6

Facial Recognition Technology and Potential for Bias and Discrimination

Marcus Smith and Monique Mann

6.1 INTRODUCTION

Facial recognition technology (FRT) is one of several data-based technologies contributing to a shift in the criminal justice system, and society more broadly, towards 'automated' decision-making processes. Related technologies include other forms of biometric identification and predictive policing tools. These technology-based applications can potentially improve investigative efficiency but raise questions about bias and discrimination.[1] It is important for designers of these systems to understand the potential for technology to operate as a tool that may can discriminate, furthering biases that are already entrenched in the criminal justice system.

This chapter examines how FRT contributes to racial discrimination in the criminal justice system, potentially exacerbating existing over-representation of racial minorities. From one perspective, this technology may be viewed by some as a value neutral, objective, decision-making tool, free from human prejudice and error. However, it is also recognised that FRT, and the associated algorithms, are dependent on datasets that influence its performance and accuracy.[2] If the input data is biased, so too is the algorithm, and consequently the eventual decisions and outputs. Moreover, this discriminatory potential inherent in the technology is compounded by existing discrimination and over-representation of minority groups.

The chapter is divided into four parts: the first discusses FRT, including current applications. The second discusses the potential for bias and discrimination in the criminal justice system in relation to FRT. The third moves away from a focus on technology and considers social and structural discrimination, integrating the relevant critical literature into our argument. Finally, we conclude that even if the technology could be designed in a way that was completely free from discrimination in a techno-determinist

[1] Avi Marciano, 'Reframing biometric surveillance: From a means of inspection to a form of control' (2019) 21 *Ethics and Information Technology* 127–136, at 134.

[2] Joy Buolamwini and Timnit Gebru, 'Gender shades: Intersectional accuracy disparities in commercial gender classification' (2018) 81 *Proceedings of Machine Learning Research Conference on Fairness, Accountability and Transparency* 1–15.

https://doi.org/10.1017/9781009321211.008 Published online by Cambridge University Press

sense, it may still be used to discriminate, given, for example, the long-standing over-policing and disproportionate representation of marginalised groups in the criminal justice system. This should be considered by governments when regulating FRT and by law enforcement and judicial officers making decisions that are informed by it.

6.2 FACIAL RECOGNITION APPLICATIONS AND ISSUES

The face is central to an individual's identity and, consequently, to identifying suspects in criminal investigations. The analysis of faces by law enforcement has progressed from descriptions and sketches of suspects to the contemporary biometric integrated closed-circuit television (CCTV) technology widely used around the world in both the public and private sectors today.[3] Although there are many applications of FRT, its fundamental process remains the same. FRT involves the automated extraction, digitisation, and comparison of the geometric distribution of facial features in a way that can identify individuals. It begins with a digital image of a subject's face, from which a contour map of their features is created and then converted into a digital template. An algorithm compares digital templates of facial images and ranks them according to similarity.[4]

There are two ways in which FRT is used. The first, and less controversial, is one-to-one matching. It is used to verify the identity of a person; for example, in a security feature granting access to a smartphone or to compare a person at an international border. The use of FRT expanded rapidly following the 9/11 terrorist attacks in 2001, when it was widely integrated into passports and international border control security systems, allowing the comparison of a facial template with a live image created using SmartGate technology.[5]

The second way it can be used is one-to-many searching: the focus of this chapter. One-to-many searching seeks to identify an unknown person, for example by scanning CCTV footage of a crowd or images gathered from social media sites or more widely on the internet. Police could search based on a photograph of an unknown suspect to identify them or search for a known person in a crowd in real time. The integration of FRT with CCTV to identify unknown persons in public spaces is a major change that has taken place progressively over the past twenty years, to the point where it is normalised and widely used today. Examples of this type of application include not only fixed cameras, but also cameras on vehicles, body worn cameras, and drones, to search public spaces for persons of interest using integrated FRT.[6]

More recently, FRT has been used to search images from the internet, including images uploaded to social media, from sites such as Twitter, Instagram, LinkedIn,

[3] Marcus Smith and Seumas Miller, *Biometric Identification Law and Ethics* (Springer, 2021).
[4] Marcus Smith, Monique Mann, and Gregor Urbas, *Biometrics, Crime and Security* (Routledge, 2018).
[5] Monique Mann and Marcus Smith, 'Automated facial recognition technology: Recent developments and regulatory options' (2017) 40 *University of New South Wales Law Journal* 121–145.
[6] *R* (*on the application of Bridges*) v. *Chief Constable of South Wales Police* (2020) EWCA Civ 1058.

