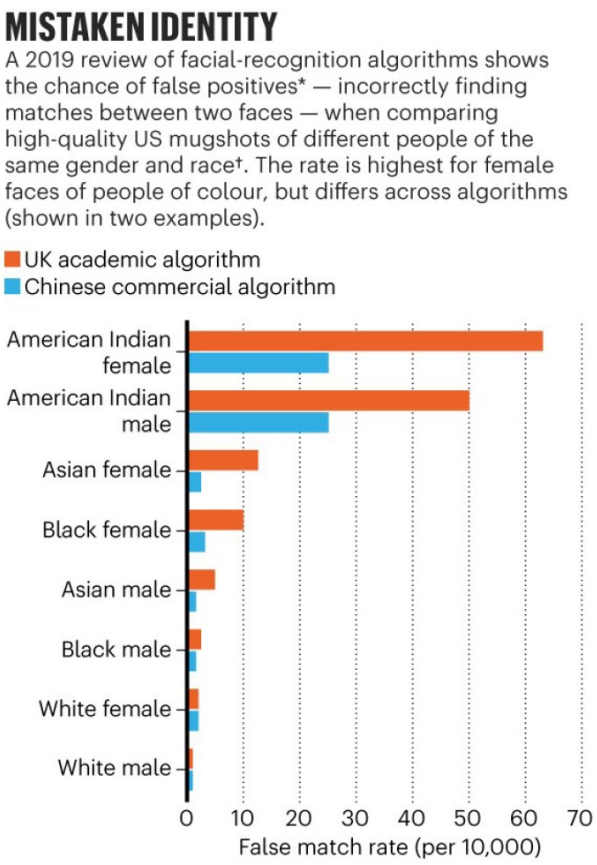

Figure 1: (Castelvecchi, 2021, p. 8)

These disparities, compounded with societal factors—such as significantly higher arrest rates for African-Americans—paint a grim picture of the potential for FRT in police work. Law enforcement is one of the largest consumers of FRT, thus driving a significant portion of the research and development of the systems. While institutional societal issues such as racism are largely beyond the scope of this paper, it is nevertheless important to emphasize that it is incumbent on programmers to consider these issues and how their work impacts the real world. As Ruha Benjamin of Princeton University said, "There is a culture in the technology industry which influences people and appears to override their

backgrounds and upbringings." Her assertions support the importance of conscientiousness in programming. Developers and researchers of FRT strive to create accurate algorithms, but few seem to ever ask themselves a simple question: just because you can develop a technology, should you? These developers and researchers must approach their work just as they do their personal lives by drawing on their life experiences and considering their own morals. The systems that they are developing do not exist purely in the sterile, academic environment that researchers work in; researchers must consider how the systems will operate as part of a larger system. One of the primary goals of this research is to further explore the culture of programming and how programmers approach their role within a broad view of society. The connection of FRT to existing, well-known societal issues provides a perfect topic to explore these issues of ethics and culture.

FRT as a Socio-Technical System:

With time, FRT continues to increase in complexity and ubiquity. For a long time, it appeared that its continued development and expansion is inevitable, but its place in the world is fluid and uncertain. In 2020, in response to the Black Lives Matter movement amplifying issues of racism in policing, several companies such as IBM and Microsoft announced that they were suspending their development of FRT until it is properly federally regulated (Gavet, 2020). This pause is a potentially massive moment in the future of FRT and provides an important opportunity to think about why companies are producing FRT systems and what society stands to gain or lose from them. Oftentimes, FRT is first developed and deployed with good intentions, such as identifying missing children, but later expands to grayer ethical areas, such as identifying peaceful anti-government

protestors or tracking certain discriminated ethnic groups (Sangomla, 2020, p. 32). For this reason, all potential uses and implications must be considered by programmers and engineers that work on FRT algorithms.

