AI & SOCIETY
https://doi.org/10.1007/s00146-023-01634-z

OPEN FORUM

When facial recognition does not 'recognise': erroneous identifications and resulting liabilities

Vera Lúcia Raposo[1,2]

Received: 17 July 2022 / Accepted: 17 January 2023

Abstract
Facial recognition is an artificial intelligence-based technology that, like many other forms of artificial intelligence, suffers from an accuracy deficit. This paper focuses on one particular use of facial recognition, namely identification, both as authentication and as recognition. Despite technological advances, facial recognition technology can still produce erroneous identifications. This paper addresses algorithmic identification failures from an upstream perspective by identifying the main causes of misidentifications (in particular, the probabilistic character of this technology, its 'black box' nature and its algorithmic bias) and from a downstream perspective, highlighting the possible legal consequences of such failures in various scenarios (namely liability lawsuits). In addition to presenting the causes and effects of such errors, the paper also presents measures that can be deployed to reduce errors and avoid liabilities.

Keywords Facial recognition · Artificial intelligence · Misidentification · Biases · Liability

1 Introduction

Facial recognition (FR) is a technology based on artificial intelligence (AI) that analyses human faces for several purposes (Raposo 2022a; The Alan Turing Institute 2019). It operates by measuring human facial features, using these measurements to create a so-called 'biometric template', which is nothing more than a mathematical formula (The Alan Turing Institute 2019).

Like many other AI-based tools—robotic agents (Sharkey 2017), chatbots (Parviainen and Rantala 2022), and dark patterns (Neuwirth 2022)—it raises legal and ethical concerns, mostly due to the blurriness of AI technology, leading to its classification as a black-box, and to the lack of a clear legal framework to address its many challenges.

This paper analyzes cases in which the use of FR to identify individuals leads to mistaken identifications, exploring its causes and legal consequences. The final aim is to provide the users of this technology with adequate knowledge about what can go wrong, why, and what legal risks they incur, so that they can protect themselves and their businesses.

Reports of dark-coloured faces not being identified as faces by Hewlett Packard's facial-tracking software or being identified as gorilla faces by Google Photos (Umoja 2018), and of Asian people being asked if they were blinking when their photos were taken by Nikon cameras, or being told that their eyes were closed by FR software in airports (Borgesius 2018), triggered this discussion.

FR has several purposes in addition to identification: analysis of human emotions, detection of genetic diseases and profiling of individuals based on their characteristics (Raposo 2022a), and inaccuracies can occur in all these domains. These other uses raise legal and ethical issues that should not be undermined, such as risks of discrimination (Castelvecchi 2020), invasion of privacy (Leong 2019), undue data processing (Raposo 2022a) and manipulation of our thoughts and actions (Neuwirth 2022). However, this paper deals exclusively with inaccuracies in FR identification.

Furthermore, there are multiple legal risks and concerns involved in using FR, specifically those related to privacy and data protection (Raposo 2022a), but also to other fundamental rights (European Union Agency for Fundamental Rights 2019) and to the rule of law itself in liberal democracies (Smith and Miller 2022). This paper, however, focuses

✉ Vera Lúcia Raposo
vera.lucia.raposo@novalaw.unl.pt

[1] Nova School of Law, Campus de Campolide 18, 1099-032 Lisbon, Portugal

[2] Centre for Research on Law and Society (CEDIS), Lisbon, Portugal

Published online: 08 February 2023

solely on FR accuracy, the possible harms arising from an erroneous identification and the associated liabilities.

2 FR for identification purposes

The process of FR for identification purposes involves several stages. First, detect a face in an image, which will then be analysed by AI software to measure the geometry of the face. Subsequently, these measurements are converted into data; that is, the picture of a face (analogue information) is converted into a template (digital information). Finally, this template is matched against other template or templates (Fu 2021).

This FR process can differ, depending on the particular form of identification applied (Raposo 2022a). In authentication/authorisation,[1] the template created from a photo of a given individual is compared with an earlier photo of that same person stored in the system (usually voluntarily submitted by that person as a precondition to receiving a benefit such as a product or service). If there is a positive match, the person is granted access to that benefit; for example, their iPhone is unlocked (Apple 2021) or they are admitted access to a restricted area (Maciel et al. 2016). The aim of such a comparison is to establish whether a person is who he/she claims to be. In contrast, when FR is used for verification/recognition, the template is compared with others saved on a database to establish whether the person is who the system operator believes him/her to be. This FR is widely used for immigration purposes in airports (Glusac 2022) and in law enforcement (Raposo 2022b).

FR operates using facial features, which are a type of biometric data (Kindt 2013). A relevant definition of biometric data results from Article 4(14) of the General Data Protection Regulation (GDPR)[2] and Article 3(33) of the AI Draft Act (European Commission 2021a, b).[3] These wide definitions cover two types of personal characteristics: those related to behaviours (actions, forms of walking, forms of writing, voice, mannerisms, habits) and those related to physical characteristics (facial features, but also DNA, iris patterns, fingerprints, palm prints). In theory, more than one method of biometric identification should be used to avoid mistaken identifications.

Biometric data are not only personal data but a special type of personal data, the so-called sensitive data (as per Article 9 of the GDPR, which as a rule forbids the processing of such data, although some exceptions are admitted in its number 2), due to its intrinsic connection to the human person, its special capacity to identify the individual and the consequent risks involved for the latter's rights and fundamental freedoms (United Nations Development Group 2017). The special 'sensitivity' of biometric data is likewise recognised in par. 74 of the UNESCO, Recommendation on the ethics of artificial intelligence (UNESCO 2021).[4]

2.1 (In)accuracy of FR

Despite FR's increasing accuracy, even the most minute inaccuracy in FR can lead to erroneous results (European Digital Rights 2019; European Union Agency for Fundamental Rights 2019; Office of the High Commissioner for Human Rights 2021). A 2016 study reported that colourful glasses frames are enough to deceive FR technology (Sharif et al. 2016). Moreover, of all possible forms of biometric identification (fingerprints, palmprints, iris, DNA, voice), FR is considered to be the least accurate (Thakkar 2017).

Assessments of FR accuracy differ. Some studies have claimed that the rate of inaccuracy is extremely high. One study analysed eight FR systems on the market and concluded that the error rate ranged between 48 and 62% (Dupr et al. 2020), and other studies report false positive rates as high as 98% (Sharman 2018). In contrast, in a study carried out in 2018, the US National Institute of Standards Technology (NIST) analysed 127 FR algorithms developed by 45 research and development laboratories (Grother et al. 2018). The study concluded that the best algorithms (according to 2018 standards) had an error rate of less than 0.2%.[5] A NIST

[1] Note that these designations are not universal. For other terminology, see, e.g., Crumpler 2020.

[2] The GDPR refers to the Regulation (EU) 2016/679 (European Parliament and Council 2016a, b). Article 4(14) defines biometric data as 'personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data'.

[3] The AI Draft Acts refers to Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final. Its Article 3(33) states that biometric data 'means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data'.

[4] 'Member States should establish their data policies or equivalent frameworks, or reinforce existing ones, to ensure full security for personal data and sensitive data, which, if disclosed, may cause exceptional damage, injury or hardship to individuals. Examples include data relating to offences, criminal proceedings and convictions, and related security measures; biometric, genetic and health data; and -personal data such as that relating to race, colour, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, or disability and any other characteristics'.