https://doi.org/10.1017/9781009321211.008 Published online by Cambridge University Press

Google, and Facebook. Facebook alone has over 250 billion images uploaded.[7] The use of Clearview AI by law enforcement agencies around the world came to light in 2020, and the company has been the subject of public debate and controversy, not least from social media and other internet companies that commenced legal action over the right to use these images. They claim its business model is in contravention of the terms of service of the websites the images were harvested from. In addition to the widespread use of the Clearview AI application by law enforcement agencies, the company also provides its services to the private sector, raising broader concerns. Clients that use the company's services for security purposes include the National Basketball Association, Bank of America, and Best Buy.[8] The use of images from the internet demonstrates how facial templates can be collected and used in ways that individuals may not be aware of and has the potential to connect many sources of data. It also provides insights into the scale of use of FRT, adding to the significance of racial discrimination and other pertinent issues in this context.[9]

There are inherent limitations in the use of FRT when deployed for the purposes of one-to-many identification that extend beyond bias and discrimination. Accuracy is impacted by factors such as the quality of images and cameras used, and the background and lighting conditions when the images were taken. Individual changes can impact on accuracy, including plastic surgery, ageing, weight gain, and facial coverings, such as the surgical masks that became commonplace during the COVID-19 pandemic.[10] In 2020, technology companies including IBM, Amazon, and Microsoft announced they would pause (or cease altogether), sales of their FRT to law enforcement and border security agencies owing to concerns around accuracy and privacy (Clearview AI was a notable exception to this position).[11] There have also been bans by some local governments in the United States – Somerville, Massachusetts, and San Francisco, California – which have outlawed any city department, including law enforcement, from using FRT.[12]

FRT has been found to be less accurate when used for the purposes of identifying people with darker skin tones, meaning that police deployment of FRT in criminal investigations can increase the likelihood that ethnic minorities will be wrongfully identified and prosecuted for crimes that they have not committed.[13] If this is not considered and addressed, it will likely increase the interaction of these individuals with police and compound their existing over-representation in the criminal justice system.

7 Marcus Smith and Gregor Urbas, *Technology Law* (Cambridge University Press, 2021).
8 Marcus Smith and Seumas Miller, 'The ethical application of biometric facial recognition technology' (2021) 37 *AI & Society* 167–175.
9 Ibid.
10 Smith, Mann, and Urbas, *Biometrics, Crime and Security*.
11 Smith and Miller, 'Ethical application of biometric facial recognition technology'.
12 Sidney Perkowitz, 'The bias in the machine: Facial recognition technology and racial disparities' (5 February 2021), MIT Schwarzman College of Computing, https://mit-serc.pubpub.org/pub/bias-in-machine/release/1?readingCollection=34db8026.
13 Laura Moy, 'A taxonomy of police technology's racial inequity problems (2021) *University of Illinois Law Review* 139–193.

https://doi.org/10.1017/9781009321211.008 Published online by Cambridge University Press

The issues we have raised in relation to racial discrimination cannot be viewed in isolation. In liberal democracies, there is ongoing tension between security, individual privacy, autonomy, and democratic accountability. The rapid growth and application of FRT in both the private and public sectors creates a power imbalance between individuals and the state (and corporations) and should be limited to specific and justified purposes (i.e., where the use of FRT is deemed to be both necessary and proportionate), with associated data and images carefully protected. As far as FRT being justified for security purposes, and privacy concerns mitigated, it must be subject to accountability mechanisms to prevent it being misused. Moreover, citizens should be informed about the potential use of their images for facial recognition and should have meaningfully consented to their use. Whether these systems are operated by public or private sector agencies or law enforcement, regulatory options should be publicly debated, and their use governed by legislation and subject to judicial review.

6.3 DATA, BIAS, AND RACIAL DISCRIMINATION

In 2020, a police investigation in Detroit involving Robert Williams received attention in the national press in the United States. Williams, an African American man, was arrested for shoplifting based on facial recognition identification. He was held for thirty hours before posting bail; but it was later established to be a false match based on his driver's licence photograph and distorted crime scene surveillance footage. The police department provided an apology and instigated a review of the use of FRT. Williams commenced litigation against the police department seeking compensation for his treatment.[14] The incident highlights the risks of inaccurate technology being used to identify suspects and relied upon in an arrest. Williams's case is one of several similar examples from across the United States that has drawn attention to the potential for racial bias to occur in relation to facial recognition, and for this to exacerbate the over-representation of minorities.

