Philosophy & Technology (2021) 34:1639–1663
https://doi.org/10.1007/s13347-021-00478-z

RESEARCH ARTICLE

Biased Face Recognition Technology Used by Government: A Problem for Liberal Democracy

Michael Gentzel[1]

Received: 7 June 2020 / Accepted: 8 September 2021 / Published online: 25 September 2021

Abstract

This paper presents a novel philosophical analysis of the problem of law enforcement's use of biased face recognition technology (FRT) in liberal democracies. FRT programs used by law enforcement in identifying crime suspects are substantially more error-prone on facial images depicting darker skin tones and females as compared to facial images depicting Caucasian males. This bias can lead to citizens being wrongfully investigated by police along racial and gender lines. The author develops and defends "A Liberal Argument Against Biased FRT," which concludes that law enforcement use of biased FRT is inconsistent with the classical liberal requirement that government treat all citizens equally before the law. Two objections to this argument are considered and shown to be unsound. The author concludes by suggesting that equality before the law should be preserved while the problem of machine bias ought to be resolved before FRT and other types of artificial intelligence (AI) are deployed by governments in liberal democracies.

Keywords Face recognition technology · Classical liberalism · Philosophy · Ethics · Ethics of technology · Ethical issues in law enforcement · Political philosophy · Machine ethics · Artificial intelligence and bias · Ethics and discrimination

1 Introduction

Artificial intelligence (AI) and the use of computer algorithms play an increasingly pervasive role in daily life (Binns, 2018; Bjerring & Busch, 2021; Crawford & Calo, 2016; de Laat, 2018; Helbing et al., 2017). The use of AI has become ever more influential in decisions made in fields as diverse as the healthcare field (Esteva et al., 2017; Martinez-Martin, 2019), employment decisions (Chamorro-Premuzic et al., 2017), money lending (Prince et al., 2019), education (Holstein et al., 2018), and the judicial system, including law enforcement (Angwin et al., 2016; Buolamwini

✉ Michael Gentzel
Lalovareigns@aol.com

[1] Pennsylvania, USA

& Gebru, 2018; Chouldechova, 2017; Garvie, 2019; Garvie et al., 2016; Lum & Isaac, 2016; Veale et al., 2018). Face recognition technology (FRT) is a type of AI, the use of which comes with both societal benefits along with moral pitfalls. On the one hand, FRT can help physicians diagnose diseases and monitor patients in the healthcare setting (Martinez-Martin, 2019), find missing and lost persons (Darsham Balar et al., 2019), and help law enforcement apprehend dangerous criminals (Eddine Lahlali et al., 2015; Garvie, 2019). On the other hand, these benefits come with associated moral risks. One such moral risk results from American law enforcement's use of FRT that expresses bias along racial and gender lines (Allyn, 2020; Angileri et al., 2019; Buolamwini & Gebru, 2018; Furl et al., 2002; Fussell, 2020; Garvie et al., 2016; Klare et al., 2012; Rhue, 2018, 2019; Wang et al., 2019).[1] According to Garvie et al.s' (2016) landmark report, the FBI and an estimated one out of every four American law enforcement agencies (state and local) make use of or have access to FRT programs and databases. This large-scale use of FRT by law enforcement makes the moral problem of biased FRT algorithms especially concerning, since there is the potential for many people to be adversely affected.

This paper addresses the moral issues that arise with use of biased FRT by law enforcement as a means for apprehending criminal suspects. More specifically, it presents an in-depth philosophical analysis, from the liberal tradition, of the moral and political consequences and problems of biased FRT used by law enforcement. The author develops and defends "A Liberal Argument Against Biased Face Recognition Technology," which concludes that biased FRT used by law enforcement is incompatible with liberal democracy because it violates the classical liberal value that all individuals deserve equal treatment before the law. While the argument of this paper does not prove that the use of this technology is immoral per se, the argument herein will provide insight into how and why the current use of such technology by law enforcement is incompatible with core principles of western liberal democracy.

To be sure, the ethical use of FRT and related AI has received an array of philosophical, legal, and public-media attention in recent years. de Laat (2018) has recently argued for greater transparency in the oversight and regulation of machine learning algorithms, recognizing that biased outcomes can result from their widespread use. Brey (2004) has provided a broad philosophical overview of the ethical issues and associated policy proposals of the use of FRT in public places. Despite the broad scope of Brey's analysis, the topic of racial bias in FRT was not included. Hale (2005) has argued that the use of FRT by law enforcement threatens self-determination by conflicting with a conception of free will that depends on social interactions. Even though Hale's astute analysis focuses on law enforcement's use

[1] A definitional distinction is in order: For the purposes of this manuscript, to call AI and FRT software "biased" is to say that that software is more likely to produce skewed or inaccurate results with respect to a particular category (in this case, the categories of race and gender). This colloquial use of "bias" is to be distinguished from a more technical use of the term which appears in the "Liberalism, Equality Before the Law, and Unjustified Bias" section of this paper. This technical use of "bias" takes the form of *unjustified bias* and refers to policies and outcomes that favor or disfavor one group over another without a morally relevant reason.

of FRT, it does not address the problem of bias. In a similar vein, Selinger and Hartzog (2020a, 2020b) have argued from a legal perspective that the risks associated with FRT surveillance used by governments and companies make the consent to its use impossible (Selinger & Hartzog, 2019). While Selinger and Hartzog discuss at length the legal aspects of FRT, racial and gender bias are not within the scope of the paper. Additionally, Selinger and Hartzog have published some popular articles in the *New York Daily News*. In their most recent piece, they argued that FRT ought not to be used in the fight against COVID-19 due to privacy concerns, but they do not consider the racial and gender bias associated with FRT technology (Selinger & Hartzog, 2020a, 2020b). In an earlier piece in the same newspaper, the authors argue that while public mask-wearing could present some complications for the use of FRT by governments to track patients infected with COVID-19, the authors note that technology firms are working on ways to improve FRT's ability to "guess" the identity of mask-wearing individuals. While the burdens on people of color are mentioned, how and why people of color will be burdened was not thoroughly explored (Selinger & Hartzog, 2020a, 2020b). In 2020, legal scholars Katelyn Ringrose and Divya Ramjee published an article in the *California Law Review* which explored the use of FRT by law enforcement to identify individuals who attend large protests. The focus of Ringrose and Ramiee centers on not only the privacy concerns of protestors and the machine bias of the FRT, but also the fact that law enforcement agencies had been working with a private company's (Clearview AI) facial image repository which yielded more than 3 billion images (Ringrose & Ramiee, 2020). This analysis raises two problems related to machine bias which call for further philosophical and legal analysis. First, an ethical and legal analysis is needed on the controversy of merging private tech companies with the law enforcement arms of the government. Secondly, an important topic for future ethical and legal analysis is the development of a regulatory framework to constrain the ubiquitous use of cameras and the associated expansion of large repositories/databases of facial images that are exploited to train FRT programs used by law enforcement.

This paper will proceed by first reviewing the current state of technological development of FRT, including its use by law enforcement and the evidence of racial and gender bias. This section will rely heavily on Garvie et al.s' report "The Perpetual Line-up: Unregulated Police Face Recognition in America". Next, the main argument of this paper, "A Liberal Argument Against Biased Face Recognition Technology," will be presented. The proceeding sections are devoted to defending the premises of the argument, including responding to objections to the most contentious premise. The final sections of the paper will define the policy implications of the argument and will mention future directions for philosophical analysis.

First, a definition of FRT is necessary. FRT is a type of AI that incorporates machine learning algorithms which identify patterns of facial features and matches a face to pictures of other faces from a large data base. Following Garvie et al., "Face recognition is the automated process of comparing two images of faces to determine whether they represent the same individual" (Garvie et al., 2016, p. 9). This report provides an excellent summary of how FRT identifies faces: "Before face recognition can identify someone, an algorithm must first find that person's face within the photo. This is called face detection. Once detected, a face is "normalized"—scaled,

rotated, and aligned so that every face that the algorithm processes is in the same position. This makes it easier to compare the faces. Next, the algorithm extracts features from the face—characteristics that can be numerically quantified, like eye position or skin texture. Finally, the algorithm examines pairs of faces and issues a numerical score reflecting the similarity of their features" (Garvie et al., 2016, p. 9). There exist many types of machine learning algorithms that can detect (Zafeiriou et al., 2015) and recognize faces (Klare et al., 2012), and their accuracy and overall performance continue to improve (Kong et al., 2006. Some FRT programs contain algorithms that go beyond merely matching a face to an image. These algorithms attempt to detect subtle facial changes that are associated with specific emotions and truth-telling/lying (Bittle, 2020; Rhue, 2018, 2019).

