# ENGR401 Assignment 1
# Case Study: Code of Ethics for AI Facial Recognition

Daniel Eisen

April 28, 2021

## 1 Introduction

Facial Recognition technology is arguably one of the most potentially dangerous applications of artificial intelligence, if its development and deployment is not kept in careful check its flaws both create and reinforce discredited categorisations around gender and race, as well as propagating bias and compromising rights [1], [2].

Facial Recognition as a technology will only improve technically with time. The rise of social media and growing online presence also enables and lowers the bar for the mass collection and analysis of photographs. The use of this technology is a popular and alluring method of biometric identification within the industry, as a person's face is a long-lasting identifier, and this roll out is already seen in wide and growing use. Key tech giants (Google, Apple, Facebook, Amazon, and Microsoft (GAFAM)) have extensively developed internal products that are already being partially deployed to user bases [3] but this tech is by no means being limited to commercial use. While China may be the only state with wide unchecked surveillance and policing deployment, the US, UK, EU and others [2], [3] all have a level of facial recognition ID in research or testing.

Facial Recognition, in these specific instances and as a whole technology, presents a host of ethical problems in its development and use. On the development side, lawful and consensual collection of large, varied datasets which by definition is likely to be personal and potentially sensitive. Closely related is the subject of access and sharing of these datasets and the products of the development (AI output, classifiers and tags) and the potential implication of privacy violation and security breaches. As mass surveillance is a likely and known target for this technology there also arises the issue of unintended bias (implement presence in datasets) and the possibility of an AI to be specially developed to discriminate persons based on bias as how these affect and are exposed to vulnerable populations.

This study will not cover whether Facial Recognition is as a whole unethical [1], but rather on the following specific ethical issues; Data without consent, Access and Privacy, and Bias and vulnerable populations. Specially focusing on use cases on the greater public (opposed to the individual, i.e. FaceID unlock etc) with the intention of discussing and backing these concerns with ethical frameworks, exploring possible solutions through the construction of a Code of Ethics to address the situation, and providing an evaluation of its limitations, gaps and likelihood of success.

## 2 Background

Before proceeding it is important to clarify definitions as this topic (facial recognition) is prone to a blurring of lines and inconsistent use. Facial scanning systems can be placed under 3 broad categorisations [4].

Detection: The most simple case, concerned on finding a face within a captured frame. Key is no Personally identifiable information (PII) is collected, nor is any derived.

Characterisation: Smile/frown detection, emotional indicators, gender/age approximating. While this does not collect PII, the process, particularly in systems to detect emotions and characteristic estimation (guessing) derives/produces a subjective assessment of a subject.

Verification and Identification: This more precisely fits the definition of Facial Recognition. This is concerned with assessing whether the subject matches a stored 'faceprint', with the purpose of answering "Is this image of a known person?", and requires the collection and storage of an explicit PII database.

These all have shared technological basis [4] and each being a superset or including the prior. Therefore the defined focus of Facial Recognition in this study is concerned with the wider use (i.e. on public) of Verification and Identification.

AI Facial Recognition is based fundamentally as the three step process: *detection*, *capture*, and *matching* [4], [3].

For this pipeline of algorithms to be successful, they require extensive training of very large ingested datasets (of human faces) with large variation in lighting and angles. In early research these datasets were consensually sourced from volunteers, but as the technology matured and grew in scope, complexity and entered the commercial market development on these algorithms has increasing relied on amassing these larger and larger datasets without permission. Often they rely taking advantage of overly lenient licences (MegaFace, based on .3 million images pulled directly from Flickr), direct internet scaping (MSCeleb, 10 million) or in the research space controversially forming datasets from university security cameras [2]. This rising common practice raises obvious ethical concerns surrounding consent and knowledge of use but also in the ongoing safety and privacy/access concerns as these datasets are shared/bought and used in future projects.

A very real development/deployment focus of the technology is integration in public surveillance and possible usage as a form of identification and tracking (at both a commercial and state level), [2], [3], [5]. Therefore there is present a very serious risk in the effect of not only unintentionally biased data and AI output, especially when the system is a decision making hierarchy, but also in the development of intentionally bias based classifiers. For example the Uyghur targeting algorithms in use in Xinjiang [5], [2].

# 3 Analysis

In order to ground the following discussion(s) in ethical theory some definitions that this study will reflect and intentions must be outlined; chiefly Virtue based, Deontological, and Utilitarian Ethics.

Virtue Ethics as an approach is that the places emphasis on what is the 'right' action for an entity to take in a situation, i.e what action would be representative of a good moral character [6]. It can be said that virtue based ethics would be less concerned with operating under (especially opposing) rules/laws or focused on an actions larger consequences if it were in contrast to the immediate moral action.

Deontological Ethics, in contrast, places emphasis on adhering to established rules and duties (hence duty ethics) [7]. Namely Deontological ethics holds that at least some acts are morally obligatory regardless of their consequences, i.e. duty is upheld, law is obeyed and rights maintained.

Utilitarian Ethics is fixated on whether an action's consequences result in the best outcome for the most amount of people, regardless if the action is of benefit to the direct participants of that action [8].

## 3.1 Training Data Collection

Facial Recognition by its nature requires the large scale acquisition of personal photographs for training and the continued storage of PII databases (identification, verification), [4], [2], [3].

