KALLENY DOS SANTOS TAVARES

BIASED FACIAL RECOGNITION TECHNOLOGIES IN THE REALM OF CORPORATE LIABILITY

Dissertation to obtain a Master's Degree in
Law, in the specialty of Business and Tech

Supervisor:
Dr. Miguel de Azevedo Moura, Professor of the NOVA School of Law

September, 2023

Declaração antiplágio

Declaro por minha honra que o trabalho que apresento é original e que todas as minhas citações estão corretamente identificadas. Tenho consciência de que a utilização de elementos alheios não identificados constitui uma grave falta ética e disciplinar.

(Nome e assinatura do/a aluno/a)

Anti plagiarism statement

I hereby declare that the work I present is my own work and that all my citations are correctly acknowledged. I am aware that the use of unacknowledged extraneous materials and sources constitutes a serious ethical and disciplinary offence.

(Student’s name and signature)

Acknowledgments

With sincere gratitude, I would like to thank my parents, Márcia Monte, and Alderico Tavares, and my sister, Maria Tavares. Their unconditional support and encouragement have been essential throughout this journey. Their love and belief in me have been my constant motivation. I'd also like to extend my thanks to my friends in Portugal, who have become the family I've chosen. Your friendship has enriched not only my academic journey but also made me feel at home. Finally, I would like to acknowledge the fundamental guidance of my supervisor, Miguel de Azevedo Moura.

NUMBER OF CHARACTERS

I declare that this dissertation's body, including spaces and footnotes, is composed of 116.901 characters.

NÚMERO DE CARACTERES

Declaro que o corpo desta dissertação, incluindo espaços e notas, ocupa um total de 116.901 caracteres.

INDEX

ABSTRACT

This dissertation explores algorithmic bias in facial recognition technology within the context of liability lawsuits. The focus is on instances where facial recognition fails to accurately authenticate and identify certain individuals. The research aims to uncover the causes of algorithmic bias and propose potential solutions. Additionally, it addresses the challenges of seeking compensation in liability cases involving misidentification due to bias. As emerging technologies challenge traditional liability laws, the question of accountability for biased facial recognition technology arises. The analysis also includes an examination of the impact of European non-discrimination law. This investigation analyzes European and national law and articles written by respected scholars and legal experts. It is concluded that creating more diverse datasets and involving underrepresented groups in technology development is crucial to mitigate bias in FRT. Moreover, while Europe's efforts to modernize laws are apparent, several gaps impede access to compensation in liability lawsuits.

Keywords: Facial recognition technology, Bias, Artificial Intelligence, Liability.

RESUMO

Esta dissertação aborda o viés da tecnologia de reconhecimento facial no contexto de ações de responsabilidade. O intuito desta investigação é analisar os casos em que o reconhecimento facial não consegue autenticar e identificar determinados indivíduos, reconhecer as causas e propor potenciais soluções. Além disso, analisa os desafios relacionados à busca de indenização em casos de responsabilidade civil. As novas tecnologias desafiam as leis tradicionais de responsabilidade, fazendo a questão da responsabilidade pela tecnologia de reconhecimento facial enviesada ganhar maior destaque nas discussões. A análise também engloba um estudo do impacto da legislação europeia de não-discriminação. Esta pesquisa analisa tanto a legislação europeia como a

nacional de países membros, bem como artigos escritos por estudiosos e jurídicos renomados. Com isso, entende-se que a necessidade de criar de uma base de dados mais inclusiva para grupos minoritários é crucial para mitigar o viés da FRT. Além disso, embora os esforços da Europa para modernizar as leis sejam evidentes, várias lacunas impedem o acesso à indenização em processos de responsabilidade civil.

Palavras-chaves: Reconhecimento Facial, Viés da tecnologia, Inteligência Artificial, Responsabilidade Civil.

I - INTRODUCTION

1.1. Initial considerations

Facial recognition technology (FRT) is central to many discussions nowadays, raising questions regarding its implications, accountability, and ethical considerations. Its use has become highly demanded during the last decades. It is more accessible to the entire population, either for simple activities like unlocking a phone or for law enforcement purposes. Over time, the technology's accuracy has improved, and error rates are lower (Grother *et al.*, 2019a, p. 6). It has captured media attention and found diverse uses, such as speeding up school queues in the U.K. and enabling subway payments in Moscow. The predictions suggest that the facial recognition market could reach over $10.15 million by 2025 (Staffer, 2021).

According to the Article 29 Data Protection Working Party (2012, p. 2), face recognition "is the automatic processing of digital images which contain the faces of individuals for the purpose of identification, authentication/verification or categorisation of those individuals". As FRT becomes increasingly prevalent, its impact on fundamental rights and its potential to perpetuate bias has been debated.

This research focuses on cases where biased FRT does not work when authenticating and identifying particular demographic groups in the realm of liability lawsuits. This requires an examination of the data-related aspects of ethical and technical issues. In this matter, the study aims to discuss algorithmic bias within and suggest possible tactics to mitigate it.

Moreover, this investigation delves into the complexity of seeking compensation in liability cases arising from misidentification due to biased technology in the context of European law. As Artificial Intelligence (AI) systems advance, the legal landscape grapples with questions of liability and responsibility. This new era of AI systems underscores the necessity to adapt and discuss the liability framework. Thus, this research will address the gap in the European legal framework when addressing biases in FRT.

Moral damages are investigated in the realm of liability lawsuits, illuminating

potential compensatory mechanisms for individuals affected by biased FRT. Moreover, the study delves into the liability of stakeholders involved in the development chain of FRT. In order to discuss this problem, several aspects need to be analyzed, among them the role of European non-discrimination law as a means to mitigate biases.

The intersection between the need for transparency in addressing biases and protecting trade secrets raises pertinent questions about accountability and the balance between innovation. To address this issue, European law, National law, and articles written by respected scholars and legal experts were analyzed.

The current landscape distinctively shows the relevance of this dissertation. Currently, the research on FRT has substantial work in the U.S., while it remains at an early stage in Europe. Despite the growing interest, academic exploration concerning the liability implications of biased FRT still needs to be expanded.

Despite efforts to regulate emerging technology, academic debate needs to be stimulated. By addressing the gap, this research hopes to advance the continuing discussion about biased FRT and liability issues by guiding the conversation toward balancing accountability with the practical aspects of real-world implementations.

1.2. Definitions

This topic initially explores the concepts essential for the discussion of biased facial recognition in corporate liability. By exploring these, it can establish a solid foundation for understanding the complexities of this research. These concepts are artificial intelligence, algorithms, and machine learning.

1.2.1. - Artificial Intelligence (AI)

AI has many concepts, some more focused on a human approach and others on more rational thinking (Stuart *et al.*, 2010, p. 1).

To be more precise, some define AI as a machine with capabilities deemed intelligent according to the standards set by human intelligence. In other words, "AI is about creating intelligent machines that think or (re)act like humans" (Coeckelbergh, 2020, p. 64). On the other hand, Margaret A. Boden proposes a different concept that relates AI to intelligence *per se* (Coeckelbergh, 2020, p. 64). According to this researcher, AI "seeks to make computers do the sorts of things that minds can do" (Boden, 2016, p.1).

More precisely, Boden (2016, p.1) argues that some of these functions, such as reasoning, are frequently referred to as 'intelligent,' others, such as vision, may not always be so identified. These activities call on psychological skills that both people and animals use to accomplish their objectives, including perception, prediction, planning, and motor control. "Intelligence isn't a single dimension, but a richly structured space of diverse information processing capacities. Accordingly, AI uses many different techniques, addressing many different tasks".

Both approaches are very significant for the guidelines of this research. For instance, if an AI is designed to (re)act or think like humans, it will inherently inherit the biases in human decision-making processes. Personal biases, societal norms, and cultural beliefs shape human intelligence.

In contrast, Boden's (2016, p.1) perspective highlights that intelligence is not merely a single factor but involves several dimensions of mental abilities and tasks. The other equally essential skills can be overlooked or underestimated when AI systems are trained and evaluated solely based on specific tasks traditionally associated with human intelligence. This could result in a biased evaluation of the abilities and performance of the AI.

Recognizing that AI cannot be equated solely with humanlike intelligence is essential to combating bias in AI. Developers and researchers should strive for a comprehensive understanding of intelligence, considering diverse information-processing capacities and avoiding replicating harmful biases.

1.2.2 - Algorithm

According to Cormen, Leiserson, Rivest and Stein (2009, p. 5) an algorithm is defined as “any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output [...] thus a sequence of computational steps that transform the input into the output”. Algorithms are therefore a set of actions taken to complete a specific goal (Barocas *et al.*, 2014, p. 3). The following sections will develop an explanatory overview of how facial recognition algorithms work.

1.2.3 - Machine Learning

The ability of software to learn is referred to as ‘machine learning’. Although some argue that this learning lacks actual cognition, unique to humans, it is rooted in statistical principles, making it a process based on statistics (Coeckelbergh, 2020, p. 84).

Machine learning can be used for many tasks and commonly involves recognizing patterns (Coeckelbergh, 2020, p. 84). In this sense, it mainly works as a form of learning through example — the machine learning algorithm is present to several examples from a database, and it extracts patterns from them, thus being able to work in cases beyond the examples provided (Barocas *et al.*, 2014, p. 4).

An illustration of this is when programmers train an algorithm to recognize photographs of cats; they do so without providing the computer with predefined guidelines for identifying what constitutes a cat. Instead, the algorithm independently constructs its model of cat photos. It will distinguish the features accessories to cat photographs and those without a cat by evaluating which specific details from the examples consistently separate cats from other entities (Barocas *et al.*, 2014, p. 4).