Many American law enforcement agencies now use or have access to FRT to aid in combating crime (Allyn, 2020; Buolamwini & Gebru, 2018; Fussell, 2020; Garvie et al., 2016; Garvie, 2019; Holstein et al., 2018). According to the Garvie report in 2016, 117 million American adults are affected by law enforcement's use of FRT, which includes the FBI's Next Generation Identification Interstate Photo System (NGIIPS), and the one out of four state and local law enforcement agencies that have access to FRT programs (Garvie et al., 2016). One way that FRT is used by law enforcement is to help identify and arrest a suspect by running an image of a suspect's face through an FRT program, which attempts to match that image to other images of faces in a large database. Facial images of crime suspects can be obtained by police due to the nearly ubiquitous use of public and private cameras. ATM machines, traffic cameras, and private security cameras around homes and businesses all provide opportunities for police to obtain digital images of suspects after or during the commission of a crime. A crime suspect's image, once obtained by police, can be entered into an FRT computer program. That FRT program would then search though a large database of faces, and then render "matches" of varying probabilities of other faces for the police to consider for further investigation. Law enforcement investigators can then make decisions about which suspects should be considered for questioning, detainment, or arrest. While law enforcement's decisions regarding which individual to arrest may not be based solely on the FRT's matches, the FRT's matches could play a crucial role in the chain of causation that determines which citizens will ultimately be investigated and arrested.

Computer scientist Joy Buolamwini and Gebru (2018) at Massachusetts Institute of Technology (MIT) have recently expressed concern about law enforcement use of FRT to identify crime suspects by writing, "…it is very likely that such software is used to identify suspects. Thus, an error in the output of a face recognition algorithm used as input for other tasks can have serious consequences. For example, someone could be wrongfully accused of a crime based on erroneous but confident misidentification of the perpetrator from security video footage analysis". In the important study led by Buolamwini, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," the researchers found that FRT algorithms are biased based on race and gender. The study examined popular commercially available FRT programs by analyzing their data sets, which is the information used to train the FRT algorithm. The study found that the data sets were mostly consist of light-skinned subjects (79.6% and 89.2%). Since the FRT data sets

incorporated predominantly light-skinned subjects, the FRT algorithms which could classify faces by gender were most accurate when the subject being identified was a white male (only a 0.8% error rate), and the algorithms were least accurate when the subject being identified was a dark-skinned female (up to 34.7% error rate). The FRT algorithms performed 11.8–19.2% worse on darker-skinned images compared with their lighter-skinned counterparts (Buolamwini & Gebru, 2018). Buolamwini et al. conclude that "urgent attention" is needed on the part of companies that produce FRT in order to maintain fairness and accountability.

While Buolamwini and her team have demonstrated in detail how and along what intersectional lines AI algorithms are biased, other researchers in the field have been aware that such built-in bias could occur (Angileri et al., 2019; Angwin et al., 2016; Chouldechova, 2017; Garvie et al., 2016; Gong et al., 2019; Klare et al., 2012; Serna et al., 2019). A 2019 study conducted by the National Institute of Standards and Technology (NIST) in the USA shared the same conclusion of Buolamwini et al., finding that FRT software varies in its accuracy depending on the race and gender of the images. The NIST study additionally found biased results when it came to East Asian, Native American, American Indian, Alaskan Indian, and Pacific Islanders faces (NIST, 2019). In the United Kingdom (UK), where FRT is being used in public places, a recent report from the University of Essex led by Professor Pete Fussey corroborates both Garvie's and Buolamwini's concerns about biased FRT being deployed by the London Metropolitan Police (Fussey & Murray, 2019). Interestingly, Furl et al. (2002) found that FRT developed and used in Western countries were more accurate for Caucasian facial images, whereas in East Asian countries, the FRT algorithms were more accurate for East Asian facial images.

Additional research suggests that due to the skewed data sets described by Buolamwini et al., some FRT programs designed to interpret emotions by facial analysis perform differently according to the race of the subject (Rhue, 2018, 2019). According to Rhue, images of black faces are more likely to be interpreted by algorithms to be expressing negative emotions (anger or contempt) at a higher rate than white faces. Since the use of these programs could be used by law enforcement on large crowds to identify individuals more likely to be threats (by associating angry or contemptuous looking faces with potential threats), any bias in these FRT programs could result in higher rates of errors when it comes to non-whites (Rhue, 2018, 2019). This same type of error could occur with FRT programs designed to analyze small changes in a person's face that have been associated with lying. The use of these "lie-detecting" FRT programs have been proposed for courtroom settings (Zhe et al., 2017), questioning suspects of crimes in the UK (Randell, 2019) and for law enforcement use to manage border crossing between countries (Bittle, 2020). Even though some of these newer programs have been tested for racial and cultural bias (Bittle, 2020), the potential for disparate effects along intersectional lines has been established by Buolamwini et al. Additionally, a more basic problem with emotion and lie-detecting algorithms comes from the underlying concept: That subtle changes in facial features can reliability predict emotions and whether someone is telling the truth. Some commentators have pointed out that such algorithms resemble the pseudo-scientific claims made by the now-debunked theories behind phrenology (the claim that skull shape predicts character) and physiognomy (the claim

that facial features predict character) (Chinoy, 2019; Spichak, 2021). Future philosophical analysis is needed to explore the validity and moral status of the use of such technology within the context of its pseudo-scientific predecessors. Despite the fact that there are many types of FRT programs that could be used by law enforcement in diverse contexts, the argument of this manuscript would apply in equal measure, *mutatis mutandis*, to any biased consequences that would be a product of all of these uses of FRT programs, should it turn out that the skewed data sets give rise to biased outcomes.

At this point, two separate claims have been established. First, American law enforcement uses FRT on a large scale. Second, AI, and more specifically, FRT algorithms and data sets demonstrate racial bias. These two claims on their own do not yet establish whether there is evidence to believe that the specific FRT algorithms in use by law enforcement are themselves biased. A review of the current research suggests that law enforcement FRT is indeed biased along racial and gender lines. The Garvie report cites Klare et al.s' 2012 study, which was co-authored by an FBI expert. This study evaluated three different commercially available FRT algorithms, which were in use by the Los Angeles County Sheriff, the Maryland Department of Public Safety, the Michigan State Police, the Pennsylvania Justice Network, and the San Diego Association of Governments (SANDAG), which runs a system used by 28 law enforcement agencies within San Diego County. The FRT algorithms were 5–10% less accurate when it came to African Americans as compared to Caucasians. According to Garvie et al. (2016, p. 54), "… this effect could lead the police to misidentify the suspect and investigate the wrong person. Many systems return the top few matches for a given suspect no matter how bad the matches themselves are. If the suspect is African American rather than Caucasian, the system is more likely to erroneously fail to identify the right person, potentially causing innocent people to be bumped up the list—and possibly even investigated. Even if the suspect is simply knocked a few spots lower on the list, it means that, according to the face recognition system, innocent people will look like better matches". The Garvie report states that in 2016, the authors of the report interviewed engineers from two of the leading FRT vendors, both of whom have contracts with law enforcement agencies. The engineers confirmed that their respective companies did not explicitly test their FRT algorithms for racial bias (Garvie et al., 2016, p. 55).

Finally, in June 2020, *Wired.com*, *NPR*, and other mainstream news outlets reported on the arrest and detainment of Robert Williams, who is the first man in the USA to be mistakenly accused and detained by police due to a racially biased FRT program used in a criminal investigation (Allyn, 2020; Fussell, 2020). According to those news reports, Williams' arrest was prompted by security footage of a theft that occurred in a retail store in Michigan being run through an FRT program by the Michigan State Police crime lab. The image of the theft suspect in the security footage was mistakenly matched by the FRT program to Williams' photo from his driver's license. Williams was detained for 30 h, released on bail, and the case was eventually dropped by the Wayne County prosecutor's office due to insufficient evidence (Allyn, 2020). This recent event provides a real-life look into the unfortunate effects of law enforcement's use of biased FRT programs, and it has prompted some state and local governments in the USA to take legislative actions that include

policy debates, regulations, and moratoriums (DeCosta-Kilp, 2020; Ryan-Mosely, 2020; Stein, 2020). Despite these state and local legislative actions, as of the year 2020, there exists no federal law in the USA that directly regulates the use of biometric technology (including FRT) (Ringrose & Ramjee, 2020).

