VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA
Wellington Faculty of Engineering
Te Wāhanga Ahunui Pūkaha
Ethics and Sustainability Symposium

Recommendations for Ethical Usage of Facial Recognition Technology

Christopher Benfell

Abstract

This paper investigates the current state of Facial Recognition Technology (FRT) and the ethical concerns surrounding it discussed in literature. Based on these concerns, a code of ethics is developed to provide recommendations on how to use FRT ethically. The essence of this code is to protect the rights of citizens and to use FRT in a way that minimizes misuse and benefits all. Finally, a specific use case of FRT is introduced to demonstrate how the proposed code of ethics can be applied effectively to a real project. The findings are then summarized, and recommendations made going forward.

Keywords: Facial recognition; Artificial intelligence; Ethics; Biometrics; Privacy.

1. Introduction

The success of Artificial Intelligence (AI) and Machine Learning (ML) has enabled the development of many technologies too complex for traditional programming. One such technology is Facial Recognition Technology (FRT). FRT has the capability to identify an individual by scanning images of them and matching them with images, or biometric maps of their face.

While FRT has many beneficial uses, such as an additional security layer or for identifying criminal suspects, there are also many ethical concerns with such systems and their development process combined with their lacking regulation [1]. These issues have placed a lot of public disapproval on FRT and have led to calls for changes to their implementation, usage, and for legal intervention [2, 3].

Many of the perceived issues with FRT are rooted in the erosion of fundamental human rights, such as privacy and individual freedom. It is also not uncommon for FRT systems to incorrectly identify and incriminate innocent citizens [4]. Finally, FRT can be used without the knowledge of an individual to collect data and build a profile of one's behaviour or interests for purposes such as surveillance or targeted advertising.

1.1. Objective

The aim of this paper is to provide recommendations for ethical usage of FRT based on an analysis of its current state. This is achieved by first identifying current ethical issues with FRT from literature in Section 2. Section 3 introduces a code of ethics to follow when developing or implementing FRT. This is based on the issues identified in Section 2 with discussions on how they are addressed. Section 4 then provides an example of how these principles are applied to a specific use case of FRT. Lastly, Section 5 summarizes the results of this study and provides direction and next steps for further ethical use of FRT.

2. Literature Review

This section focusses on identifying and contextualising the ethical concerns with FRT expressed in literature. Each concern is described and discussed regarding why it is problematic and what should be done to reduce the effects of unethical FRT.

2.1. Erosion of Privacy and Rights.

The issue most prevalent within literature is the erosion of privacy and other human rights through the widespread use of FRT. This occurs throughout all phases of FRT including research, development, and deployment of such systems in differing ways [1, 5]. However, the most egregious violations of privacy are performed by governments, law enforcement, and corporations involved in FRT.

The main use case of FRT within government and law enforcement is public surveillance with the goal of identifying or locating criminal suspects to reduce or prevent crime [6]. Such surveillance is done with the goal of improving public security and safety, but this comes at the cost of reduced public freedom and privacy of citizens. Public use of FRT gives law enforcement the power to effectively track and monitor any individual without restraint as most countries lack FRT regulation [4]. Additionally, this infringes on inherent rights to privacy, autonomy, and freedom. For such reasons, public use of FRT was temporarily banned within the European Union in 2021 to allow laws and regulations to catch up [7]. Corporations can also use FRT in a similar way to monitor customers within stores or through their devices in the case of technology companies. This can be used to benefit the customer but can also be used maliciously or to collect data without risk of being held accountable.

Despite, resistance against the adoption of the technology, early users of FRT have claimed the benefits of increased security outweigh the loss of privacy [6]. While FRT is undeniably valuable towards reducing crime and increasing public safety, the invasiveness of the practice can conversely make citizens feel less safe and trusting of law enforcement. FRT will be unwelcome due to the many numbers of innocent citizens that must be monitored and scanned regularly for FRT to be successful. Such a system has the potential to be easily abused by governments if not properly maintained. Examples of this have already occurred in China where citizens have been publicly shamed for simply wearing pyjamas outside [8].

However, with enough regulation and oversight, FRT can become more acceptable in public spaces [6]. For example, citizens could be provided notification of public FRT use so they are aware of the implications of being in the area. However, public FRT will always be a breach of privacy, so for FRT to benefit all, it must be used in a way that respects privacy laws and is unobstructive of ordinary citizens. Such as holding footage for no longer than necessary and facilitating significantly more justified arrests than unjustified.