Thus, the computer takes on a different role by autonomously generating models that align with the data. Unlike traditional approaches, the starting point in machine learning is the data itself, guiding the decision-making process and

determining the next steps (Coeckelbergh, 2020, p. 84). In the realm of face recognition algorithms, it is the tool that allows it to recognize faces (Sharma *et al.*, 2020, p. 1163).

II - FACIAL RECOGNITION TECHNOLOGIES

2.1. A historical overview

Nowadays, face recognition is one of the most common methods of biometrics identification (Mallon, 2003, p. 957). During the last few years, its use has developed quickly and revolutionized how individuals and technology interact. It is present in the everyday life of citizens, from the moment someone unlocks their cell phone to assist law enforcement officers, to enter buildings, or even to find a missing person.

A face recognition system is a computer program that uses specific facial traits and a facial database to automatically identify or confirm a person from a video frame or digital image (Thorat, 2010, p. 325). This biometric technology uses automated algorithms based on a person's physiological characteristics (Tolba, 2014, p. 88). It is similar to other biometrics-identifying techniques, such as fingerprint, and it is frequently associated with secure applications and systems (Thorat, 2010, p. 325).

To understand the workings of facial recognition technology, Kaur *et al.* (2020, p. 2) outline five basic steps: (i) first, the image is captured by a camera with or without the knowledge of the subject — this image capture can also be known as probe image; (ii) secondly there is a face detection from the entire picture captured; (iii) third it happens the feature extractions, in this phase the unique and specific features are collected to match the images in the database — a face template is generated; (iv) fourth there is the matching of the face template generated with the images in the database; and (v) fifth is the verification or identification of the person.

Even though it’s easy for humans to identify faces in a crowd, machine recognition is a much more difficult operation (Chellappa *et al.*, 1995, p. 705). The pioneers of this technology were Woodrow Wilson Bledsoe, Charles Bisson and Helen Chan Wolf during the 1960s — they worked on a computer program that was

able to identify human faces with the financial support of an unknown intelligence agency, thus not much information was released (Thorat, 2010, p. 325).

Two reports, both written by Woodrow Wilson Bledsoe for Panoramic Research Incorporated, were published as the firsts to be the attempt to develop this technology. The first one was “A proposal for a study to determine the feasibility of a simplified face recognition machine” on January 30th, 1963 had the scope of developing a machine that solves a simplified face recognition problem using high resolution and single view pictures. The second one was entitled “Face Recognition Project Report” dated March 6th 1964, this one aimed to increase project goals to address the more challenging facial recognition issue (Bledsoe, 1963).

Bledsoe developed a system that could identify an unknown face by comparing it to data points from previously uploaded pictures. Despite the fact that it was in its infancy, the idea had great value and even aroused a lot of interest, especially of law enforcement agents (Klosowski, 2020). In the early and mid 1970s, typical pattern classification algorithms were employed. Despite the stagnation of research on FRT in the 1980s, interest in this technology increased significantly in the 1990s (Chellappa *et al.*, 1995, p. 705). According to Chellappa *at el*. (1995, p. 705), many factors contributed to this, such as the rising demand for surveillance applications and even the availability of hardware that allows the development of this technology.

In 1991, Matthew Turk and Alex Pentland (1991, p. 71) created a computer system that can monitor and locate a person's head, and then identify them by comparing their facial traits to those of known people. Upon implementing the Eigenfaces method, they discovered that the residual error could be applied to facial identification, making an automated facial recognition system possible, albeit there may be certain limitations due to the surroundings (Kaur *et al.*, 2020, p. 2).

The National Institute of Standards and Technology (NIST)[1] together with the Department of Defense (DoD) Counterdrug Technology Development Program Office, sponsored the Face Recognition Technology (FERET) program in September of 1993. The FERET program's scope was to develop automated facial recognition

[1] A non-regulatory government agency of the U.S. Department of Commerce's Technology Administration.

technology to help intelligence, security and law enforcement agents to carry out their duties (NIST, 2017).

This program established a large and independently collected facial image database of 1,199 individuals with a total of 14,126 photographs. Previous to this program, limited-size datasets had claimed strong recognition performance, however FERET offered a shared database and standard testing procedure that enabled direct quantitative assessments and comparisons of various methodologies (NIST, 2017).

The FERET was crucial to the expansion and advancement of the commercial facial recognition sector. It presented clear insight into the facial recognition state of the art at the time and a direction for further study. Furthermore, it made it easier to evaluate the benefits and drawbacks of various algorithms, revealing any technical issues that needed to be fixed (NIST, 2017). Since then, this technology has developed a lot and it has become increasingly prevalent in various industries.

As it gained popularity, this technology was introduced into the social networking sphere by Facebook in December of 2010, as a feature to help users save time. More specifically, the main goal was to ‘tag’ a user who appeared in their photo albums by simply clicking on them, connecting their accounts to the pictures. As a result, Facebook created one of the biggest collections of digital photo archives in the world (Kashmir and Mac, 2021).

Five years later, Facebook took its facial recognition algorithm to a new level. Led by Yann LeCun, Facebook’s artificial intelligence lab, developed an algorithm capable of recognizing individuals in photos, even when their faces are not visible — similar to humans. It was able to recognise individual people’s identities with 83% accuracy (Rutkin, 2015).

Another moment that highlighted the utilization of FRT occurred when the Tampa Police Department employed this technology at the Raymond James Stadium during Super Bowl XXXV on January 28, 2001, to identify potential terrorists and criminals (López, 2001). The face recognition system, FaceTrac, originally developed by the Massachusetts Institute of Technology (MIT), was facilitated by Graphco

Technologies and operated using software from Viisage Technology Inc (Grossman, 2001).

The discourse surrounding the application of FRT is garnering increasing attention, with extensive academic research and discussions taking place, particularly in the United States. In the European context, the controversial case involving Clearview AI served as a starting point for raising awareness and initiating debates (Ragazzi *et al*., 2021, p. 8).

In January 2020, a press article entitled “The Secretive Company That Might End Privacy as We Know It” was published by Kashmir Hill. Clearview AI developed a facial recognition application that consists in taking a picture of a person and after uploading it, all the public photos of that individual becomes available, as well as the link to where you can find them. According to this company, the database was composed of YouTube, Vimeo, Facebook and dozens of other sites. It is estimated that they had about three billion images (Kashmir, 2020).

After a month the client list of Clearview AI was leaked by BuzzFeed News and included over 2,200 clients, of which 26 countries were outside the United States and mostly within the European Union (Mac *et al.*, 2020). The discovery of this has led to a significant opposition against the development of this technology by companies, such as Clearview AI, and especially its use by law enforcement agencies to universities (Ragazzi *et al*., 2021, p. 8)[2].

Although the companies developing FRT promise a high accuracy rate, this technology does not work the same way for all individuals. This research will discuss what can lead to algorithm bias and will examine this issue from the liability perspective.

[2] See more: The Cube. (2021). Facial recognition: Clearview AI breaks EU data privacy rules, says French watchdog. Euro News .https://www.euronews.com/my-europe/2021/12/16/facial-recognition-clearview-ai-breaks-eu-data-privacy-rules-says-french-watchdog; Jasserand, C. (2022, May 5). Clearview AI: illegally collecting and selling our faces in total impunity? (Part II). CitTip. https://www.law.kuleuven.be/citip/blog/clearview-ai-illegally-collecting-and-selling-our-faces-in-total-impunity-part-ii/; DW. (2021). Privacy activists challenge Clearview AI in EU. https://learngerman.dw.com/en/privacy-activists-challenge-clearview-ai-in-eu/a-57691756;

2.2. Authentication and identification

Facial recognition systems employ artificial intelligence techniques to identify and recognize faces within images, video or still format by utilizing people's face image that are biometric data (Fernandez *et al.*, 2020, p.5). This biometric data has different qualities such as uniqueness, immutability, and difficulties in concealment. Compared to other types of biometric identification such as fingerprints or DNA, obtaining facial photos is more easy and simple (European Union Agency for Fundamental Rights, 2019, p. 5).

According to the Article 4(14) of the General Data Protection Regulation (GDPR) (EU, 2016), biometric data is defined 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".

In the EU, the processing of biometric data for uniquely identifying someone is not allowed under the GDPR (Article 9(1))[3]. However, there are exceptions that can be applied (Article 9 (2)). For example, processing biometric data may be permitted if the data subject gives explicit consent or if it is necessary for authentication or security purposes, provided that there is a significant public interest involved (KPMG, 2021, p. 2).

According to the Recital 51, photographs will be considered biometric data "only when processed through a specific technical means allowing the unique identification or authentication of a natural person" (EU, 2016). Due to their sensitive nature, they are categorized as special categories of personal data or sensitive data. Thus, EU data protection law offers heightened protection and additional safeguards distinguishing it from other types of personal data (European Union Agency for Fundamental Rights, 2019, p. 5).

[3] Article 9(1) GDPR: "*Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited*".

FRT is extensively utilized in the private and public sectors. Can be utilized for advertising and targeted marketing in the private sector[4], Or even filters in social media[5], mobile phones[6] and automatically tagging people on photos[7]. In the public sector for finding missing person[8], identifying suspects or solving crimes[9].

This research will only delve into cases in which FRT is employed for identification and verification/authentication purposes.

Authentication aims to confirm a person's identity. In order to do this, an authentication process using a 1:1 matching system is used, comparing the collected face features with a template that has been stored for confirmation (Sullivan, 2021, p. 2).

To illustrate this, consider a scenario where a user creates a profile for an application, such as a banking app. To set up the profile, the user must upload an identification document, take a selfie, and enter identification information. The user is requested to take a selfie whenever they access the app. The software then creates a biometric template from the collected image and matches it to the person's previously stored image. Thus, the system proceeds with the user's authentication and authorizes access to the application if there is a good match based on an accuracy score (Sullivan, 2021, p. 2).