Given the Garvie report and Buolamwini et al.s' recent study on racial and gender bias in FRT algorithms, along with the Robert Williams case, the following propositions are true:

American law enforcement agencies have used and currently use FRT on a large scale to fight crime.
The specific FRT programs in use by American law enforcement agencies demonstrate racial and gender bias.

2 A Liberal Argument Against Biased Face Recognition Technology

Given the truth of propositions A and B, the author of this paper develops and defends the following argument:

A Liberal Argument Against Biased Face Recognition Technology (FRT)

(1) Classical liberalism requires that all individuals be treated equally before the law.
(2) Government participating in unjustified bias is incompatible with treating all individuals equally before the law.
(3) Government participating in unjustified bias is incompatible with classical liberalism (from 1 and 2).
(4) Law enforcement use of biased face recognition technology is a case of government participating in unjustified bias.
(5) Therefore, law enforcement use of biased face recognition technology is incompatible with treating all individuals equally before the law (from 2, 3, and 4).
(6) Therefore, law enforcement use of biased face recognition technology is incompatible with classical liberalism (from 3, 4, and 5).

This argument, as stated, is logically valid because it is impossible for the conclusion to be false while the premises are assumed to be true.[2] The next task is to demonstrate that the argument is sound by establishing that each premise is true.

[2] Just to be sure no logical fallacy is being committed in my argument, here is the argument, in more formal prose: Let P = "Classical liberalism is true"; let Q = "Equality before the law is true"; Let A = "Government participates in unjustified bias"; let B = "Law enforcement uses biased FRT".
Premise 1P Q.
Premise 2A ~ Q.
Premise 3A ~ P (from premises 1 and 2).
Premise 4B A.
Premise 5B ~ Q (from premises 2, 3, and 4).
Therefore: B ~ P (from premises 3, 4, and 5).

3 Liberalism, Equality Before the Law, and Unjustified Bias

Premise one states that, "Classical liberalism requires that all individuals be treated equally before the law." Classical liberalism is a political philosophy with origins in the seventeenth and eighteenth centuries (Vincent, 2009) that challenged the prior centuries marked by political authoritarian rule by monarchs and the nobility class. Classical liberalism was a new system of political thought, whereby the concepts of the exclusive rights of the king and nobility were replaced by a theory positing the natural rights of all individuals. Consequently, classical liberalism "flips" the prior notion of rights upside-down: away from belonging only to members of an elite class, and instead toward belonging to all individuals who share a set of equal basic rights. Individual liberty becomes the default societal assumption, and any restriction thereof by government requires good reasons in its favor. For example, one classically liberal "good reason" that would justify restriction on individual liberty would be to prevent person A from harming person B (the harm principle, as defended by John Stuart Mill[3]).[4]

With classical liberalism's emphasis on individualism comes the basic idea that each citizen must be treated equally before the law. While modern liberal democracies differ with respect to their adherence to all aspects of classical liberalism, they all share in the essential aim of *equality*. As Tommie Shelby (2004) writes, "It is a central if not defining tenet of liberalism that all persons are to be regarded as free and equal in a just society". Similarly, Friedrich Hayek (1960, p. 85) famously wrote that "The great aim of the struggle for liberty has been equality before the law". While liberalism acknowledges the empirical fact that individuals can differ (in talents, abilities, physical size, and strength), and therefore be factually unequal in many different respects, this does not justify the government treating individuals or classes of people differently from a political, and rights, point of view. Locke (1689), one of the founders of classical liberal thought, famously wrote on equality before the law in *Second Treatise of Government*, declaring that, "…freedom of men under government is to have a standing rule to live by, common to every one of that society, and made by the legislative power erected in it; a liberty to follow my own will in all things, where the rule prescribes not; and not to be subject to the inconstant, uncertain, unknown, arbitrary will of another man". Equality before the law, for the liberal, prevents the government from, on the one hand, granting arbitrary exemptions to the law to favored groups while on the other hand capriciously oppressing some disfavored group of people or individuals in society. Arneson (1999, p. 103) astutely captures this notion of equal treatment before the law thusly: "All humans have an equal basic moral status. They possess the same fundamental rights, and the comparable interests of each person should count the same in calculations that determine social policy. Neither supposed racial differences, nor skin color, sex, sexual

[3] See Mill (1859).

[4] For a thorough analysis of the harm principle in the liberal tradition, see Feinberg's (1984) "Harm to Others: The Moral Limits of the Criminal Law".

orientation, ethnicity, intelligence, nor any other differences among humans negate their fundamental equal worth and dignity".

While classical liberalism, as described in this section, can provide a theoretical bulwark against governments treating individuals unequally before the law, it is not the only means of protecting this liberal tenet. Basic human rights frameworks have arisen in the wake of liberalism's defense of equality before the law and have been codified in national and international human rights treatises and laws. Indeed, Amendment 14 of The United States Constitution was passed after the American Civil War. It guarantees equal treatment under the law for all citizens and broadly prohibits discrimination based on race (Legal Information Institute). On the international scale, Article 1 of The United Nations Universal Declaration of Human Rights states, "All human beings are born free and equal in dignity and rights.". Most notably, Article 7 states, "All are equal before the law and are entitled without any discrimination to equal protection of the law. All are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination." (United Nations, 1948). Similarly, the Charter of Fundamental Rights of the European Union stipulates in Title III, Articles 20 and 21, equality before the law and non-discrimination based on sex, race, and other immutable traits, respectively (European Union, 2012).

More specific to FRT and data collection, liberal democracies and associated policy advisors around the world have weighed in on FRT as it pertains to human rights. The European Union Agency for Fundamental Rights (FRA) has released several recent reports; one that highlights the problem of racial and gender bias in FRT data sets (FRA Focus, 2019) and another that prohibits discrimination and profiling in law enforcement and border management (FRA, 2018). The European Commission also released a Joint Research Center study in 2019 that recognized the problem of racial and gender bias in the FRT used in law enforcement and border management across Europe (Galbally et al., 2019). The French Data Protection Authority (CNIL) published a report in 2019 that provided guidelines and best practices for the use of FRT in public places (CNIL, 2019). In the UK, The Information Commissioner's Office (ICO) released a 2019 report that issued policy recommendations for the use of FRT by law enforcement in the UK, a practice which is, at the time of this paper, already in widespread use (ICO, 2019). In July of 2020, the Council of Europe held an online data protection webinar which featured a session devoted to the ethical use of facial recognition technology. Some of the presentations recognized the problem of biased FRT (Council of Europe, 2020).

From a purely pragmatic perspective, mere appeals to these human rights frameworks and policy proposals would be sufficient to condemn the use of FRT technology by law enforcement that results in unequal treatment of some citizens. However, from a methodological and philosophical perspective, such appeals are not sufficient on their own because they need to be justified by the theoretical foundations from which those human rights frameworks derived. Modern liberal democracy, replete with these crucial national and international treatises and laws that form the frameworks of human rights, can trace its emphasis on equal treatment before the law to the classical liberal tradition and its historical roots described in this section. It was classical liberalism and the subsequent liberal tradition that gave rise to the notion of

moral and social equality, and thereby provided the theoretical common ground that all liberal democracies now share as a lens through which to judge the normative status of particular uses of AI-related technology and their effects on society. Demonstrating that biased FRT used by law enforcement contradicts a specific human rights framework or policy guideline would only be the beginning of an important normative analysis. This paper takes the analysis further by demonstrating how and why biased FRT used by law enforcement is incompatible with an essential value shared by all liberal democracies. It is for this reason that this paper argues from these first principles, and not merely from appeal to the resulting human rights frameworks.

Premise two of the Liberal Argument Against Biased FRT states: "Government participating in unjustified bias is incompatible with treating all individuals equally before the law." This premise introduces the concept of unjustified bias. To be unjustifiably biased is to favor or disfavor one group or individual over another group or individual, when there exists no good reason to do so. In the context of this argument, a "good reason" is to be understood as a morally relevant reason that would justify discrimination. Accordingly, people, intentions, actions, policies, and outcomes can be biased in the manner so defined. Furthermore, discrimination involves treating one individual or group differently than another individual or group. Following Joel Feinberg, when two individuals are the same in all relevant respects, then those two individuals deserve to be treated equally. Discrimination between two individuals is arbitrary, and thereby unjust, when two individuals are the same in all relevant respects and one is treated differently than the other. Discrimination is non-arbitrary, and hence justified, when two individuals are different in some relevant respect, and one is treated differently than the other based on the relevant difference. As such, a good reason in this context is a morally relevant reason that would justify treating Jones differently from Smith (Feinberg, 1973). James Rachels illustrates the point concretely in *The Elements of Moral Philosophy* with an example: An employer who discriminates due to her bias against blind applicants for a job, where visual acuity is not a relevant factor, discriminates arbitrarily, and therefore unjustly, against the blind. Her bias against the blind is thereby unjustified. The same analysis applies to cases in which the employer discriminates against applicants who are black or Jewish where those traits are not morally relevant factors in considering qualifications for a job. Conversely, an employer who discriminates against blind applicants when visual acuity is a relevant factor (perhaps a job in air-traffic control) does not discriminate arbitrarily, and therefore discriminates based on a relevant, or good reason. Such non-arbitrary discrimination is therefore not unjust (Rachels, 2004). Her biased policy in favor of applicants who are not visually impaired is not unjustified. Various forms of invidious discrimination and biases like racism, sexism, ableism, ageism, and the like are unjust because they are committed when there exists no relevant reason to justify them.