[5] However, some studies pointed a caveat a caveat in those results: fit seems that in at least 10% of the images analysed, the algorithm produced a correct identification but with a low confidence level (Renaissance Numérique Renaissance Numérique 2020).

report from 2020 confirmed FR's accuracy, by concluding that 'the best facial recognition algorithms in the world are highly accurate and have vanishingly small differences in their rates of false-positive or false-negative readings across demographic groups' (McLaughlin, Castro 2020).

Regardless of these different results, the fact remains that when the volume of people being 'scanned' is very high, even a narrow margin of error allows a concerning number of misidentifications. 'A system deployed on the scale of the population of the European Union would have to achieve an error rate of 0.00000224% (i.e., an accuracy rate of 99.99999776%) to commit less than 10 errors for a total of 446 million individuals. We are still a long way from such performances' (Renaissance Numérique 2020, 23). In absolute terms, the percentage of error grows when the number of identifications grows, but moreover, the size of the target population also affects accuracy. The more individuals that the system must 'scan', the higher the percentage of error.

For instance, suppose FR is being used at the UK's Heathrow Airport, where an average of 219,458 people pass through daily (Heathrow Airport n.d.). Assuming a percentage of false positives of 0.01% (in the case of extremely precise technology), this equates to 22 people per day being erroneously tagged and stopped. This is clearly problematic. However, considering that the more modern FR software has an accuracy rate of 99% (McLaughlin, Castro 2020) and that the human eye reaches, at best, an accuracy of 85% (White et al. 2015), a properly developed FR system is more effective than human operators.

Several factors may hamper the accuracy of an FR identification,such as the passage of time or the use of medication or recreational drugs (Renaissance Numérique 2020). If a person being scanned has a disability that affects the bone structure or involves facial paralysis, this may undermine the system's precision (European Union Agency for Fundamental Rights 2019). Age is another obstructive factor, as the level of FR accuracy diminishes when current images are compared with photos taken a long time ago (Boussaad and Boucetta 2020). Facial resemblance remains a problem, as many algorithms cannot distinguish between twins (Noroozi and Toygar 2017).The use of face masks–recurrent during the pandemic–also detracts from accuracy (Hariri 2022). Actions taken deliberately to trick the system are an obvious source of frustration, a clear example being data poisoning.[6]

The main challenges addressed in this paper are as follows: (i) the nature of FR, which operates through probabilities, and the consequent definition of the threshold, for a positive result (ii) AI unintelligibility and (iii) AI biases.

[6] Data poisoning software operates by manipulating individual pixels in an image in such a way that the change is invisible to the human eye but can indeed fool the FR, although the most developed system of FR can be trained to ignore data poisoning (Radiya-Dixit and Tramer 2021).

2.2 FR as a matter of probability

As stated by the European Union Agency for Fundamental Rights (2019, 34), 'Facial recognition technology algorithms never provide a definitive result, but only probabilities'. FR as identification operates as a face-comparison model. The system compares pictures–either one-to-one (authorisation) or one-to-many (recognition)–to identify a match. The final result is not expressed as a 'yes' or 'no' answer but as a probability (European Union Agency for Fundamental Rights 2019). The so-called 'confidence level' or 'confidence score' shows the probability, as a percentage, of the image being correctly detected by the algorithm (European Data Protection Board 2022). The software cannot determine whether two templates belong to the same person (an exact match) but only how likely it is that they belong to the same person (Commission Nationale de l'Informatique et des Libertés 2019; Institute and International Association of Chiefs of Police 2019). Such probabilities depend on how accurate the software is (although, as noted above, FR is more accurate than eyewitnesses (Bambauer 2021).

In a one-to-one comparison, such as when FR is used to unlock an iPhone, the level of accuracy is higher, because the matching operates between only two photos instead of comparing one photo with an entire database of potentially millions of photos (Thompson 2020). Apple claims, 'The probability that a random person in the population could look at your iPhone or iPad Pro and unlock it using Face ID is less than 1 in 1,000,000 with a single enrolled appearance whether or not you're wearing a mask' (Apple 2022). If that is the case, the level of accuracy is extremely reliable.

In a one-to-many comparison, however, the percentage of error increases, as the chance of error rises with the number of photos compared. This is, therefore, the context that raises the most concerns.

When comparing two templates, FR software indicates the level of probability that the templates match each other. The system declares a match when the probability meets an established threshold (Commission Nationale de l'Informatique et des Libertés 2019). The higher the threshold is set, the higher the number of matches the system misses. For instance, if the algorithm is set to only provide results with 99% confidence, it may miss many positive matches, which are reported as (false) negatives. However, the matches the system does provide have a very high confidence level, minimising false positives (Crumpler 2020). It is especially important to set a high threshold when the process is conducted without human overview (European Data Protection Board 2022), as human intervention is necessary to diminish the number of false positives. Although setting a high threshold necessarily increases the number of false negatives, this result is deemed to be less legally problematic than the alternative (Crumpler 2020).

Therefore, the determination of the threshold will depend on the risks arising from false positives and false negatives. For instance, a false positive that grants someone access to another individual's iPhone is less problematic than an erroneous mismatch in a criminal case, risking a wrongful conviction. Thus, in the former, the threshold for a match may be lowered, whereas the latter requires a very high threshold (Umoja 2018). An error in the identification of a criminal is particularly concerning, as certain ramifications, such as detention (which, in the case of erroneous identification, constitutes unlawful detention) and public disclosure of the person's identity, may damage the individual's reputation (European Union Agency for Fundamental Rights 2019).

2.3 AI unintelligibility

There are two types of AI: black box and white box (Loyola-González 2019). 'Black box' (Smith 2020)is a term used to describe the opaque nature of this technology. In its 2020 White Paper on Artificial Intelligence, the European Commission (2020, 12) made note of AI's 'opacity ("black box-effect"), complexity, unpredictability and partially autonomous behaviour'. Complex AI applications involve significant opacity in terms of the visibility of and justifications for the algorithms and data they use.

FR has been described as a 'black box' AI. Researchers have yet to determine how FR algorithms operate. The fact that modern FR is based on deep neural networks contributes to this phenomenon. In this form of AI, not even the programmers know how the AI is developed and which elements it uses to form its determinations (Bathaee 2018). This opacity makes it difficult to correct an FR system's biases and malfunctions.

The nebulous nature of FR is exacerbated by the secrecy involved in the performing technology, protected by laws governing intellectual property rights and commercial secrets (European Digital Rights 2019). The novelty of this technology, which strongly relies on business secrecy, probably discourages developers and manufacturers from disclosing relevant information. This is not a hypothetical situation. In a British case related to the use of FR for law enforcement purposes, an audit of the FR system was required to establish potential biases, but the provider of the AI refused to disclose information invoking 'business secrecy', and the court upheld this argument (R [Bridges] v Chief Constable of South Wales Police [2020] EWCA Civ 1058, §199).