Additionally, one-to-many technology (1:many) is used for face identification to match an unknown face from a photo, video, or security camera with a database of recognized faces. Is the system capable of correctly predicting this person's identity regarding identification? By comparing the person's facial traits to those in the database with known faces, the software determines whether a "match" exists. Face

---

[4] See more: Nasrullah, Q. (2022). Explainer: Are Australian retailers using facial recognition software on their unknowing customers? Cosmos. https://cosmosmagazine.com/technology/facial-recognition-technology-australian-retailers/.

[5] See more: Ruggeri, A. (2023). The problems with TikTok's controversial 'beauty filters'. BBC. https://www.bbc.com/future/article/20230301-the-problems-with-tiktoks-controversial-beauty-filters.

[6] See more: Kelion, L. (2017). Apple iPhone X adopts facial recognition and OLED screen. *BBC*. https://www.bbc.com/news/technology-41228126.

[7] See more: Simonite, T. (2017). Facebook Can Now Find Your Face, Even When It's Not Tagged. *Wired*. https://www.wired.com/story/facebook-will-find-your-face-even-when-its-not-tagged/.

[8] See more: Global Times. (2015). AI technology used to find missing child after 19 years. URL. https://www.globaltimes.cn/content/1165558.shtml.

[9] See more: Clayton, J., & Derico, B. (2023). Clearview AI used nearly 1m times by US police, it tells the BBC. BBC. https://www.bbc.com/news/technology-65057011.

verification is widely utilized by law enforcement, retailers, schools, casinos, and in crowded places (Sullivan, 2021, p. 2).

2.3. FRT as a Probabilistic Technology and Opacity

Although authentication and identification are two separate functions, they share a probabilistic nature. Both facial recognition methods rely on estimating the similarity between the template being compared and the baseline(s). A method of comparison determines the likelihood that the person being authenticated or identified is indeed the same person. If this probability rises beyond a predetermined threshold set by the user or system developer, it's a match. This likelihood is referred to as the 'confidence score'. For example, when comparing two templates, a confidence score of 90% indicates a high likelihood that they belong to the same person (European Data Protection Board, 2023, pp. 10-13).

The use of confidence thresholds is critical in cases where algorithmic matches are not verified by humans. In this scenario identification errors can have serious effects for people who are misidentified. Prioritizing higher miss rates over false positives is preferable, and strict confidence thresholds must be enforced to limit negative consequences (Crumpler, 2020).

If this technology is employed for investigation purposes, where a list of potential candidates is generated for human evaluation, confidence levels are often lowered. Humans are involved in evaluating the results and making final decisions about how to use the returned data. While the number of false matches should be no more than without the technology, many concerns have been raised about potential biases of human operators toward trusting algorithmic findings when certain matches obtain higher confidence levels than others (Crumpler, 2020).

The threshold is usually determined by the potential risks posed by false positives or negatives. For instance, a false positive that unintentionally grants access to someone's phone or apps is less damaging than a false mismatch in a criminal case that might result in a wrongful conviction. Therefore, the threshold for a

match can be adjusted accordingly, with a lower threshold for access to services or applications and a higher threshold to minimize the risk of incorrect identifications in criminal cases (Raposo, 2023, p. 4).

According to Loyola-González (2019, pp. 154097 - 154101) there are two types of AI systems: white-box and black-box. The first one are models which are easily comprehensible to experts in the relevant application domain. The black box are “mainly used for labeling all those machine learning models that are (from a mathematical point of view) very hard to explain and to be understood by experts in practical domains”.

The opacity in an AI system is often referred to as the black-box effect (Dignum, 2018). FRT is considered black-box AI system. As such, the operational mechanisms of this type of algorithm are still not fully understood by researchers, posing challenges in rectifying biases and malfunctions within FRT systems (Raposo, 2023, p. 4).

Bias is exacerbated by the opacity of the system. If bias cannot be understood due to this opacity, even experts cannot address or control it effectively (Alturis, 2022). While removing the algorithmic black-box may offer some insight into the algorithm's behavior, it won't eliminate the biased patterns it learns from the data. Despite these challenges, addressing bias comprehensively remains essential in AI development (Dignum, 2018).

Facial recognition technology has the potential to exhibit bias and inflict harm upon individuals, consequently posing challenges for both companies and individuals. The forthcoming chapters will delve into the mechanisms behind these issues and explore the factors contributing to these inaccuracy outcomes.

III - ADDRESSING BIASES IN FACIAL RECOGNITION TECHNOLOGIES

3.1. Understanding bias

Initially, it is important to define the concept of bias and distinguish it from the

concept of discrimination. Although these two concepts are often perceived as the same, they have different meanings. The term ‘bias’ “simply refers to deviation from a standard [...] there are many types of bias depending on the type of standard being used” e.g. moral bias, legal bias, social bias and others (Danks, 2017, p. 2).

Given the broad and varied usage of the term, bias can sometimes be employed in a relatively neutral sense, such as when a grocery shopper avoids purchasing damaged fruit. However, bias can also carry significant moral implications, for instance if an individual owns a building and decides that he is not renting any of the apartments to women (Friedman and Nissenbaum, 1996, p. 332).

Even though bias does not necessarily always lead to discrimination, for the purposes of this research it will be focused on the cases that this type of AI ends up violating the European principle of non-discrimination against individuals or groups of individuals.

The word ‘discrimination’ has not been historically at the center of philosophical debates. This is most due to the fact that most of them had a very narrow view of the meaning and scope of this term. According to the Canadian philosopher Sophia Moreau (2020, p. 173), ‘discrimination' was linked to an individual's intention to cause disadvantage to a certain group[10].

However, even if there is no aim to discriminate against a certain group or thoughts that they are less deserving than others, there are still a number of ways that actions and regulations might operate to deny others equal status. Following this line of thought many social scientists and many lawmakers “now accept that much of the discrimination that occurs in our societies involves ‘implicit bias’ against certain groups, rather than contempt for them or a belief in their inferiority” (Moreau, 2020, p. 174)[11]. This line of reasoning is crucial when discussing biased FRT.

Algorithms are being integrated into various aspects of society and influencing

[10] In Moreau’s words (2020, p. 173): “on this view, an agent wrongfully discriminates when he disadvantages a certain group of people because he holds an objectionable attitude toward them or an objectionable belief about them. What makes discrimination wrong, on this view, is the mental state that motivates it. Some of the first accounts that philosophers offered of discrimination were essentially just a more precise articulation of this lay conception, tracing the moral wrongness of discrimination back to certain objectionable mental states”.

[11] See more: BAGENSTOS, Samuel. Implicit Bias, Science, and Antidiscrimination Law. Harvard Law & Policy Review, Vol. 1, Issue 2. 2007. pp. 477-494; BANKS, Richard *et al*. Discrimination and Implicit Bias in a Racially Unequal Society. 94 California Law Review. 2006. p.1169-1190.

several decisions, such as prison sentences, loans approvals or even hiring (Buolamwini, 2017, p. 13). Prior to the prevalence of algorithms, the onus of decision-making rested upon humans (Lee *et al.*, 2019). Individuals often resorted to biases to make decisions in everyday life, resulting in negative consequences such as discriminatory practices. Similarly, algorithms are now reflecting and exacerbating these human biases, presenting a similar challenge.

In this context, Virginia Eubanks' (2018, pp. 15-16) book 'Automating Inequality' provides valuable insights into how disruptive new technologies can perpetuate and exacerbate biases in society. Eubanks highlights how these technologies, often operating within what she terms a "digital poorhouse," disproportionately impact financially disadvantaged individuals, especially people of color, across the United States. These systems may unintentionally perpetuate biases by relying on historical data that reflect societal disparities and prejudices.

The sociologist Ruha Benjamin (2019) also addresses the reflection of human inequalities in new technologies as “the New Jim Code”, *in verbis*: “*the employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory system of a previous era*”.

In this matter, Cathay O’Neil points out the same, in her words “*racism is the most slovenly of predictive models. It is powered by haphazard data gathering and spurious correlations, reinforced by institutional inequities, and polluted by confirmation bias*” (O’Neil, 2017, p. 26).

Thus, bias is an ethical and societal issue (Coeckelbergh, 2020, p. 124). Rather than breaking the cycle of unfairness, bias in algorithms are perpetuating and even intensifying it (Benjamin, 2019). The scope and character of the bias problem are frequently kept from view. Take for example, if the training data contains biases that reflect societal disparities, these biases may be unwittingly ingrained in the models that are developed (Buolamwini, 2017, p. 13).

A bias algorithm will discriminate when it denies an opportunity or a good to an individual or group of individuals on unjustified or unsuitable reasons. Building upon this premise, two crucial points must be considered. Firstly, (i) the algorithm must

consistently exhibit discriminatory behavior towards the specific individual or group of individuals, an occasional occurrence would not qualify as bias but rather an error. Secondly, (ii) the biased behavior must lead to an unfair outcome (Friedman and Nissenbaum, 1996, p. 333).

Bias is frequently unintentional in AI systems. Developers, users, and others involved in company management frequently fail to anticipate discriminatory impacts directed at certain groups or individuals (Coeckelbergh, 2020, p. 128). They pose ethical challenges due to their extensive analysis and complexity (Mittelstadt *et al.*, 2016, p 3).

This issue is not new, in 1988, the UK Commission for Racial Equality found St. George's Hospital Medical School to have engaged in discrimination. This because the computer program used to scan applicants for school places was prejudiced against women and individuals with non-European names. “The program was written after careful analysis of the way in which the staff were making these choices” (Lowry and Macpherson, 1988, p. 657).

Coeckelbergh (2020, p. 128) argues that bias may develop in a variety of ways during all stages of the design, testing or application processes. It can arise because the data does not represent the diversity in the society or it’s incomplete; or either because the data set contains mostly images that represent the populations of a certain country, but it’ll be used in other countries that have different features — leading to a cultural bias.

