1

Facial Recognition Technology

Key Issues and Emerging Concerns

Neil Selwyn, Mark Andrejevic, Chris O'Neill, Xin Gu, and Gavin Smith

1.1 INTRODUCTION

Facial recognition technology (FRT) is fast becoming a defining technology of our times. The prospect of widespread automated facial recognition is currently provoking a range of polarised responses – from fears over the rise of authoritarian control through to enthusiasm over the individual conveniences that might arise from being instantly recognised by machines. In this sense, FRT is a much talked about, but poorly understood, topic of contemporary social, political, and legal importance. As such, we need to think carefully about exactly what 'facial recognition' is, what facial recognition does, and, most importantly, what we as a society want facial recognition to become.

Before this chapter progresses further into the claims and controversies surrounding FRT, a few basic definitions and distinctions are required. While various forms of technology fall under the broad aegis of 'facial recognition', we are essentially talking about technology that can detect and extract a human face from a digital image and then match this face against a database of pre-identified faces. Beyond this, it is useful to distinguish three distinct forms of facial technologies that are currently being developed and implemented. First, and most widespread to date, are relatively constrained forms of FRT that work to match a human face extracted from a digital image against one pre-identified face. This 'one-to-one' matching will be familiar to the many smartphone users who have opted for the 'Face-ID' feature. The goal of one-to-one matching (sometimes termed 'verification' or 'authentication') is to verify that someone is who they purport to be. A smartphone, for example, is programmed to ascertain if a face in front of the camera belongs to its registered user (or not) and then unlock itself accordingly (or not).

In this manner, one-to-one facial recognition makes no further judgements beyond these repeated one-off acts of attempted identification. Crucially, the software is not capable of identifying who *else* might be attempting to unlock the device. In contrast, a second 'one-to-many' form of FRT is capable of picking a face out of a crowd and matching it to an identity by comparing the captured face to a

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

database containing thousands (or even millions) of faces. This form of isolating any face from a crowd and making an identification has more scope for mass surveillance and tracking. Alongside these forms of facial recognition technologies designed to either verify or ascertain *who* someone is, is a third form of 'facial processing' technologies, ones that seek to infer *what* someone is like, or even *how* someone is feeling. This is technology that extracts faces from digital images and looks for matches against databases of facial expressions and specific characteristics associated with gender, race, and age, or in some cases even emotional state, personality type, and behavioural intention. This form of facial scanning has prompted much interest of late, leading to all manner of applications. During the height of the COVID-19 pandemic, for example, we saw the development of facial processing technology designed to recognise high body temperature and thus infer symptoms of virality through the medium of the face.

All told, considerable time, investment, and effort is now being directed towards these different areas of facial research and development. For computer scientists and software developers working in the fields of computer vision and pattern matching, developing a system that can scan and map the contours and landmarks of a human face is seen as a significant computational challenge. From this technical perspective, facial recognition is conceived as a complex exercise in object recognition, with the face just one of many different real-life objects that computer systems are being trained to identify (such as stop signs on freeways and boxes in warehouses). However, from a broader point of view, the capacity to remotely identify faces en masse is obviously of considerable social significance. For example, from a personal standpoint, most people would consider the process of being seen and scrutinised by another to be a deeply intimate act. Similarly, the promise of knowing who *anyone* is at any time has an understandable appeal to a large number of social actors and authorities for a range of different reasons. A society where one is always recognised might be seen as a convenience by some, but as a threat by others. While some people might welcome the end of obscurity, others might rightfully bemoan the death of privacy. In all these ways, then, the social, cultural, and political questions that surround FRT should be seen as even more complex and contestable than the algorithms, geometric models, and image enhancement techniques that drive them.

1.2 THE INCREASING CAPABILITIES AND CONTROVERSIES OF FACIAL RECOGNITION TECHNOLOGY

Facial recognition has come a long way since the initial breakthroughs made by Woody Bledsoe's Panoramic Research lab in Palo Alto nearly sixty years ago. By 1967 Bledsoe's team had already developed advanced pointillistic methods that could assign scores to faces and make matches with a mugshot database of what was described as 400 'adult male Caucasians'. Despite steady subsequent technical advances throughout the 1970s and onwards, FRT became practicable on a

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

genuinely large scale only during the 2010s, with official testing by the US National Institute of Standards and Technology, reporting accuracy rates for mass-installed systems in excess of 99 per cent by 2018.

As with all forms of AI and automated decision-making, FRT development over the past ten years has benefited from general advances in computational processing power, especially deep learning techniques, and the data storage capabilities required to develop and train large-scale machine learning models. However, more specifically, the forms of FRT that we are now seeing in the 2020s have also benefited from advances in cheap and powerful camera hardware throughout the 2010s (with high-definition cameras installed in public places, objects, and personal devices), alongside the collation of massive sets of pre-labelled photographed faces harvested from publicly accessible social media accounts.

Thus, while the technical 'proof of concept' for FRT has been long established, the society-wide acceleration of this technology during the 2020s has been spurred primarily by recent 'visual turns' in consumer digital electronics *and* popular culture towards video and photo content creation, and the rising popularity of self-documenting everyday life. But, equally, it has been stimulated by the desire of organisations to find automated solutions for managing the problem of distancing and anonymity that networked digital technologies have effected, as well as by vendors who market the virtues of the technology as a means to improve security, convenience, and efficiency, while eliminating the perceived fallibilities of human-mediated recognition systems. Thus, a combination of cultural factors, alongside exceptional societal events such as the COVID-19 pandemic, and the wider political economic will to propose and embrace techno-solutions for redressing social issues and to increasingly automate access to various spaces and services, has fashioned receptive conditions for an expansion in FRT and its concurrent normalisation.

