**Ethical Framework for a Facial Recognition AI used in Identification**

**1.0 Introduction**

In recent years, the integration of artificial intelligence (AI) into our everyday activities has seemingly become commonplace. This is especially so for facial recognition AI, which is now present in a variety of consumer-focused products, including phone/device unlocking, and automated photo sorting. Further development and implementation of this software can have some great advantages, namely greater personal and state-level security. This however does not come without consequences, and to date, the ethical implications of this technology have not been effectively addressed.

With the recent widespread adoptions of this newly emerging and under-regulated technology into consumer electronics, there have been discussions on what the ethical practices regarding these technologies will look like. In particular, the use of facial recognition AI for identification and surveillance raises a collection of ethical concerns. These concerns often include but are not limited to the protection of user privacy, the potential disenfranchising of minorities, and the issue of accountability for the decision made by the technology.

This case study aims to identify and consolidate the key issues outlined in the ongoing discussions on the development and implementation of facial recognition AI for identification within New Zealand. It will then present a practical code of ethics designed to resolve these issues by providing an ethical framework for their development and implementation. This will be followed by a discussion on the possible enforcement methods of this framework, with suggestions on implementation for both current and future development and usage.

**2.0 Ethical Considerations of Facial Recognition**

This study focuses on three ethical issues, privacy, bias, and accountability. The following section of this study aims to discuss the wide range of ethical implications that both the development and implementation of a facial recognition AI may have regarding these issues. It should be noted that the ethical issues discussed in this study have been selected due to their repeated occurrence in existing literature discussing ethical AI [1] [2] [3] [4], and are by no means the only issued faced by this technology.

**2.1 Privacy**

The issue of privacy in the most basic sense is allowing the user to maintain the right to access and control their own data [1] [2]. This issue is one that is embedded in every aspect of facial recognition AI, including the collection of data, the sharing of data, and user access to data.

The underlying ethical issue to this definition of privacy in its current state is the idea of providing consent [1] [2]. For an entity to gain access to a user’s data there must be an expression of consent regarding the aforementioned collection of data, sharing of data, and user access to data. Currently, this consent is obtained with privacy agreements, in the form of long and obfuscated walls of text that encourage the reader to simply accept the agreement to access the provided service. It is said in IEEE’s Ethically Aligned Design that “General obfuscation regarding privacy policies … demonstrate that even when individuals provide consent, the understanding of the value regarding their data and its safety is out of an individual’s control” [1]. This implies that although legislatively the required consent for the access of data may have been met, the ethical duties of the collecting party have not. As a result of this, the users right to know and have control over how their data is being used has been infringed.

The lack of meaningful consent is a key ethical issue in the collection of the data to produce AI training datasets. The prevalence of social media and other public file and photo sharing websites in society has allowed for the mass collection of image-based data sets. These datasets are created by ‘scraping’ these social media platforms for public images and the data associated with them [5]. As these platforms tend to require the acceptance of some form of privacy agreement as discussed prior, these collections are within legal bounds. This collection method does not notify the users of these platforms that the collection is taking place, and as such removes the agency of the individuals to be able to deny the collection. This loss of agency means that the users are no longer capable of controlling their own data and have inherently lost their right to control it. Examples of this are seen in the vast number of datasets produced from the social media site Flickr [5] [6]. For this particular social media, consent is granted for the collection of data solely by having an active account. This has led to many people being unaware and unable to opt-out of the collection and widespread sharing of their data [5].