Sometimes factual differences between groups are used as reasons to try to justify treating certain individuals differently, when those differences turn out to be based on unjustified bias. The author of this paper has argued elsewhere that, "… there are many poor 'reasons' that some people generate to try to justify pernicious discrimination. Perceived disparities between groups of people—including perceived

differences between genders, age groups, cultures, and ethnic groups—have been marshaled as reasons to treat members of one group differently than members of another group. This is a typical ploy used by racists and sexists. But one must not conflate perceived differences between groups with moral or political differences at the individual level. Even if such differences are considered at a social level, they should not result in differentiation between individuals" (Gentzel, 2020). Just because groups might be thought to be different does not automatically justify treating individuals differently.

It is not always easy to discern whether cases of discriminatory outcomes are the result of unjustified bias or are justified by morally relevant reasons. For example, discrimination occurs in the medical context of deciding which patients get priority consideration for organ transplantation. Young patients, particularly those under the age of 12 years, are much less likely than older patients to receive lung transplants. This might, *prima facie*, appear to be a policy of unjustified ageism, where medical authorities are unjustly discriminating against younger patients by promoting an age-biased policy that favors older patients who will receive cadaver lungs before younger patients. However, a closer look at the reasons that are given by medical specialists to justify such a biased policy reveals a morally relevant reason that justify this policy. Patients younger than 12 years old have much smaller bodies than older patients, and it is a physiological reality that adult-sized cadaver lungs do not fit inside the smaller bodies of younger patients.[5] Patients under 12 are less likely to receive lung transplants before older patients because of a larger dearth in the available donor organs for that demographic. To try to fit lungs that are too large into the bodies of younger patients would harm those patients, so the policy that *prima facie* appeared to be ageist is backed by morally relevant reasons, and therefore is not unjustifiably biased against young patients.

Despite recognizing the goal of equality before the law, liberal societies have at times fallen short of achieving it. Government can violate the equality before the law principle by passing or enforcing laws that could be used to pick out one group of people in society, and then treat that group differently than other people, when there is no good moral reason to do so. A historical example would be the legal (yet immoral) internment of Japanese American citizens during the Second World War. In the immediate aftermath of Japan's attack on Pearl Harbor on December 7, 1941, the US government became suspicious that Japanese American citizens could be enemy spies, despite the lack of any corroborating evidence. President Franklin D. Roosevelt signed Executive Order 9066, which gave the US military the power to exclude anyone from any designated area. On March 18, 1942, The Federal War Relocation Authority was commissioned to arrest and surround all people of Japanese ancestry with troops and prevent them from buying land. Later that month, all Japanese people living along the West Coast of the USA were ordered to report to military stations and register all family members. They were then made to report to

[5] See the case of Sara Murnaghan, who was a 12-year-old girl in need of a lung transplant. Her parents challenged the policy of the United Network of Organ Sharing, which they claimed unfairly discriminated against her daughter (De Sante et al., 2014).

internment camps, where they would be forced to stay for the duration of the war (Britannica, 2020).

The internment of Japanese Americans is just one example of how laws can be enforced to violate the classically liberal value of equality before the law. In this case, the law (Executive Order 9066) did not explicitly mention Japanese Americans, but it was applied and enforced in a biased manner and without good reason, which violated the rights of Japanese Americans. In 1988, more than 40 years after the Japanese American's internment by the US government, President Ronald Reagan signed The Civil Liberties Act of 1988, which formally apologized for the internment and issued reparation payments to survivors of the camps.

Liberal democracy, and its essential value of equality before the law, exemplify the founding principles of the American democratic constitutional republic. As is obvious from the Japanese internment and other historical misfortunes, these principles were not always upheld. These liberal principles are also embraced by many other democracies throughout the world. It is therefore important that the enforcement of laws be applied equally without unjustified bias; that the actions of government agents performed on behalf of the state reflect this requirement of equality before the law; that all new laws that are passed protect equality before the law; and that government's use of new technology is consistent with treating all citizens equally before the law.[6]

To sum up the preceding analysis of bias, not all instances of bias and its resulting discrimination are unjustified. Unjustified bias, and its resulting arbitrary discrimination whereby individuals and groups are treated differently, is characterized by the absence of a morally relevant reason in its favor. Premises two and three deal specifically with cases in which such bias is exercised by the government. As was argued previously in this section, equality before the law entails treating people and groups equally unless there is a morally relevant reason against doing so. Equality before the law is an essential value of classical liberalism and modern liberal democracy. Therefore, when government treats groups or individuals differently when no morally relevant reason supports that disparity, that government fails to treat everyone equally before the law, and that government's actions are incompatible with classical liberalism. Premises one, two, and three of the Liberal Argument Against Biased FRT are therefore true.

[6] Strictly speaking, classical liberalism requires that all citizens be treated equally by government at all times (all laws and enforcement thereof, policies, and actions). There is considerable debate about whether government programs and laws that apply only to specific groups within the population are consistent with equality before the law. Programs like Medicare, Medicaid, Supplemental Nutrition Assistance, and laws that prohibit convicted felons from voting all treat one group of citizens differently from another group. Moreover, the actions of government agents (i.e., judges determining sentences of the convicted) can vary depending on group membership of the individual in question (those with prior criminal histories are treated differently than those without a criminal record). Much of the debate related to these issues will revolve around whether such bias is backed by morally relevant reasons, and are therefore, justified. These details, while important, are beyond the scope of this manuscript. Nonetheless, classical liberalism's principle of equality before the law establishes the narrower focus of this manuscript: that racial and gender bias is unjustified when law enforcement uses biased FRT.

4 Premise 4: Law Enforcement Use of Biased Face Recognition Technology is a Case of Government Participating in Unjustified Bias

The main premise that requires analysis and defense is premise four. One assumption contained in this premise is that law enforcement agencies are government entities. While this is the least controversial claim contained in premise four, its truth is worth establishing. In the USA, law enforcement agencies are publicly funded by government (tax revenue) and are therefore government entities. The high cost of such public spending has recently garnered some public criticism in the US media. Indeed, according to a recent piece from *The Washington Post*, "The United States spends more than twice as much on law and order as it does on cash welfare programs," with up to 40% of major cities' budgets going to law enforcement funding (Ingraham, 2020). Law enforcement agencies in Europe are also publicly funded by government and are therefore government entities (Eurostat, 2020).

It was demonstrated in the Sect. 1 that the FRT programs used by American law enforcement agencies are significantly biased against images that are not of Caucasian males (Buolamwini & Gebru, 2018; Garvie et al., 2016). It was also established that bias is unjustified in the absence of a morally relevant reason. Unless a morally relevant reason can be marshaled in favor of treating people differently based on race and gender, such racist and sexist examples of bias are unjustified and are incompatible with treating people equally. These conditions, applied to the classical liberal value of equality before the law, dictate that everyone deserves equal treatment before the law unless a morally relevant reason justifies unequal treatment. FRT that is biased along racial and gender lines used by law enforcement is a case of unjustified bias committed by law enforcement because the biased outcomes are not justified (i.e., not backed by morally relevant reasons), since the use of FRT that happens to contain bias does not morally justify disparate outcomes along racial and gender lines in the context of law enforcement. There does not exist a morally relevant reason to use biased FRT in law enforcement, so the biased outcomes are cases of unjust bias committed by government.

The remainder of this section will consider two counterarguments to premise four. The first counterargument challenges premise four with respect to FRT's bias being unjustified, and the second counterargument challenges premise four's idea that government is the one responsible for committing the bias. Replies in defense of premise four will be offered, respectively.