2.4 Bias and discrimination

FR algorithms differ in their accuracy rates between genders and ethnic groups. The 2018 Gender Shades project analysed three gender classification algorithms and concluded that they all performed poorly with regard to dark-skinned females, with an error rate 34% higher than that for light-skinned females (Buolamwini and Gebru 2018). A 2019 report from the NIST analysed 189 FR algorithms (covering almost the entire market), and it concluded that most of them had some type of bias, especially against Black and Asian faces (for these groups, the frequency of error was about 10–100 times higher than for white faces). Grother et al. (2019, 2) stated, 'Our main result is that false positive differentials are much larger than those related to false negatives and exist broadly, across many, but not all, algorithms tested. Across demographics, false positive rates often vary by factors of 10 to beyond 100 times'. They also revealed discrepancies regarding gender, stating, 'We found false positives to be higher in women than men, and this is consistent across algorithms and datasets. This effect is smaller than that due to race' (Grother et al. 2019, 2). A recent report from the US General Services Administration (US General Services Administration 2022) also pointed out racial biases against African Americans. Overall, FR has proven to be discriminatory towards certain groups, namely females (Buolamwini and Gebru 2018)and Black people (Najibi 2020).

Two main causes of such biases have been proposed: biased databases and biased programmers (Cowgill et al. 2020).

Databases are biased when they predominantly include photos from a specific group of people, resulting in algorithms trained to recognise mostly individuals from that group, usually white males (Umoja 2018). The algorithm is not able to identify other types of individuals with the same accuracy, as its training lacks adequate representation of these types.

Programmers are also said to be biased because, like all humans, they have conscious and unconscious preconceptions that they transfer to the codes they write, for instance, by selecting the variables to be used or ignored (Donald 2019). The so-called 'algorithm bias' refers to cases in which algorithms 'inherit' the prejudices of their creators, namely towards ethnic groups and females, as most programmers are white males (Borgesius 2018). The algorithms on which AI is based tend to emulate the social biases of the human operators who created them and, voluntarily or inadvertently, transposed such biases to the algorithms (Chattopadhyay et al. 2020; Cofone 2019; Donald 2019; European Digital Rights 2019; European Union Agency for Fundamental Rights 2018). These two factors are probably interrelated, as studies show that 'engineers exert greater effort and are more responsive to incentives when given better training data' (Cowgill et al. 2020).

The tendency is for such flaws to be overcome. Recent studies show advances in this regard (US Department of Homeland Security 2019) as compared with studies from one decade ago (Klare et al. 2012), not only due to technological improvements but also because developers realised

that richer and more diverse datasets provide less biased results. A 2020 NIST report concluded that the best FR software does not show any significant difference in recognising people of different genders and ethnic groups: false-negative rates of 0.49% or less for Black females and no more than 0.85% for white males (McLaughlin, Castro 2020). Modern FR software has an accuracy rate of 99% (McLaughlin, Castro 2020) and studies continue to provide proposals to prevent discriminatory results (Serna et al. 2022).

3 Legal consequences: liability related to FR inaccuracies

3.1 Harm and compensation for moral damages

Public or private entities using FR may be held liable for damage caused by the use of this technology, either under the rules of contract law or, most commonly, tort law. Errors can, at best, lead to awkward or compromising situations; however, they can also impose severe harm on the individuals involved (as in the case of a man arrested and detained for six days because of a mismatch in FR (Infobae 2019), for which the authorities can be held accountable. Moreover, an erroneous identification may result in discrimination and harm to the individual's dignity. 'This misrecognition on the basis of rights is ethically problematic not just as a violation of moral principles regarding equal treatment, but also because it can have long-lasting damaging effects on a person's self-development, in particular on their sense of self-respect' (Waelen 2022).

Discrimination—banned by Article 21 of the European Charter of Fundamental Rights, Article 14 of the European Convention of Human Rights and Protocol 12 to the European Convention of Human Rights (European Digital Rights 2019)—involves mostly non-white individuals (Najibi 2020) and females (Buolamwini and Gebru 2018). Other vulnerable individuals are also at great risk, namely, people with facial deformities or that suffer from any disease capable of altering facial traits. In a sense, every individual is vulnerable in face of this technology, as even minor changes in look (bigger glasses, a different haircut) can lead to erroneous identifications. However, some individuals—the ones whose ethnic features or other distinctive traits are less known to the machine—are especially at stake.

It is at least theoretically conceivable that a person mistakenly identified—and thus subject to public embracement and humiliation—files a compensation request for moral damages against the manufacturer and/or the user of the FR system.[7] This possibility, however, might be more theoretical than real. The reason is that in common law jurisdictions compensation for moral damages solely does not usually take place, unless intentionally inflicted or when physical or patrimonial losses are also at stake (Ben-Shahar and Porat 2018), requirements that hardly will happen in our scenario. In contrast, in European continental law, compensation for purely moral damages remains more flexible (Basenko, Avanesian and Strilko 2022). Based on the roman tradition, civil law distinguishes between patrimonial damages, which involve both damages to property and the body, and moral damages. Recovery for moral damages exclusively (i.e., independently of the existence of any other damage) in tort law is fully recognised in civil law jurisdictions, which can potentially include the distress and anguish caused by an erroneous identification by an FR system. The problem might be, however, that not every moral damage is to be compensated. It is common in civil law jurisdictions to restrict compensation for moral damages, a limitation sometimes established by statute.[8] If that is the case, it would remain at the discretion of national courts to decide whether the moral harm argued by a victim of an erroneous FR identification is serious enough to receive compensation under tort law rules.

3.2 Liability of various stakeholders

3.2.1 Manufacturer and user liability

The developers of FR algorithms and the manufacturers of AI systems may be held accountable, based on laws governing defective products. In the context of such a complex device, several players may be involved in its development and manufacture, raising the question of the proper distribution of liability.

Liability may also fall on the AI user, that is, the natural or legal person using the FR. As FR algorithms provide only probabilities in the form of percentages and not definitive matches or identifications, it is up to the natural or legal person using the algorithm to decide what do to with such percentages. A rush decision, taken without sufficient evidence, may lead to errors and thus to liability. For instance, in 2021, a man sued the Detroit police department that had arrested him for an alleged robbery based on algorithmic

[7] Considering the kind of harm resulting from mistaken identifications by FRT–the scope of this paper–it is difficult to envisage a case of criminal liability, which does not deal with purely moral harm, except in the case of crimes against the honour.

[8] See, for instance, Article 2059 of the Italian Civil Code ('Il danno non patrimoniale deve essere risarcito solo nei casi determinati dalla legge') and Paragraph 823(2) of the German Civil Code ('Die gleiche Verpflichtung trifft denjenigen, welcher gegen ein den Schutz eines anderen bezweckendes Gesetz verstößt'),

identification. FR was used on images collected by camera surveillance, and the system matched the man in the images with the plaintiff's driving licence photo, taken from the driving licence database. Subsequently, the match was proven to be wrong (Williams v. City of Detroit, Michigan, A Municipal Corporation et al. (2:21-cv-10827), Michigan Eastern District Court). However, the user's liability is complicated, as users may not be able to identify, much less monitor, the training data used to develop the algorithm because developers of FR software are generally unwilling to disclose the data used for copyright and trade secret protection purposes (European Union Agency for Fundamental Rights 2019).