The lack of meaningful consent is not an issue that is limited to social media and extends well beyond the bounds of the internet. Surveillance using CCTV cameras is often considered to be a norm in todays society, and an individual would not be alarmed to see themselves being filmed as they enter private property. This filming within private property remains ethical from a deontological perspective as there is no loss of agency. This is due to the individual’s ability to choose whether they will enter the observed property, allowing them to preserve the right to control their own data. This however cannot be held true for the issue of the surveillance within public spaces, and although it is agreed that there is a certain loss of privacy within these spaces, few believe there should be no expectation of privacy [7]. When looking at the New Zealand privacy act of 2020, there are no principles directly relating to the collection of information within a public setting [8]. It can be noted however that Information privacy principle 3 – “*Collection of information from subject*” requires an agency to ensure that all individuals are aware that their data is being collected [8] [9]. Regarding the use of CCTV cameras in public places, there is often little to no indication that data is being collected, leaving people unaware that they are being surveyed. This presents an ethical issue regarding the beforementioned right for individuals to control their own data. If a given individual is not provided with the knowledge that their data is being collected, they have lost the agency to act on that information, and therefore no longer control their own data.

Equally important is the user’s ability to access and modify their collected data. In the New Zealand privacy act of 2020, privacy principle 6 - “Access to personal information*”* and principle 7 - “*Correction of personal information”* it is stated that a user must be able to request access to their data, and have the ability to correct or remove data that has been collected [8]. However due to the nature of AI datasets this is not always possible. According to ‘exposing.ai’, individuals can request to be removed from further distributions of a dataset; however, they cannot retroactively be removed from the copies already distributed [5]. This again brings forwards the ethical issues around the loss of both access and control of your own data.

**2.2 Bias**

The ability to hold a pre-conceived opinion about someone and have that affect how we perceive that person is inherently human. Because of this, the concept of a bias held by an inherently inhuman agent is foreign to us. This however is a well-known issue for currently existing and developing facial recognition AI’s [1] [2].

This internal bias found within facial recognition AIs can often lead to discrimination against minority groups. This discrimination can manifest in many ways but will typically be seen as an inability to recognise, or a misidentification of the given minority [10]. This can often lead to disastrous effects depending on the situation, with some cases of misidentification leading to false and unlawful arrests.

Research suggests that AI are subject to bias based on the location of their development, with AI’s performing more accurately on the populations that developed them [11]. This inherent bias is also not limited to race, and these same studies show that gender disparities in the development of facial recognition AI will lead to inherent bias as well [10]. This issue becomes more apparent when we look at the demographics that generally tend to hold the role software developer. From the 2020 Google Diversity Report, it can be seen that a large majority of Google’s employees identify as either Asian or White in race/ethnicity, and over two-thirds of employees are men [12].

When a facial recognition AI has an inherent bias, it means that it will be less effective at correctly identifying the racial or gender groups that it is biased against [11]. This leads to a range of possible ethical issues dealing with this unequal representation and treatment [1]. When looking at this from a deontological perspective it can be said that all people have a right to equality irrespective of their race or their gender. With this bias present in facial recognition, AI’s this right to equality has been broken, and it is the duty of the developers and those using the AI to ensure that these biases are removed to protect this right. This same conclusion can be drawn by looking at this issue from a utilitarianist perspective, in which the greatest good can be achieved by reducing the effects of bias, thereby preventing discrimination.

Another area of bias within facial recognition AI is within the datasets being used to train them. A biased dataset will disproportionately provide unfair and unethical results when used on the groups that the dataset is biased against [13]. This is currently a key issue regarding the training of these AI, as it has historically been very difficult to produce a dataset that is absent of these biases. One such example of a biased dataset is the IMDB-WIKI dataset, which is predominantly white at 81.8% [14]. Due to the disproportionate number of white people compared to other racial groups, if used for training this dataset is unlikely to produce a fair and unbiased AI. This draws on the same ethical issues as discussed previously, meaning that developers and those using the AI have the duty to ensure that these biases are removed to protect the right to fairness.

**2.3 Accountability and Responsibility**

The ideas of accountability and responsibility regarding facial recognition AI are based on the knowledge that the system will inevitably fail at some point. It is this expectation of failure that allows for those developing the Ai to describe the steps that must be taken to mitigate and prevent ethical harms [1]. Based on this idea that the system will fail at some point, the ethical implication of this to both the users and the developers can be identified and discussed.

