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

Ethics of Facial Recognition Technology in Law Enforcement: A Case Study

Oscar Camplin

Abstract

Facial Recognition Technology (FRT) has promising applications in law enforcement due to its efficiency and cost-effectiveness. However, this technology poses significant ethical concerns that overshadow its benefits. Responsible use of FRT requires consideration of these ethical concerns that legislation fails to cover. This study investigates the ethical issues of FRT use and relevant ethical frameworks and principles designed to combat these issues. Drawing on this, we propose and discuss a code of ethics for FRT to ensure its ethical use in the context of New Zealand law enforcement.

Keywords: Ethics; Facial recognition technology; Law enforcement.

1. Introduction

Facial Recognition Technology (FRT) is becoming increasingly widespread, showing promise for various applications. FRT involves analysing and comparing facial features between a captured image and existing facial images to find a match. First, features are extracted from the image by analysing attributes such as size, shape, or relative position of facial parts such as the nose or eyes. This is then processed using Artificial Intelligence or other algorithms to match the features to a facial image contained within a database.

As this technology improves, it has attracted global interest from law enforcement agencies as a cost-effective and efficient solution to assist crime-fighting efforts. Its use by law enforcement can be broadly categorised into two types: static image or dynamic video analysis, also known as Live Facial Recognition (LFR) [1]. In both categories, FRT has mainly been used for verification and identification. An example is in the United Kingdom, where border security uses FRT to verify identities against passport pictures [2]. In 2016, the Metropolitan Police Service deployed LFR at a carnival. The South Wales Police followed suit in 2017, deploying LFR to identify people on a watchlist at the UEFA Champions League final [1].

Although seemingly beneficial, this technology has the potential to do more harm than good. The improvement to public safety that FRT may deliver hangs in the balance of the pressing ethical concerns it presents, such as privacy, bias, and function creep discussed in [3–5]. Therefore an ethical code of conduct for FRT should be carefully developed and used to address these concerns.

1.1 Objective

The objective of this paper is to develop a code of conduct that guides ethical use of FRT. The framework's principles draw from understanding of ethical issues and existing ethical frameworks for FRT found in a literature review. The framework is applied to the specific case of FRT use in NZ law enforcement to assess its strengths and limitations.

2. Literature Review

The following section will analyse relevant literature to identify ethical issues of FRT use. We aim in this section to extract key ethical implications that we can use to apply to a code of ethics in the context of FRT for NZ Law Enforcement.

2.1. Ethical Issues of FRT

The concern surrounding the use of FRT have prompted many research efforts to investigate ethical issues. In [4], Smith and Miller analyse ethical issues drawing from applications in Australia, the United States, and the United Kingdom. Privacy is discussed as one of the primary issues. The study highlights that while FRT indeed threatens a person's fundamental right to privacy, the notion of privacy does not have clear boundaries. One similar study [6] suggests that public places do not have a strong privacy expectation, while [7] argues that dragnet FRT use violates reasonable privacy expectations, illustrating a divide in the perception of privacy. This makes it difficult to draw the line between ethical and what is not, perhaps indicating that ethical principles that cover grey areas of privacy should be implemented regardless. On the other hand, many privacy issues may already be considered protected by the International Bill of Human Rights [8]. This begs the question of whether complete transparency (e.g. providing public notice of FRT use) may adequately cover legal grey areas of privacy, as people could avoid being subject to FRT systems where they may feel their privacy is invaded.

Smith and Miller [4] also draw from the scenario of metadata sharing between government agencies to highlight that data gathered could be linked with other data collected for another purpose without proper justification. A lack of transparency and consent poses significant ethical concerns in these situations. Like [4], [6] discusses the idea of function creep, pointing out that widening the scope of FRT use could be unethical. This is certainly something to be considered for law enforcement, where FRT could be used for a myriad of purposes, ranging from identifying persons with a warrant for arrest or, more broadly, identifying persons of interest. While the former may be acceptable, the latter could be considered harassment.

FRT has faced heavy criticism over the ethical concerns that arise from bias. Bacchini and Lorusso [5] explore these concerns and discuss the contribution of FRT to the perpetuation of racial discrimination. The study sheds light on how black people in the United States are overrepresented in many facial databases used in FRT due to higher stop, arrest and incarceration rates. This leads to disproportionate numbers of matches. The paper also points out that FRT has difficulties identifying faces with darker features due to a lack of contrast. This is demonstrated in [9], which discovered noticeably lower matching accuracy on black people across six different FRT algorithms. Even with balanced databases and highly accurate FRT algorithms, there is always some bias present, illustrating the need for ethical principles that actively mitigates it. One of these potential mitigations is to add human oversight to FRT processes. However, [10] argues that this is a false comfort, suggesting that humans are also a source of bias when overseeing FRT processes.