2.2. Algorithmic Bias and Inaccuracy

Another concern with FRT which is common in literature is the discrepancies in accuracy rates between different races and false positives. Many studies have reviewed the efficacies of various FRT algorithms and have always found a higher rate of false positives amongst people of colour (POC), most notably black and Asian people [4]. It was found that 35% of errors occur on female women of colour as opposed to 1% on white males amongst US FRT systems. This discrepancy indicates an inherent racial bias with FRT algorithms resulting in further discrimination against these people whether intended or not. This has severe implications of individuals being falsely identified and becoming the victim of an unjustified arrest, another violation of rights.

The main cause of this algorithmic bias is the dataset on which the FRT is trained upon. As FRT is simply an ML algorithm, it requires an incredibly large dataset of people's faces and biometric data for it to learn how to recognize and identify faces. A lower proportion of POC within this dataset will result in the FRT being less accurate at identifying individuals within these groups. Furthermore, a separate dataset may be used during operation for identifying individuals which may increase the false positive rate based on its distribution. This problem is exacerbated further when combined with racist policing strategies which may fill these databases with POC, causing the system to have more data it cannot classify accurately [1].

Beyond racial biases, FRTs do not hold perfect accuracy in general. Anyone can be falsely identified and implicated in crimes. FRT algorithms are said to ensure a classification accuracy of 90%, which is not equal between ethnicities, but this drops down to 50% among mask wearers which is common during the spread of Covid-19 [1]. Additionally, FRT errors can direct more harm or disregard towards minority groups such as transgender people which systems will fail to identify [3].

The issue of algorithmic bias and inaccuracy are difficult to resolve given the obfuscated reasoning of ML algorithms., however, several steps can be taken to reduce such errors. Firstly, a representative dataset can be built to train new FRT systems and reduce the issue. Secondly, the results of identification can be further scrutinized to ensure the correct individual is found. Lastly, law enforcement can be transparent about the qualities and usage of their FR system to be accountable for mistakes and provide public understanding of the issue.

2.3. Data Collection and Lack of Transparency

FRT systems require incredibly large datasets to develop and evaluate their efficacy, which is typical of AI, however, the methods by which this data is obtained is a point of concern. In most cases, these databases are built by scraping photos of faces from various internet sources and building biometric data [3, 5]. Image aggregation sites, such as Flickr, or social media sites, like Facebook, are prime targets due to large numbers of publicly available images. These databases are often shared between research groups and corporations. One such example is MSCeleb from Microsoft where 10 million images were scrapped from the internet and distributed [9].

Data can also be collected through smartphones with cameras equipped with FRT as a security layer. The developers of these devices could easily obtain large amounts of biometric data by taking secretly from their users. Similarly, governments and law enforcement often have access to large national services containing facial data, such as a driver's license database, from which an extraordinary amount of data can be obtained. Countries such as Zimbabwe have already done this with CCTV cameras, financial systems, national databases, etc…, to build a facial database of millions of its citizens [10]. Additionally, there is often no regulation against such actions unless governments or corporations place these upon themselves.

Most of these methods of building facial databases are done without the consent or knowledge of those involved. While many images are publicly available or under creative commons licenses, the individuals involved have likely not consented for their data to be used for FRT [3]. This also breaches privacy and, more specifically, the right to control one's personal information. There are also cases when consent is obtained, but the researcher is not transparent about the intended use of such data [11].

Many researchers involved in FRT are calling for regulation to be put in place to ensure that the collection or usage of large datasets is managed through ethical bodies [2, 5]. However, these don't necessarily apply to corporations or governments. Instead, these issues should be remedied through the adoption of ethical practices. Corporations should strive to receive informed consent from each participant, including transparency on the intended use or sharing of data, to provide the owner full control over their information.

2.4. Data Storage and Security

One concern less discussed in literature is the way facial and biometric data is stored. Given that facial and biometric data are unique identifiers for each individual, it is critical that these are stored to the highest data protection standards to ensure data privacy and security [4]. Data breaches are a big privacy concern for the public and government as poor storage practices can lead to leaks of massive amounts of sensitive biometric data and loss of privacy. Leaks of such information also has the potential for identity fraud when using biometric security systems. For example, Apple's iPhone FaceID was previously breakable in under 120 seconds [12]. In many countries, there is also no legal framework to claim compensation for such a violation leaving individuals on their own. Therefore, all data facial and biometric data collected should be stored in a trustworthy cloud storage system with high-security standards and proper data encryption [1]. Protecting the data in this way is likely the best way to ensure integrity for those in the dataset.

2.5. Misuse of Facial Recognition Technology

The final issue discussed in this paper is the misuse of FRT. FRT can be developed to do a variety of things, and this includes malicious actions. For example, corporations can use FRT to inform hiring decisions beyond or to collect data and build profiles on individuals [3]. Another issue is function-creep, where the system is modified to perform tasks beyond those originally approved [6]. The nature of these additions could result in malicious use of the FRT, such as sensitive data collection or tracking of civilians.

