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

Case Study: Code of Ethics for Facial Recognition Technology

Isabella Tomaz Ketley

Abstract

Facial Recognition Technology (FRT) is a way of identifying an individual by comparing their facial signature to a database of known faces. It has many current uses, and this is only expected to grow due to its many potential and promising applications. However, this technology presents a lot of ethical issues that need to be addressed. This paper examines current literature to find the current ethical issues of FRT such as its usage to infringe on one's privacy, use an individual's data without their consent and misidentify individuals. Based on these issues, a code of ethics will be created that will ensure an individual's privacy and data security and minimize the biases and misidentifications in FRT. This code of ethics is applied and assessed via a case study about the usage of live facial recognition technology in the New Zealand Police.

Keywords: Facial Recognition; Ethics; Ethical Framework.

1. Introduction

Facial Recognition is a form of Artificial Intelligence which uses biometric data to recognise faces. There are a range of current applications of facial recognition technology (FRT) ranging from unlocking phones to identifying criminals. These uses and any potential uses of FRT create a range of ethical issues which need to be addressed. This paper will aim to determine these ethical issues via a literature review and provide a code of ethics to combat these issues. Furthermore, a case study about using live facial recognition technology within the New Zealand Police will discuss how the code of ethics can be applied to ensure its ethical usage.

1.1. Background

FRT is a way to identify an individual's face via the use of biometrics. It does this by mapping facial features from images or videos and comparing them to a database of known faces [1]. It aims to find a match between the scanned face and someone in the database. This allows for the identification and verification of individuals, which has many useful applications. FRT is currently widely used, and this usage is only expected to grow [1]. For example, FRT is currently used when unlocking phones, boarding flights, and searching for missing people [1]. However, due to the intrusive nature of this technology, FRT raises some ethical concerns.

1.2. Objective

The objective of this paper is to design a code of ethics for FRT to ensure its ethical usage. The code of ethics will be created based on the ethical issues FRT is faced with and proposed solutions to these issues outlined in current literature. It will then be applied to a case study to understand its effectiveness and any limitations it may have.

2. Literature review

This section reviews and analyses current literature to determine current ethical issues with FRT and ways to combat these issues. These findings will be used to create a code of ethics for FRT. The literature used in this review is selected based on its relevance to FRT and publish date. It is important to review recent literature as FRT is constantly changing. Due to this, only literature from the past four years is used.

2.1. Ethical Issues

2.1.1. Privacy

Information privacy is easily breached by FRT, presenting an ethical issue. Information privacy gives people the right to control how their personal information is collected and used [2]. However, as FRT doesn't require any physical interaction, people easily lose the ability to give consent to this [2]. This results in people's facial information being collected involuntarily, which is unethical. For example, without people's consent, Clearview AI captured billions of images from online platforms to create a facial recognition database for law enforcement agencies [2]. Furthermore, Facebook was sued due to its use of FRT to automatically suggest photo tags [2]. Hence, the collection and usage of people's data, without their consent, is a major ethical issue in FRT.

Additionally, FRT can easily breach an individual's personal privacy. People have the right to be free from undue surveillance [3]. However, using FRT to monitor people, can be done against one's consent and without one's knowledge. This can cause FRT to be used in an unethical manner and infringe on people's privacy. An example of this is when China implemented a mass surveillance system to target ethnic monitories [4, 5]. This usage is unethical and should be prevented. However, the use of FRT for surveillance isn't always unethical, for example when it's used to track down terrorists. This presents the issue of determining when using FRT with surveillance is ethical [4, 6]. The use of FRT for surveillance, without sufficient justification, is another significant issue that needs to be prevented.

2.1.2. Data Protection / Security

Another ethical concern FRT presents is the potential for people's biometric data to be misused or accessed without authorisation. Facial data is sensitive information, which is unique to one's identity, and unlike a password, it cannot be easily changed. Hence, any unauthorised access or usage of this data would not only be unethical but could also have negative effects on individuals [7, 8]. Data breaches and confidentiality breaches occur due to a lack of data protection and can result in identity theft, stalking and harassment [8]. Hence, the lack of data protection poses an ethical concern to the usage of FRT.