These incidents took place around the same time as the murder of George Floyd by a police officer, and the subsequent attention on the issue of racial discrimination through the Black Lives Matter movement.

The existing over-representation of minority groups in police databases will mean that they are more likely to be identified using facial recognition. Brian Jefferson notes that in the United States more than three-quarters of the black male population is listed in criminal justice databases.[15] Because facial images are included in these databases, they can also be used by analysis by FRT. Depending on the specific use cases (i.e., how the technology is deployed and the watchlists used), it is reasonable to suggest that FRT directs police towards those individuals who are already known to them.

[14] Drew Harwell, 'Wrongfully arrested man sues Detroit police over false facial recognition match' (13 April 2021), *Washington Post*, www.washingtonpost.com/technology/2021/04/13/facial-recognition-false-arrest-lawsuit/.

[15] Brian Jefferson, *Digitize and Punish: Racial Criminalization in the Digital Age* (University of Minnesota Press, 2020), p. 11.

https://doi.org/10.1017/9781009321211.008 Published online by Cambridge University Press

There are also data-based reasons why minority groups may be subjected to mis-identification, or over-identification, in relation to FRT, as established by empirical studies on the issue of racial bias associated with FRT. In 2019, a National Institute of Standards and Technology (NIST) report indicated that the technology achieved significantly lower rates of accuracy in African American and Asian faces – in fact, it found that faces of these races were between 10 and 100 times more likely to be mis-identified, when compared with white male faces.[16] This is supported by other research which has found that the mis-identification rate for dark-skinned women is about 35 per cent, fifty times higher than white males.[17]

The reason for this rate of mis-identification is the data inputs that the algorithms undertaking the matching rely upon. It has been established that, on average, the datasets used to train the algorithms comprise approximately 80 per cent 'lighter skinned' subjects.[18] The issues with accuracy are therefore likely to be caused by ethnic representation in datasets used to create and train the matching algorithms. Designers of the technology need to consider the racial representation in the datasets used to train facial recognition algorithms. Failing to rectify this issue, by not proactively taking steps to include representative representation in the FRT datasets, could constitute a form of racism, whether that is intended or an oversight.[19]

This is especially concerning given that ethnic minorities are already disproportionately scrutinised by law enforcement and over-represented in the criminal justice system. Increased error rates and mis-identification by facial recognition and other new technologies may compound this serious existing problem. This should be a focus for those building facial recognition systems – designing out the potential for racial discrimination by embedding racial equality in the data used to train the algorithms. Beyond this issue, any form of identification technology should not be relied upon in isolation, but only ever used in the context of other circumstantial evidence in an investigation. However, addressing the technology will only ever be part of the solution. As Damien Patrick Williams notes, 'merely putting more Black faces in the training data will not change the fact that, at base, these systems themselves will be most often deployed within a framework of racialised and gendered carceral justice'.[20]

[16] Patrick Grother, Mei Ngan, and Kayee Hanaoka, *Face Recognition Vendor Test (FRVT) Part 2: Identification* (NIST, 2019).

[17] Joy Buolamwini and Timnit Gebru, 'Gender shades: Intersectional accuracy disparities in commercial gender classification' (2018) 81 *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* 77–79.

[18] Ibid.

[19] Clare Garvie, Alvaro Bedoya, and Jonathan Frankle, 'The perpetual line-up: Unregulated police face recognition in America' (18 October 2016), Georgetown Law Center on Privacy and Technology, www.perpetuallineup.org/.

[20] Damien Patrick Williams, 'Fitting the description: Historical and sociotechnical elements of facial recognition and anti-black surveillance' (2020) 7 *Journal of Responsible Innovation* 74–83.

6.4 SOCIAL AND STRUCTURAL DISCRIMINATION

Police attention is not equally applied across the population; racial minorities are subject to disproportionate criminal justice system intervention. The consequences of this are most clearly seen in the disproportionate over-representation of minority groups in prisons around the world. This context is a necessary consideration when thinking about FRT and discrimination, because as we have described, technology can potentially perpetuate racial inequality. The following part of this chapter moves on from technical or technologically deterministic sources of bias and discrimination introduced above (i.e., those within the data or algorithms underpinning the technology) and adopts a broader structural and social view. It considers facial recognition as a socio-technical phenomenon and argues there is a need to dis-aggregate the technical and social dimensions to discrimination, as well as understand their interaction, and to do so it is necessary to clearly define and evaluate the use cases of technology vis-à-vis specific social and institutional contexts.