On the acquisition and training side it is a uncomfortable truth that a not insignificant number of the datasets currently in use are obtained and/or constructed data collected without complete or any informed concert [2]. An example that is particularly representative is a dataset compiled and distributed by Duke University in Durham, North Carolina, which contains more than 2 million video frames of footage of students walking on the university campus [2], [9]. Examining this case as a typical procedure of dataset acquisition and resulted danger of such is insightful as to why this practice with AI development is a concern. The MegaPixel project [9] tracked the usage and citations associated with this dataset and it was not only used for the express purpose of the development of re-identification algorithms, "Tracking Social Groups Within and Across Cameras", and with citations in papers linked to the Chinese Military and surveillance companies active in the Xinjiang region of China.

From the perspective of virtue ethics, as an individual should have the ability to exercise free choice and maintain control of themselves, informed consent as well and full transparency of the use of the data and the option to opt out of the usage of PII is a requirement for that collection to ethical and right. The collected data should then be treated as the 'property' of of the person it identifies and thus a failure in the above is a violation.

A utilitarian approach could be more lenient as an argument could be made that the illicit collection of the data was outweighed by the good it created via for example a Law Enforcement application [4]. This however is narrow and has major assumptions about assumed intent of the enforcing entity, the security of the collected and produced data and equality concerns of the distribution of that system. So what could be allowed in a narrow applications of utilitarian values may in reality result in undue harm.

Deontological concerns arise from the potential violation of the defined rights of the citizens the data is collected from (rights to privacy etc) and not just a developments adherence to laws and agreements.

### 3.2 Bias and Vulnerable Populations

Particularly in the application of facial recognition and characterisation based surveillance systems, the affect of bias within a training dataset can be disproportionate in its impact on vulnerable populations [2]. A theoretical system that is within the decision making hierarchy could very well amplify bias and unguided could result in the unfair persecution (whether intended or not) of individuals.

Since the introduction of 'deep learning' techniques into the field about a decade ago, the accuracy of facial recognition has improved dramatically. However, whether or not this means it can be used on lower-quality, "in the wild" photos is a hotly debated subject. And there are also unanswered concerns about how to test facial-recognition devices in a straightforward manner [5].

Timnit Gebru and other key tech players published a ground breaking paper that found that leading facial-recognition software packages performed much worse at recognising the gender of women and people of colour than male, white faces. [5].

## 4 Recommendations - Code of Ethics

### Guiding Principal

Mass usage of facial recognition should only be undertaken to make the world safer, more secure and more convenient while ensuring the harm is minimised and that steps are taken to identify and mitigate misuse. If violation of this principle inhibits and or limits the capability if a system, then that is the intended consequence of this CoE.

### Regarding Data Collection, Use, and Privacy

- The entity collecting data, and or enrolling an individual in a facial recognition database must receive full informed specific consent from the individual.
- At any point an individual must have the ability to fully opt-out of the process with **PII** being fully transferred and or destroyed as per request.
- The entity must hold security at the highest priority and be held fully accountable for breaches
- Access by 3rd party entities must be in full accordance to the law and be included in accordance with the previous points

### Regarding Bias and Vulnerable Populations

- Based on the history and inherent limitations of the technological facial recognition should never be used in real-time surveillance of lawful activity
- Have its access completely restricted from those under the age of consent
- Never be used as the sole evidence in court proceeding
- Characterisation based systems should never be used in the distinguishing of a population or be the basis of any control or external monitoring of a person.

# 5 Evaluation

This code of ethics is by design narrow in scope and hence has gaps and limitations. It is then only useable in providing the guiding principle for the development and deployment of an AI driven facial recognition system as well as targeted solutions to problems of consensual PII data collection and usage in avoiding bias and disproportionate affects on vulnerable populations. On the topic of enforceability it may be used to guide state and corporate policy but it itself is not law but voluntary policy.

## References


[1] L. Stark, "Facial recognition is the plutonium of ai," *XRDS: Crossroads, The ACM Magazine for Students*, vol. 25, pp. 50–55, 04 2019. [Online]. Available: http://dx.doi.org/10.1145/3313129

[2] R. Noorden, "The ethical questions that haunt facial-recognition research," *Nature*, vol. 587, pp. 354–358, 11 2020. [Online]. Available: http://dx.doi.org/10.1038/d41586-020-03187-3

[3] "Facial recognition: (tech, vendors, markets, use cases and latest news)," Apr 2021. [Online]. Available: https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition

[4] E. Selinger and B. Leong, "The ethics of facial recognition technology," *SSRN Electronic Journal*, 01 2021. [Online]. Available: http://dx.doi.org/10.2139/ssrn.3762185

[5] D. Castelvecchi, "Is facial recognition too biased to be let loose?" *Nature*, vol. 587, pp. 347–349, 11 2020. [Online]. Available: http://dx.doi.org/10.1038/d41586-020-03186-4

[6] R. Hursthouse and G. Pettigrove, "Virtue Ethics," in *The Stanford Encyclopedia of Philosophy*, winter 2018 ed., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2018. [Online]. Available: https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/

[7] L. Alexander and M. Moore, "Deontological Ethics," in *The Stanford Encyclopedia of Philosophy*, winter 2020 ed., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2020. [Online]. Available: https://plato.stanford.edu/archives/win2020/entries/ethics-deontological/

[8] W. Sinnott-Armstrong, "Consequentialism," in *The Stanford Encyclopedia of Philosophy*, summer 2019 ed., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2019. [Online]. Available: https://plato.stanford.edu/archives/sum2019/entries/consequentialism/

[9] "Exposing ai: Duke mtmc dataset," 2018. [Online]. Available: https://exposing.ai/duke_mtmc/