In the same vein, Friedman and Nissenbaum (1996, p. 330) noted that bias can develop in three distinct categories: preexisting, technical, and emergent. More specifically, “preexisting bias has its roots in social institutions, practices, and attitudes. Technical bias arises from technical constraints or considerations. Emergent bias arises in a context of use”.

To provide a more comprehensive understanding of each type of bias, preexisting bias is deeply rooted in practices, social institutions, and attitudes. It happens when prejudices that already exist independently and from before the system's establishment are reflected in computer systems. Even if there is no aim, preexisting bias has the potential to permeate a system. It may have originated from

individuals who wield considerable influence over system design or from broader societal influences that shape the development of technology, such as organizations, cultural norms, institutions (Friedman and Nissenbaum, 1996, pp. 333-335).

Additionally, technical bias arises from technical constraints and considerations within computer systems. It encompasses several key aspects. Firstly, can originate from limitations in computer technology, including hardware, software, and peripherals. It can also emerge from the use of decontextualized algorithms that do not treat all groups equally in a variety of important circumstances. Furthermore, this type of bias can arise from the imperfections in random number generation or in attempts to formalize human constructs such as judgments, intuitions or discourse to make them compatible with computers (Friedman and Nissenbaum, 1996, pp. 333-335).

Lastly, emergent bias emerges within the context of use. Unlike the other ones, this type of bias arises after the design phase, influenced by evolving societal knowledge, shifts in population dynamics, or changes in cultural values. Within emergent bias, two primary sources can be highlighted. Firstly, bias can originate from the emergence of new societal knowledge that cannot be or is not incorporated into the system design. Or it can also arise from a mismatch between the users of the system and the assumptions made during the design phase — different values or expertise from the place that was designed to the population that was implemented (Friedman and Nissenbaum, 1996, pp. 333-336).

Inspired by Coeckelbergh (2020) and Friedman and Nissenbaum's (1996) classification of bias, this study will examine three primary sources of bias in FRT: programmers, data, and human. Programmers may unintentionally embed their own biases into the code, influenced by societal factors that shape technology development. Moreover, the lack of diversity within development teams, primarily consisting of white males, can perpetuate these biases within the systems they create. Data bias originates from the limitations of datasets, often failing to comprehensively represent certain demographic groups. For instance, an FRT system may demonstrate higher accuracy in identifying white males but struggle with recognizing black females due to insufficient data from the latter group. Furthermore, human bias can manifest when operators excessively rely on AI-generated

outcomes, potentially sidelining external information and placing undue trust in the AI system's outputs. This issue gains particular relevance given the probabilistic nature of such AI systems.

3.2. Accuracy and Bias in Facial Recognition Technology

Facial recognition systems employ artificial intelligence techniques to identify and recognize faces within images, video, or still formats by utilizing people's face images that are biometric data (Fernandez *et al.*, 2020, p. 5). This biometric data has different qualities, such as uniqueness, immutability, and difficulties in concealment. Obtaining facial photos is more accessible and more straightforward than other types of biometric identification, such as fingerprints or DNA (European Union Agency for Fundamental Rights, 2019, p. 5).

Over the past few years, FRT accuracy has significantly increased. These improvements in accuracy can be attributed mostly to the increased processing capacity, big volumes of data, and more sophisticated machine-learning techniques. The most accurate face identification algorithm had an error rate of only 0.08% by April 2020, which is a huge improvement over the top algorithm error rate of 4.1% in 2014 (Crumpler, 2020).

However, a number of studies have shown that FRT has bias against specific demographic groups, which results in disparities in performance and a higher risk of misidentification. For instance, individuals with darker skin tones or Asian faces have been found to experience higher error rates and misidentifications than those with Caucasian faces. Further highlighting the differences in performance across various age groups, FRT has demonstrated lesser accuracy when detecting the faces of elderly people or children. Moreover, gender disparities have been observed, with some algorithms exhibiting higher error rates when identifying women's faces compared to men (NIST, 2019).

3.2.1. Academia

The issue of bias in FRT gained significant attention due to, *inter alia*, an MIT researcher, Joy Buolamwini (2017, p. 3) on her thesis entitled “Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers”. The research, published in 2017, focuses on studying the gender and skin type distribution in different facial recognition datasets and classification benchmarks. The thesis evaluates the accuracy of gender classifiers from Adience, IBM, Microsoft, and Face++ considering gender, skin type, and the intersection of gender and skin type.

The research found that the evaluated datasets are predominantly composed of lighter-skinned individuals, ranging from 79.6% to 86.24%. For example, one of the companies analyzed had a dataset with low representation of females (24.6%) and even lower representation of darker-skinned females (4.4%), while featuring a majority of lighter-skinned males (59.4%). Another company had achieved roughly equal gender distribution (52.0% female), but has a limited representation of darker-skinned individuals (13.76%) (Buolamwini, 2017, p. 3).

When evaluating the four gender classifiers, a significant disparity is observed in the classification accuracy across different groups. Females are generally classified less accurately compared to males (with differences ranging from 9% to 20%), and individuals with darker skin are also classified less accurately compared to those with lighter skin (differences ranging from 10% to 21%). Notably, darker-skinned females are the most poorly classified group, contributing to 37% to 83% of the classification errors. Conversely, lighter-skinned males are the least error-prone group, accounting for only 0.4% to 3% of the overall classification errors (Buolamwini, 2017, p. 3).

Buolamwini’s thesis highlights the implications of misclassification and emphasizes the importance of constructing inclusive training sets and benchmarks for facial recognition systems. Those datasets, heavily male and white, provide the false impression of progress but are incorporating bias into machine learning algorithms (Buolamwini, 2017, p. 98).

In the following year, Joy Buolamwini and Timnit Gebru (2018, p.1) published another research entitled “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”. The research uncovered large gender racial bias

in FRT sold by giant companies — it didn't have the same level of performance accuracy in people of color, especially women.

Through their research, it was revealed the existence of profound gender and racial bias within AI systems. Buolamwini and Gebru examined the distribution of gender and skin type in two facial analysis benchmarks, IJB-A and Adience. After that, they evaluated the accuracy of gender classification in International Business Machines (IBM), Microsoft, and Face++ gender classifiers (Buolamwini and Gebru, 2018, p. 8).

The study showed disparities in performance based on gender and skin tone. Male faces were classified more accurately than female faces, while lighter faces were classified more accurately than darker faces. Darker female faces had the highest error rate. Microsoft and IBM classifiers performed best on lighter male faces, while Face++ classifiers performed best on darker male faces (Buolamwini and Gebru, 2018, p. 8). Those findings triggered a widespread outcry that gained further momentum when Joy Buolamwini shared YouTube videos demonstrating the technology's misclassification of Michelle Obama, as a man (Singer, 2019).

Buolamwini's research played an essential role in bringing the issue of bias FRT to the public's awareness. She coined the term 'Conde Graze' for algorithms that exhibit bias (Buolamwini, 2016). Her work received extensive media coverage[12] and was the subject of a documentary on Netflix entitled 'Coded Bias'[13] that featured her research.

Buolamwini was not the only one to address this concern. The academic community has extensively researched and published numerous studies addressing cases of bias, particularly concerning racial and gender issues in AI. These works

[12] See: Johnson, A. (2023). Racism And AI: Here's How It's Been Criticized For Amplifying Bias. Forbes. https://www.forbes.com/sites/ariannajohnson/2023/05/25/racism-and-ai-heres-how-its-been-criticized-for-amplifying-bias/?sh=3673acc3269d; Tucker, I. (2017). A white mask worked better': why algorithms are not colour blind. The Guardian. https://www.theguardian.com/technology/2017/may/28/joy-buolamwini-when-algorithms-are-racist-facial-recognition-bias; Lohr, S. (2018, February 9). Facial Recognition Is Accurate, if You're a White Guy. The New York Times. https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html; Girish, D. (2020). Coded Bias' Review: When the Bots Are Racist. The New York Times. https://www.nytimes.com/2020/11/11/movies/coded-bias-review.html.

[13] See: Kantayya, S. (2020). *Coded Bias*. Brooklyn, New York City, New York, USA.

shed light on the concerning presence of bias in AI systems and various FRT applications.

In this matter, Klare *et al.* (2012, p. 1) also investigated demographic differentials in face recognition accuracy across algorithms, providing valuable insights into disparities based on age, gender, and race. In this research it was examined three commercial face recognition algorithms'[14] performance for different demographic groups (gender, race/ethnicity, age).

One of Klare's *et al.* (2012, p. 13) important findings was that females, black, and younger people are harder to recognize for all types of face recognition algorithms tested. Also, when face recognition systems are trained exclusively on a specific demographic group, their performance improves for that group's race/ethnicity and age.

> "[...] the results in this study should motivate the design of algorithms that specifically target different demographic cohorts within the race/ethnicity, gender and age demographics. By focusing on improving the recognition ac- curacy on such confounding cohorts (i.e., females, Blacks, and younger subjects), researchers should be able to further reduce the error rates of state of the art face recognition algorithms and reduce the vulnerabilities of such systems used in operational environments" (Klare *et al.,* 2012, p. 13).

Additionally, Nkonde (2020, pp. 30-36) conducted an analysis of the effects on people of color by the use of bias FRT by law enforcement agents. Drozdowski *et al.* (2020, pp. 98-99) delved into the impact of demographic bias in biometric systems and emphasizes the importance of developing proper frameworks and rules to navigate the evolving landscape of biometric technologies. Similarly to others researchers concluded that demographic factors impact biometric algorithms, leading to biases towards certain groups. Studies show lower performance for females, younger individuals, and dark-skinned females in biometric recognition systems.