The first counterargument to premise four would be the following: While it is true that the FRT used by law enforcement is biased against non-Caucasians, the bias (and resulting disparate treatments of non-white suspects compared with white suspects) is not unjustified because law enforcement agencies are not intentionally committing bias against non-Caucasians. The bias originates from unknown faults in the FRT software program, and not from an unfair policy intentionally adopted by law enforcement agencies. Since the bias is not due to

intentional actions by police that target one specific group for disparate treatment, as was the case in the internment of Japanese Americans during World War II, the bias under question is not unjustified in the morally relevant sense. The morally relevant reason that could justify the disparate outcomes would be the variable margins of error across demographics inherent in the FRT programs, and the intentions of the police agencies and officers cannot be blamed for blunders caused by machines or other such products. Therefore, law enforcement's use of biased FRT is not a genuine case of government participating in *unjustified* bias, and premise four is false. Therefore, the author's Liberal Argument Against Biased FRT is unsound. In its formal iteration, this counterargument takes the following form. Call it the "Counterargument from Intentions".

(1) Government participating in bias is unjustified only when the agents participating in bias intend to favor one party over another without a morally relevant reason.
(2) Law enforcement use of biased FRT is not a case in which the agents participating in bias intend to favor one party over another without a morally relevant reason (because the bias is caused by a program error).
(3) Therefore, it is not the case that law enforcement use of biased FRT is a case of government participating in unjustified bias.

This Counterargument from Intentions against premise four is logically valid. However, one of the premises is false, so it is not sound. The false premise is premise one. It is not the case that bias committed by government is unjustified only when agents intend to design a policy or take actions that result in biased outcomes. Granted, many historical cases in which the American government treated citizens unequally before the law were the result of the government passing and enforcing laws that had as their goals the premeditated and deliberate biased treatment of a targeted group of individuals. But this historical observation is only contingently true. Indeed, there exist cases in which government policies resulted in unjustifiably biased outcomes without their corresponding intentions, and only after extensive research was it discovered that the explanation for such outcomes were not the result of deliberate bias. Nonetheless, such cases of bias are still considered to be unjustified, because morally relevant reasons cannot be marshaled on their behalf, and so they require a remedy to restore equality before the law.

An example will illustrate this point. The cross-race effect is the psychological phenomenon that people are better at recognizing and distinguishing between faces that belong to their own race than faces that belong to a different race. This is one of the most highly replicated phenomena in social and cognitive psychology (Hourihan et al., 2012; Meissner & Brigham, 2001; Smith et al., 2004). An important consequence of the cross-race effect is that the accuracy of eyewitness testimony in court trials is weakened when the race of the crime suspect is different from the eyewitness. When an eyewitness to a crime is a different race from the suspect, the eyewitness will be more likely to make an error in recognizing

the suspect's face in a line-up, compared to a case under which the suspect and the witness shared the same race. Hourihan et al. (2012) summarize the problem thusly: "This finding is particularly important for legal and psychological scholars who study eyewitness memory, as it indicates that we are more likely to falsely identify an innocent suspect if he or she is from a different race".

The police policy of placing a high degree of confidence in the use of eyewitness testimony to identify suspects can result in individuals being wrongfully convicted and punished for crimes they did not commit. An innocent person being falsely accused and convicted for a crime because he or she was of a different race from the eyewitness is a case of unjustified bias, despite the psychological explanation (the cross-race effect) having nothing to do with malicious intentions on the part of the eyewitness or police. Moreover, racially biased outcomes could result if eyewitnesses tend to be a different race from those who are more likely to experience police encounters. Furthermore, false convictions due to the cross-race effect would be government failing to treat all individuals equally before the law, especially if law enforcement agencies continue to place high confidence in eyewitness identification of suspects when it is known that the cross-race effect significantly contaminates that body of evidence. The classical liberal value of treating all citizens equally before the law applies not only to the policies and their enforcement, but most importantly, to how people are actually treated by government, and whether outcomes wherein people are treated differently can be justified by morally relevant reasons. If one group is favored by a policy over another, whether a law or policy is intended to do so, that policy is unjustifiably biased unless a morally relevant reason in its favor can be produced. Merely citing the cross-race effect for why the disparities occur is not a morally relevant reason to support the police using cross-racial eyewitness testimony in the same way that FRT that has biased algorithms due to a non-representative data set is not a morally relevant reason for police to use that biased technology. Just because biased outcomes of government actions were not intended does not make the resulting bias justified. Since bias due to the cross-race effect in eyewitness testimony is a case of unjustified government bias even though intentions did not play a role, premise one of the Counterargument from Intentions is false, making the counterargument unsound. Therefore, premise four of the Argument Against Biased FRT remains true.[7]

There is a second objection against premise four: One can grant that the FRT used by law enforcement is unjustifiably biased against non-Caucasians, but this does not logically entail that the unjustified bias is being committed by government, which

[7] This response to the Counterargument from Intentions would apply, *mutatis mutandis*, to various situations in which deploying FRT could result in biased outcomes. Some examples include the following: (a) The FRT training data sets, whether created by the police department or a third-party, might be biased. For example, a third-party could produce a biased FRT due to a dearth of sufficient diversity in the data set. Or a police department could produce a biased FRT due to employing biased policing methods to collect the data sets. (b) The FRT itself could be implemented by police departments in a biased fashion. For example, a police department might target specific areas of a jurisdiction for deploying FRT while leaving others unaffected by the technology (and presumably its potential errors). In all these scenarios, intensions are unnecessary for bias to be unjustified.

is what premise four claims. Police officers, who are acting on the government's behalf, are not the agents committing the unjustified bias. The unjustified bias is being perpetrated by the private companies that develop, manufacture, and distribute the biased FRT programs, the use of which brings about biased outcomes when law enforcement agencies use these commercially available products in the field. The objection would continue by stating that classical liberalism requires that the government treat everyone equally before the law, but it does not necessarily require that private companies, including the individuals acting as private citizens developing FRT software programs, treat everyone equally before the law. As a result of the bias originating from the private companies and not from the government agents themselves, law enforcement use of biased FRT is not a case of *government* committing unjustified biased. Therefore, premise four of the Liberal Argument Against Biased FRT is false. Call this the "Private Sector Counterargument".

There are two replies to the Private Sector Counterargument to premise four. The immediate reply is that this counterargument raises a different and equally important question: Is classical liberalism, as defined here, consistent with a government that permits private citizens to make choices that could lead to unjustified bias and discrimination? Equality before the law, as defended in this manuscript, requires that the government refrain from engaging in unjustified bias and discrimination, but it does not necessarily require that private individuals refrain from discriminating. For example, most reasonable people have no moral scruples with the government allowing private individuals to arbitrarily discriminate when choosing a romantic partner or with whom to be friends. While certain bases for discrimination in the private and personal sector might seem uncouth or even condemnable, the freedom to choose one's spouse, friends, and associates, based on considerations of race, age, political affiliation, religion, sexual orientation, and gender, seems to be consistent with the freedom essential to classical liberalism embodied in the value of individualism. Freedom of association among private citizens seems to be a classical liberal value on a par with equality before the law. While the bases upon which some people make choices regarding with whom to associate in the context of private companies and individuals might smack of the abhorrent, such freedom to make these choices has an independent value.

At the same time, there is a contrary position that argues that government should prevent private citizens and companies from practicing unjustified bias and its associated discrimination. This might be called the strong version of equality before the law. For example, there exist laws in the USA that prohibit a private employer from discriminating against people based on certain character traits, including race, gender, and religion. Anti-discrimination laws like these extend the classical liberal value of equality before the law to include equality before private business transactions. According to this position, it is not morally sufficient for the government to remain silent with respect to private individuals committing discrimination, and the law should prevent not only government discrimination, but also private discrimination. If one adopts the strong version of equality before the law, then one can use it as a means to respond to the larger issue raised in the Private Sector Counterargument, which is the problem of a private company selling a commercial product that is biased on racial and gender grounds. The strong version of equality before the law

could make it incompatible with classical liberalism to sell such a biased product, even when the companies are doing so as private companies. While this is an interesting path to pursue, its implications are beyond the scope of this paper. Ultimately, defending the strong version of equality before the law will not be the position of this paper.