And yet the recent rise to prominence of FRT has also led to a fast-growing and forceful counter-commentary around the possible social harms of this technology being further developed and implemented. Growing numbers of critics contend that this is technology that is profoundly discriminatory and biased, and is something that inevitably will be used to reinforce power asymmetries and leverage unfair ends. Such push-back is grounded in a litany of controversies and misuses of FRT over the past few years. For example, the United States has seen regular instances of FRT-driven racialised discrimination by law enforcement and security agencies – not least repeated instances of US police using facial recognition to initiate unwarranted arrests, false imprisonment, and other miscarriages of justice towards minoritised social groups. Similar concerns have been raised over FRT eroding civil liberties and human rights – constituting what Knutson describes as conditions of 'suspicionless surveillance', with state authorities emboldened to embark on delimited 'fishing expeditions' for all kinds of information about individuals.[1]

[1] A. Knutson, 'Saving face' (2021) 10(1) *IP Theory*, www.repository.law.indiana.edu/ipt/vol10/iss1/2/.

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

Elsewhere, FRT has proven a key element of Chinese authorities' suppression of Muslim Uyghur populations, as well as in illegal targeting of political protesters by authorities in Myanmar and Russia. Moreover, for others, FRT represents a further stage in the body's progressive colonisation by capital, as the technology has enabled the capture of increasingly detailed information about individuals' activities as they move through public and shared spaces. This data can be used to sort and manipulate consumers according to commercial imperatives, tailoring the provision of products and services so that consumption behaviours are maximised. All told, many commentators contend that there have already been sufficient examples of egregious, discriminatory, and harmful uses of FRT in everyday contexts to warrant the cessation of its future development.

Indeed, as far as many critics are concerned, there is already ample justification for the outright banning of facial recognition technologies. According to Hartzog and Selinger, 'the future of human flourishing depends on facial recognition technology being banned before the systems become too entrenched in our lives'.[2] Similarly, Luke Stark's thesis that 'facial recognition is the plutonium of AI' advocates the shutdown of FRT applications in all but the most controlled circumstances.[3] In Stark's view, the potential harms of using FRT for any purpose in public settings are sufficient reason to render its use too risky – akin to using a nuclear weapon to demolish a building. Such calls for the total suppression of FRTs have been growing in prominence. As noted scholar-activist Albert Fox Cahn recently put it: 'Facial recognition is biased, broken, and antithetical to democracy. … Banning facial recognition won't just protect civil rights: it's a matter of life and death.'[4]

1.3 JUSTIFICATIONS FOR FACIAL RECOGNITION AS PART OF EVERYDAY LIFE

While some readers might well feel sympathetic to such arguments, there are also many practical reasons to raise doubts that such bans could ever be practically feasible, even with sufficient political and public support. Proponents of FRT counter that it is not possible to simply 'dis-invent' this technology. They argue that FRTs are now deeply woven throughout the fabric of our digital ecosystems and that commercial imperatives for the information technology and surveillance industries to continue developing FRT products remain too lucrative to give up. Indeed, the technology is already becoming a standard option for closed-circuit television

[2] W. Hartzog and E. Selinger, 'Facial recognition is the perfect tool for oppression' (2 August 2018), *Medium*, https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66.

[3] L. Stark, 'Facial recognition is the plutonium of AI' (2019) 25 (3) *XRDS – Crossroads, The ACM Magazine for Students* 50–55, https://doi.org/10.1145/3313129.

[4] Cited in A. Hern, 'Human rights group urges New York to ban police use of facial recognition' (25 January 2021), *The Guardian*, www.theguardian.com/technology/2021/jan/25/new-york-facial-recognition-technology-police.

(CCTV) equipment and is regularly used by police even in jurisdictions without any formal rules governing its deployment. The industry-led and practitioner-backed promissory discourse that propagates the various virtues of FRT is already so deeply entrenched in organisational thinking and practice that it would seem highly unlikely for systems and applications to be withdrawn in the various social contexts where they now operate. In this sense, we perhaps need to look beyond polarised discussions over the fundamental need (or not) for the existence of such technology and instead pay closer attention to the everyday implications of FRT as it gets increasingly rolled out across various domains of everyday life to transform how people, things, and processes are governed.

Proponents of FRT – especially those with a commercial interest in encouraging public and political acceptance of the technology – will often point to a number of compelling 'use cases' that even the staunchest opponents of FRT will find difficult to refute. One common example is the use of FRT to reunite kidnapped, lost, or otherwise missing children with their families. The controversial face recognition company Clearview AI, which has scraped billions of face images from online sources, has highlighted the use of the app to identify victims and perpetrators of child sexual abuse.[5] Other pro-social use cases include the use of face recognition to identify people whose documentation has been lost or destroyed during natural disasters, as well as the development of specialised facial recognition software to identify the victims of war and disaster, providing some sense of closure to loved ones and avoiding the time and cost of alternative methods (such as DNA analysis or dental records). Even critics such as Luke Stark concede that FRT might have merit as a specialised accessibility tool for visually impaired people. Indeed, given the fundamental human need to know who other people are, it is always possible to think of potential applications of this technology that seemingly make intuitive or empathetic sense.