In certain cases, the user's liability is relatively clear. Consider an FR system whose components are not initially defective but become so through misuse. For instance, suppose the software is sold with a high-resolution camera, which is damaged by the user and replaced with a lower-quality camera that provides images in which small details of people's faces cannot be distinguished. Erroneous identifications cannot, in this scenario, be blamed on the manufacturer. Similarly, consider an FR system intended to be placed and used in a brightly lit area (and sold with clear statements of this information). If, instead, the user installs the camera in a dimly lit area, this may result in low-quality images, causing erroneous identifications. Such an FR system cannot be considered defective, even if the percentage of errors is above the established threshold, as the malfunction is due to the contravention of the manufacturers' instructions.

3.2.2 The new norms on manufacturer's liability

The Product Liability Directive (PLD) (Council of the European Communities 1985) is a European law that provides for strict liability for defective products. Its application exempts plaintiffs from the duty to demonstrate in court the culpability of the manufacturer.

The PLD applies to a very restrictive set of 'defects', all related to malfunctions that jeopardise the user's safety. As stated in Article 6(1) of the PLD, 'A product is defective when it does not provide the safety which a person is entitled to expect'. A defect, under the PLD, is associated with a lack of safety. As such, a misidentification by an FR system would be out of its scope, as it is a non–safety-related malfunction. A safety issue may exist regarding FR hardware, such as when an FR camera burns a person's face when he/she comes too close. However, it is difficult to envisage such harm resulting from FR software. Even the moral harm suffered by a misidentified individual–which is especially likely to happen with identification activities carried out by police forces–does not equate to a lack of safety under the PLD, as it states in the introductory notes that its aim is 'to protect the physical well-being and property of the consumer'. In an extremely flexible interpretation of this concept, it could be said that in some circumstances an erroneous identification can put a person at risk (for example, in cases of detention and in which the person is physically attacked); however, it is far-fetched in terms of practical application.

In any case, even considering a defect as described in the PLD, in its current (not revised) version, it is not clear whether the PLD applies to FR technology and, if so, in which terms. It has long been accepted that the PLD only applies to products, not to services, as recently underlined in the Krone case (Vi v Krone,–Verlag Gesellschaft mbH & Co KG, 10 June 2021, Case C-65/20, ECLI:EU:C:2021:471). However, the Advocate General's opinion in this case went one step further, stating that the PLD applies 'only [to] the physical properties of the product' (Advocate General 2021). The European Commission itself did not go that far; it simply underlined the difficulties of applying the PLD to AI products (European Commission 2021a).

However, a draft revised version of the PLD was recently disclosed, in which AI software, including FR software, is firmly placed within the scope of the PLD (European Commission 2022b). Therefore, the ones harmed by the hardware parts in FR systems can sue the manufacturer under the strict liability regime therein established, but also the ones harmed by the software used (though, within FR, harm due to the software's lack of safety is hardly foreseeable).

Moreover, when it comes to a software defect, an additional set of norms will intervene, the AI Liability Directive (European Commission 2022a), recently proposed by the EU. According to its current draft, people harmed by the AI technology underneath an FR system will be able to require the AI provider to have access to information ('evidence') relevant to ground a lawsuit for damages,[9] which in this case could involve information about the quantity and quality of images used to train and test the FR system or about the code used. If such evidence is not provided, a presumption of fault (i.e., a presumption of non-compliance with a relevant duty of care) will operate to give as demonstrated the very same events that the missing evidence was trying to prove, as set forth in Article 3(5) of the AI Liability Directive. Moreover, a presumption of causation between the violation of the duty of care and the output might also apply (Article 4 of the AI Liability Directive).

Despite all these additional tools to address liability, the fact that this particular AI technology works with

[9] Article 3 of the AI Liability Directive establishes the right to access evidence when high-risk AI are at stake, as these are defined under Article 6 of the AI Draft Act. Considering that FR technology poses a high risk of harm to the health and safety or the fundamental rights of persons (recital 32, Article 7(1)(b), Article 7(2)(c) of the AI Draft Act), it will be considered a high-risk system (European Parliament 2021).

probabilities will very likely difficult the assessment of malfunctioning. As a correct result cannot be guaranteed in FR, it all comes to the definition of the threshold to accept its result and the put in place of adequate mechanism of confirmation (e.g., other forms of biometric identification) and human oversight. These measures consubstantiate the standard of care, as defined in Article 2(9) of the AI Liability Directive, both for AI developers and AI users (as this Directive applies to both) in what regards this technology.

4 Measures to avoid erroneous results (and exclude liability)

4.1 State of the art technique

Accuracy strongly relates to the technique used, which must be state of the art (European Union Agency for Fundamental Rights 2019), a typical requirement in domains characterised by extreme technological complexity and huge investment (Boyd and Ingberman 1995). This observation, which seems obvious, immediately faces obstacles in its implementation. First, the difficulty in defining the concept of 'state of the art' and which technology can be classified as such. Secondly, even assuming that the state of the art criterion is well-defined, it is difficult to have access to such technologies, usually protected by strong IP rights. Thirdly, a concern related to the costs of using highly developed technologies, which might overcome the expected profit.

The first concern is based on two difficulties, one legal and the other technical. Legally, there is no consensus about the exact content of the concept 'state-of-the-art': does it refer to the respect for common (standard) practices in a given industry or does it imply the inexistence in the market of a safer/better product (Boyd and Ingberman 1995)? Technically, it is difficult to reach an agreement about which technology should be considered 'state of the art' at any given moment (Boyd & Ingberman 1995) mostly because FR technology is continuously developing. Within FR it is commonly agreed that computation power and developed algorithms (such as deep convolutional neural networks) should be used to achieve more accurate results. In recent years, a significant improvement in FR was achieved with the introduction of neural networks, a form of machine learning able to recognise individuals from a large dataset of images (Welinder and Palmer 2018), in particular convolutional neural networks (Wright 2019), considered the state of the art in this technology (Hancock et al. 2020), but it is not known until when. Moreover, the more developed the technology is, the more skills are required of the operators; thus, technological development must come with proper technological training (Sarabdeen 2022). The fact that AI is able to learn can contribute to improvements in its level of accuracy. 'The algorithm then adjusts and fits itself for accuracy, enabling it to make predictions about a new photograph with increased precision' (Office for Product Safety and Standards 2021, 15). For instance, in an FR authentication mechanism, the system learns how to detect small changes in a person's face, such as from the use of makeup or from ageing (Jesdanun 2017; Apple 2022). In the context of the face ID used to unlock the iPhone, Apple states that '[t]his data [the mathematical representations of the face] will be refined and updated as you use Face ID to improve your experience, including when you successfully authenticate' (Apple 2022). Hardware also plays an important role (Roth 2009). For instance, many of the cameras used in FR do not adequately capture darker skin, resulting in images of lower quality for persons of colour (The Alan Turing Institute 2019) but a switch to high-resolution cameras may thus help to increase accuracy.

The obstacle to the use of state-of-the-art technology resulting from IP rights must also be considered, as novel technologies will most likely be protected by patents. Even though the protection conferred to patents is not absolute—the protected AI system can still be used for interoperability purposes, scientific research purposes, private use or under a compulsory license (Dent et al. 2006)—the most viable solution is a licence agreement.

However, eventually, we might have to rethink the balance between the protection of IP rights and the diffusion of knowledge. More than a decade ago, it was already acknowledged that '[f]rom an economic perspective, it is to be considered that IP monopolies, while spurring investment in new creations, also impede follow-on innovation requiring the use of pre-existing, protected material. Hence, there is a delicate balance inherent in all IP protection regimes' (Senftleben 2011: 4). Nowadays, considering the vertiginous technological development and the frequent inability of the law to keep up with it, this statement makes even more sense. The legal regime applicable to patents must accommodate the growing demands for more knowledge in digital societies. Those demands might force a flexibilization of patent law and patent rights.