One of the biggest examples of intentional malicious use of FRT is currently happening in China. The Chinese government has employed FRT specifically tuned to identify Uyghurs and enable mass surveillance and detentions of these minority groups to re-education camps [5, 13]. Developing FRT to specifically target minority groups is likely one of the most harmful use cases of the technology and is simply unethical.

The examples in this section are all ways that FRT should not be used. The best possible recommendation to avoid abuse of the technology is to simply not engage in such actions and to hold accountable those who do.

3. Code of Ethics

This section will detail and justify principles for ethical usage of FRT. These principles are based on the issues identified through literature and ways to combat them.

3.1. Uphold Privacy Rights

This principle protects the rights of civilians in relation to usage of FRT in public spaces. FRT should be used in a way that is beneficial to all. It should improve public safety for citizens without eroding rights and should assist law enforcement with identifying suspects. Human rights should still be prioritised with the implementation of public FRT. The principles of necessity and proportionality can be implemented to decide whether FRT would be beneficial for a given location and outweigh the loss of privacy.

3.2. Promote Fairness

Due to inaccuracies in FRT and the lower accuracy rates for POC, human oversight of FRT systems is important for minimizing the risk of falsely arresting innocent people. To ensure few false positives are acted upon, a human should be analysing the results of the FRT to determine if it has found a valid match. Proper oversight of FRT will help reduce inherent biases in the algorithm and contribute to a more fair and effective system.

3.3. Obtain Informed Consent

When collecting an individual's data, FRT users should receive written consent from the individual to the intended usage. The intent should be clearly stated and unambiguous. The data should not be used beyond these consented terms. This is important for upholding privacy rights as it gives the individual control over their data and how it is used. This principle is difficult to uphold but should be followed as much as possible. Civilians can additionality be given notification of public FRT usage to implicitly obtain consent. However written forms of consent should be preferred.

3.4. Be Transparent and Accountable

Users of FRT should be transparent about how, where, and why they are using FRT and data. This applies to both obtaining consent and deployments of FRT. Additionally, there should transparency around how the system works and the issues with it along with users of FRT being properly accountable for mistakes.

3.5. Ensure Data Security

Given that FRT requires the use of sensitive facial and biometric data, the chosen storage method should comply with robust data protection standards stored to the highest data protection standard. This is important for preventing data breaches of sensitive and unique information. Ideally, this could be with a trustworthy cloud provider with a dedicated security team to ensure no data breaches occur. Data should further be encrypted to make it harder for unauthorized people to access.

3.6. Avoid Misuse

FRT should not be used in ways that are directly harmful to others. Examples of this include targeting minority groups or harvesting data from users. The effects of all FRT usage should be carefully considered to identify potential issues with the system.

4. Case Study

Recently, the New Zealand (NZ) Police have done an investigation into the benefits and risks involved in using FRT [14]. So far, NZ Police has chosen to delay implementing FRT into their work and instead follow recommendations to develop a robust system first. The main goal of the NZ Police was to use FRT to improve their ability to identify and capture suspects, similar to other police forces. This section will discuss how the code of ethics developed here can apply to future use of FRT by NZ Police and how this will contribute to a more ethical approach to FRT in law enforcement.

4.1. Uphold Privacy Rights

NZ Police can follow this guideline through compliance with NZ Privacy Laws and other relevant legislation when operating FRT. Furthermore, NZ Police should carefully analyse each deployment of FRT to measure how effective it is at stopping crime in the area. The NZ Police should determine whether it is justified to continue surveillance in each area, or if the benefits fail to outweigh the erosion of privacy and human rights. Additionally, NZ Police aims to develop policies for regulating the use of FRT in public places to further uphold the rights of citizens.

4.2. Promote Fairness

This principle means that the NZ Police should have people carefully overseeing the use of FRT and monitoring the results it produces. Their goal will to be determine whether the system has correctly identified an individual before notifying other officers in order to prevent disruptions and unjust consequences to those unrelated. If this is implemented effectively, it should result in most arrests through FRT systems being justified and of wanted suspects. Furthermore, with a human checking the results, NZ Police will be able to better manage the bias in the FRT algorithms, leading to a much fairer and better performing system.

4.3. Obtain Informed Consent

Regarding the usage of FRT for surveillance, it would be unfeasible for the NZ Police to gain the consent of every individual within the area of operations. This makes it harder to follow the recommended consent guidelines, however, they can at least make citizens aware of the fact by posting notifications of FRT usage in the area. Policy development by the NZ Police on public FRT usage also suggests regulations for when capturing and storing public images of citizens are acceptable to ensure the public is well informed on how FRT is used.

4.4. Be Transparent and Accountable

One thing the NZ Police should do to maintain public support for their use of FRT is to be transparent. This includes how they are using it, how their system works, and where they are using it. The report from the NZ Police suggests they intend to follow this route by providing reports on developments of their FRT system and implementing ways to understand the collective public opinion on the matter [14].