Of course, were FRT to remain restricted to such exceptional 'potential limited use cases',[6] then most people would rarely – if ever – come into contact with the technology, and therefore the concerns raised earlier over society-wide discrimination, biases, and harms would be of little significance. Nevertheless, we already live in times where a much wider range of actual applications of FRT have proven to be largely ignored or presumed uncontentious by a majority of the general public. These 'everyday' uses of FRT, we would argue, already mark the normalisation of a technology that is elsewhere perceived as controversial when in the hands of police, security services, the military, and other authorities.

These 'pro-social' uses span a diverse range of everyday contexts and settings. Perhaps one of the most established installations of facial recognition can be found at airports. FRT is a key component of 'paperless boarding' procedures, allowing

[5] K. Hill and G. Dance, 'Clearview's facial recognition app is identifying child victims of abuse' (7 February 2020), *New York Times*.

[6] Stark, 'Facial recognition', p. 55.

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

airline travellers to use this one-to-one biometric matching capacity between their e-passport photo and their physical face to check in, register their bag-drop, and then proceed through the departure and arrival gates. A major rationale for this automated infrastructure is that it makes travel processes more seamless, lessening queues and cutting costs, while also enhancing the recognition capacities (and thus organisational efficiency) of the airport authority. For instance, various studies on recognition have illustrated that the technology outperforms human recognisers, in this case, the security officials and airline clerks stationed at passport control or check-in counters. Another public setting with a long history of FRT is the casino industry. Most large casinos now operate some form of FRT. For example, the technology is strategically used to enforce blocklists of banned patrons, to enforce 'responsible gaming' by identifying under-age and 'impaired' players, and to support the exclusion of self-identified problem gamblers, as well as for recognising VIP guests and other high spending customers at the door who can then be quickly escorted to private areas and given preferential treatment.

Various forms of facial recognition and facial processing technology are also being deployed in retail settings. The most obvious application is to augment retail stores' use of CCTV to identify known shoplifters or troublemakers before they gain entry to the premises. Yet, as is the case with casinos, a range of other retail uses have also come to the fore – such as using FRT to recognise repeat customers; target screen-based advertising to particular demographics; collect information on how different customers use retail space and engage with particular arrangements of goods; and gauge satisfaction levels by monitoring the facial expressions of shoppers waiting in checkout lines or engaging with particular advertisements. Another major retail development is the use of 'facial authentication' technology to facilitate payment for goods – replacing the need to present a card and then tap in a four-digit PIN with so-called 'Pay By Face' systems, and thus lessening the 'friction' that stems from a customer forgetting or wrongly entering their code on the EFTPOS terminal, while also reducing opportunities for fraudulent activity to occur.

Alongside these cases, there are other instances of FRT being used in the realms of work, education, and healthcare. For example, the growth of FRT in schools, universities, and other educational settings encompasses a growing range of activities, including students using 'face ID' to pay for canteen meals and to check out library books; the detection of unauthorised campus incursions; the automated proctoring of online exams; and even gauging students' emotions, moods, and levels of concentration as they engage with content from the curriculum and different modes of teaching delivery. Similarly, FRT is finding a place in various work settings – often for 'facial access control' into buildings and for governing the floors and areas that employees and contractors can (and cannot) enter, as well as for registering who is in the building and where people are in the case of emergency. Other facial recognition applications also allow factory and construction employees to clock in for work via contactless 'facial time attendance' applications, and – in a more

disciplinary sense – can be utilised to monitor the productivity and activities of office staff who are working from home. Similarly, in healthcare contexts, FRT is being used for multiple purposes, from more efficient recognition of patients' identities as they enter clinical facilities so that the need for documentation is reduced (a handy administrative feature in the case of a medical emergency or to support those suffering from mental conditions such as dementia or psychosis), to improving knowledge on wait times and thus better targeting resources and services. FRT is also used to enhance facility security by controlling access to clinical facilities and identifying visitors who have previously caused trouble, as well as for patient monitoring and diagnosis, even to the point of purportedly being able to 'detect pain, monitor patients' health status, or even identify symptoms of some illnesses'.[7]

These workplace technologies are complemented by the rise of domestic forms of FRT – with various products now being sold to homeowners and landlords. One growing market is home security, with various manufacturers producing low-cost security systems with facial recognition capabilities. For example, homeowners are now using Wi-Fi-enabled, high-definition camera systems that can send 'familiar face alerts' when a person arrives on their doorstep. Anyone with an inclination towards low-cost total surveillance can run up to a dozen separate facial recognition cameras inside a house and its surrounding outside spaces. Facial recognition capabilities are also being enrolled into other 'smart living' products, such as the rise of in-car facial processing. Here, some high-end models are beginning to feature in-car cameras and facial analysis technology to infer driver fatigue and trigger 'drowsiness alerts'. Some systems also promise to recognise the faces of different drivers and adjust seating, mirror, lighting, and in-car temperatures to fit the personal preferences of whoever is sitting behind the wheel.