There is a second and more direct reply that undoes the Private Sector Counterargument. The core problem with the Private Sector Counterargument is that it would permit many unacceptable cases of unjustified bias resulting from government policy or procedure under classical liberalism. This would turn equality before the law into a nearly meaningless standard, capable of preventing only the most conspicuous forms of government-initiated bias. If the Private Sector Counterargument is taken to its logical limits, then *any* case in which government treats citizens unequally through the use of a private sector product or contractor can be deemed to be consistent with treating everyone equally before the law, as long as the private sector can bear the blame for the bias. But this is clearly problematic because it violates the very spirit of and aims of liberalism; to ensure that citizens are treated equally by their government. Imagine the following hypothetical example that illustrates the problem: Let us assume that law enforcement uses K-9 dogs to assist police officers in fighting crime. Some of the tasks that K-9 dogs perform involve detecting the presence of illicit drugs in a motorist's vehicle by smell. Such dogs are trained to give an indication to the K-9 officer handler when trace amounts of illicit drugs are detected by the dog (Jezierski et al., 2014). If the dog signals to its handler that it smells drugs (by either sitting or digging), that signal can give police probable cause to search the vehicle for contraband (Hinkel et al., 2011). Suppose that all the K-9 dogs used for police operations in American law enforcement were bred, trained, and commercially available from one specific private company: Apex Company. Suppose further that Apex Company, unbeknownst to the law enforcement purchasers of this advertised "top-of-the-line police dogs," made a crucial error while training their dogs to detect illicit drugs by associating only persons of color (blacks and Hispanics) with the scent of narcotics. Law enforcement agencies all around the nation go on to purchase and use K-9 dogs from Apex Company. As a result, black and Hispanic motorists' vehicles are searched at a much higher rate than white motorists' vehicles because the dogs have been improperly trained by Apex Company. White motorists who are trafficking drugs evade detection while black and Hispanic motorists, many of whom are innocent, become detained and subjected to searches at significantly higher rates.

In this hypothetical thought experiment, it seems clear that the actions of law enforcement, and by extension government, have been unjustifiably biased against black and Hispanic motorists in their use of Apex Company's improperly trained K-9 dogs. To be sure, Apex Company cannot physically commit biased actions against black and Hispanic Americans without law enforcement agencies (or other purchasers of the dogs) using the dogs that have been trained in an unintentionally biased manner. Improperly trained dogs must be deployed for the biased outcomes to occur, and law enforcement (the government) are the facilitators in this racial bias and subsequent unjust discrimination. The case of biased FRT mirrors this hypothetical case of biased K-9's. Both cases involve private companies that create a product

that when used in the field, produces unjustified bias and discrimination. When government is the purchaser and user of these products, and the American citizens are the ones who are treated in an unjustifiably biased manner by government's use of these products, then the government (and not merely a private company) has violated the people's right to be treated equally before the law. Both cases represent unjustified bias committed by the government. Therefore, the Private Sector Counterargument fails, and premise four of the Liberal Argument Against Biased FRT remains true.

Now that premises one through four of the Liberal Argument Against Biased FRT have been established, premise five of the Liberal Argument follows logically from premises two, three, and four. Therefore, the Liberal Argument Against Biased FRT is both valid and sound, and the conclusion is true*: Law enforcement use of biased face recognition technology is incompatible with classical liberalism.*

5 The Argument's Policy Implications and Future Philosophical Analysis

The Liberal Argument Against Biased FRT presented and defended in this paper does not, on its own, prove that biased FRT used by law enforcement is immoral. Instead, this argument demonstrates that law enforcement's use of biased FRT is incompatible with the set of values that are essential to western liberal democracy, namely, the classical liberal value of equality before the law. In addressing the problem presented in this paper, there are two distinct paths forward that can be taken:

(1) Law enforcement agencies stop using FRT programs that are biased (either by using programs that are not biased or by ending the use of FRT altogether).
(2) Liberal democracy, along with its central value of equality before the law, is ultimately rejected.[8]

This section will briefly outline the consequences of both pathways. This section will conclude by mentioning future directions for philosophical analysis of AI.

5.1 Path 1

The most favorable path forward would be for the companies that manufacture and distribute the FRT programs that are used by law enforcement to eliminate the bias

[8] A third alternative could reject *both* biased FRT and classical liberalism/liberal democracy. In this situation, there would be many possible iterations regarding the details of such a society, and to explore these possibilities would be beyond the scope of this paper. Moreover, whether it is conceptually possible to separate equality before the law (and the foundations of human rights of liberal democracy) from classical liberalism is also beyond the scope of this paper. As such, the author considers only those possible paths forward that are consistent with either rejecting liberal democracy or rejecting biased FRT (Thanks to an anonymous reviewer for raising this important point).

in the algorithms and data sets. There is evidence that some companies that produce FRT are both aware of the bias and are currently seeking solutions to improve FRT's performance across all racial categories and genders. In 2018, Microsoft announced that their commercially available FRT had undergone updates to improve performance across race and gender categories by expanding the data sets used to train the machine learning algorithms and by improving the face classifier for greater accuracy. These improvements resulted in, according to the company, 20 times lower error rates for darker skin tones and 9 times lower error rates for women (Roach, 2018).

Buolamwini, the MIT researcher whose recent work identified and measured the ways in which commercially available FRT is biased along race and gender categories, has created the Algorithmic Justice League, whose mission is to promote awareness of and solutions for bias and harm caused by AI. Computer science researchers, now increasingly aware of the demographic bias in FRT, are beginning to develop programs designed to specifically undo the bias. For example, Gong et al. (2019) have created a program called "DebFace," which "learns" to control for race, gender, and age to better distinguish and identify facial feature across these demographics, thus increasing accuracy of an FRT program.

Improvements along these lines would diminish, and perhaps eventually eliminate the bias in the FRT programs used by law enforcement. Until this occurs, not only would FRT programs need to be adjusted to control for bias, law enforcement agencies should be aware of the currently biased algorithms and data sets in AI programs and be willing to require testing that would screen for such biases before mistakes in the field lead to violations of the principle of treating everyone equally before the law. Once bias is confirmed in FRT programs currently in use by law enforcement, policymakers should consider suspending the use of such programs until the bias is eliminated. This, according to the argument presented in this manuscript, is what the values of a classically liberal polity require.

5.2 Path 2

Given the conclusion of the Liberal Argument Against Biased FRT, the alternative path forward would be to accept the biased FRT and reject liberalism, and in particular, the principle that requires that government treat all citizens equally before the law. While this might smack of the absurd to modern western sensibilities, it should be recognized that not all societies, at present nor in the past, accept liberalism's commitment to equality before the law. For example, various societies specific to the Asian continent had political systems which strictly adhered to rigid class and caste systems in which heredity determines one's social class for life. The caste system of India, which persists in modern times, is one such example. Another example of a political theory which could reject the principle of equality before the law would be the meritocracy proposed by Plato in his *Republic*, whereby the most qualified philosopher kings would rule society for the well-being of everyone. Nevertheless, even though some societies and political theories reject the liberal ideal of equality before the law, this comes with serious drawbacks. Without the basic

requirement for government to treat everyone equally, the widespread oppression of minority groups and the underrepresented can become commonplace. Any society that values liberty and equality would do well to retain the classical liberal value of equality before the law as a protection against the oppression of the minority by the majority; the weak by the strong; the disabled by the abled; the disfavored by the favored; and the underprivileged by the privileged. It is therefore recommended that the value of equality before the law be preserved while the bias is removed from the FRT used by government.

There are several related pathways for future philosophical research in this area. FRT is not the only type of AI that expresses racial and gender bias. For example, Cave and Dihal (2020) have argued that AI, both genuine and in works of fiction, has been depicted as being predominantly white (Caucasian), which suggests that a more widespread issue of bias might pervade western society. Along this research pathway, bias against non-whites through AI and FRT could represent a more generalized problem of societal bias against non-whites in western democracies.[9] The previously cited NIST study from 2019 found that some algorithms developed in East Asian countries did not display biases between East Asian and Caucasian facial images, whereas some western FRT did contain such discrepancies (NIST, 2019). While the NIST study did not directly investigate this difference, more research is needed to shed light on how a particular society's underlying biases could be reflected in their uses of AI. Future work along these lines could investigate the general tendency for technological biases to favor and reflect those groups who happen to be in positions of power within a societal structure. Such technologically based biases might be situated within a broader context, whereby many other societal structures often (technology being one among many) contain biases in favor of those in power and against those who are underrepresented.

Even if it is found that bias in technology does indeed represent a particular society's broader underlying (and perhaps systemic) bias, this finding would not undermine the conclusion of this paper. If the classical liberal value of equality before the law is to be upheld, then such biases, wherever they are found to be committed by government actions, ought to be identified and removed.

More specific to AI used in the law enforcement context, algorithms used by the American judicial system to assess the recidivism risk of individual convicts have been found to express unjustified (as defined in this manuscript) bias on racial grounds, leading to unjustified racial disparities in sentencing and police surveillance (Angwin et al., 2016). Another related example would be predictive policing algorithms, which utilize large sets of data based on geographic location of past crimes and arrest information, to provide police departments with a "heat list" of individuals who would be "forecast" to commit future crimes. This has led to racially biased police investigations on both the community and the individual level (Lum & Isaac, 2016). These cases, and others like them, suggest a need for additional philosophical analysis of these types of AI and the bias involved.