4.5. Ensure Data Security

All data used by the NZ Police, whether that be training data or data obtained through operation, should be stored in compliance with robust data protection standards. Dedicated cloud storage providers are recommended along with strong data encryption. This is important to protect the privacy of NZ citizens and prevent the leakage of sensitive information.

4.6. Avoid Misuse

NZ Police has already laid out suggestions for ensuring FRT systems are not subject to abuse or scope creep [14]. This will be achieved by continuous governance and oversight of deployed FRT systems to assure they are well maintained and are fit for purpose with no functionality past their original scope. Any additional features to FRT capabilities must be approved before deployment. The suggested NZ Police framework will be effective for ensuring that FRT is used appropriately and is not at risk of being used maliciously.

5. Conclusion and recommendations

This study investigated the current ethical concerns surrounding the use of FRT and developed ways to approach these problems through a code of ethics. The main issues found with FRT were the lack of respect for privacy, erosion of public freedom, the algorithmic bias, the data collection practices, lack of transparency, and misuse of FRT. Each point has been addressed within the code of ethics with recommendations on how to ethically handle each one. Lastly, an example use case of public FRT by the NZ Police was used to demonstrate how these principles can be applied.

The suggested code of ethics only provides a starting point for encouraging ethical usage of FRT. In the future, more concrete measures for ensuring FRT is used fairly and to benefit all need to be put in place. Things such as laws and regulations will be necessary going forward to prevent abuse and exploitation of such technology by governments and corporations against the common people.

References

[1] K. R. Gangarapu, “Ethics of Facial Recognition: Key Issues and Solutions,” G2, 25 January 2022. [Online]. Available: https://learn.g2.com/ethics-of-facial-recognition. [Accessed 25 April 2022].

[2] nature, “Facial-recognition research needs an ethical reckoning,” *nature,* vol. 587, p. 330, 2020.

[3] I. D. Raji, T. Gebru, M. Mitchell, J. Buolamwini, J. Lee and E. Denton, “Saving Face: Investigating the Ethical Concerns of Facial,” in *AAAI/ACM Conference on AI, Ethics, and Society*, New York, USA, 2020.

[4] J. Sarabdeen, “Protection of the rights of the individual when using facial,” *Heliyon,* vol. 8, no. 3, p. e09086, 2022.

[5] R. V. Noorden, “The ethical questions that haunt facial-recognition research,” *nature,* vol. 587, pp. 354-358, 18 November 2020.

[6] P. Brey, “Ethical Aspects of Facial Recognition Systems in Public Places,” *Journal of Information, Communication & Ethics in Society,* vol. 2, no. 2, pp. 97-109, 2004.

[7] T. Madiega and H. Mildebrath, “Regulating facial recongition in the EU,” September 2021. [Online]. Available: https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/698021/EPRS_IDA(2021)698021_EN.pdf. [Accessed 3 May 2022].

[8] D. V. Boom, “Chinese city uses surveillance tech to shame citizens for wearing pajamas outside,” CNET, 22 January 2020. [Online]. Available: https://www.cnet.com/culture/chinese-city-uses-surveillance-tech-to-shame-citizens-for-wearing-pajamas-outside/. [Accessed 3 May 2022].

[9] Y. Guo, L. Zhang, Y. Hu, X. He and J. Gao, “MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition,” in *Computer Vision - ECCV 2016*, Amsterdam, 2016.

[10] A. Hawkins, “Beijing’s Big Brother Tech Needs African Faces,” Foreign Policy, 24 July 2018. [Online]. Available: https://foreignpolicy.com/2018/07/24/beijings-big-brother-tech-needs-african-faces/. [Accessed 3 May 2022].

[11] C. Wang, Q. Zhang, W. Liu, Y. Liu and L. Miao, “Expression of Concern: Facial feature discovery for ethnicity recognition,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 2018.

[12] D. Winder, “Apple's iPhone FaceID Hacked In Less Than 120 Seconds,” Forbes, 10 August 2019. [Online]. Available: https://www.forbes.com/sites/daveywinder/2019/08/10/apples-iphone-faceid-hacked-in-less-than-120-seconds/?sh=b407a3a21bc3. [Accessed 3 5 2022].

[13] P. Mozur, “One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority,” The New York Times, 14 April 2019. [Online]. Available: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html. [Accessed 3 5 2022].

[14] New Zealand Police, “Police release findings from independent expert review of Facial Recognition Technology,” New Zealand Police, 9 December 2021. [Online]. Available: https://www.police.govt.nz/news/release/police-release-findings-independent-expert-review-facial-recognition-technology. [Accessed 3 May 2021].