1.4 THE LIMITS OF FACIAL RECOGNITION 'FOR GOOD': EMERGING CONCERNS

Each of these 'everyday' forms of FRT might appear innocuous enough, but taken as a whole, they mark a societal turn towards facial technologies underpinned by a growing ecosystem of FRT, perhaps even biometric consciousness, that is becoming woven into the infrastructural fabric of our urban environments, our social relations, and our everyday lives. Most importantly, it could be argued that these growing everyday uses of FRT distract from the various latent and more overt harms that many people consider this technology to perpetuate, specifically in a landscape where the technology and its diverse applications remain either under-regulated or not regulated at all. Thus, in contrast to the seemingly steady acceptance and practical take-up of FRT throughout our public spaces, public institutions, and private

[7] M. Johnson, 'Face recognition in healthcare: Key use cases' (21 January 2022), Visage Technologies, https://visagetechnologies.com/face-recognition-in-healthcare/.

lives, there is a pressing need to pay renewed attention to the everyday implications of these technologies in situ, especially to temper some of the political rhetoric and industry hyperbole being pushed by various proponents of these systems.

1.4.1 *Function Creep*

A first point of contention is the tendency of FRT to be adopted for an ever-expanding range of purposes in any of these settings – in what might be described as processes of 'function creep'. The argument here is that even ostensibly benign implementations of FRT introduce logics of automated monitoring, tracking, sorting, and blocking into everyday public and private spaces that can then lead quickly onto further (and initially unanticipated) applications – what Andrejevic describes as a cascading logic of automation.[8] For example, scanning the faces of casino guests to identify self-excluded problem gamblers in real time may seem like a virtuous use of the technology. Yet the introduction of the technology fits with other uses that casino-owners and marketers might also welcome. As noted earlier, facial recognition can be a discreet way of recognising VIP guests and other lucrative 'high rollers' at the door who can quickly be whisked away from the general melee and then provided with personalised services to capture, or manipulate, their loyalty to (and thus expenditure in) the venue. This logic can then easily be extended into recognising and deterring repeat customers who spend only small amounts of money or whose appearance is not in keeping with the desired aesthetic of the premise, or to identify croupiers whose tables are not particularly profitable.

This cascading logic soon extends to various other applications. To continue the casino example, face recognition could be used to identify and prey on excessive gamblers, using incentives to entice them to spend beyond their means – thereby contributing to the ongoing toll the industry takes on those with gambling addictions. What if every vending machine in a casino could recognise customers through the medium of their faces before displaying prices? A vending machine that could adjust the prices based on information about customers' casino spending patterns and winnings might be programmed to serve as Robin Hood and to charge the wealthy more to subsidise the less fortunate. The more likely impulse and outcome, however, would be for casino operators to attempt to extract from every consumer as much as they would be willing to pay over the standardised price at any given moment. It is easy to envision systems that gauge the motivation of a purchaser at a particular moment, subject to environmental conditions ('how thirsty do they appear to be?', 'what kind of mood do they seem to express?', 'with whom are they associating?', and so on).

This tendency for function creep is already evident in the implementation of facial recognition by governments and state authorities. For example, the development

[8] M. Andrejevic, *Automated Media* (Routledge, 2020).

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

of facial recognition 'check-in' systems during the pandemic lockdowns to monitor COVID-19 cases undergoing home quarantine have since been repurposed by police forces in regions of India to enforce periods of house arrest. Similarly, in 2022 the UK government contracted a tech company specialising in monitoring devices for vulnerable older adults to produce facial recognition watches capable of tracking the location of migrants who have been charged with criminal offences. This technology is now being used to require migrants to scan their faces and log their geolocation on a smartwatch device up to five times a day.[9] Similarly, Moscow authorities' use of the city's network of over 175,000 facial recognition-enabled cameras to identify anti-war protesters also drew criticism from commentators upset at the re-appropriation of a system that was previously introduced under the guise of ensuring visitor safety for the 2018 FIFA World Cup and then expanded to help track COVID-19 quarantine regulations. All these examples illustrate the concern that the logics of monitoring, recording, tracking, and profiling – and the intensified forms of surveillance that result – are likely to exacerbate (and certainly not mitigate) the manipulative, controlling, or authoritarian tendencies of the places within which they are implemented.

1.4.2 *The Many Breakdowns, Errors, and Technical Failures of FRT*

A second category of harms are those of error and misrecognition – whether this is misrecognition of people's presumed identities and/or misrecognition of their inferred characteristics and attributes. In this sense, one fundamental problem is the fact that many implementations of FRT simply do not work in the ways promised. In terms of simple bald numbers, while reported levels of 'false positives' and 'false negatives' remain encouraging in statistical terms, they still involve large numbers of people being erroneously 'recognised' by these systems in real life. Even implementations of FRT to quicken the process of airport boarding only report success rates 'well in excess' of 99 per cent (i.e., wrongly preventing one in every few hundred passengers boarding the plane). Airports boast the ideal conditions for FRT in terms of well-lit settings, high-quality passport photographs, high-spec cameras, and compliant passengers wanting to be recognised by the camera to authenticate their identity and thus mobility. Unsurprisingly, error rates are considerably higher for FRT systems that are not located within similar ideal conditions. More egregious still is the actual capacity of facial processing systems to infer personal characteristics and affective states. As Crawford and many others have pointed out,[10] the idea of automated facial analysis and inference is highly flawed – in short, it is simply not possible to accurately infer someone's gender, race, or age through a face, let alone

[9] N. Kelly, 'Facial recognition smartwatches to be used to monitor foreign offenders in UK' (5 August 2022), *The Guardian*, www.theguardian.com/politics/2022/aug/05/facial-recognition-smartwatches-to-be-used-to-monitor-foreign-offenders-in-uk.