[9] The author gives thanks to an anonymous reviewer for raising this important consideration.

6 Conclusion

This paper presented a detailed philosophical analysis of the problem of law enforcement use of biased FRT within liberal democracies. After establishing the existence of bias in the FRT programs used by law enforcement, the author presented and defended "A Liberal Argument Against Biased FRT". This argument concluded that law enforcement's use of biased FRT is incompatible with the classical liberal value that requires that all citizens deserve equal treatment before the law. Two counterarguments were considered, and both were shown to be unsound. In light of the Liberal Argument Against Biased FRT, two possible paths forward were examined: Eliminate the bias in the FRT programs used by law enforcement or reject the classical liberal principle that citizens deserve equality before the law. The author has argued that any society that values liberty and equality will have enough reason to maintain the classical liberal value of equality before the law as a guard against the oppression of minorities and the disadvantaged, so the bias must be eliminated from FRT programs. This analysis provides a valuable example of how the development and deployment of emerging technology by government must be tempered by a continual commitment to and awareness of the values essential to western liberal democracy.

References

Allyn, B. (June 24, 2020). 'The computer got it wrong': How facial recognition led to false arrest of black man. *NPR*. Retrieved January 8, 2021, from https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig

Angileri, J., Brown, M., Dipalma, J., Ma, Z., & Dancy, C. L. (2019). Ethical considerations of facial classification: Reducing racial bias in AI. Retrieved February 21, 2020, from https://doi.org/10.13140/RG.2.2.28601.11368

Angwin, J., Larson, J., Mattu, S., & Kirschner, L. (May 23, 2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. Retrieved February 21, 2020, from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Arneson, R. (1999). What, if anything, renders all humans morally equal? In D. Jamieson (Ed.), *Peter singer and his critics* (pp. 103–128). Blackwell.

Balar, B. D., Kavya, D. S., Chandana, M., Anush, E., & Hulipalled, V. R. (2019). Efficient face recognition system for identifying lost people. *International Journal of Engineering and Advanced Technology (IJEAT), 8*(5S). ISSN: 2249 – 8958.

Binns, R. (2018). Algorithmic accountability and public reason. *Philosophy & Technology, 31*, 543–556. https://doi.org/10.1007/s13347-017-0263-5

Bittle, J. (March 13, 2020). Lie detectors have always been suspect. AI has made the problem worse. *MIT Technology Review*. Retrieved January 6, 2021, from https://www.technologyreview.com/2020/03/13/905323/ai-lie-detectors-polygraph-silent-talker-iborderctrl-converus-neuroid/

Bjerring, J. C., & Busch, J. (2021). Artificial intelligence and patient-centered decision-making. *Philosophy & Technology., 34*, 349–371. https://doi.org/10.1007/s13347-019-00391-6

Brey, P. A. E. (2004). Ethical aspects of face recognition systems in public places. *Journal of Information, Communication and Ethics in Society, 2*(2), 97–109. https://doi.org/10.1108/14779960480000246

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research, 81*, 1–15.

Cave, S., & Dihal, K. (2020). The whiteness of AI. *Philosophy & Technology., 33*, 685–703. https://doi.org/10.1007/s13347-020-00415-6

Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & Sherman, R. A. (December 2017). The datafication of talent: How technology is advancing the science of human potential at work. *Current Opinion in Behavioral Sciences, 18*, 13–16. https://doi.org/10.1016/j.cobeha.2017.04.007

Chinoy, S. (July 10, 2019). The racist history behind facial recognition. *The New York Times, Opinion*. Retrieved June 21, 2021, from https://www.nytimes.com/2019/07/10/opinion/facial-recognition-race.html

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data, 5*(2), 153–163.

CNIL. (December 19, 2019). Facial recognition: For a debate living up to the challenges. Retrieved January 7, 2021, from https://www.cnil.fr/sites/default/files/atoms/files/facial-recognition.pdf

Council of Europe. Data protection views from Strasbourg in Visio (1–3 July 2020). Session 5: Facial Recognition. Retrieved January 4, 2021, from https://www.coe.int/en/web/data-protection/facial-recognition

Crawford, K., Calo, R. (October 13, 2016). There is a blind spot in AI research. *Nature: International Weekly Journal of Science*. Retrieved February 22, 2020, from Commentary.

de Laat, P. B. (2018). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & Technology, 31*, 525–541. https://doi.org/10.1007/s13347-017-0293-z

De Sante, J., Caplan, A., Hippen, B., Testa, G., & Lantos, J. D. (2014). Was Sarah Murnaghan treated justly? *Pediatrics, 134*(1), 155–162. https://doi.org/10.1542/peds.2013-4189

DeCosta-Kilp, N. (December 21, 2020). How the Massachusetts police reform bill would actually affect law enforcement use of facial recognition technology. *Boston.com*. Retrieved January 8, 2021, from https://www.boston.com/news/politics/2020/12/21/massachusetts-police-reform-bill-facial-recognition-technology

Eddine, L. S., Sadiq, A., & Mbarki, S. (2015) A review of face sketch recognition systems. *Journal of Theoretical and Applied Information Technology, 81*(2), 255–265.

Encyclopedia Britannica. Japanese American internment. UNITED STATES HISTORY January 2020. Retrieved February 14, 2020, from https://www.britannica.com/event/Japanese-American-internment/Life-in-the-camps

Esteva, A., Kuprel, B., Novoa, R., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*, 115–118. https://doi.org/10.1038/nature21056

European Union. (2012). Charter of Fundamental Rights of the European Union. Available at: https://www.refworld.org/docid/3ae6b3b70.html. Accessed 4 Jan 2021.

European Union Agency of Fundamental Human Rights (FRA). (2018). Preventing unlawful profiling today and in the future: A guide. Retrieved January 4, 2021, from https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-preventing-unlawful-profiling-guide_en.pdf

European Union Agency of Fundamental Human Rights (FRA). (2019). FRA focus: Facial recognition technology: Fundamental rights considerations in the context of law enforcement. 2019. Retrieved January 4, 2021, from https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf

Eurostat. (March 25, 2020). Government expenditure on public order and safety. Eurostat: Statistics Explained. Retrieved January 13, 2021, from https://ec.europa.eu/eurostat/statistics-explained/index.php/Government_expenditure_on_public_order_and_safety

Feinberg, J. (1973). *Social philosophy*. Prentice Hall.

Feinberg, J. (1984). *Harm to others: The moral limits of the criminal law*. Oxford University Press.

Furl, N., Phillips, P. J., & O'Toole, A. J. (2002). Face recognition algorithms and the other-race effect: Computational mechanisms for a developmental contact hypothesis. *Cognitive Science, 26*, 797–815. https://doi.org/10.1016/S0364-0213(02)00084-8

Fussell, S. (June 24, 2020). A flawed facial-recognition system sent this man to jail. *Wired.com*. Retrieved January 8, 2021, from https://www.wired.com/story/flawed-facial-recognition-system-sent-man-jail/

Fussey, P, & Murray, D. (2019). Independent report on the London Metropolitan. The Human Rights, Big Data, and Technology Project. The University of Essex. Retrieved January 9, 2021, from https://48ba3m4eh2bf2sksp43rq8kk-wpengine.netdna-ssl.com/wp-content/uploads/2019/07/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report.pdf

Galbally, J., Ferrara, P., Haraksim, R., Psyllos, A., & Beslay, L. (2019). Study on face identification technology for its implementation in the Schengen Information System. JRC Science for

Policy Report. European Commission. Retrieved January 6, 2021, from https://publications.jrc.ec.europa.eu/repository/bitstream/JRC116530/sis_face-jrc_science_for_policy_report_22.07.2019_final.pdf

Garvie, C. (May 16, 2019). Garbage in, garbage out: Face recognition on flawed data. *Georgetown Center on Privacy and Technology*. Retrieved February 10, 2020, from https://www.flawedfacedata.com/#footnote49_8ujkx6a

Garvie, C., Bedoya, A., & Frankle, J. (2016). The perpetual line-up: Unregulated police face recognition in America. *Georgetown Law, Center on Privacy & Technology*. https://www.perpetuallineup.org/sites/default/files/2016-12/The%20Perpetual%20Line-Up%20-%20Center%20on%20Privacy%20and%20Technology%20at%20Georgetown%20Law%20-%20121616.pdf. Accessed 12 Feb 2020.