According to Drozdowski *et al.* (2020, p. 10), biased systems can cause harm, and there is a need for algorithmic fairness, independent assessments, transparency, and accountability. Despite the academic attention there are limited legal provisions in existence. *In verbis*:

[14] Cognitec's FaceVACS v8.2, PittPatt v5.2.2, and Neurotechnology's MegaMatcher v3.1.

> “Biased automated decision systems can be detrimental to their users, with issues ranging from simple inconveniences, through disadvantages, to lasting serious harms. This relevance notwithstanding, the topic of algorithmic fairness is still relatively new, with many unexplored areas and few legal and practical provisions in existence. Recently, a growing academic and media coverage has emerged, where the overwhelming consensus appears to be that such systems need to be properly assessed (e.g., through independent benchmarks), compelled to some degree of transparency, accountability, and explainability in addition to guaranteeing some fairness def- initions. Furthermore, it appears that, in certain cases, legal provisions might need to be introduced to regulate these technologies” (Drozdowski *et al.,* 2020, p. 10).

These works exemplify the growing attention and efforts to address the bias challenges in FRT and underscore the importance of developing fair and equitable facial recognition technologies.

This issue also gained a lot of significant attention in the media. Numerous reports and articles have highlighted the potential implications of biased FRT systems on individuals and society as a whole. For example, prominent news publications like The New York Times, The Guardian, Washington Post and BBC News have extensively covered stories on biased facial recognition algorithms leading to misidentifications and potential discrimination.

Recent controversies surrounding high-profile cases of misidentification and racial bias in FRT have also sparked public debate and activism. One of them is the Algorithmic Justice League (AJL), founded by Joy Buolamwini (Algorithmic Justice League, 2023).

3.2.1. NIST's Face Recognition Vendor Test (FRVT)

Since the beginning of the development of this technology, NIST's Face Recognition Vendor Test (FRVT) program has been the main examiner of face recognition algorithms, serving as an important benchmark (Parker, 2020, p. 1). According to the latest report, published in 2019, the 'Ongoing Face Recognition Vendor Test (FRVT)' — parts one and two, which analyze verification and identification respectively — there was a significant increase in accuracy compared to the previous report from 2014 (Grother *et al.*, 2019a, p. 6).

These reports were created using four datasets: domestic mugshots, application pictures, visa pictures, and border crossing photos. All of these datasets

were provided by the U.S. government applications, resulting in a total of 18.27 million photos of 8.49 million individuals. The evaluation utilized 189 primarily commercial algorithms developed by 99 developers (Grother *et al.*, 2019b, p. 6).

Even though it has made great progress in accuracy, it is important to note that even a small error rate (e.g., 0.1%) can lead to significant issues if it's used in a significant number of individuals and lead to significant consequences, public embarrassment or humiliation (European Union Agency for Fundamental Rights, 2019, p. 9).

Determining the level of accuracy of FRT is not an easy task. There are multiple approaches available to evaluate and assess accuracy, which may differ depending on the specific task, purpose, and context in which they are employed (Grother *et al.*, 2019a, pp. 16-17).

According to Grother *et al.* (2019a, pp. 16-17) several aspects can interfere with accuracy, such as the quality of the image, especially for specific demographics like young children or tall individuals. Demographic differences during image capture may arise from camera inadequacies, environmental factors, or client-side detection without algorithmic culpability. Photographic standards also play a role, and aging is a factor for everyone, but children undergo this process faster, which can make them more vulnerable to inaccuracies in FRT.

The lastlest NIST report concluded that “across demographics, false positives rates often vary by factors of 10 to beyond 100 times” (Grother *et al.*, 2019b, p. 2). To be more precise, false negative error rates vary significantly among face recognition algorithms, from below 0.5% to above 10%. The most accurate algorithms have lower false negative rates and smaller demographic differences (Grother *et al.*, 2019b, p. 7).

Domestic mugshots show higher false negatives in Asian and American Indian individuals compared to white and black faces. High-quality application photos also play a role to decrease this issues, having very low error rates, while lower-quality border crossing images result in higher false negatives, especially for individuals born in Africa and the Caribbean, particularly in older age groups (Grother *et al.*, 2019b, p. 7).

False positive demographic differentials in verification algorithms showed higher rates in women 2 to 5 times compared to men, especially in West and East African and East Asian individuals, while East European individuals had the lowest rates. Algorithms that were made in Asian countries perform better on people with Asian faces than those developed in the West, in this regard the NIST report observed that the ones developed in China had lower false positives on East Asian faces than on Caucasian faces (Grother *et al.*, 2019b, pp. 7-8).

The impact of these differences on error rates will vary depending on the application of this algorithm. See for example in cases of verification where the is to determine if two images belong to the same person or different individuals. When images from the same person are compared, it should indicate a match. Conversely, when comparing images from different people, it should indicate a mismatch. In ideal conditions, the mismatch scores should be low, clearly distinguishing different individuals, while the match should be high, representing accurate matches. However, in practice, errors can occur. Some mismatch scores might exceed a specific threshold, leading to false positives, where the system incorrectly identifies a person as someone else. Additionally, some genuine comparisons may fall below the threshold, resulting in false negatives, where the system fails to recognize the correct identity of a person. These false positives and false negatives can impact the overall accuracy and reliability of the facial recognition system (Grother *et al.*, 2019b, pp. 2-4).

In identification, it compares features from a search image with enrolled gallery images, returning similar candidates above a preset threshold. In this case, the impact of demographic differences depends on the situation. False positives happen when the system incorrectly identifies someone or shows a candidate match for human review. For example, during visa or passport fraud checks, a false positive could lead to wrongful accusations or detentions. On the other hand, higher false negatives may benefit someone trying to deceive the system, but it weakens the system's security goals (Grother *et al.*, 2019b, p. 5).

However, it is strongly emphasized during the report that error rates vary widely among the algorithms tested, and those with lower error rates also exacerbate extremely low disparities across demographic groups. To substantiate this claim,

they present a technical analysis of all the factors utilized in reaching the indicated results (Grother *et al.*, 2019b, pp. 2-4). *In verbis:*

> "We found empirical evidence for the existence of demographic differentials in the majority of contemporary face recognition algorithms that we evaluated. The false positive differentials are much larger than those related to false negatives. False positive rates often vary by one or two orders of magnitude (i.e., 10x, 100x). False negative effects vary by factors usually much less than 3. The false positive differentials exist broadly, across many, but not all, algorithms. The false negatives tend to be more algorithm-specific [...]
>
> The accuracy of algorithms used in this report has been documented in recent FRVT evaluation reports [16, 17]. **These show a wide range in accuracy across algorithm developers, with the most accurate algorithms producing many fewer errors than lower-performing variants. More accurate algorithms produce fewer errors, and will be expected therefore to have smaller demographic differentials**" (Grother *et al.*, 2019b, p. 6).

Because practical face recognition systems frequently rely on a single algorithm, assessing its performance is critical. To make judgments and improve future performance, policymakers, developers, and consumers should be aware of algorithm-specific variances in accuracy among demographics groups. Understanding these distinctions can lead to more equitable and effective face recognition algorithms (Grother *et al.*, 2019b, p. 3).

Besides that, it is also argued that "*reporting of demographic effects has been incomplete, in both academic papers and in media coverage*" and suggest the inclusion of information on the purpose of the system, the stage of differential occurrence, relevant metrics, process duration, recognition threshold value, affected demographic groups, and consequences of errors and error remediation procedures (Grother *et al.*, 2019b, p. 10).

The NIST's report directly addressed the research published by Joy Buolamwini. It clarified that Buolamwini's research did not assess face recognition algorithms but rather focused on face analysis algorithms, which produce outputs related to age or emotional state. Face recognition algorithms, on the other hand, do not have a built-in notion of a specific person; they utilize face detection and feature extraction to convert images into identity-related vectors. It is essential to distinguish between these types of algorithms when discussing bias in face recognition and reporting research findings accurately (Grother *et al.*, 2019b, p. 10).

> “Much of the discussion of face recognition bias in recent years cites two studies showing poor accuracy of face gender classification algorithms on black women. Those studies did not evaluate face recognition algorithms, yet the results have been widely cited to indict their accuracy” (Grother *et al.*, 2019, p. 4).

The NIST report brings attention, in a technical matter, to the considerable improvements made in the accuracy of FRT. While acknowledging variations in accuracy across different demographic groups, the report stresses that these disparities are influenced by the specific algorithms being used.

Prioritizing algorithms with lower error rates among various demographic groups is a constructive path forward as a way to promote a more equitable and inclusive technological landscape for the future.

3.2.3. Corporate

Academia and the media are strongly critical of the bias of facial recognition technology, especially on issues related to gender and race. However, they were not the only ones to address this concern. In fact, Big Tech companies have acknowledged this concern and taken steps to remove the use of this technology in mass surveillance or racial profiling, IBM was one of them. The company announced that AI systems need to be tested for bias (BBC, 2020).

In this matter, instances of biased face recognition technology have been observed in various popular platforms, leading to significant controversies and concerns. In 2019, researchers at the M.I.T. Media Lab also conducted a study revealing that Amazon's facial recognition system, Rekognition, exhibited significant error rates in accurately determining the gender of female faces and faces with darker skin tones in comparison to comparable services offered by IBM and Microsoft (Singer, 2019).

Amazon has actively promoted Rekognition to law enforcement agencies, positioning it as a valuable tool to aid identifying potential suspects (Singer, 2019). This led to the drafting of an open letter addressed to Amazon, expressing

opposition to the utilization of Rekognition by police departments and government entities due to its biased nature towards people of color (Singer and Metz, 2019).

Facebook, as mentioned before, is also a prominent company when it comes to facial recognition technology and it was a key player in popularizing FRT. However, Facebook's automated prompt was found biased when asked users if the users wanted to "keep seeing videos about primates" after watching a video that featured a black man that had nothing to do with primates (Mac, 2021).