2.1.3. Accuracy and Biases

The lack of accuracy and the biases in FRT technology poses a large ethical issue. Facial recognition algorithms have high misidentification rates and show some sort of bias. An evaluation performed in [9] showed that on average, facial recognition APIs perform worse on darker skin tones and that it is least accurate for dark skinned females. This creates ethical issues due to FRT misidentifying individuals, which can lead to false arrests and exacerbate existing racial biases in society [10, 11]. For example, FRT in Detroit was used to identify a thief, and this resulted in a black man being falsely arrested [11, 12]. The inaccuracy of FRT, especially for dark-skinned people, has huge ethical issues which need to be resolved.

2.2. Proposed Solutions of Ethical Issues

2.2.1. Privacy

The non-consensual collection and usage of facial data is an ethical issue which must be addressed. Facebook combatted this issue by agreeing to get informed consent for specific usage of their FRT [13]. This solution is consistent with what a survey, conducted by Nature, found researchers believed to be the ethical solution [10]. This survey showed that people most commonly believe that informed consent must be obtained before using an individual's facial data. Furthermore, principles outlined in [14] and [15] specify that to ethically collect and use data, informed, written, and specific consent must be given. Hence, violating one's information privacy can be combatted by obtaining informed consent before the collection and usage of data.

Violating the personal privacy of individuals is another ethical issue which needs to be combatted. People have the right to be free from undue surveillance, however, how can we determine when surveillance is excessive or disproportionate. This is considered in [4] which explains that using FRT to track down a terrorist can be ethically justified. However, using FRT to track down a petty thief is not ethically justified. Hence, to determine what is considered undue surveillance, and what is ethical, it is proposed that the principles of necessity and proportionality are used [4, 7]. These principles relate to determining whether FRT is needed, or whether less invasive methods would suffice and whether the use of FRT is a proportionate response [4]. From this, we can conclude that the principles of necessity and proportionality help to ensure people's rights aren't unjustly violated.

Another commonly proposed solution to ensure the privacy of individuals is for FRT be transparent. There is a consensus that FRT must be transparent about when it is being used, how people's data is stored, and how long their data is stored for [9, 16]. This is especially important when FRT is being used in public places, so that individuals know if they are subject to FRT. By ensuring FRT is transparent, it can help people to retain their privacy.

2.2.2. Data Security

A proposed solution to combat data misuse and data leaks in FRT is to implement data security. Data misuse occurs due to insufficient security and accountability measures and data breaches occur due to vulnerabilities in software or insufficient protection of data. Due to this, FRT needs to have appropriate security measures in place. To mitigate vulnerabilities in software, [18] outlines that the security measures should involve the identification and remediation of any security vulnerabilities and the resilience to malicious attacks. Furthermore, there is a consensus that there should be protocols to prevent the unintended use and unauthorized access of data [14, 15, 18]. More specifically [15] states that access controls and data auditing should be implemented. By ensuring FRT implements sufficient data protection measures, data misuse and leaks can be mitigated.

2.2.3. Accuracy and Biases

The inaccuracy and biases of FRT need to be resolved via more testing and training. The most common way to combat this issue is to ensure FRT is trained on a database which is large enough and diverse enough to correctly represent the population [19, 20]. It is also suggested that standards be implemented for both the accuracy rates of FRT, and the quality of images used [19]. It has been shown that facial recognition algorithms can be extremely accurate, as a study performed by NIST showed many tested algorithms had an accuracy rate exceeding 99% [19]. From this, we can conclude that by ensuring certain standards are met, the issue of lack of accuracy can be resolved.

Furthermore, a proposed solution to minimize the consequences of any FRT misidentifications is to ensure human oversight. [16] writes that FRT does not need to be 100% statistically accurate, so long as any FRT outcomes are treated as predications rather than facts. FRT should only be used to inform human decision making and its results should be reviewed and validated by a human, before acting upon them [7, 16]. This human oversight ensures that FRT does not make decisions autonomously, to reduce any misidentifications of FRT being acted upon.

3. Code of Ethics

This section outlines a code of ethics designed from the findings of the literature review. The goal of this code is to ensure that FRT is used in an ethical manner. FRT is becoming more widely used despite its ease of unethical use. Hence, it is essential to create this code to ensure the growing use of FRT is used appropriately.