Finally, the costs. Innovation has a cost and the more complex and innovative (and thus, the more precise) the FR system is, the higher the price. Unless the company manages to internalise the cost, the financial burden might be unbearable and some companies might prefer to take the risk of using a 'not so good' technology, hoping that no negative outcome will take place. A possible solution could be the establish a technological partnership with the developing institution (a university or a private company) to have access to such technology at a lower price, an option especially adequate for public bodies or other non-profit entities. In exchange, the user (i.e., the receiver) of this technology could be available for trials to identify drawbacks of the

technology or marketing actions. Traditionally, this kind of partnership takes place to develop new technologies, joining the resources (human, capital) of two or more companies (Santangelo 2000), but they can also be employed to facilitate access to new technologies (Water 2022).

From a policy perspective, economic incentives could be provided to companies willing to acquire developed–and thus expensive–technology, such as tax benefits and/or other financial incentives. These policies are already in place in several countries (OCDE 2018). It remains to be seen, however, if national governments see the investment in more secure and precise FT systems as a type of R&D investment able to justify the use of such incentives.

4.2 Ethical dimension

The technical dimension of FR should also incorporate ethical concerns, following the principles of good governance (Mökander et al. 2022), that embrace ethics, policy, and law (Raposo 2023). This is crucial to building systems that are ethical by design, that is, whose source codes take into consideration ethical dimensions (Renaissance Numérique 2020). In this regard, Mutale Nkonde (2020, 33) refers to the 'design justice framework', a theory first developed by Sascha Costanza-Chock (2020). The design justice framework centres the AI design on the groups that suffer particular negative effects of its use—in this case, as a result of the higher percentage of error—and develops around that focus.

Racial biases are a major ethical concern. Despite improvements in this domain, there are still some disconformities in results. Therefore, the FR system must be based on data that matches end-user characteristics (World Economic Forum 2020), meaning that different parts of the world might have to resort to different databases to train their FR systems, according to the ethical composition of the respective population (see next section). A plan to identify potential biases during the testing phase and correct them must be in place. In a 2020 White Paper on FR, the World Economic Forum (2020) recommended to entities using this technology an impact assessment regarding the possible biases and their impact on the targeted individuals. The assessment would involve, among other measures, the identification of the possible discrimination risks and the way the organization deals with them and mitigates their effects.

Transparency is also part of the ethical (and legal) measures to consider. Information shall be provided to the team using this technology and to the ones affected by it, covering the way the FR system operates, the principles of good governance that have been adopted, the possible negative outcomes, and the expected percentage of false positives and false negatives. This shall be communicated in a language that is accessible to laypeople to guarantee comprehension (World Economic Forum 2020).

Data protection is another stringent concern.[10] It seems to raise a different set of issues than those related to inaccurate recognitions, but the principle of data accuracy (Article 5(1)(d) GDPR) connects them both. Personal data shall be accurate and when an incorrect identification is tagged to a biometric template this principle is compromised, leading to a violation of the GDPR.

4.3 Quantity and quality of images used

As one of the causes of 'algorithm bias' (Borgesius 2018) relates to the relevant datasets. Therefore, its resolution demands an improvement in the quality and quantity of the images used (Veale and Binns 2017). FR training requires a large amount of data in the form of photos. Different types of AI require different amounts of data, and 'deep learning' (the kind of AI that now forms the basis of FR) requires more data than other types of AI (Office for Product Safety and Standards 2021). As mentioned above, a persistent problem in FR is the predominance of white male images in such data, leading FR to present very high inaccuracy rates for females and people of other ethnic groups (European Union Agency for Fundamental Rights 2019). Solving the problem of the lack of diversity in the data would solve a substantial part of the FR accuracy problem.

It is worth noting that FR is not intended to discriminate against ethnic minorities (e.g., Asians, and African Americans in the West). However, when groups are underrepresented in the community, they are likely to also be underrepresented in the databases, exactly as other minorities are in other geographical areas. For instance, there is evidence that an FR system developed in Asia is less able to recognise white people than Asians (Lunter 2020) and, therefore, is biased against whites.

The quality of the images (both those used for training and testing and those stored in the database) also affects the accuracy of the technique. Several factors must be considered: differences in hair and skin colour[11]; differences between compared images, including ageing and other individual transformations, and emotions; lighting conditions; camera distance; background; head orientation; and the size of the face in the image (European Union Agency for

[10] Privacy and self-determination are also at stake. Apart from the cases in which FR technology in embodied in devices that are voluntary used (an iPhone, for instance), its use by law enforcement authorities is a legal challenge. The same is valid when FR is employed by other public or private entities in such a way that the data subject has no real power to decide whether to be 'scanned' or not (Raposo 2022a).

[11] 'Reflection of light affects the quality of facial images of very fair-skinned persons, and not enough light affects the quality for very dark-skinned persons' (European Union Agency for Fundamental Rights 2019, 27).

Fundamental Rights 2019). The images to train and test the system shall be collected in a condition similar to the one in which the FT cameras will operate. Another mitigation measure is the use of cropping photos, disregarding peoples' hair, to improve accuracy for the individuals that have their heads covered (a turban or a hijab) (World Economic Forum 2020).

Moreover, the use of FR differs significantly between a controlled environment, such as police stations and airports, in which the lighting and orientation of the subject are controlled, and a non-controlled environment, such as random images from CCTV cameras, especially live footage of people passing in a street (Crumpler 2020; European Union Agency for Fundamental Rights 2019; Harwell 2019). The law must distinguish between these two types of matches, for instance, by demanding additional methods of identification for non-controlled environments, such as other forms of biometric identification (e.g., fingerprints, DNA), before taking any action (Renaissance Numérique 2020). Preferably, images should be taken in controlled environments, where the accuracy of FR is very high (European Union Agency for Fundamental Rights 2019). A typical example of a controlled environment is that of the airport boarding gate, for which a NIST report from 2021 confirmed an accuracy rate of 99.5% (Grother et al. 2021).

The criteria regarding the quantity and quality of images used for FR should cover not only the training images but also the testing images. The reason that some tests fail to detect bias is that the data used to test the system are usually as biased as those used for training because data scientists generally divide a single set of data into two parts and use one part for training and the other for testing. When the same dataset is used for training and testing, testing cannot detect a problem with bias (Hao 2019).

If FR technology manages to solve the problem of misidentifying people from less common ethnicities or females, FR may become less discriminatory than human-operated methods of identification. Humans (e.g., law enforcement agents) are frequently accused of discrimination and even racism (Schwartz 2020), whereas properly working FR software should be free of such biases.

The data used to train FR algorithms must be accurate and up-to-date to avoid not only liability due to misidentification but also liabilities related to the violation of data processing norms (Information Commissioner's Office 2019; European Data Protection Board 2022).Although this perspective is not developed in this paper, it should be noted that the principle of accuracy[12] is stated in Article 5(1)(d) of the Law Enforcement Directive (European Parliament and Council 2016a), Article 5(1)(d) of the General Data Protection Regulation (European Parliament and Council 2016b), and Article 5/3/d of the Convention 108 (Council of Europe 1981).