[10] K. Crawford, *Atlas of AI* (Yale University Press, 2021).

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

anticipate and thus modulate their emotions or future behaviours. As a consequence of technological limitations, as well as flaws regarding the knowability of human cognition and controllability of futures, this imaginary remains better off situated in the science fiction genre than as a plausible part of current policy and practice.

Whether or not one is perturbed by not being allowed on a plane at the first attempt or correctly recognised as feeling happy (or sad) probably depends on how often this inconvenience occurs – and what its consequences are. An erroneous emotion inference might simply result in a misdirected advertising appeal. However, in another instance it could jeopardise one's job prospects, or might even lead to someone being placed under police suspicion. System failures can have more alarming consequences – as reflected in the false arrests of innocent misrecognised individuals, people being denied access to social welfare benefits or Uber drivers being refused access to their work-shift and thereby their income. When a face recognition system fails or makes erroneous decisions, it can be onerous and time-consuming to prove that the machine (and its complex coding script) is wrong. Moreover, trial programs and test-cases continue to show the propensity of FRT to misrecognise certain groups of people more frequently than others. In particular, trials of FRT continue to show racial bias and a particular propensity to mis-recognise women of colour.[11] Similarly, these systems continue to work less successfully with people wearing head-coverings and veils, and those with facial tattoos – in other words, people who do not conform to the 'majority' appearance in many parts of the world.[12]

Of course, not being immediately recognised as a frequent flyer or a regular casino customer is unlikely to lead to serious inconvenience or long-term harm in the same way that being the victim of false arrest can generate trauma and distrust – or even ruin someone's life. Yet even these 'minor' misrecognitions and denials might well constitute further micro-aggressions in a day already replete with them. In celebrating the conveniences of contactless payments and skipping queues, we need to remember that FRTs are not experienced by every 'user' as making everyday life smoother, frictionless, and more convenient. These systems are layered on long histories of oppression and inequity, and often add further technological weight or a superficial technological veneer to already existing processes of social division and differentiation.

1.4.3 *The Circumstantial Nature of Facial Recognition 'Benefits'*

As these previous points suggest, it is important to recognise how the nature and extent of these harms is experienced disproportionately – with already minoritised populations bearing the worst effects. Indeed, the diverging personal

[11] See J. Buolamwini and T. Gebru, 'Gender shades', Conference on Fairness, Accountability and Transparency (January 2018), *Proceedings of Machine Learning Research*, pp. 77–91.

[12] See S. Magnet, *When Biometrics Fail* (Duke University Press, 2011).

experiences of technology (what Ruha Benjamin describes as 'vertical realities' of how different groups encounter the same technology) go some way to explaining why FRT is still being welcomed and embraced by many people.[13] While many groups experience facial recognition as a technology of surveillance and control, the same technologies are experienced as sources of convenience and security by others. As Benjamin reminds us, 'power is, if anything, relational. If someone is experiencing the underside of an unjust system, others, then, are experiencing its upside'.[14]

In this sense, much of what might appear as seemingly innocuous examples of FRT are apt examples of what Chris Gilliard and David Golumbia term 'luxury surveillance' – the willingness of middle-class consumers to pay a premium for tracking and monitoring technologies (such as personal GPS devices and home smart camera systems) that get imposed unwillingly in alternative guises on marginalised groups. This asymmetry highlights the complicated nature of debates over the benefits and harms of the insertion of FRT into public spaces and into the weave of everyday social relations. Indeed, 'smart door-bells', sentient cars, and 'Pay By Face' kiosks are all examples of how seemingly innocuous facial recognition features are being quietly added to some of the most familiar and intimate settings of middle-class lives, *at the same time* as major push-back occurs against the broader use of this technology in public spaces and by police and security forces, where the stakes are perceived to be higher or much less certain. At the moment, many middle-class people seem willing to accept two different modes of the same technology. On the one hand is the 'smart' convenience of being able to use one's face to unlock a smartphone, pay for a coffee, open a bank account, or drive to work in comfort. On the other hand is the general unease at the 'intrusive' and largely unregulated use of FRT in their child's school, in their local shopping centre, or by their local police force.

Yet this ambiguity could be seen as a slippery slope – weakening protections for how the same technology might be used on less privileged populations in more constrained circumstances. The more that FRT is integrated into everyday objects such as cars, phones, watches, and doorbells, the more difficult it is to argue for the complete banning of the technology on grounds of human rights or racial discrimination. Even requesting limitations on application gets harder the more diversified, hard-wired, and normalised the technology becomes. Thus the downside of middle-class consumers continuing to engage with forms of facial recognition that they personally feel 'work for them' is the decreased opportunities to initiate meaningful conversations about whether this is technology that we collectively want to have in our societies and, if so, under what kinds of conditions. As Gilliard and Golumbia conclude:

[13] R. Benjamin, *Race after Technology* (Polity, 2019).
[14] Ibid., p. 65.