3.1. Principle 1. Be Accurate and Impartial

FRT must be sufficiently statistically accurate for their purposes and provide reliable results across all demographic groups. There should be minimal if any, error rates. An accuracy threshold for the FRT must be set with a clear justification for why this threshold was chosen. Any facial recognition algorithms used must undergo robust testing to ensure it matches the threshold.

3.2. Principle 2. Be Transparent

There must be transparency regarding when and why FRT is being used, how an individual's data will be used and stored and how long their data will be retained. Additionally, when FRT is being used in public areas, there must be clear and visible signage stating that FRT is in use, why this is, and how to access more information.

3.3. Principle 3. Incorporate Human Oversight

Human oversight is required to review, validate, and act on any decisions made by FRT. Any outcomes of FRT need to be treated as an informed estimate to be inspected by a human, rather than a fact to be acted upon instantly.

3.4. Principle 4. Obtain Informed Consent

An entity must obtain informed, explicit, and written consent from an individual prior to collecting, using, and storing their facial data. This consent must include what specific purpose the data will be used for and anyone the data will be shared with. Informed consent must be re-acquired before an individual's information is used for or shared with anyone not specified in the original agreement.

3.5. Principle 5. Ensure Data Security

FRT must have ample security measures and access controls in place to prevent the leakage or misuse of data. The technology must employ audit logs and access controls which comply with the principle of least privilege. Firewalls should also be deployed to prevent outside access to any information. Furthermore, penetration testing should be performed yearly, to ensure the system remains secure. The vulnerabilities or risks identified in any penetration tests must be acted upon to ensure they are prevented or mitigated.

3.6. Principle 6. Be Necessary and Proportionate

When FRT is used, impacts on individuals, groups and wider communities must be considered and any potential infringements must be proportionate to the need and the benefits of its use. If FRT could infringe on people's rights, privacy, or beliefs, then it should only be used if it is necessary, and if the outcome is proportional to the cost.

4. Case Study Discussion

The New Zealand Police (NZP) currently have an image management system called 'Photo Manager', which only has limited facial recognition capability [7]. They do not use any live facial recognition (LFR) technology, nor do they own any public CCTV cameras [17]. However, the usage of LFR could be very useful in different policing situations, such as locating high-risk suspects or monitoring certain spaces for offenders [17]. Despite the potential benefits LFR can have for the NZP, a lot of these usages are ethically questionable. The biggest ethical issues are that LFR can infringe on people's privacy and result in false arrests. Due to this, this case study will discuss and apply the code of ethics outlined in section three, to the use of LFR in the NZP. This will aim to provide an ethical framework for LFR technology in the NZP.

4.1. Principle 1. Be Accurate and Impartial

LFR technology within the NZP must be statistically accurate and impartial. The algorithms used in the FRT must have very minimal error rates and provide reliable results across all demographic groups. The accuracy threshold of the NZP's FRT should be high due to the severe consequences misidentifying an individual can have. The NZP will also need to justify this accuracy threshold and ensure their technology regularly undergoes robust testing to ensure this threshold is met.

4.2. Principle 2. Be Transparent

The NZP needs to be transparent with its usage of LFR, especially as the technology will be deployed in public spaces. People must know where they are subject to LFR and how their information is being used. Hence, the NZP will need to provide clear and visible signage in the locations where LFR technology is being used.
Furthermore, people must be able to easily find further information about the NZP's usage of LFR. The NZP must make the public aware, and document how an individual's information is being used and stored and how long their information will be retained.

This is essential in ensuring the ethical usage of LFR technology and ensuring trust with the public. However, it can also negatively impact the efficacy of LFR. If the NZP make it known which areas are being monitored, then criminals will avoid these areas. Hence, this presents a limitation of this principle.

4.3. Principle 3. Incorporate Human Oversight

To adhere to this principle, the NZP must have employees responsible for reviewing, validating, and acting on any outcomes made by LFR technology. They should also have policies to ensure that no final decisions are made without human oversight. Any LFR used by the NZP must be treated as assistance, and any outcomes it makes must be treated as a prediction which needs to be inspected by a human.

This principle is important in the NZP due to the severe consequences misidentifications can have, such as falsely arresting an innocent person. However, it also increases the necessary workload when using LFR technology.