2.2. Ethical Frameworks and Principles for FRT use

One report [11] reviews FRT use in NZ and its legal and ethical implications. The report analyses FRT using existing NZ ethical frameworks [12–14] for FRT or similar technologies. The set of principles developed [12] by the Privacy Commissioner for the safe and effective use of data analytics is undoubtedly relevant to the ethical use of FRT. These principles include transparency, treaty partnership, people focus, fit for purpose data, privacy, and retaining human oversight. Although this standard briefly recognises the relevance of the Treaty of Waitangi, there is no depth provided on what this entails. In [14], Māori Data Sovereignty principles such as Manaakitanaga and Rangatiratanga can be used to give this depth to ethical principles that ensure that FRT use upholds and respects the treaty.

In [3], the ethics of FRT use in law enforcement are discussed in consideration of human rights frameworks. The paper poses ten ethical questions that should be addressed for any use of FRT. Privacy, accountability, and function creep are the primary ethical concerns these considerations aim to address. The paper suggests that who develops, procures, tests, and management of FRT should come under question to challenge any biases. In the context of NZ, this could be applied to the Treaty of Waitangi to ensure collaboration on FRT implementation occurs between Māori stakeholders and the law enforcement agency. Although this paper does not present an ethical framework or code, its questions provide a solid foundation for developing a framework for FRT.

In the MPS LFR Policy Document [15], we can observe the above ethical considerations implemented into a framework specific to law enforcement. Although these guidelines include measures required by legislation, stipulations for ethical deployment are also included in the framework. One important stipulation is that the technology must not "result in bringing unacceptable gender or racial bias into policing operations" however, it is unclear what is unacceptable in this context. To counter potential injustices brought about by LFR use, such as above, the policy also stipulates that LFR operators should be trained to understand risks and limitations. The three clauses a, c, and e aim to enforce proper justification of LFR use and ensure that its benefits are not outweighed by any harm the technology brings. These policies arguably address the foremost significant ethical concerns for FRT deployment, however, they may not be comprehensive enough in all contexts, such as NZ.

3. Code of Ethics

This section presents a code of ethics for FRT use based upon the ethical issues and existing frameworks explored in the previous section. Ethical issues can be used to form fundamental principles that should be included in this code. Existing frameworks are used to identify best practice regarding how the stipulations of these principles are defined.

3.1. Non-discrimination

I. FRT and surrounding processes must not result in unfair outcomes to people based on attributes including but not limited to race, gender, and age.

II. The agency must take measures to actively understand and mitigate bias present in the system. These measures include:
   a. Training of any personnel that have oversight over FRT processes or operation to understand the technology, its limitations, and awareness of personal biases.
   b. The FRT system, including its processes and policies, must be designed, developed, delivered and governed in partnership with stakeholders representative of the community.
   c. Data used must not represent groups disproportionately unless justified.

III. The technology must be tested periodically to measure its accuracy. This must provide evidence that the technology maintains an agreed level of accuracy, which is to be independently reviewed.

3.2. Transparency

I. Information regarding the FRT system and its use must be readily available. This information must include:
   a. The purpose and justification of FRT use for each deployment.
   b. Policies on FRT use, including how data is collected, how long it is retained for, how it is used, how accountability is ensured, and how the technology is reviewed.
   c. Time and place of FRT deployment.

d. Technical documentation describing how the technology works and its limitations.

II. Information regarding FRT must be disclosed before any deployment of the technology. This notice period must be agreed upon by relevant stakeholders.

III. Disclosure of information must be active rather than on request. The information must be easy to access.

3.3. Accountability

I. The technology, data, and processes are traceable, auditable and explainable.

II. An audit trail must provide evidence of compliance with these ethical principles and relevant legislation.

III. The FRT system must not make decisions autonomously. It must only be used as an assistive tool to make informed and reviewed decisions that a human operator is accountable for.

3.4. Purpose and Scope

I. The FRT system must have a clearly defined purpose. This purpose must be justified and accepted by relevant stakeholders before deploying the technology.

II. FRT use must have a clearly defined scope that aligns with the purpose. Relevant stakeholders must agree upon this scope. The scope must not be widened during deployment. FRT use must not fall outside of the defined scope.