Facebook changed its name for Meta and announced the end of the use of FRT on its platform (Dwoskin and Harwell, 2021). Meta's (former Facebook) announcement highlighted the significance of facial recognition technology as a powerful tool for identity verification and fraud prevention, as well as the impact of its shutdown on the Automatic Alt Text (AAT) — an AI that generates descriptions of images for people who are visually impaired or blind (Pesenti, 2021)

Even though it was disabled, the company made it clear that it still had an interest in developing this technology and would continue to explore its use for other company platforms (Pesenti, 2021). Shortly after announcing the removal of the facial recognition system on Facebook, Meta stated that it would use the same mechanism for its metaverse products (Ians, 2021).

Similarly, a software engineer, Jacky Alciné, noticed that Google Photos' image recognition algorithms were incorrectly classifying his black friends as "gorillas". Although Google said it was going to fix the problem, a report from Wired found out that only the company only blocked the words, "gorilla," "chimp," "chimpanzee," and "monkey" (Vincent, 2018).

In a post published on the website of Nudest, a company founded at Harvard Business School, the founder, Atima Lui, shared a notable instance of biased face recognition technology. The example involved Snapchat, where during a reunion with a college friend, the Snapchat bunny filter worked perfectly on her friend's face but failed to apply the filter on her own face, seemingly disregarding her presence entirely. As Atima Lui described, even in well-lit conditions with clear contrasts between her dark face, white teeth, and eyes, Snapchat still failed to recognize her as a human — "Even in bright lighting where the whites of my eyes and white teeth

contrast against my dark face, Snapchat concludes that no human is present” (Lui, 2023).

According to Apple, “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, 2023). Although, there have been some media reports that indicate otherwise. In 2017, a Chinese woman named Yan from Nanjing encountered an unusual issue with her iPhone X's facial recognition function. Her colleague managed to unlock not just one, but two different devices using her face. Even after returning the first iPhone and getting a refund, the same problem persisted with the second device. Despite configuring the facial recognition software to recognize only her face, her colleague could still unlock the phone effortlessly; both women were unrelated (Wong, 2017).

During the pandemic, universities started the use of proctoring systems that utilize FRT to remotely monitor students during online exams. However, these proctoring systems also have raised concerns. Instances have been reported where students of color have not been detected by the algorithm, thus damaging students[15].

One notable case involves a student, who raised the concern that the Proctorio exam software required her to use a lamp near her face to enable the algorithm to detect it properly. There have been several legal cases in the USA and the Netherlands discussing the use of proctoring, but most have centered around privacy rather than addressing racial discrimination (Meaker, 2023).

To be more specific, during the 2020/21 academic year, she was pursuing a master's degree at the Vrije Universiteit Amsterdam Foundation. With exams moving predominantly online due to the pandemic, an "anti-cheating software" was

[15] See more: Proctor Ninja. (2021). Proctorio's facial recognition is racist.https://web.archive.org/web/20220112203737/https:/proctor.ninja/proctorios-facial-recognition-is-racist; Meaker, M. (2023). This Student Is Taking On 'Biased' Exam Software. Wired. https://www.wired.co.uk/article/student-exam-software-bias-proctorio; Clark, M. (2021). Students of color are getting flagged to their teachers because testing software can't see them. The Verge. https://www.theverge.com/2021/4/8/22374386/proctorio-racial-bias-issues-opencv-facial-detection-schools-tests-remote-learning;

implemented to prevent fraud. The student had to install this software, which included a webcam check (DHRC, 2022).

However, she frequently experienced issues with the software's facial recognition. Messages like "face not found" or "room too dark" appeared, causing stress and insecurity. The software repeatedly removed her from the online exam environment, after which she had to re-register (DHRC, 2022).

In this specific case, the student initiated a legal case against the University in the Dutch Human Rights Court to ascertain whether the university had engaged in racial discrimination by employing the Proctorio software for online exam monitoring, as well as the University's negligence in responding to her discrimination complaint in this context (DHRC, 2022).

The University denies that the software discriminated, stating that it has undergone an independent audit confirming its proper functioning, and refutes any claims that the software operates less effectively for individuals with darker skin tones. Emphasizing that the issues experienced by the student appear to stem from an unstable internet connection. The legal case is ongoing, and an interim ruling was issued, indicating strong indications that the software used by VU Amsterdam was discriminatory based on race (DHRC, 2022).

Moreover, Uber's implementation of facial recognition technology to verify driver identities also raised concerns related to bias and accuracy. Reports indicated that the system encountered difficulties in accurately recognizing non-white drivers (Barry, 2021; Tomas, 2022).

A former Uber driver, Pa Edrissa Manjang, has initiated legal proceedings concerning allegedly bias facial recognition software, provided by Microsoft (Employment tribunal, 2022, p. 1-2)[16]. When he attempted to sign in for work, UberEats' face recognition software failed to recognize him, as a result, he was fired (Worker Info Exchange, 2022).

[16] Mr P E Manjang v. Uber Eats UK Ltd and others, Case n. 3206212/202.

The Claimant's case centers around an incident on April 30, 2021 and “claims are for harassment related to race under s.26 Equality Act 2010, victimisation under s.27 Equality Act 2010 and indirect race discrimination under s.19 Equality Act 2021”. The company first informed him that his account was suspended by email due to alleged account sharing. Subsequently, he was informed that he had failed a facial recognition check, with the software failing to verify his submitted photo (Employment tribunal, 2022, pp. 2-4).

Even after the former driver complained, there was no change in the company's decision — “*your algorithm by the looks of things is racist and this needs to be addressed as it is not able to recognise and verify my photos which is probably why I get asked to take photos of myself multiple times a day”* (Employment tribunal, 2022, p. 4).

Despite conversations with an agent, the Claimant's account deactivation decision remained unchanged. Even after having a human review, the decision for the app was maintained[17]. In this point it was unclear to what extent the human review was made, in verbis: “*it is wholly unclear what ‘review’ process was undertaken by the Respondents and the extent to which the review involved a human comparison between the submitted photograph and the image of the Claimant on their records*”. A recent judgment declined Uber's application to strike out the claim, but no decision on the merits of the case has been rendered (Employment tribunal, 2022, p. 4).

Uber India announced the implementation of Real-Time ID Check, powered by Microsoft Cognitive Services, specifically Microsoft's Azure Face service in 2017. The company stated that this technology would improve safety for both riders and drivers, as well as prevent issues before they occur (Uber India, 2022).

Despite several benefits that Uber has highlighted in implementing this technology, a recent MIT study has drawn attention to the company locking drivers out of the app due to facial recognition failures. A total of 150 drivers took part in the survey. Some of them blame insufficient lighting and most presume the issues

[17] As previously discussed, there are indications that many concerns have been raised about potential biases of human operators toward trusting algorithmic findings.

comes from alterations in the appearance, such as haircut or facial hair (Bansal, 2022).

Uber India has highlighted that their facial recognition technology is designed to accommodate changes in hair style. Moreover, they have clarified that account deactivations based solely on facial recognition are not a possibility. Instead, their Real-Time ID Check feature operates by identifying cases where facial recognition does not yield a match. In such scenarios, these instances are then elevated for manual review, involving the assessment of at least two human evaluators (Bansal, 2022).

On the other hand, a study conducted by Gaurav Jain and Smriti Parsheera in 2021 examined the performance of various commercial facial recognition systems, including Microsoft Azure's Face, on Indian faces. The research concluded that Microsoft had the highest detection error rate, reaching 3.17%. This suggests that the system struggled to accurately identify more than a thousand faces within the dataset (Jain and Parsheera, 2021, p. 16).

In June 2022, Microsoft announced a shift in its approach to ethics, leading to restrictions on access to its FRT. This change is particularly relevant to the use of Microsoft's Azure Face service, which is employed by companies like Uber for identity verification purposes, as highlighted in the cases mentioned earlier. Under the updated guidelines, companies are required to actively apply for permission and adhere to Microsoft's AI ethics standards in order to utilize the facial recognition features (Hern, 2022).

This technology has the potential to be utilized in various fields but can lead to significant discrimination when it fails to function properly, particularly concerning an individual's gender and skin tone. It is crucial to expose the flaws in these systems and advocate for an active role within companies to improve and foster more accountability and transparency.

These examples demonstrate the pressing need for careful consideration and continuous improvement in face recognition algorithms to mitigate biases and ensure equitable treatment for all individuals. As FRT becomes increasingly prevalent in

various applications, ranging from social media platforms to identity verification systems, it is crucial to thoroughly assess and mitigate any potential biases.

In today's technologically advanced society, the use of FRT is already widespread, making it unrealistic and impractical to consider banning or avoiding its use. However, it cannot be ignored that FRT does exhibit biases and can result in harm to specific gender or race groups.

Incidents of misidentification have occurred and are happening, leading to infringements on citizens' rights, causing emotional and psychological harm. European law should guarantee a robust path for seeking compensation in such cases and establish more stringent regulations to prevent the use of low-performing algorithms in the market.

Striking a balance between technology's benefits and safeguarding individual rights should be a primary goal as society navigates the complexities of this technology-driven era.

IV - CORPORATE LIABILITY FOR BIASES IN FACE RECOGNITION TECHNOLOGY

Liability is an extremely important concept in everyday life, as it ensures that individuals who have suffered damage or loss have the right to obtain compensation and reparation from the responsible party. Moreover, it also fosters pecuniary incentives for both individuals and companies/organizations to proactively prevent such damage or loss from occurring (European Parliament, 2020, p. 2).

Despite its immense benefits, FRT can be biased and cause harm to individuals thereby placing companies and individuals in challenging positions. It is therefore of utmost importance to analyze how companies can be held liable in a manner that preserves innovation when such incidents occur, considering the perspective of European legislation.