4.4. Principle 4. Obtain Informed Consent

The NZP needs to obtain informed, explicit, and written consent from an individual prior to collecting, using, and storing their facial data. This includes when collecting people's data for the NZP's LFR database. This ensures that people have a say in whether their biometric data is collected and stored by the NZP. The consent that the NZP must obtain, must be specific about what an individual's data will be used for, and anyone this data will be shared with.

This principle is very limiting, and it is not always practical to follow, especially in this usage within the NZP. It is likely that there may arise some situations where the NZP will need to search for a high-risk criminal who has not given consent for their data to be collected and stored. This is something that this principle does not consider, however, principle six addresses this.

4.5. Principle 5. Ensure Data Security

The NZP must ensure that it has acceptable security measures in place to prevent any data breaches or misuse. The NZP's LFR technology will store sensitive information about lots of people and due to this, this data must be protected. To prevent any malicious attacks or data breaches, the NZP must have policies in place to ensure at least yearly penetration tests are performed on their FRT and databases. The policies should also ensure any findings are followed up on and mitigated. Furthermore, the NZP must deploy an internal and external firewall to further protect people's data.

The NZP will need to put in place certain security measures and policies, so that people have limited privileges. This means the NZP will need to employ access controls to ensure the principle of least privilege is followed. Furthermore, there should be audit logs to record any access and changes to data. These implemented policies must ensure the confidentiality and the integrity of any facial recognition data the NZP has.

4.6. Principle 6. Be Necessary and Proportionate

Within the NZP, there may be situations where they want to or need to infringe on one's rights or beliefs such as privacy. One of the ways that the NZP wants to use LFR is to monitor certain spaces for offenders, however, this violates people's right to privacy. Due to this, to ensure LFR is only used in ethical situations, any potential usage of LFR will need to be assessed to determine if it infringes on anything. Hence, the NZP will need to implement a set of policies and procedures to ensure each different proposed usage of LFR is assessed and evaluated. These policies also need to ensure that if there are any potential infringements, LFR is only used when it is necessary and proportionate.

New Zealand has a Treaty partnership with the Māori and so it is necessary to consider this when assessing any potential infringements caused by LFR. Any violations of the Treaty of Waitangi, and the Māori Data Sovereignty poses a negative effect. Hence, the NZP must consider these infringements, as well as the infringements on people's rights.

5. Conclusion and Recommendations

The possible applications of FRT present many ethical concerns which need to be addressed. The use of FRT can result in infringements on an individual's rights, exposure and misuse of very personal information, and negative consequences on people due to its lack of accuracy and biases. These ethical concerns need to be addressed and prevented when using FRT. Hence, section three outlines a code of ethics to address these issues. It is recommended that businesses using FRT follow this code and implement specific obligatory policies and procedures to ensure the code is followed. Furthermore, there should be sufficient training on these policies and procedures so people can follow them correctly. The proposed code of ethics outlined in this document provides a good basis to combat the main ethics issues FRT is faced with. However, it is recommended that this code be applied, tested, and likely built upon in the future. FRT has many beneficial applications, and by following this code of ethics and expanding upon it, we can help to ensure it is used ethically.

References

[1] S. Symanovich and NortonLifeLock, "What is facial recognition? How facial recognition works," Norton, 20 August 2021. [Online]. Available: https://us.norton.com/internetsecurity-iot-how-facial-recognition-software-works.html. [Accessed Apr. 29, 2022].

[2] X. Lai and P.-L. P. Rau, "Has facial recognition technology been misused? A public perception model of facial recognition scenarios," Computers in Human Behavior, vol. 124, Science Direct, 2021.

[3] "Right to Privacy and Freedom From Surveillance," Liberty Victoria, [Online]. Available: https://libertyvictoria.org.au/content/right-privacy-and-freedom-surveillance. [Accessed Apr. 19, 2022].

[4] M. Smith and S. Miller, "Facial Recognition and Privacy Rights," in Biometric Identification, Law and Ethics, Springer, Cham, 2021, p. 21–38.

[5] C. Buckley and P. Mozur, "How China Uses High-Tech Surveillance to Subdue Minorities," New York Times, 22 May 2019. [Online]. Available: https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html. [Accessed Apr. 19, 2022].