III. Data collection, processing or analysis part of FRT must not fall outside the defined scope or purpose.

4. Case Study Discussion

Live Facial Recognition technology is unprecedented in NZ for law enforcement. Current and potential uses of FRT for the NZ Police remain limited to recognising static images [7]. Live FRT poses similar ethical concerns to static FRT, however, its element of public surveillance adds additional factors. This provokes the question of how the technology might be deployed ethically by NZ law enforcement. The following section will analyse the code of ethics proposed in section 4 for LFR use by the NZ Police.

4.1. Non-discrimination

One of the challenges with LFR is bias, which can result in discrimination. Some level of bias will always be present in LFR systems, whether it be in the technology itself or the processes and people surrounding it. The Non-discrimination principle aims to minimise this presence of bias as much as possible. Even though, as argued in [12], human oversight is a false comfort for minimising bias, clause 2a addresses this. Clause 2b is essential for this case because it aligns with the Treaty of Waitangi principle of partnership. Any partnership will also help in the effort to uphold Manaakitanga (upholding the dignity of Māori). The mitigation of bias that clause 2 overall aims to achieve is not only crucial in other jurisdictions to comply with human rights but is especially important for upholding Manaakitanga in NZ.

Another challenge of LFR use is the unequal representation of data used in the system, which causes bias. Clause 2c aims to remedy this but is potentially undermined by data sources the NZ Police may use (e.g. from the criminal justice system). This presents a problem because of the significant overrepresentation of Māori people in the system [16]. If such data sources are used, the potential for discrimination of Māori naturally increases, therefore violating Manaakitanaga. Since equality may be unattainable in all use cases, the clause also aims to ensure that proper justification is provided for such disproportionate levels of representation.

Even though FRT systems cannot be 100% accurate, it is essential that the system's accuracy is tested, understood, and maintained to minimise bias. Clause 3 addresses this by ensuring that an agreed level of accuracy is maintained and independently reviewed to promote trust.

4.2. Transparency

An essential aspect of deploying FRT is transparency. This is especially true for LFR, where the technology is deployed in public areas. In such cases, people must be aware that surveillance is taking place, which is addressed by clause 2. This allows people to choose to be present in areas where LFR is deployed, offering some form of consent. In the context of NZ, Manaakitanga stipulates that Māori must be given free, prior and informed consent, which this clause may address to some degree. It is also important that the public is made aware of why the technology is being used; thus, clause 1 enforces that clear justification is made. Furthermore, clause 1 ensures that the public is aware and educated about the technology, processes, and policies. Finally, clause 3 aims to address the nature of transparency. Failure to be actively transparent regarding all such cases could incite public discomfort and mistrust, especially due to the public nature of LFR.

4.3. Accountability

The accountability principle aims firstly to ensure that any use of LFR is traceable, auditable and explainable. This is important to ensure that any failures of the system to comply with policy can be identified, understood, explained, and learnt from to improve LFR and its surrounding processes. Clause 2 stipulates that an audit trail must be kept, which will help to ensure this traceability. An essential aspect of this is that information that is part of this trail must be able to provide evidence of compliance. If compliance is not kept, then the part of the system where non-compliance occurred can be traced to a point where an aspect of the system can be held accountable. While this can be used to identify non-compliant elements of the LFR system, 'blaming' the technology may not always be an adequate response.

In some cases, a person must be held accountable for decisions resulting from LFR. Clause 3 addresses this by ensuring that human oversight of the system is guaranteed. By enforcing that the technology must not make any final decisions and that it must be used in an assistive capacity, a human operator can be held accountable for non-compliance with ethical and legal policies.

4.4. Purpose and Scope

This is a principle designed to address the ethical issues of function creep and misuse. The adaptability of LFR enables it to be used for various purposes, some of which may be unethical. Clause 1 tries to address this by ensuring that a purpose is clearly defined, justified, and agreed upon by relevant stakeholders. In this case, relevant stakeholders may include representatives of the Māori community, which would help to uphold Treaty partnership. Treaty partnership can also be applied to Clause 2, which stipulates that the scope of FRT use must be defined and agreed upon before use. Once the technology is deployed, the clause states that the scope mustn't be widened, which aims to combat function creep of the technology. However, not allowing the scope to be widened would limit the potential uses of the technology. Therefore the principle stipulates that the scope must be agreed upon **before** deployment and should not be widened **during** deployment. An important aspect of LFR is data collection (for facial images and metadata), processing, and analysis. Clause 3 ensures that data collection must also fall within the scope and purpose of the LFR use. This is important for LFR use in NZ to protect against data misuse and uphold Manaakitanga (respectful use of the data). Although this clause may cover

data misuse in other jurisdictions, the NZ Privacy Act [17] contains legislation that overlaps with this clause, thus rendering it somewhat unnecessary in this case.