As previously mentioned, this technology tends to perform less accurately within certain demographic groups. Several examples have been discussed earlier,

illustrating how this can lead to various challenges in the daily lives of these individuals. These challenges range from delays in accessing exams due to anti-cheating software issues or encountering difficulties with social media filters. More significantly, these biases can even extend to more serious consequences, such as wrongful arrests within the law enforcement use.

Considering the impact of bias in this technology, this research will focus on its legal ramifications, specifically within liability lawsuits. It will explore the legal consequences that arise when FRT biased performance in specific demographic groups leads to discriminatory and harm.

The advancement of emerging technologies has spurred the debate about the effectiveness of traditional liability laws in such cases. Thus, who is to be liable when bias FRT does work probably in an individual or a group of individuals?

4.1. Legal Personhood of AI

Firstly, It is essential to emphasize that AI systems do not possess legal personality. Recent judicial decisions in the United Kingdom, United States, and European Union in the Device for the Autonomous Bootstrapping of Unified Sentience (DABUS) case confirm that this trend will persist, and there is no indication of the judiciary moving towards alternative interpretations (Bide, 2021; IPWatchDog, 2021; European Patent Office, 2021).

These recent judicial decisions were related to patent applications applied for the inventions of DABUS, an AI. Although the creations of this AI were eligible for patent, the registration was rejected in the EU, the UK, and the US, specifically because the term ‘inventor’ is only applied to humans (Bide, 2021; IPWatchDog, 2021; European Patent Office, 2021).

Nonetheless, the patent was granted by South Africa’s patent office — being the first to grant a patent for a product by an AI inventor. However, South Africa lacks a substantive patent examination system, with applicants only required to file their

inventions. Thus, the significance of the acceptance does not hold the same weight as it does in other jurisdictions with more specific and strict legislations (Bide, 2021).

The concept of conferring legal personhood upon an AI system was dismissed by the European Parliament (2020, p. 4). This decision stems from the understanding that any harm caused by an AI system, whether directly or indirectly, can be traced back to human involvement in constructing, implementing, or intervening with these systems[18].

The same position was stated in Liability for Artificial Intelligence and other Emerging Digital Technologies made by the Expert Group on Liability and New Technologies by the European Commission in 2019. According to them (2019, p. 34), giving legal status to emerging digital technologies is unnecessary. Harm caused by these technologies can usually be linked to existing individuals or entities. Creating new laws for individuals is a better approach than creating a new legal category and could cause ethical issues.

> "Still, the experts believe there is currently **no need to give a legal personality** to emerging digital technologies. Harm caused by even fully autonomous technologies is generally reducible to risks attributable to natural persons or existing categories of legal persons, and where this is not the case, new laws directed at individuals are a better response than creating a new category of legal person. Any sort of legal personality for emerging digital technologies may raise a number of ethical issues. More importantly, it would only make sense to go down that road if it helps legal systems to tackle the challenges of emerging digital technologies. 103 Any additional personality should go hand-in-hand with funds assigned to such electronic persons, so that claims can be effectively brought against them. This would amount to putting a cap on liability and – as experience with corporations has shown – subsequent attempts to circumvent such restrictions by pursuing claims against natural or legal persons to whom electronic persons can be attributed, effectively 'piercing the electronic veil'. In addition, in order to give a real dimension to liability, electronic agents would have to be able to acquire assets on their own. This would require the resolution of several legislative problems related to their legal capacity and how they act when performing legal transactions" (Expert Group on Liability and New Technologies, 2019, p. 38).

Therefore, granting legal personality to AI systems is not deemed necessary (European Parliament, 2020, p. 4). The rights and responsibilities of these systems

[18] "[...] *7. Notes that all physical or virtual activities, devices or processes that are driven by AI-systems may technically be the direct or indirect cause of harm or damage, yet are nearly always the result of someone building, deploying or interfering with the systems; notes in this respect that it is not necessary to give legal personality to AI-systems; is of the opinion that the opacity, connectivity and autonomy of AI-systems could make it in practice very difficult or even impossible to trace back specific harmful actions of AI-systems to specific human input or to decisions in the design; recalls that, in accordance with widely accepted liability concepts, one is nevertheless able to circumvent this obstacle by making the different persons in the whole value chain who create, maintain or control the risk associated with the AI-system liable*" (European Parliament, 2020, p. 4).

will not be discussed, but rather the responsibility of the humans who own and control such systems (Lai, 2021, p. 12).

4.2. Liability In the era of AI systems

AI systems in general present challenges when it comes to liability. The main concerns revolve around certain features, such as opacity, transparency, autonomous behavior, and explainability (European Commission, 2022a, p. 5). These pose significant obstacles for identifying and proving the fault of a potentially responsible party, demonstrating the existence of a defect, and establishing a causal link between the fault/defect and the resulting damage. As a result, obtaining compensation becomes difficult (Madiega, 2023, p. 3).

Opacity in FRT systems often makes it challenging to discern where the responsibility for bias and potential harm lies. Operators of such AI systems might argue that certain actions were beyond their control due to the system's autonomous operation. This could lead to situations where assigning liability becomes challenging, potentially resulting in unfair or inefficient outcomes. When a system exhibits discriminatory behavior, it's unclear whether the fault lies with the developer, the operator, or even the data used to train the model (European Parliament, 2020, p. 10-11).

Existing liability rules may not always yield suitable outcomes when it comes to addressing risks associated with emerging digital technologies. It may lead to an unfair distribution of losses, especially when it is unclear who is responsible for causing the damage, who benefited from it, who controlled the risk, or who could have taken the most cost-effective preventive measures. While existing liability laws outline compensation options, they are not comprehensive enough to adequately address harms arising from emerging digital technologies due to their distinctive challenges. Compared to traditional technologies and usual liability issues, the application of the current liability laws affects both fair compensation and access to justice, making litigation complex and costly for victims. To tackle this issue, it becomes imperative to make adjustments and revisions to liability frameworks,

taking into account the diverse array of risks posed by different technologies (Expert Group on Liability and New Technologies, 2019, p. 34).

It is necessary to ensure people receive the same level of protection as in situations without AI, along with fair compensation that effectively tackle these legal issues and reduce the likelihood of individuals hesitating to embrace new technology. Users should feel secure knowing that potential AI-related damage is adequately insured and that there are established legal channels for pursuing compensation (European Parliament, 2020, p. 3).

4.2. Liability Framework in the EU

The current legal framework in the European Union regarding liability is partially harmonized and comprises concurrent national liability rules, the Product Liability Directive 85/374/EEC (PLD), Article 82 of the General Data Protection Regulation (GDPR)[19] for compensating damages concerning data protection infringements, and liability arising from damages related to competition law under the Antitrust Damages Directive 2014/104/EU.

In general, national liability laws do not have explicit liability rules for damage caused by the use of new digital technologies. The harm caused "can be compensated under existing ('traditional') laws on damages in contract and in tort in

[19] Article 82 GDPR: "Right to compensation and liability 1. Any person who has suffered material or non-material damage as a result of an infringement of this Regulation shall have the right to receive compensation from the controller or processor for the damage suffered. 2. Any controller involved in processing shall be liable for the damage caused by processing which infringes this Regulation. A processor shall be liable for the damage caused by processing only where it has not complied with obligations of this Regulation specifically directed to processors or where it has acted outside or contrary to lawful instructions of the controller. 3. A controller or processor shall be exempt from liability under paragraph 2 if it proves that it is not in any way responsible for the event giving rise to the damage. 4. Where more than one controller or processor, or both a controller and a processor, are involved in the same processing and where they are, under paragraphs 2 and 3, responsible for any damage caused by processing, each controller or processor shall be held liable for the entire damage in order to ensure effective compensation of the data subject. 5. Where a controller or processor has, in accordance with paragraph 4, paid full compensation for the damage suffered, that controller or processor shall be entitled to claim back from the other controllers or processors involved in the same processing that part of the compensation corresponding to their part of responsibility for the damage, in accordance with the conditions set out in paragraph 2" (EU, 2016).

each Member State" (Expert Group on Liability and New Technologies, 2019, p. 16). However, it encounters gaps and barriers.

The PLD establishes liability rules within the European Union for damage caused by defective products and states that the producer of a product is liable for any damage caused by defects in the product[20]. It was adopted in 1985 establishing a no-fault liability regime, i.e. the producer may be liable even if they were not negligent or at fault when a defective product[21] hurts a customer (European Commission, 2016b).

The term 'producer' includes the manufacturer of the finished product, the producer of raw materials, the manufacturer of a component part, and anyone who presents themselves as the producer through branding. This directive also considered that importers of products into the community for commercial purposes are producers and share liability. In cases where the original producer cannot be identified, each supplier is treated as the producer unless they disclose the producer's identity to the injured party. Thus, more than one individual can be held jointly liable (EU, 1985)[22].

A product is considered defective if it fails to meet the safety expectations that a person is entitled to have, considering three factors: how the product is presented; the intended use of the product; and the timing of when the product is introduced to the market (EU, 1985). Under the PLD (EU, 1985), rights of injured person expire 10 years after the product was put on the market[23].