[6] C. Fontes and C. Perrone, "Ethics of surveillance: harnessing the use of live facial recognition technologies in public spaces for law enforcement," Technical University of Munich, 2021.

[7] N. Lynch, L. Campbell, J. Purshouse and M. Betkier, "Facial Recognition Technology in New Zealand: Towards a Legal and Ethical Framework," Open Access Te Herenga Waka-Victoria University of Wellington, 1 Dec 2020. [Online]. Available: https://openaccess.wgtn.ac.nz/articles/report/Facial_Recognition_Technology_in_New_Zealand_Towards_a_Legal_and_Ethical_Framework/17204078/1. [Accessed Apr. 19, 2022].

[8] T. K. Lively, "Facial Recognition in the United States: Privacy Concerns and Legal Developments," ASIS, 1 December 2021. [Online]. Available: https://www.asisonline.org/security-management-magazine/monthly-issues/security-technology/archive/2021/december/facial-recognition-in-the-us-privacy-concerns-and-legal-developments/. [Accessed Apr. 18, 2022].

[9] I. Raji, T. Gebru, M. Mitchell, J. Buolamwini, J. Lee and E. Denton, "Saving Face: Investigating the Ethical Concerns of Facial," in AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, New York, 2020.

[10] R. V. Noorden, “The ethical questions that haunt facial-recognition research,” Nature, 18 November 2020. [Online]. Available: https://www.nature.com/articles/d41586-020-03187-3. [Accessed Apr. 17, 2022].

[11] J. Borenstein and A. Howard, “Emerging challenges in AI and the need for AI ethics education,” in AI Ethics, 2021, p. 61–65.

[12] B. Allyn, “'The Computer Got It Wrong': How Facial Recognition Led To False Arrest Of Black Man,” National Public Radio, 24 June 2020. [Online]. Available: https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig. [Accessed Apr. 20, 2022].

[13] N. Singer and M. Isaac, “Facebook to pay $550 million to settle facial recognition suit,” New York Times, 29 Jan 2020. [Online]. Available: https://www.nytimes.com/2020/01/29/technology/facebook-privacy-lawsuit-earnings.html. [Accessed Apr. 17, 2022].

[14] ACLU, “An Ethical Framework for Facial Recognition,” National Telecommunications and Information Administration. [Online]. Available: https://www.ntia.doc.gov/files/ntia/publications/aclu_an_ethical_framework_for_face_recognition.pdf. [Accessed Apr. 17, 2022].

[15] Y. Zeng, E. Lu, Y. Sun and R. Tian, “Responsible Facial Recognition and Beyond,” ArXiv, 2019. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1909/1909.12935.pdf. [Accessed Apr. 17, 2022].

[16] Information Commissioner’s Office, “Information Commissioner’s Opinion: The use of live facial recognition technology in public places,” Information Commissioner’s Office, 18 June 2021. [Online]. Available: https://ico.org.uk/media/2619985/ico-opinion-the-use-of-lfr-in-public-places-20210618.pdf. [Accessed Apr. 28, 2022].

[17] N. Lynch and A. Chen, “Facial Recognition Technology: Considerations for use in Policing,” New Zealand Police, November 2021. [Online]. Available: https://www.police.govt.nz/sites/default/files/publications/facial-recognition-technology-considerations-for-use-policing.pdf. [Accessed Apr. 28, 2022].

[18] Department of Industry, “Australia’s Artificial Intelligence Ethics Framework,” Department of Industry, Science, Energy and Resources. [Online]. Available: https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles. [Accessed Apr. 20, 2022].

[19] M. McLaughlin and D. Castro, “The Critics Were Wrong: NIST Data Shows the Best Facial Recognition Algorithms Are Neither Racist Nor Sexist,” Information Technology and Innovation Foundation, 27 January 2020. [Online]. Available: https://itif.org/publications/2020/01/27/critics-were-wrong-nist-data-shows-best-facial-recognition-algorithms. [Accessed Apr. 17, 2022].

[20] A. Najibi, “Racial Discrimination in Face Recognition Technology,” Harvard University, 24 October 2020. [Online]. Available: https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/. [Accessed Apr. 17, 2022].