5. Conclusion

FRT will undoubtedly be commonplace for law enforcement agencies in years to come. The significant ethical issues of FRT explored in this paper, such as privacy, bias, and function creep, most definitely highlight the need for a comprehensive ethical framework governing its use. Ethical frameworks must address these issues while not being so restrictive that FRT use does not benefit the community's safety. While the general code of ethics proposed in this paper certainly addresses some of these issues, there are still some gaps when applied to LFR use by the NZ Police, especially regarding upholding the Treaty of Waitangi principles. While the proposed code may not comprehensively address all cases, it provides foundations for future iterations. Therefore, we recommend that this code of ethics be built upon through trials and stakeholder co-design before finalising it for actual deployment.

References

[1] Information Commissioners Office, "ICO investigation into how the police use facial recognition technology in public places," 2019. [Online]. Available: https://ico.org.uk/media/about-the-ico/ documents/2616185/live-frt-law-enforcement-report-20191031.pdf (accessed: 2022-04-07).

[2] J. Sanchez del Rio, D. Moctezuma, C. Conde, I. Martin de Diego, and E. Cabello, "Automated border control e-gates and facial recognition systems," Computers & Security, vol. 62, pp. 49–72, 2016.

[3] D. Almeida, K. Shmarko, and E. Lomas, "The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of us, eu, and uk regulatory frameworks," AI and Ethics, pp. 1–11, 2021.

[4] M. Smith and S. Miller, "The ethical application of biometric facial recognition technology," Ai & Society, vol. 37, no. 1, pp. 167–175, 2022.

[5] F. Bacchini and L. Lorusso, "Race, again: how face recognition technology reinforces racial discrimination," Journal of Information, Communication and Ethics in Society, vol. 17, no. 3, pp. 321-335, 2019.

[6] P. Brey, "Ethical aspects of facial recognition systems in public places," Journal of Information, Communication and Ethics in Society, vol. 2, no. 2, pp. 97-109, 2004.

[7] M. Hirose, "Privacy in public spaces: The reasonable expectation of privacy against the dragnet use of facial recognition technology," Conn. L. Rev., vol. 49, p. 1591, 2016.

[8] United Nations, "The International Bill of Human Rights," 1948. [Online]. Available: https://www.ohchr.org/sites/default/files/Documents/Publications/Compilation1.1en.pdf (accessed: 2022-04-16).

[9] B. F. Klare, M. J. Burge, J. C. Klontz, R. W. V. Bruegge, and A. K. Jain, "Face recognition performance: Role of demographic information," IEEE Transactions on Information Forensics and Security, vol. 7, no. 6, pp. 1789–1801, 2012.

[10] B. Green and A. Kak, "The false comfort of human oversight as an antidote to AI harm," Organisation 2021. [Online]. Available: https://slate.com/technology/2021/06/human-oversight-artificial-intelligence-laws.html (accessed: 2022-04-05).

[11] N. Lynch, L. Campbell, J. Purshouse, and M. Betkier, "Facial recognition technology in New Zealand: Towards a legal and ethical framework," New Zealand Law Foundation, 2020.

[12] Privacy Commissioner, "Principles for the safe and effective use of data and analytics," Organisation 2018. [Online]. Available: https://www.privacy.org.nz/assets/New-order/Resources- /Publications/Guidance-resources/Principles-for-the-safe-and-effective- use-of-data-and-analytics-guidance3.pdf (accessed: 2022-04-05)

[13] F. Tweedie and S. Welsh, "Trustworthy AI in Aotearoa," AI Forum of New Zealand, 2020. [Online]. Available: https://data.govt.nz/assets/data-ethics/algorithm/Trustworthy-AI-in-Aotearoa-March-2020.pdf (accessed: 2022-04-05).

[14] Te Mana Raraunga, "Principles of Māori data sovereignty," 2018. [Online]. Available: https://www.temanararaunga.Māori.nz/s/TMR-Māori-Data-Sovereignty- Principles-Oct-2018.pdf (accessed: 2022-04-05).

[15] Metropolitan Police Service, "MPS LFR Policy Document," United Kingdom 2020. [Online]. Available: https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/advice/lfr/policy-documents/lfr-policy-document.pdf (accessed: 2022-04-12).

[16] Ministry of Justice, "Safe and effective justice," 2021. [Online]. Available: https://www.justice.govt.nz/justice-sector-policy/key-initiatives/hapaitia-te-oranga-tangata/ (accessed: 2022-04-13).

[17] Ministry of Justice, "Privacy Act 2020," 2020. [Online]. Available: https://www.legislation.govt.nz/act/public/2020/0031/latest/whole.html#LMS23223 (accessed: 2022-04-16).