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

> We need to develop a much deeper way of talking about surveillance technology and a much richer set of measures with which to regulate their use. Just as much, we need to recognize that voluntarily adopting surveillance isn't an isolated choice we make only for ourselves but one that impacts others in a variety of ways we may not recognize. We need always to be asking what exactly it is that we are enthusiastically paying for, who 'we' are and who is 'them' on the outside, and what all of us are being made subject to when we allow (and even demand) surveillance technology to proliferate as wildly as it does today.[15]

1.4.4 *The Harms of FRT Cannot Be 'Fixed'*

A fourth point of contention are the ways in which discussion of the harms of FRT in political, industry, and academic circles continues to be limited by a fundamental mismatch between computational and societal understandings around issues of 'bias'. The idea that FRT can be 'fixed' by better data practices and technical rigour conveys a particular mindset – that algorithms and AI models are not biased in and of themselves. Instead, algorithms and AI models simply amplify bias that might have crept into the datasets that they are trained in and by, and/or through the data that they are fed. As such, it might appear that any data-driven bias is ultimately correctable with better data. Nevertheless, as Deb Raji describes, this is not the case.[16] Of course, it is right to acknowledge that the initial generation of data can reflect historical bias and that the datasets used to develop algorithmic models will often contain representation and measurement bias. However, every aspect of an algorithmic system is a result of programming and design decisions and can therefore contain additional biases. These include decisions about how tasks are conceived and codified, as well as how choices are modelled. In particular, algorithmic models are also subject to what are termed aggregation and evaluation biases. All told, any outcome of an algorithmic model is shaped by subjective human judgements, interpretations, and discretionary decisions along the way, and these are reflected in how the algorithm then autonomously performs its work and acts on the world. In this sense, many critics argue that FRT developers are best advised to focus on increasing the diversity of their research and development teams, rather than merely the diversity of their training datasets.

Yet increasing the diversity of AI development teams will do little to improve how the algorithmic outputs and predictions of FRTs are then used in practice – by, for example, racist police officers, profit-seeking casino owners, and suspicious employers. Ultimately, concerns over the bias and discriminatory dimensions of FRT relate to the harms that an FRT system can do. As many of the examples outlined in previous sections of this chapter suggest, there are a lot of harms that are initiated and amplified through the use of FRT. While many of these are existing harms, the bottom line

[15] C. Gilliard and D. Golumbia, 'Luxury surveillance' (6 July 2021), Real Life, https://reallifemag.com/luxury-surveillance/.

[16] D. Raji, Post, Twitter (24 April 2021), https://twitter.com/rajiinio/status/1385935151981420557.

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

remains that FRT used in a biased and divided society will result in biased outcomes, which will then result in the exacerbation of harm already being disproportionally experienced by socially marginalised groups. Thus, as Alex Allbright puts it, rather than focussing on the biases of predictive tools in isolation, we also need to consider how they are used in different contexts – not least social settings and institutional systems that are 'chock-full' of human judgements, human discretions, and human biases.[17]

In this sense, all of the harms of FRT discussed so far in this chapter need to be seen in terms of biased datasets, biased models, *and* the biased contexts and uneven social relations within which any algorithmic system is situated and used. To the extent that it concentrates new forms of monitoring and surveillance power in the hands of commercial and state entities, the deployment of facial recognition contributes to these asymmetries. This means that algorithmic 'bias' is not simply a technical data problem, but a sociotechnical problem constituted both by human relations and the ensuing human–data relations that seek to represent and organise the former (and therefore not something that can ever be 'fixed'). Humans will always act in subjective ways, our societies will always be unequal and discriminatory. As such, our data-driven tools will inevitably be at least as flawed as the worldviews of the people who make and use them. Moreover, our data-driven tools are most likely to amplify existing differences and unfairness, and to do so in opaque ways, unless they are deliberately designed to be biased towards more inclusive outcomes and 'positive' discrimination.

All told, there cannot be a completely objective, neutral, and value-free facial recognition system – our societies and our technologies simply do not and cannot work along such lines. The danger, of course, is not that FRT will reproduce existing biases and inequalities but that, as an efficient and powerful tool, it will exacerbate them – and create new ones. As such, the development of a more 'effective' or 'accurate' means of oppression is not one to be welcomed. Instead, many applications of FRT can be accused of bolstering what Ruha Benjamin terms 'engineered inequality' by entrenching injustices and disadvantage but in ways that may superficially appear as more objective and scientific, especially given their design and implementation 'in a society structured by interlocking forms of domination'.[18] Thus, as far as Benjamin is concerned, more inclusive datasets 'is not a straightforward good but is often a form of unwanted exposure'.[19]

1.5 FUTURE DIRECTIONS AND CONCERNS

The development of FRT to date clearly raises a host of important and challenging issues for regulators and legislators to address. Before we consider the prospects for what this handbook describes as 'possible future directions in regulating

[17] A. Albright, 'If you give a judge a risk score' (29 May 2019), www.law.harvard.edu/programs/olin_center/Prizes/2019-1.pdf.

[18] Benjamin, *Race after Technology*.

[19] Ibid., p. 125.

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

governments' use of FRT at national, regional and international levels', it is also worth considering the broader logics and emerging forms of FRT and facial processing that have been put into train by the development of FRT to date, and the further issues, concerns, and imperatives that this raises.

One obvious emerging application of concern is the growing use of facial processing to attempt to discern internal mental states. Thus, for example, face recognition has been used by job screeners to evaluate the stress levels and even the veracity of interviewees. While these inferences are without scientific basis, this does not necessarily stop them from being put to use in ways that affect people's life chances. This raises the human rights issue of protecting the so-called forum internum – that is, control over the disclosure of one's thoughts, attitudes, and beliefs. Inferential technologies seek to bypass the ability of individuals to control the disclosure of the innermost sentiments and thoughts by reading these directly from visible external signs. We are familiar with the attempt to 'read' sentiment through non-verbal cues during the course of interpersonal interactions, but automated systems provide these hunches with the patina of (false) scientific accuracy and machinic neutrality in potentially dangerous and misleading ways. The inferential use of this type of automated inference for any type of decision making that affects people's life chances should be strictly limited.