It has been recently argued that 'assisted' (rather than 'automated') facial recognition is a more suitable descriptor for the technology given the way that it is used to inform and direct police activities and operations (rather than truly 'automate' them).[21] Pete Fussey and colleagues' research examines a range of organisational, system, and operator factors, including the processes of human–computer interaction, and demonstrates how technical and environmental influences impact on the operation of facial recognition systems deployed by police. Fussey argues that 'while practitioners shape and condition the application and potential of their technological instruments, these practices, forms of action and ways of thinking are simultaneously shaped and conditioned by these technologies and the affordances they bring'.[22] They conclude that 'operator decision-making activities involving discretionary and suspicious judgements over who should be stopped once a possible identification has been articulated by the algorithm' and that 'technological capability is conditioned by police discretion, but police discretion itself is also contingent on the operational and technical environment'.[23] These are important considerations, because the roots of discrimination in policing do not stem entirely from the use of new technology in and of itself, but rather the institutions of policing and the actions of police officers in discretionary and discriminatory enforcement of the law.

Work by Simon Egbert and Monique Mann on discrimination and predictive policing technologies also draws attention to the socio-technical interactions between the inputs/outputs of predictive technologies and the street level decisions made by police.[24]

[21] Pete Fussey, Bethan Davies, and Martin Innes, '"Assisted" facial recognition and the reinvention of suspicion and discretion in digital policing' (2021) 61 *British Journal of Criminology* 325–344.

[22] Ibid.

[23] Ibid.

[24] Simon Egbert and Monique Mann, 'Discrimination in predictive policing: The dangerous myth of impartiality and the need for STS-analysis' in V. Badalic (ed.), *Automating Crime Prevention, Surveillance and Military Operations* (Springer, 2021), pp. 25–46.

Egbert and Mann argue that predictive policing is 'a socio-technical assemblage, encompassing not only the technical predictions themselves, but also the enactment of the predictions on the street level police – which can also have serious ramifications including discrimination'.[25] Connecting this argument to the work by Fussey, we argue that like predictive policing technologies, facial recognition technologies operate within a wider socio-technical assemblage that is shaped by the technology and wider social and structural factors such as police discretion and long-standing discrimination by police and criminal justice institutions. We contend that more attention needs to be directed to the social and structural contexts of technologies to understand their discriminatory potential when examining discrimination in policing, including in the application and use of facial recognition technologies.

Even if FRT could be designed to be perfectly 'bias free' from a technological perspective, it may still be targeted specifically against racial minorities or deployed in contexts that control and oppress them. An example of the relevance of such contextual considerations in which technology is deployed with discriminatory potential and impacts are the Smart City developments in Darwin, Australia. Pat O'Malley and Gavin Smith examine the deployment of this programme to improve public safety and public spaces, which involved the deployment of an extensive network of CCTV cameras.[26] While administrators assert that the video analytics do not include facial recognition software, there is nothing to prevent police from using facial recognition software on the CCTV footage collected. This is significant given the stark over-representation of Indigenous people in the criminal justice system in this part of Australia. For example, in 2016–2017, Indigenous people comprised 84 per cent of the prison population, and Indigenous youth comprised almost 95 per cent of those in youth detention, in addition to many other forms of disadvantage demonstrating social-economic inequality and injustice.

O'Malley and Smith argue that the Smart City technologies deployed in Darwin are 'directed at the monitoring and control of [Indigenous] people in public places' and draw attention to the 'very real prospect of the system being used to sharpen a criminalising gaze on the predominantly marginalised and excluded bodies of the Indigenous people living in and around the city'.[27] The risk is that the surveillant capabilities of the Smart City in Darwin will create negative and disproportionate impacts for Indigenous people, not only because they are already the focus on a racialized criminal justice system, but also by virtue of their daily presence in public spaces in Darwin, which is connected to social factors including unemployment and homelessness, which is in turn a consequence of Australia's colonial past and the dispossession of Indigenous people from their lands. O'Malley and Smith conclude that 'the impacts of Smart City programmes on crime control cannot be read

[25] Ibid., p. 25.
[26] Pat O'Malley and Gavin Smith, '"Smart" crime prevention? Digitization and racialized crime control in a smart city' (2022) 26(1) *Theoretical Criminology* 40–56, at 40.
[27] Ibid.

https://doi.org/10.1017/9781009321211.008 Published online by Cambridge University Press

off in a technocratically deterministic fashion … but must be situated and analysed in specific contexts' and that the 'enduring legacies of colonialism have done much to shape the nature and implications of Smart Cities projects'.[28]