One such possible mode of failure is the disparity between the requirements and expectations of software engineers to their company and the ideas of professionalism. The current software engineering environment values and incentivises the meeting of functional requirements, deadlines, and financial constraints [1]. This environment often has the engineer placing the interests of the company before the ethical requirements of the end-users [15]. This conflict of interests between software development companies and the end-users transforms the ethical decision-making process into one of competition rather than collaboration [15]. This can be seen as innately unethical; the developers are expected to pursue the interests of the company over the ethical interests of the users.

**3.0 Recommendations**

The following section of this report will outline the practical code of ethics that has been designed to resolve the issues discussed in section 2 and loosely based upon existing AI codes of ethics [1] [2] [16]. This code of ethics will focus on each of the three topics discussed aiming to present a set of enforceable rules that should be followed to maintain ethical practices around the production of facial recognition AI. There will also be recommendations given as to the enforcement of this code of ethics, to help ensure that the ethical framework is upheld.

The following code of ethics is written to meet the ethical requirements laid out by deontological and teleological frameworks. These frameworks have been chosen as they are most suited to the implementation of ethics in practical design and programming [1]. This is due to these ethical frameworks being centred around the idea of a duty to preserve rights, allowing for them to easily be deconstructed into a collection of implementable rules.

**3.1.0 Code of Ethics**

**3.1.1 Privacy**

* **Legal Requirements** – An entity designing a facial recognition AI to be trained or used on datasets that include New Zealand citizens must conform to the standards set out by the 2020 Privacy Act [8].
* **Collection** – When collecting data from an individual in any form (whether that be for training or active usage), meaningful consent must be obtained. This requires that the individuals be aware that data is being collected and are aware of the purpose and use of the collected data.
* **Access &** Control– Individuals must be capable of accessing all their data and have control over its use. This includes the ability to request the removal and correction of data.
* **Sharing** – An entity must inform and receive meaningful consent from an individual regarding the sharing of data. It is the responsibly of the collecting entity to amend any edited or removed data from copies of the dataset that have then been shared.
* **Public Spaces Collection ­**– Entities wishing to collect data from public spaces must inform individuals entering the area that the collection is occurring. This requires that the individuals be aware that data is being collected and are aware of the purpose and use of the collected data.

**3.1.2 Bias**

* **AI Development** – An entity that develops an AI for the purpose of facial recognition must ensure that all internal biases are removed.
* **AI Usage –** An entityusingAI for the purpose of facial recognition must regularly monitor and test for any internal biases that become present within the system. This process should integrate members of the potentially disenfranchised group to help the diagnosis of issues.

**3.1.3 Accountability and Responsibility**

* **Professionalism** – When developing an AI for the purpose of facial recognition, the concept of professionalism within the workplace must include and encourage active discussion on ethical implications.
* **External Interests** – Entities developing AI for the purpose of facial recognition must hold a fiduciary duty towards the ethical requirements of end users.

**3.2 Enforcement**

Seeing as this code of ethics has mostly been written to conform with a deontological and teleological framework, the enforcement of the code will rely mostly on legislation. This section will seek to briefly discuss the legislative changes that should take place to help enforce this code, and any other methods of enforcement that be employed.

**3.2.1 Privacy**

It is the recommendation of this study that the section pertaining to privacy be enforced legislatively. This will result in breaches of privacy as it has been defined in this code, to result in legal ramifications. This enforcement will require substantial additions to the New Zealand privacy act of 2020 [8], emphasising and redefining consent. It is also suggested that amendments are made to existing privacy principles 3, 6, and 7, to establish specific requirements for the mass collection of data using an online medium.

It is also the recommendation of this study that drafting of entirely new legislation is begun, solely regarding the use of AI within New Zealand. This will allow for substantial changes to the regulations directly surrounding AI, without the complications generated by changing an already standing privacy act.

**3.2.2 Bias**

Similarly, it is the recommendation of this study that the section pertaining to bias be enforced legislatively. This will result in breaches pertaining to bias and its mitigation to result in legal ramifications.