The concept of technological determinism provides a framework for discussing these issues of intent and inevitability in engineering with regard to FRT. The development of FRT does not exist in a vacuum and, due to its inherent connections to socio-technical issues, must be considered by developers as a part of a larger ethical system. The developers themselves must also reflect on their roles as agents in this system and the agency they carry in their work. But how much agency do developers truly have when they do not necessarily always have a say in their project assignments? Perhaps changing the way that developers think about their work can influence their feelings of agency in their work.

The example of the ambitious FRT program in India provides a very good basis for establishing a narrative of intent in the field. The technology and system were designed under the narrative that it would be used to identify missing children and dead bodies (Gent, 2018). Under these assumptions, the intentions of the government, police, and developers seem positive and uncontroversial. When the technology was first put into use by the police, it did not perform particularly well at identifying children, but the error rate was acceptable for the purposes. It is troubling, then, to consider the same error rate applied to the same technology when it is used to identify and prosecute protestors. The intentions of the socio-technical system, and therefore the narrative, have changed.

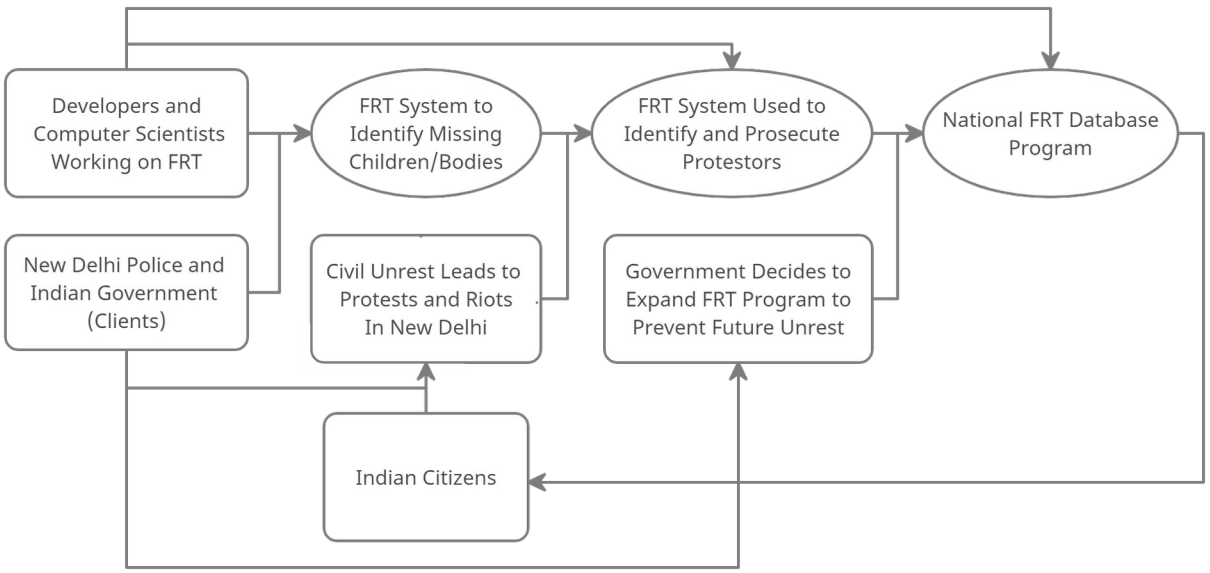

Figure 2: An attempt to visualize the actors, processes, and consequences in the socio-technical system of FRT in New Delhi, India (Sangomla, 2020).

Figure 2 attempts to visualize the actors and influences that impacted the development and implementation of India's National Facial Recognition Database. It is important to note that, in the lifetime of the program, citizens have no influence or consent. At every level of development, the only direct influence comes from the development and implementation. The initial motivations and intentions that kick started the project were perverted and expanded to a larger system unrecognizable from the original implementation to identify missing children.

So where does FRT go from here?:

Establishing a culture of ethics in programming is essential in ensuring a better future for powerful, controversial technologies such as facial recognition. This begins with

continuing ethics education throughout developers' careers. Ethics is a constant pursuit rather than a college course requirement.