Gentzel, M. (2020). Classical liberalism, discrimination, and the problem of autonomous cars. *Science and Engineering Ethics, 26*, 931–946. https://doi.org/10.1007/s11948-019-00155-7

Gong, S., Liu, X., & Jain, A.K. (2019). DebFace: De-biasing face recognition. ArXiv. https://arxiv.org/abs/1911.08080

Hale, B. (2005). Identity crisis: Face recognition technology and freedom of the will. *Ethics, Place & Environment, 8*, 141–158.

Hayek, F. A. (1960). *The constitution of liberty*. University of Chicago Press.

Hinkel, D., & Mahr, J. (6 January 2011). Tribune analysis: Drug-sniffing dogs in traffic stops often wrong. *Chicago Tribune*. Retrieved February 29, 2020, from https://www.chicagotribune.com/news/ct-xpm-2011-01-06-ct-met-canine-officers-20110105-story.html

Holstein, K., McLaren, B. M., & Vincent, A. (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In *Proceedings of the International Conference on Artificial Intelligence in Education* (AIED 2018) (pp. 154–168). Springer.

Hourihan, K. L., Benjamin, A. S., & Liu, X. (2012). A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race. *Journal of Applied Research in Memory and Cognition, 1*(3), 158–162. https://doi.org/10.1016/j.jarmac.2012.06.004

Information Commissioner's Office (ICO). (October 21, 2019). ICO investigation into how the police use facial recognition technology in public places. Retrieved January 6, 2021, from https://ico.org.uk/media/about-the-ico/documents/2616185/live-frt-law-enforcement-report-20191031.pdf

Ingraham, C. (June 4, 2020). U.S. spends twice as much on law and order as it does on cash welfare, data show. *The Washington Post*. Retrieved January 13, 2020, from https://www.washingtonpost.com/business/2020/06/04/us-spends-twice-much-law-order-it-does-social-welfare-data-show/

Jezierski, T., Adamkiewicz, E., Walczak, M., Sobczyńska, M., Gorecka-Bruzda, A., Ensminger, J., & Papet, L. E. (2014). Efficacy of drug detection by fully-trained police dogs varies by breed, training level, type of drug and search environment. *Forensic Science International*. https://doi.org/10.1016/j.forsciint.2014.01.013

Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V., & Jain, A. K. (2012). Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security, 7*(6), 1789–1801.

Kong, S. G., Heo, J., Abidi, B. R., Paik, J., & Abidi, M. A. (2006). Recent advances in visual and infrared face recognition—A review. *Computer Vision and Image Understanding*. https://doi.org/10.1016/j.cviu.2004.04.001

Legal Information Institute, Cornell Law School. Retrieved January 4, 2021, from https://www.law.cornell.edu/constitution/amendmentxiv

Locke, J. (1689). *Second treatise of government* 13 (C.B. Macpherson ed., Hackett Publ'g Co. 1980).

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance, 13*(5), 14–19.

Martinez-Martin, N. (2019). What are important ethical implications of using facial recognition technology in health care? *AMA Journal of Ethics, 21*(2), E180-187. https://doi.org/10.1001/amajethics.2019.180

Meissner, C., & Brigham, J. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*, 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Mill, J. S. (1859). *On liberty*. Penguin Books.

National Institute of Standards and Technology (NIST). (December 19, 2019). Retrieved January 9, 2021, from https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software

Plato, (c. 375 BC). *Republic* (3rd ed.). Translated by Allan Bloom. Basic Books; (November 22, 2016).

Prince, A., & Schwarcz, D. B. (August 5, 2019). Proxy discrimination in the age of artificial intelligence and big data (August 5, 2019). *Iowa Law Review, Forthcoming*. Available at SSRN: https://ssrn.com/abstract=3347959

Rachels, J. (2004). *The elements of moral philosophy*. McGraw Hill.

Randell, I. (July 1, 2019). Could a face-reading AI 'lie detector' tell police when suspects aren't telling the truth? UK start up is in talks with Indian and British police for trials. *Daily Mail. com*. Retrieved January 6, 2021, from https://www.dailymail.co.uk/sciencetech/article-7200315/Could-face-reading-AI-lie-detector-tell-police-suspects-arent-telling-truth.html

Rhue, L. (November 9, 2018). Racial influence on automated perceptions of emotions. Available at SSRN: https://ssrn.com/abstract=3281765 or https://doi.org/10.2139/ssrn.3281765

Rhue, L. (January 3, 2019). Emotion-reading tech fails the racial bias test. *The Conversation*. Retrieved January 8, 2021, from https://theconversation.com/emotion-reading-tech-fails-the-racial-bias-test-108404

Ringrose, K., & Ramjee, D. (September 2020). Watch where you walk: Law enforcement surveillance and protester privacy, 11 *Calif. L. Rev*. Online 349 (Sept. 2020), Retrieved January 14, 2021, from https://www.californialawreview.org/law-enforcement-surveillance-protester-privacy

Roach, J. (June 26, 2018). Microsoft improves facial recognition technology to perform well across all skin tones, genders. *Microsoft Blog*. Retrieved February 29, 2021, from https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/

Ryan-Mosely, T. (December 29, 2020). Why 2020 was a pivotal, contradictory year for facial recognition. *MIT Technology Review*. Retrieved January 8, 2021, from https://www.technologyreview.com/2020/12/29/1015563/why-2020-was-a-pivotal-contradictory-year-for-facial-recognition/

Selinger, E., & Hartzog, W. (2019). The inconsentability of facial surveillance (March 19, 2020). 66 *Loyola Law Review* 101. Available at SSRN: https://ssrn.com/abstract=3557508. Accessed 1/5/2021.

Selinger, E., & Hartzog, W. (May 11, 2020a). Don't use face recognition to fight COVID: We need disease surveillance, not a surveillance state. *New York Daily News*, opinion. Retrieved November 12, 2020, from https://www.nydailynews.com/opinion/ny-oped-dont-use-face-recognition-to-fight-covid-20200511-jt53lyz6mrbztjvvcai626m5be-story.html

Selinger, E., & Hartzog, W. (April 6, 2020b). Masks and our face-recognition future: How coronavirus (slightly) clouds the picture painted by tech firms. *New York Daily News*, opinion. Retrieved January 5, 2021, from https://www.nydailynews.com/opinion/ny-oped-our-complicated-face-recognition-future-20200406-ukkhwmnr4faxpbfgalfa66eumy-story.html

Serna, I., Morales, A., Fierrez, J., Cebrian, M., Obradovich, N., & Rahwan, I. (2019). Algorithmic discrimination: Formulation and exploration in deep learning-based face biometrics. *Association for the Advancement of Artificial Intelligence* (www.aaai.org). Retrieved February 23, 2020, from GitHub: https://github.com/BiDAlab/DiveFace

Shelby, T. (2004). Race and ethnicity, race and social justice: Rawlsian considerations. *Fordham Law Review, 72*, 1697.

Smith, S. M., Stinson, V., & Prossor, M. A. (2004). Do they all look alike? An exploration of decision-making strategies in cross-race facial identification. *Canadian Journal of Behavioural Science, 36*, 146–154.

Spichak, S. (May 29, 2021). Facial recognition is regurgitating racist pseudoscience from the past. *Medium*. Retrieved June 21, 2021, from https://medium.com/age-of-awareness/facial-recognition-is-regurgitating-racist-pseudoscience-from-the-past-76ed0a28747c

Stein, M. I. (December 18, 2020). New Orleans City Council bans facial recognition, predictive policing and other surveillance tech. *MIT Technology Review*. Retrieved January 8, 2021, from https://thelensnola.org/2020/12/18/new-orleans-city-council-approves-ban-on-facial-recognition-predictive-policing-and-other-surveillance-tech/

United Nations. (December 10, 1948). Universal Declaration of Human Rights. Signed December 10, 1948. Retrieved January 4, 2021, from https://www.un.org/en/universal-declaration-human-rights/

van den Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., Hoven, J., Zicari, R., & Zwitter, A. (2017). Will democracy survive big data and artificial intelligence? *Scientific American*. Retrieved March 1, 2020, from https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/

Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in highstakes public sector decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI 2018). ACM.

Vincent, A. (2009). *Modern political ideologies* (3rd ed.). Wiley-Blackwell.

Wang, M., Deng, W., Hu, J., Tao, X., & Huang, Y. (2019). Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 692–702).

Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding, 138*, 1–24.

Zhe, W., Singh, B., Davis, L. S., & Subrahmanian, V. S. (December 12, 2017). Deception detection in videos. Retrieved January 6, 2021, from https://arxiv.org/abs/1712.04415 [cs.AI].