It is the recommendation of this study that entities developing facial recognition AI for use on the public be required to perform bias identification tests. This will enforce the requirement of the code of ethics that there be no internal biases within facial recognition systems.

**3.2.3 Accountability and Responsibility**

Lastly, it is the recommendation of this study that the section pertaining to accountability and responsibility regarding the management of external interests is enforced legislatively. This can be implemented by requiring entities that wish to develop facial recognition AIs to legally be identified as a fiduciary agent towards the best interests of the public.

Regarding professionalism, the enforcement of this cannot be implemented legislatively as this is inherently not an issue of rights and duty. To account for this it is the recommendation of this study that accreditations from institutes such as IEEE are instated to represent compliance. This accreditation will be enforced by IEEE and will allow users of the AI to differentiate between companies that follow the outlined code of ethics.

**4.0 Evaluation**

The suggested code of ethics and the enforcements that have been laid out in this report look to prevent issues of ethics within the areas of privacy, bias, accountability, and responsibility for the development and implementation of facial recognition AI. This code of ethics is however very limited, as the issues surrounding facial recognition Ai are not limited to those three discussed in this study. The results of this study should be taken as a small component of a much larger study that is yet to be conducted.

# **5.0** **References**

|  |  |
| --- | --- |
| [1] | The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2, IEEE, 2017. |
| [2] | Université de Montréal, "MONTRÉAL DECLARATION FOR A RESPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE," Université de Montréal, Montréal, 2018. |
| [3] | L. Floridi, J. Cowls, M. Beltrametti and et al, "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds & Machines,* vol. 28, p. 689–707, 2018. |
| [4] | J. Borenstein and A. Howard, "Emerging challenges in AI and the need for AI ethics education," *AI Ethics,* vol. 1, pp. 61-65, 2021. |
| [5] | A. Harvey and J. LaPlace, "Exposing.ai," 2021. [Online]. Available: https://exposing.a. |
| [6] | O. Solon, "Facial recognition's 'dirty little secret': Millions of online photos scraped without consent," NBC News, 12 March 2019. [Online]. Available: https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921. |
| [7] | B. J. Goold, " Privacy rights and public spaces: CCTV," *Criminal Justice Ethics,* vol. 21, no. 1, pp. 21-27, 2002. |
| [8] | Ministry of Justice New Zealand, *Privacy Act,* Ministry of Justice, 2020. |
| [9] | Privacy Commissioner New Zealand, "Can I take photos or make recordings of people in public places?," December 2020. [Online]. Available: https://www.privacy.org.nz/tools/knowledge-base/view/322?t=83158\_108710. |
| [10] | Georgetown Law - Centre on Privacy & Technology, "The Perpetual Line-up - UNREGULATED POLICE FACE RECOGNITION IN AMERICA," 18 October 2016. [Online]. Available: https://www.perpetuallineup.org/findings/racial-bias. |
| [11] | P. J. Phillips, F. Jiang, A. Narvekar, J. Ayyad and A. J. 'Toole, "An other-race effect for face recognition algorithms," *ACM Transactions on Applied Perception,* vol. 8, no. 2, 2011. |
| [12] | Google, "Google Diversity Annual Report 2020," Alphabet Inc, 2020. |
| [13] | N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman and A. Galstyan, "A Survey on Bias and Fairness in Machine Learning," Cornell University, 2019. |
| [14] | I. D. Raji, T. Gebru, M. Mitchell, J. Buolamwini, J. Lee and E. Denton, "Saving Face: Investigating the Ethical Concerns of FacialRecognition Auditing," in *AAAI/ACM Conference on AI, Ethics, and Society*, 2020. |
| [15] | B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nat Mach Intell,* vol. 1, pp. 501-507, 2019. |
| [16] | American Civil Liberties Union, "Privacy Multistakeholder Process: Facial Recognition Technology," 15 June 2016. [Online]. Available: https://www.ntia.doc.gov/other-publication/2016/privacy-multistakeholder-process-facial-recognition-technology. |