This demonstrates the importance of a focus on social, political, and historical context when thinking about how technology might be 'biased' or 'discriminatory', and the need to understand the specific use cases of policing technologies, including but not limited to FRT. Even if technologically 'bias free' forms of facial recognition were indeed available, we could assume that they will be deployed in ways that are not 'neutral' and, rather, would operate to further marginalise, discriminate against, and control certain groups, especially those that are already the most marginalised and oppressed. This is pertinent given critiques by Sara Yates that 'the narrative that [FRTs] are problematic only due to their lack of transparency and inaccuracy is faulty'.[29] Yates argues that 'if these tools are allowed to be used by law enforcement, whether they have been reformed to address the accuracy and transparency issues … they will still be used disproportionally against marginalized groups and people of colour…'.[30] A focus on addressing discrimination in FRT through only technologically deterministic approaches will not remedy broader historical social injustices and harm done by police institutions and the criminal justice system, nor will banning or outlawing facial recognition. As Yates acknowledges, 'the greatest harm from these systems does not come from these tools themselves, but instead from the unjust institutions that use them'.[31] While calling for bans on FRT may be intuitively appealing, they will not resolve institutional and systemic racism and injustices perpetrated by such institutions.

The task must be to first address these fundamental injustices, or they will recur in the guise of objective technology.[32] There is a need to disaggregate the technical and social dimensions to bias and discrimination and seek to better understand the specific use cases of technology within specific institutional and social contexts. It is necessary to understand these various sources of bias and discrimination, for example those that arise from individuals (i.e., police/operator discretion), the way the system is designed (i.e., in public places that racial minorities tend to frequent), and the wider system objectives (i.e., the reason supporting the deployment of technology in that context). Analyses of the interactive effect of social and technological factors are required in order to evaluate whether the objectives and applications of certain technologies in specific contexts are necessary and proportionate, while ensuring that individual rights are upheld (including privacy, anti-discrimination, and equality). Regulatory strategies to address this issue could be targeted according to the level of risk presented in specific contexts and specific use cases of technology.

[28] Ibid.

[29] Sara Yates, 'The digitalization of the carceral state: The troubling narrative around police usage of facial recognition technology' (2022) 19 *Colorado Technology Law Journal* 483–508.

[30] Ibid., p. 505.

[31] Ibid., p. 506.

[32] Damien Patrick Williams, 'Fitting the description: Historical and sociotechnical elements of facial recognition and anti-black surveillance' (2020) 7 *Journal of Responsible Innovation* 74–83.

Moving forward, there is a need to consider, implement, and evaluate measures that aim to reduce discrimination and harm in existing systems (including the criminal justice system) and design better systems. In doing so, the structural discrimination that is a feature of many systems must be addressed to ensure that existing inequalities are not perpetuated by new technologies such as facial recognition.

6.5 CONCLUSION

The use of FRT in the criminal justice system and its association with racial discrimination is an important issue for society, given the rapidly expanding application of the technology and the limited regulation in many jurisdictions. This technology may operate to further historical forms of oppression, discrimination, bias, and over-representation of minority groups in the criminal justice system. There is evidence that FRT may contribute to racial discrimination by operating with reduced accuracy, owing to the fact that the data used to inform the operation of the technology does not include sufficient representation, leading to inaccuracy and misidentification. While this issue must be dealt with, addressing it in isolation will not be sufficient. The disproportionate focus on minorities is a far bigger problem in the criminal justice system, and the extent to which FRT perpetuates this is a subset of a much bigger, complex and historically entrenched problem. Along with the data problem, this context must be considered by those operating the technology, and by law enforcement organisations and governments, and they should not over-deploy it in areas where these minority groups are concentrated.

Rather than ban the technology altogether, we need to focus on structural discrimination and inequality – calling for a widespread ban of technologies altogether, while it may be appealing to some, is not going to be productive in the long term, nor is it realistic. While there are data-based issues here that can be addressed, this step alone will not be sufficient, and there is a need to address the social issues if we are to achieve meaningful change. Technology is not the problem, nor is it the solution. In conclusion, there are two perspectives to take account of: a data perspective and a social perspective. Although they are inter-related, they need to be disaggregated, and their socio-technical interaction better understood. First, we can see that when technology is based on datasets skewed towards white populations, it does not function as accurately on minorities. Second, technology may further existing bias and racism inherent in the individuals and organisations deploying and operating it, and in terms of inequality within the criminal justice system and society more broadly. We need to ensure that there is representative racial representation in datasets (the technical issue), and ensure that it is not over-used it in areas where racial minorities are concentrated (the social issue).