As important as personal ethics are in programming, there must be accountability at higher levels with regard to FRT, particularly government entities. Less discriminatory and more accurate algorithms does not necessarily correlate to less discriminatory and exploitative applications. If anything, more accurate algorithms are potentially an even more dangerous weapon if used for the wrong purposes. The most obvious approach to handling this issue is federal regulation of the development and application of FRT by both private and public entities.

Technological determinism and the narrative of intent provide a useful framework for analyzing FRT as a socio-technological system. Since FRT is often presented as an inevitable development and "the next big thing," it is easy to ignore the negative ramifications of its implementation on society as a larger system. Some implementations of FRT, such as Apple's iPhone face-scanning unlocking or Facebook's automatic tagging in photographs, present themselves as novel features on popular platforms that enhance the user experience. This narrative encourages consumers to ignore the inherent risks in the trade-off between convenience and privacy.

Conclusion:

Developers of facial recognition technology face myriad dilemmas beyond the accuracy of their technology; the current global climate of government surveillance, social media data collection, and systemic racial bias creates a hotbed of issues that will define the

future of FRT and its relationship with society. Conscientiousness must take a forefront in the culture of FRT development in order to ensure that algorithms do not contribute to problematic ethical dilemmas such as racial bias in criminal justice. Researchers and developers must consider the end-goals of the system and the stakes of its implementation. The intentions of the system and its developers often fail to consider the endless complexities of real-world implementation as a part of a larger system. The example of the FRT system in India illustrates this issue well, since the system began with the intention to help police identify missing children, but eventually was used to identify and track peaceful protestors. The noble intentions of the system were perverted into a much more morally fraught implementation.

FRT will likely continue to be a controversial topic for the foreseeable future. Developers must consider their role as a part of a larger socio-technical system when working on projects that potentially enforce systemic issues or contribute to human rights issues. Creating a perfect FRT algorithm would be as monstrous as it is laudable if there is no oversight into its use to violate human rights and perpetuate unjust systems in society.

References:

Bromberg, D. E., Charbonneau, Ã., & Smith, A. (2020, January 1). Public support for facial recognition via police body-worn cameras: Findings from a list experiment. *Government Information Quarterly, 37*(1).

Castelvecchi D (2020, November 1). Is facial recognition too biased to be let loose?. *Nature*, 587(7834), 347 – 346.

Gates, K. (2011). Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance. New York: New York University Press.

Gavet, M. (2020, July 1). Facial recognition technology is inevitable—it's time we make it human-centered. Fast Company. https://www.fastcompany.com/90522885/facial-recognition-technology-is-inevitable-its-time-we-make-it-human-centered.

Gent, E. (2018, May 5). Face recognition spots missing Indian children. *New Scientist*, 238(3176), 8 - 7.

Katti H, & Arun SP (2019, July 1). Are you from North or South India? A hard face-classification task reveals systematic representational differences between humans and machines. *Journal of Vision*, 19(7), 1 – 0.

Kortli, Y., Jridi, M., Falou, A. A., & Atri, M. (2020, January 15). Face Recognition Systems: A Survey. *Sensors (14248220), 20*(2), 1 - 36.

Leong, S., Phan, R. C.-W., Baskaran, V. M., & Ooi, C. (2020, November 1). Privacy-preserving facial recognition based on temporal features. *Applied Soft Computing Journal*, 96.

Payal, P., & Goyani, M. M. (2020, March 1). A comprehensive study on face recognition: methods and challenges. *Imaging Science Journal, 68*(2), 114 - 127.

Sangomla, A. (2020, October 16). FACE OFF: Facial recognition has become a frontline policing tool in India amid fears that it is prone to errors and allows the government to expand surveillance without much oversight. *Down to Earth, 29*(11), 28 – 37.

Saran, C. (2019, May 28). Are software algorithms inadvertently racist?. Computer Weekly, 4 - 6.

Stinson, C. (2021, January 1). The Dark Past of Algorithms That Associate Appearance and Criminality. *American Scientist, 109*(1), 26 – 29.

United States Government Accountability Office. (2020). Report to Congressional Requesters: Facial Recognition. *GAO-20-568*. 1 Sep. 2020, pp. 1 – 101.