4.4 Constant monitoring

FR must be subject to constant monitoring and sporadic auditing to analyse the performance of its algorithms, immediately detect biases and other failures, and amend them (Renaissance Numérique 2020). Towards this aim, the law should require that records be kept of the programming of the algorithm, the training methodologies applied to the AI system and the data used to train that system.

In the US, the 2022 Algorithmic Accountability Act, in its Sect. 3(1), imposes an obligation on companies to regularly assess the accuracy of their FR algorithms, by performing an impact assessment (Senate and House of Representatives 2019). In the European Union, the Draft Proposal on AI (European Commission 2021b) proposes a post-commercialisation monitoring mechanism, which involves two main dimensions: AI systems providers are required to report serious incidents and malfunctions of their systems (Article 62 of the AI Draft Regulation); and market surveillance authorities are asked to control the AI systems already put on the market, assumingly also to control their accuracy (Article 63 of the AI Draft Regulation).

It has also been suggested that vendors should make their datasets and algorithms available to the general public, to permit independent auditing (Ho et al. 2020). However, the feasibility of this proposal is questionable for reasons related to IP rights. The novelty of this technology will most likely hamper public disclosure of this material, to prevent business rivals to develop competing products.

Likewise, in its 2022 guidelines on FR in law enforcement, the European Data Protection Board (2022) recommended certain measures to prevent inaccurate results, which, although specifically addressing data protection issues, are generalisable to other sources of legal liability. Specifically, the European Data Protection Board recommended the implementation of procedures to supervise algorithmic accuracy, such as logging and reporting mistakes (European Data Protection Board 2022). When a problem is identified (e.g., data poising, spoofing), measures should be immediately taken to address it. If necessary, the algorithm must be retrained, using an enriched version of the same dataset or an entirely new one (European Data Protection Board 2022).

Monitoring is only efficient if a maximum limit of error is established, above which the FR system must be cancelled or at least revised. A problem that remains to be settled is how to determine a reasonable level of accuracy to establish such an error limit. An intuitive answer is that FR should

[12] The principle of accuracy refers to the obligation to use error free data and to immediately correct any mistakes detected on the data.

only be allowed when it causes the same number as, or fewer errors than, human recognisers in the same setting (Sarabdeen 2022). However, it must be clarified whether this refers to average humans or so-called 'super-recognisers', who are particularly good at remembering faces and recognising them in crowds and who account for only one to two per cent of the population. This would obviously constitute a much more demanding comparative criterion to establish the limit of error.

5 Preliminary conclusions

Despite recent advances in technology, FR still presents flaws in its identification abilities. This is a weakness that manufacturers and potential uses of FR must consider and address, as the most common types of FR misidentifications can result in legal discrimination and lawsuits against these actors.

Most of the flaws associated with FR, however, can be controlled and even amended. Recent studies analysing the development of this technology reveal improvements in FR. When using modern software trained with a panoply of images, FR accuracy goes up to 99% (McLaughlin, Castro 2020). In the future, it may even become the standard of care for identification. At present, however, prudence is recommended, and appropriate measures and safeguards should be put in place.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Funding Open access funding provided by FCT|FCCN (b-on).

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The author has no financial, personal, academic, or other conflicts of interest in the subject matter discussed in this manuscript.

References

Apple (2021) About face ID advanced technology. https://support.apple.com/en-us/HT208108. Accessed 3 Jul 2021

Apple (2022) About Face ID advanced technology. https://support.apple.com/en-us/HT208108#:~:text=The%20probability%20that%20a%20random,you're%20wearing%20a%20mask. Accessed 27 Jul 2022

Bambauer J (2021) Facial recognition as a less bad option, Aegis Series Paper No. 2107. https://www.hoover.org/sites/default/files/research/docs/bambauer_webreadypdf.pdf. Accessed 22 Jun 2022

Basenko R, Hennadii A, Dmytro S (2022) Institute of compensation for moral damage: international legal experience and legislative innovations. Entrepreneurship, Econ Law 1:5–10. https://doi.org/10.32849/2663-5313/2022.1.01

Bathaee Y (2018) The artificial intelligence black box and the failure of intent and causation. Harvard J Law Technol 31(2):890–938

Ben-Shahar O, Porat A (2018) The restoration remedy in private law. Columbia Law Rev 118(6):1901–1952

Boyd J, Ingberman D (1995) Should 'state of the art' safety be a defense against liability?, Discussion Papers dp-96–01, Resources For the Future

Borgesius FZ (2018) Discrimination, artificial intelligence, and algorithmic decision-making. Council of Europe, Strasbourg

Boussaad L, Boucetta A (2020) Deep-learning based descriptors in application to ageing problem in face recognition. J King Saud Univ—Computer Inf Sci. https://doi.org/10.1016/j.jksuci.2020.10.002

Buolamwini J, Gebru T (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. Proceed Machine Learn Res 81:1–15

Castelvecchi D (2020) Is facial recognition too biased to be let loose? Nature 587:347–349. https://doi.org/10.1038/d41586-020-03186-4

Chattopadhyay S, Nelson N, Au A et al (2020) A tale from the trenches: cognitive biases and software development. Assoc Computing Machinery. https://doi.org/10.1145/33778113380330

Cofone IN (2019) Algorithmic discrimination is an information problem. Hastings L.J. 79:1389–1444

Commission Nationale de l'Informatique et des Libertés (CNIL) (2019) Reconnaissance faciale - Pour un debat à la hauteur des enjeux, https://www.cnil.fr/fr/reconnaissance-faciale-pour-un-debat-la-hauteur-des-enjeux. Accessed 22 Jun 2022

Costanza-Chock S (2020) Design justice: community-led practices to build the worlds we need. MIT Press, Cambridge, MA

Council of Europe (1981) Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108). https://rm.coe.int/1680078b37. Accessed 28 Jan 2022

Council of the European Communities (1985) Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective product. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31985L0374. Accessed 9 May 2022

Cowgill B, Dell'Acqua F, Deng S et al. (2020) Biased programmers? Or biased data? A field experiment in operationalizing AI ethics. Proceedings of the 21st ACM Conference on Economics and Computation 679–681. https://doi.org/10.2139/ssrn.3615404

Dent C, Jensen P, Waller S, et al. (2006) STI Working Paper 2006/2: Research use of patented knowledge: A review, OECD - Organisation for Economic Co-Operation and Development.