Second is the prospect of the remote, continuous, passive collection of facial biometric data at scale, and across all public, semi-public and private spaces. At stake is not simply the diminishment of individual privacy, but also the space for democratic participation and deliberation. Unleashed on the world, such technology has a very high potential for a host of new forms of social sorting and stalking. Marketers would like to be able to identify individuals in order to target and manipulate them more effectively, and to implement customised offers and pricing. Employers, health insurers, and security officials would be interested in using it for the purposes of background checking and forensic investigations. With such technology in hand, a range of entities could create their own proprietary databases of big spenders, poor tippers, potential troublemakers, and a proliferating array of more and less desirable customers, patients, employees, tenants, clients, students, and more.

Indeed, the continued integration of facial processing capabilities into urban CCTV systems with automated facial recognition also marks a fundamental shift in how surveillance in public space operates. Standard 'dumb' forms of CCTV see the same thing and record what people already see in public and shared space – but do not add extra information. The ability to add face detection and recognition enables new strategies of surveillance and control that are familiar from the online world. For example, with facial recognition, the target of CCTV surveillance can shift from particular individuals or groups to overall patterns. Cameras that track all the individuals within their reach enable so-called pattern of life analysis, looking for different patterns of activity that facilitate social sorting and predictive analytics. For example, the system might learn that particular patterns of movement or

interaction with others correlate with the likelihood of an individual making a purchase, getting into a fight, or committing a crime. This type of analysis does not necessarily require identifying individuals, merely recognising and tracking them over time and across space.

Finally, then, there are concerns over how FRT is part of an increasing turn towards surveillance as a replacement of trust. As the philosopher Byung-Chul Han puts it, 'Whenever information is very easy to obtain, as is the case today, the social system switches from trust to control.'[20] No amount of surveillance can ever fully replace trust, but it can undermine it, leading to an unfillable gap that serves as an alibi for ever more comprehensive and ubiquitous data collection. Han describes a resulting imperative to collect data about everything, all the time, in terms of the rise of 'the society of transparency'. It is not hard to trace the symptoms of this society across the realms of social practice: the collection of increasingly comprehensive data in the workplace, the home, the marketing realm, and public spaces. As sensors and network connections along with data storage and processing become cheaper and more powerful, more data can be collected with respect to everything and anything. Face recognition makes it possible to link data collected about our activities in shared and public spaces to our specific identities – and thus to link it with all the other data troves that have been accumulating both online and offline. All told, the concern here is that the technology addresses broader tendencies towards the automated forms of control that characterise social acceleration and the crisis of social trust associated with the changing information environment.[21]

1.6 THE NEED FOR (AND PROSPECTS OF) REGULATION AND OVERSIGHT

With all these issues in mind, it seems reasonable to conclude that FRT requires to be subject to heightened scrutiny and accountability. For many commentators, this scrutiny should involve increased regulatory control, government oversight, and increased public understanding of the issues arising from what is set to be a defining technology of the next decade and beyond. That said, as this chapter's brief overview of the sociotechnical complexity of the technology suggests, any efforts to regulate and hold FRT to account will not be easy. We therefore conclude by briefly considering a number of important concerns regarding the philosophical and regulatory implications of FRT, issues that will be developed and refined further in the remainder of the book.

As with most discussions of technology and society, many of the main concerns over FRT relate to issues of power. Of course, it is possible to imagine uses of FRT

[20] B. Han, *The Transparency Society* (Stanford University Press, 2015), p. vii.

[21] R. Garland, 'Trust in democratic government in a post-truth age' in R. Garland (ed.), *Government Communications and the Crisis of Trust* (Palgrave Macmillan, 2021), pp. 155–169.

that redress existing power imbalances, and provide otherwise marginalised and disempowered populations with a means of resisting authoritarian control and to hold power accountable. For example, during the 2020 Black Lives Matter protests, activists in Portland developed FRT to allow street protesters to identify and expose violent police officers. Nevertheless, while it can be used for sousveillance, the mainstream roll-out of FRT across society looks set to deepen asymmetry of power in favour of institutions. Indeed, there is an inherent asymmetry in both power and knowledge associated with these processes of datafication. Only those with access to the databases and the processing power can collect, store, and put this information to use. In practice, therefore, face recognition is likely to become one more tool used primarily by well-resourced organisations and agencies that can afford the necessary processing power and monitoring infrastructure.

As such, any efforts to regulate FRT need to focus on issues of civil rights and democracy, the potential misuse of institutional power, and resulting harms to marginalised and minoritised groups. In this sense, one of the profound shifts envisioned by the widespread use of automated facial recognition is the loss of the ability to opt-out. When public spaces we need to access for the conduct of our daily lives – such as the shops where we get our food, or the sidewalks and streets we travel – become equipped with face recognition, we do not have a meaningful choice of whether to consent to the use of the technology. In many cases we may have no idea that the technology is in place, since it can operate passively at a distance. The prevalence of existing CCTV networks makes it possible to implement facial recognition in many spaces without significantly transforming the visible physical infrastructure.