Donald SJ (2019) Don't blame the AI, it's the humans who are biased. Towards Data Science. https://towardsdatascience.com/dont-blame-the-ai-it-s-the-humans-who-are-biased-d01a3b876d58. Accessed 2 May 2022

Dupr D, Krumhuber EG, Küster D, McKeown GJ (2020) A performance comparison of eight commercially available automatic classifiers for facial affect recognition. PLoS ONE 15:4. https://doi.org/10.1371/journal.pone.0231968

European Commission (2020) White paper on artificial intelligence - A European approach to excellence and trust. Brussels, 19.2.2020 COM(2020) 65 final. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf. Accessed 15 July 2022

European Commission (2021a) Inception Impact Assessment, Ref. Ares(2021a)4266516—30/06/2021a. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12979-Civil-liability-adapting-liability-rules-to-the-digital-age-and-artificial-intelligence_en. Accessed 17 Jul 2022

European Commission (2021b) Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021b/206 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021bPC0206. Accessed 3 May 2022

European Commission (2022a) Proposal for a directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), Brussels, 28.9.2022a COM(2022a) 496 final 2022a/0303 (COD)

European Commission (2022b) Proposal for a directive of the European Parliament and of the Council on liability for defective product, Brussels, 28.9.2022b COM(2022b) 495 final 2022b/0302 (COD)

European Data Protection Board, Guidelines (2022) 05/2022 on the use of facial recognition technology in the area of law enforcement, Version 1.0, Adopted on 12 May 2022. https://edpb.europa.eu/system/files/2022-05/edpb-guidelines_202205_frtlawenforcement_en_1.pdf. Accessed 1 Jul 2022

European Digital Rights (EDRi) (2019) Facial recognition and fundamental rights. https://edri.org/our-work/facial-recognition-and-fundamental-rights-101/. Accessed 1 Jul 2022

European Parliament (2021) Regulating facial recognition in the EU, EPRS - European Parliamentary Research Service, https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/698021/EPRS_IDA(2021)698021_EN.pdf. Accessed 1 Nov 2022

European Parliament and Council (2016a) Directive (EU) 2016a/680 of the European Parliament and of the Council of 27 April 2016a on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA. Accessed 28 Jan 2022

European Parliament and Council (2016b) Regulation (EU) 2016b/679 of the European Parliament and of the Council of 27 April 2016b on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Accessed 28 Jan 2022

European Union Agency for Fundamental Rights (2018). #BigData. Discrimination in data-supported decision making. Luxembourg Publications Office, Luxembourg

European Union Agency for Fundamental Rights (2019). Facial recognition technology: fundamental rights considerations in the context of law enforcement. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf. Accessed 4 Jul 2022

Fu X (2021) Design of facial recognition system based on visual communication effect. Computational Intell Neurosci. https://doi.org/10.1155/2021/1539596

Glusac E (2022) What you need to know about facial recognition at airports. The New York Times. https://www.nytimes.com/2022/02/26/travel/facial-recognition-airports-customs.html. Accessed 27 Jul 2022

Grother P, Ngan M, Hanaoka K (2019) Face recognition vendor test (Frvt), Nat'l Inst of Standardsm & Tech., Part 3: demographic effects 1. https://doi.org/10.6028/NIST.IR.8280

Grother P, Ngan M, Hanaoka K (2018) Information access division information technology laboratory, Ongoing face recognition vendor test (FRVT) Part 1: Verification,https://www.nist.gov/system/files/documents/2018/02/15/frvt_report_2018_02_15.pdf

Grother P, Hom A, Ngan M, Hanaoka K (2021) Information access division information technology laboratory, Face recognition vendor test (FRVT) Part 7: Identification for paperless travel andimmigration, https://doi.org/10.6028/NIST.IR.8381

Hancock JB, Somai RS, Mileva VR (2020) Convolutional neural net face recognition works in non-human-like ways. R Soc Open Sci 7:200595200595. https://doi.org/10.1098/rsos.200595

Hao K (2019) This is how AI bias really happens—and why it's so hard to fix. MIT technology review. https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/. Accessed 29 May 2022

Hariri W (2022) Efficient masked face recognition method during the COVID-19 pandemic. SIViP 16:605–612. https://doi.org/10.1007/s11760-021-02050-w

Harwell D (2019) A face-scanning algorithm increasingly decides whether you deserve the job. The Washington Post. https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/. Accessed 10 Jul 2022

Heathrow airport (nd) Facts and Figures, https://www.heathrow.com/company/about-heathrow/company-information/facts-and-figures. Accessed 4 Jan 2022

Ho DE, Black E, Agrawala M et al. (2020) How regulators can get facial recognition technology right, Brookings. https://www.brookings.edu/techstream/how-regulators-can-get-facial-recognition-technology-right/. Accessed 16 Jul 2022

Infobae (2019) Un hombre estuvo seis días preso por un error policial, Infobae. https://www.infobae.com/sociedad/policiales/2019/08/02/un-hombre-estuvo-seis-dias-preso-por-un-error-del-sistema-de-reconocimiento-facial/. Accessed 17 Jul 2022

Information Commissioner's Office (2019) Information Commissioner's Opinion: The use of live facial recognition technology by law enforcement in public places, Reference: 2019/01. https://jerseyoic.org/media/moqjayy1/live-frt-law-enforcement-opinion-20191031.pdf. Accessed 17 Jul 2022

Institute and International Association of Chiefs of Police (IACP) (2019) Law Enforcement - Facial Recognition Use Case Catalog. https://www.theiacp.org/resources/document/law-enforcement-facial-recognition-use-case-catalog. Accessed 22 June 2022.

Jesdanun A (2017) Apple's face ID technology can learn, but it takes time. Las Vegas Review-Journal. https://www.reviewjournal.com/news/science-and-technology/apples-face-id-technology-can-learn-but-it-takes-time/. Accessed 22 Jun 2022

Kindt E (2013) Privacy and data protection issues of biometric application. Springer

Klare BF, Burge MJ, Klontz JC et al (2012) Face recognition performance: role of demographic information. IEEE Trans Inf Forensics Secur 7(6):1789–2180. https://doi.org/10.1109/TIFS.2012.2214212

Leong B (2019) Facial recognition and the future of privacy: I always feel like somebody's watching me. Bullet Atomic Scientists 75(3):109–115. https://doi.org/10.1080/00963402.2019.1604886

Loyola-González O (2019) Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. IEEE Access 7:154096–154113. https://doi.org/10.1109/ACCESS.2019.2949286

Lunter J (2020) Beating the bias in facial recognition technology. Biometric Technol Today 2020(9):5–7. https://doi.org/10.1016/S0969-4765(20)30122-3

Maciel HS, Cardoso I, Silva D et al. (2016) An embedded access control system for restricted areas in smart buildings. 2016 International Multidisciplinary Conference on Computer and Energy Science (SpliTech) 1–6. https://doi.org/10.1109/SpliTech.2016.7555926

McLaughlin M, Castro D (2020) The critics were wrong: NIST data shows the best facial recognition algorithms are neither racist nor sexist. ITIF, https://itif.org/publications/2020/01/27/critics-were-wrong-nist-data-shows-best-facial-recognition-algorithms. Accessed 5 May 2022

Mökander J, Juneja P, Watson DS et al. (2022) The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: what can they learn from each other? Minds & Machines, https://doi.org/10.1007/s11023-022-09612-y

Najibi A (2020) Racial discrimination in face recognition technology. Harvard GSAS Science Policy Group. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/. Accessed 15 Jul 2021

Neuwirth RJ (2022) The EU artificial intelligence act book regulating subliminal AI systems. Taylor and Francis

Nkonde M (2020) Automated anti-blackness: facial recognition in Brooklyn, New York. Harvard Kennedy School J African Am Policy 2019–2020:30–36

Noroozi F, Toygar Ö (2017) Recognition of identical twins using fusion of various facial feature extractors. J Image Video Proc. https://doi.org/10.1186/s13640-017-0231-0

OCDE (2018) OECD Review of national R&D tax incentives and estimates of R&D tax subsidy rates, https://www.oecd.org/sti/rd-tax-stats-design-subsidy.pdf. Accessed 12 Nov 2022

Office for Product Safety and Standards (2021) Study on the impact of artificial intelligence on product safety final report (released on 23 May 2022). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1077630/impact-of-ai-on-product-safety.pdf. Accessed 12 Jul 2022

Office of the High Commissioner for Human Rights (OHCHR) (2021) The right to privacy in the digital age: Report. https://www.ohchr.org/en/calls-for-input/calls-input/2021/right-privacy-digital-age-report-2021. Accessed 1 Jul 2022

Parviainen J, Rantala J (2022) Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. Med Health Care and Philos 25:61–71. https://doi.org/10.1007/s11019-021-10049-w

Radiya-Dixit E, Tramer F (2021) Data poisoning won't save you from facial recognition. ICML 2021 Workshop on Adversarial Machine Learning. https://arxiv.org/pdf/2106.14851.pdf

Raposo VL (2022a) (Do not) remember my face: uses of facial recognition technology in light of the general data protection regulation. Inf Commun Technol Law. https://doi.org/10.1080/13600834.2022.2054076

Raposo VL (2022b) The use of facial recognition technology by law enforcement in Europe: a non-orwellian draft proposal. Eur J Crim Policy Res. https://doi.org/10.1007/s10610-022-09512-y

Raposo VL (2023) Digital governance na proposta de regulamento da Comissão Europeia relativa à inteligência artificial: Breve périplo sobre good governance e direitos fundamentais.' In: Estudos em Homenagem Presidente Costa Andrade (forthcoming)

Renaissance Numerique (2020) Facial recognition: Embodying European values. https://www.renaissancenumerique.org/publications/facial-recognition-embodying-european-values. Accessed 17 Jul 2022

Roth L (2009) Looking at Shirley, the ultimate norm: colour balance, image technologies, and cognitive equity. Can J Commun 34(1):111–136

Santangelo GD (2000) Corporate strategic technological partnerships in the European information and communications technology industry. Res Policy 29(9):1015–1031

Sarabdeen J (2022) Protection of the rights of the individual when using facial recognition technology. Heliyon 8(3):e09086. Doi: https://doi.org/10.1016/j.heliyon.2022.e09086

Schwartz SA (2020) Police brutality and racism in America. Explore 16(5):280–282. https://doi.org/10.1016/j.explore.2020.06.010

Sharkey A (2017) Can robots be responsible moral agents? And why should we care? Connect Sci 29(3):210–216. https://doi.org/10.1080/09540091.2017.1313815

Senate and House of Representatives (2019) Algorithmic Accountability Act. https://www.congress.gov/bill/116th-congress/house-bill/2231/text. Accessed 1 Jul 2022

Senftleben M (2011) Chapter 8: Overprotection and protection overlaps in intellectual property law–the need for horizontal fair use defences. In: The structure of intellectual property law. Edward Elgar Publishing, Cheltenham UK

Serna I, Morales A, Fierrez J (2022) Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep learning. Artif Intell 305:103682. https://doi.org/10.1016/j.artint.2022.103682

Sharif M, Bhagavatula S, Bauer L et al. (2016) Accessorize to a crime: real and stealthy attacks on state-of-the-art face recognition. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security 1528–1540

Sharma J (2018) Metropolitan police's facial recognition technology 98% inaccurate, figures show. https://www.independent.co.uk/news/uk/home-news/met-police-facial-recognition-success-south-wales-trial-home-office-false-positive-a8345036.html. Accessed 15 Jul 2022

Smith GJ (2020) The politics of algorithmic governance in the black box city. Big Data Soc. https://doi.org/10.1177/2053951720933989

Smith M, Miller S (2022) The ethical application of biometric facial recognition technology. AI & Soc 37:167–175. https://doi.org/10.1007/s00146-021-01199-9

Thakkar D (2017) Top five biometrics (face, fingerprint, iris, palm and voice) modalities comparison. Bayometric. https://www.bayometric.com/biometrics-face-finger-iris-palm-voice/. Accessed 15 Jul 2021

The Alan Turing Institute (2019) Understanding bias in facial recognition technologies. https://www.turing.ac.uk/sites/default/files/2020-10/understanding_bias_in_facial_recognition_technology.pdf. Accessed 3 Jun 2022

Thompson KA (2020) Countenancing employment discrimination: Facial recognition in background checks. Tex. a&m l. Rev. 8(1):64–88. https://doi.org/10.37419/LR.V8.I1.2

UNESCO (2021) Recommendation on the ethics of artificial intelligence, https://unesdoc.unesco.org/ark:/48223/pf0000380455. Accessed 2 Nov 2022

United Nations Development Group (2017) Data privacy, ethics and protection guidance note on big data for achievement of the 2030 agenda, https://unsdg.un.org/sites/default/files/UNDG_BigData_final_web.pdf. Accessed 23 Oct 2022

US Department of Homeland Security, (2019) Transportation security administration and US Customs and border protection:

Deployment of biometric technologies report to Congress. https://www.tsa.gov/sites/default/files/biometricsreport.pdf#:~:text=This%20Report%20to%20Congress%20was%20compiled%20pursuant%20to,on%20October%205%2C%202018%2C%20which%20states%20in%20part%3A. Accessed 14 May 2022

US General Services Administration (GSA) (2022) Executive Order 13985—equity Action Plan, January 20, 2022. https://www.gsa.gov/cdnstatic/GSA_Equity_Action_Plan_2022.pdf. Accessed 17 Jul 2022

Umoja S (2018) Algorithms of oppression: How search engines reinforce racism. Noble NYU Press, 2018. 256 pp. Science. 2021 374(6567):542. https://doi.org/10.1126/science.abm5861

Veale M, Binns R (2017) Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. Big Data Soc. https://doi.org/10.1177/2053951717743530

Waelen RA (2022) The struggle for recognition in the age of facial recognition technology. AI Ethics. https://doi.org/10.1007/s43681-022-00146-8

Waters L (2022) Technology partnerships: What they look like and why they're important, Hubspot, https://blog.hubspot.com/sales/technology-partnerships. Accessed 11 Nov 2022

Welinder Y, Palmer A (2018) Face recognition, real-time identification, and beyond. In: Selinger E, Polonetsky J, Tene O (eds) The Cambridge handbook of consumer privacy. Cambridge University Press, Cambridge, pp 102–124. https://doi.org/10.1017/9781316831960.006

White D, Dunn JD, Schmid AC et al (2015) Error rates in users of automatic face recognition software. PLoS ONE. https://doi.org/10.1371/journal.pone.0139827

William C (2020) How Accurate are facial recognition systems—and why does it matter?, Centre for Strategic & International Studies, https://www.csis.org/blogs/technology-policy-blog/how-accurate-are-facial-recognition-systems-%E2%80%93-and-why-does-it-matter. Accessed 28 Oct 2022

World Economic Forum (2020) A Framework for responsible limits on facial recognition use case: flow management, White Paper, https://www3.weforum.org/docs/WEF_Framework_for_action_Facial_recognition_2020.pdf. Accessed 1 Nov 2022

Wright E (2019) The future of facial recognition is not fully known: Developing privacy and security regulatory mechanisms for facial recognition in the retail sector. Fordham Intell Prop Media & Ent L.J. 29:611–685