Following this logic, then, it is likely that automated face recognition in the near future will become a standard feature of existing CCTV surveillance systems. Regulatory regimes that rely on public notification are ineffective if they do not offer genuine opt-out provisions – and such provisions are all but impossible in shared and public spaces that people need to access. When face recognition is installed in public parks or squares – or in commercial locations such as shopping centres, the only choice will be to submit to their monitoring gaze or avoid those spaces. Under such conditions, their decision to use those spaces cannot be construed as a meaningful form of consent. In many cities CCTV has become so ubiquitous that its use passes without public notification. Without specific restrictions on its use, facial recognition is likely to follow the same trajectory. Seen in this light, there are many reasons why regulation and other attempts to hold FRT to account faces an uphill battle (if not the prospect of being thwarted altogether). This is not to say that regulation is not possible. For example, more than two dozen municipalities in the United States banned government use of one-to-many face recognition during the first few years of the 2020s, and the European Union continues to moot strict regulation of its use in public spaces. Nevertheless, the use of the technology by private entities for security and marketing and by government agencies for policing continues apace.

https://doi.org/10.1017/9781009321211.003 Published online by Cambridge University Press

All our future discussions of possible FRT regulation and legislation therefore need to remain mindful of the strong factors driving continued demand for FRT and its uptake. For example, the promise of convenience and security combined with increasing accuracy and lower cost all serve as strong drivers for the uptake of the technology. There are also sustained commercial imperatives to continue this technology – not least the emergence of a $5 billion FRT industry that is estimated to grow to $50 billion by 2030. At the same time, we are living in a world where there are a number of powerful authoritarian drivers to continue the uptake of FRT regardless of pushback from civil society. As discussed earlier in this chapter, universal automated access comes at the expense of perpetual tracking and identification. In addition to the pathologies of bias and the danger of data breaches and hacking, there is also the threat of authoritarian levels of control. Widespread facial recognition creates the prospect of a tool that could, in the wrong hands, be used to stifle political opposition and chill speech and legitimate forms of protest. It can also be used to extract detailed information about people's private lives, further shifting control over personal information into the hands of those who own and control the monitoring infrastructure.

Regardless of such impediments and adversaries, many people contend that the time to develop clear regulations in keeping with commitments to democracy and human rights is now. Building support for such regulation will require concerted public education programmes that focus on the capabilities and potential harms of the technology. At the moment, its potential uses and capabilities are not understood widely and are often framed in terms of personal privacy invasion rather than its potentially deleterious effects on democracy and civic life. Developing appropriate regulation will also require negotiating the tension between the commercial pressures of the data-driven surveillance economy, the security imperatives of law enforcement, and civic values of freedom of expression, movement, and personal autonomy. The outcome we need to avoid is the one towards which we seem to be headed: a situation in which the widespread deployment of the technology takes place in a regulatory vacuum without public scrutiny or accountability.

The legal challenge of FRT lies in the fact that the consent scheme is not the best approach to protect individual rights as discussed earlier. And in some contexts, preventing its uses based on individual rights' argument may not be in the interest of the general public. In this complex situation, we should not be forced into making a choice between protecting the individuals and protecting the society at large (an argument that Chinese lawmakers are now working on through the introduction of a revised data protection law effective in 2021). Instead, we need to develop laws that will not obscure self-governance (individual rights protection) in relation to the promotion of the application of FRT as public interests. The boundaries of legal application of FRT need to be established. In it, the liability of those who are collecting, collating, and analysing facial data should be a key consideration. For

example, if the use of FRT is permitted, the re-use of such information without individual authorisation should be prohibited. The emphasis should also be about how to prevent harms resulting from public interest exceptions.

1.7 CONCLUSIONS

These are just a few opening observations and points in what needs to be a prolonged society-wide discussion over the next decade and beyond. While it is unlikely that a consensus will ever be reached, it is possible to develop a clear sense of the boundaries that we want to see established around this fast-changing set of technologies. That said, such is the pace of change within biometrics and AI, it might well be that *facial* recognition technology is only a passing phase – researchers and developers are already getting enthused over the potential scanning of various other bodily features as a route to individual identification and inference. Yet many of the logics highlighted in this chapter apply to whatever other part of the human body this technology's gaze is next trained on – be it gait, voice, heartbeat, or other.

Of course, many of the issues raised in this chapter are not unique to FRT per se – as McQuillan reminds us, every instance of 'socially applied AI has a tendency to punch down: that is, the collateral damage that comes from its statistical fragility ends up hurting the less privileged'.[22] Nevertheless, it is worth spending time unpacking what is peculiar about the computational processing of one's face as the focal point for this punching down and cascading harm. This chapter has therefore presented a selection of issues that we identify from the perspective of sociology as well as culture, media, and surveillance studies. There are many other disciplines also scrutinising these issues from across the humanities and social sciences – all of which are worth engaging with as bringing a valuable context to legal discussions of FRT. Yet we hope that the law and legal disciplines can bring an important and distinctive set of insights in taking these issues and conversations forward. Legal discussions of technology bring a valuable pragmatism to the otherwise ambiguous social science portrayals of problematic technologies such as FRT – striving to develop 'a legitimate and pragmatic agenda for channelling technology in the public interest'.[23] We look forward to these conversations continuing across the rest of this handbook and beyond.

[22] D. McQuillan, *Resisting AI* (University of Bristol Press, 2022), p. 35.

[23] R. Calo, 'The scale and the reactor' (9 April 2022), SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4079851, p. 3.