---

[20] Article 1 PLD: "The producer shall be liable for damage caused by a defect in his product".
[21] Article 2 PLD: "For the purpose of this Directive 'product' means all movables, with the exception of primary agricultural products and game, even though incorporated into another movable or into an immovable. 'Primary agricultural products' means the products of the soil, of stock-farming and of fisheries, excluding products which have undergone initial processing. 'Product' includes electricity".
[22]Article 3 PLD: "1. 'Producer' means the manufacturer of a finished product, the producer of any raw material or the manufacturer of a component part and any person who, by putting his name, trade mark or other distinguishing feature on the product presents himself as its producer. 2. Without prejudice to the liability of the producer, any person who imports into the Community a product for sale, hire, leasing or any form of distribution in the course of his business shall be deemed to be a producer within the meaning of this Directive and shall be responsible as a producer. 3. Where the producer of the product cannot be identified, each supplier of the product shall be treated as its producer unless he informs the injured person, within a reasonable time, of the identity of the producer or of the person who supplied him with the product. The same shall apply, in the case of an imported product, if this product does not indicate the identity of the importer referred to in paragraph 2, even if the name of the producer is indicated".
[23] Article 1 PLD.

The article 7 of this directive addresses cases in which the producer will not be liable for damages, these being (i) did not has not responsible to put the product into circulation; (ii) when the issue that caused the damage did not exist at the time the product was put into circulation or only subsequently surfaced; (iii) when he did not produce the product to obtain economic means; (iv) when the defect occurs due to needing to comply with public standards; (v) when scientific and technical knowledge at the time of creation of the product did not allow such a defect to be detected; and (vi) the defect is attributed to the design of the product[24] (EU, 1985).

Thus, within the PLD, there are three avenues for liability claims. The fault-based liability claim that requires proof of damage, fault, and causality. Also the strict liability claim that does not depend on fault. At last, a claim against the producer of a defective product, being necessary to demonstrate the product's defect and the causal nexus to the damage (Madiega, 2023, pp. 2-3).

There are no cases of bias face recognition systems in which individuals were given compensation in the EU under the PLD. However, numerous business concerns emerge as a result of legal uncertainty stemming from outdated and ambiguous liability rules at both the EU and national levels (Madiega, 2023, p. 3).

Examining liability laws across Member States reveals significant variations and responses to AI-related concerns. In the absence of consistent EU actions, diverse practices and interpretations could arise, possibly impacting the efficient operation of the internal market.

[24] Article 7 PLD: “The producer shall not be liable as a result of this Directive if he proves: (a) that he did not put the product into circulation; or (b) that, having regard to the circumstances, it is probable that the defect which caused the damage did not exist at the time when the product was put into circulation by him or that this defect came into being afterwards; or (c) that the product was neither manufactured by him for sale or any form of distribution for economic purpose nor manufactured or distributed by him in the course of his business; or (d) that the defect is due to compliance of the product with mandatory regulations issued by the public authorities; or (e) that the state of scientific and technical knowledge at the time when he put the product into circulation was not such as to enable the existence of the defect to be discovered; or (f) in the case of a manufacturer of a component, that the defect is attributable to the design of the product in which the component has been fitted or to the instructions given by the manufacturer of the product”.

4.3. Addressing the gap in the Liability Framework for Bias in Facial Recognition Technology in the EU

4.3.1. European non-discrimination law

The potential consequences of misidentification in FRT are far-reaching and can have significant impacts on individuals and society as a whole. Misidentification occurs when the technology wrongly matches an individual's face with another person's or fails to accurately identify an individual, leading to a range of negative effects.

In cases such as the Proctorio incident, where a student initiated a legal case against her university in the Dutch Human Rights Court, the principle of non-discrimination was at the forefront. The student's lawsuit sought to determine whether the university had engaged in racial discrimination by utilizing the Proctorio software for online exam monitoring (DHRC, 2022).

The cases presented in this research previously highlight the potential consequences of inaccurate FRT implementation, leading to discrimination outcomes. Overall, misidentification in FRT can unintentionally result in discriminatory outcomes for certain demographic groups[25]. In the EU, discrimination is protected by both primary and secondary law[26](Madiega and Hendrik, 2021, p. 17), underlining the significance of addressing these issues.

Such outcomes are an infringement of the principle of non-discrimination, safeguarded by the laws of Europe, which holds a central role in advancing equality and justice across diverse sectors of society and prohibits any discrimination based on race and gender, among others. The EU anti-discrimination framework it is entrenched in various treaties and legislations of the European Union, including Article 14 and Protocol 12 of the European Convention on Human Rights (1950)[27]

[25] See III.3.2.
[26] Racial Equality Directive (2000/43/EC), the Employment Equality Directive (2000/78/EC), the Gender Goods and Services Directive (2004/113/EC), and the Gender Equality Directive (2006/54/EC).
[27] Article 14 ECHR: “Prohibition of discrimination: The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status [...] Protocol No. 12: Article 1: General prohibition of discrimination. 1. The enjoyment of any right set forth by law shall be secured without discrimination

and Article 21 of the European Charter of Fundamental Rights (2012)[28]. According to Hacker (2018, p. 8), besides being a fundamental right, it is also an essential principle within EU market law, exemplified by several anti-discrimination directives.

There is a notable gap within the EU anti-discrimination framework when it comes to addressing cases involving AI systems. This is mostly due to the fact that harm done by an FRT is challenging to address due to its complexity. Lack of algorithm knowledge makes it tough for victims to establish comparators, prove disparities, or counter justifications and are difficult to understand for non-experts like judges and victims (Madiega and Hendrik, 2021, p. 17).

4.3.2. Moral Damages

Amidst the growing concerns surrounding FRT, a gap in the existing legal framework has become evident, particularly in addressing bias and its potential consequences. The path to get compensation measures for people impacted by biased FRT applications has shown to be difficult and complicated as the EU tries to adapt traditional laws to the changing scenario.

The term moral damages refers to “non-property loses due to moral or physical suffering or other adverse phenomena caused to a natural or legal person by illegal actions or omissions of other persons”. Compensating for moral harm is challenging due to the inability to determine the precise compensation amount that corresponds to the inflicted damage. Therefore, it becomes essential for courts to exercise their judgment based on the principle of reasonableness (Basenko, 2022, p. 6).

The path to pursue compensation for moral damages stemming from biased FRT predominantly involves contractual liability and tortious liability (European

on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status. 2. No one shall be discriminated against by any public authority on any ground such as those mentioned in paragraph 1”.

[28] Article 21 ECFR: “Non-discrimination. 1. Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited. 2. Within the scope of application of the Treaties and without prejudice to any of their specific provisions, any discrimination on grounds of nationality shall be prohibited”.

Commision, 2019, p. 6). *As per* Raposo (2023, p. 5), there exists a theoretical possibility for individuals who are inaccurately identified by FRT, resulting in public embarrassment and humiliation, to potentially pursue compensation for moral damages from the manufacturer and/or user. Although, this possibility may be more theoretical than practical depending on the legal tradition from common law to civil law.

The foundation of the legal framework in the common law tradition is a set of unwritten laws established through legal precedents by the courts (Chen, 2022). This system is present in countries like Australia, England, and the United States (Federation of Law Societies of Canada, 2021, p. 1). Similar to the concept of moral harm presented, Ben-Shahar and Porat (2018, p. 1902) address the complexities of seeking compensation for emotional harm within common law jurisdictions. According to them, emotional harm is challenging to assess and quantify; therefore, traditional remedies in private law such as monetary compensation may not effectively address this issue.

It is hard to receive compensation for this type of harm arising from contract law. Most justifications are based on the parties' intentions before the event. A few explanations for this include the unforeseeable nature of this kind of harm, the speculative character, and the possibility of avoiding it by the other party purchasing the service from someone else (Ben-Shahar and Porat, 2018, pp. 1907-1908).

In cases arising from tort law, compensation is generally more feasible than in contract law. However, only when physical injury is involved — "stand-alone emotional harm, not accompanied by physical injury, is generally uncompensated under tort law unless intentionally inflicted" (Ben-Shahar and Porat, 2018, p. 1912).

Therefore, according to Ben-Shahar and Porat (2018, pp. 1902-1912), obtaining compensation solely for non-physically or non-pecuniarily inflicted emotional harm is challenging to achieve, when intentional physical or pecuniary harm is not involved.

Moving beyond the compensation mechanisms for moral damages in common law jurisdictions, it is crucial to explore how this process is approached in European continental law, also known as civil law jurisdiction. These jurisdictions have distinct

legal landscapes and principles compared to common law countries, resulting in different avenues and considerations when seeking compensation for moral damages.

Based on Roman legal traditions, European continental law distinguishes between moral damages and patrimonial damages, with a more flexible approach towards compensating for moral harm (Magnus, 2012). According to the Principles of European Tort Law (PETL)[29], moral damages are defined as follows:

> "Art. 10:301. Non-pecuniary damage
>
> (1) Considering the scope of its protection (Article 2:102), the violation of an interest may justify compensation of non-pecuniary damage. This is the case in particular where the victim has suffered personal injury; or injury to human dignity, liberty, or other personality rights. Non-pecuniary damage can also be the subject of compensation for persons having a close relationship with a victim suffering a fatal or very serious non-fatal injury.
> (2) In general, in the assessment of such damages, all circumstances of the case, including the gravity, duration and consequences of the grievance, have to be taken into account. The degree of the tortfeasor's fault is to be taken into account only where it significantly contributes to the grievance of the victim.
> (3) In cases of personal injury, non-pecuniary damage corresponds to the suffering of the victim and the impairment of his bodily or mental health. In assessing damages (including damages for persons having a close relationship to deceased or seriously injured victims) similar sums should be awarded for objectively similar losses" (European Group on Tort Law, 2023, p.13).

In tort law, an individual can seek compensation solely based on moral damages independently of any other harm, such as when a biased FRT causes instances of distress from a misidentification (Raposo, 2023, p. 5).

However, not all forms of moral damages will be compensated. This variation occurs within civil law jurisdictions, which often impose restrictions on this type of compensation. Therefore, the success of an individual in obtaining compensation for a case of misidentification will depend on the national law and the court's determination of whether the harm was serious enough to warrant compensation under the rules of tort law (Raposo, 2023, p. 5).

In Europe, it is now widely recognized that compensation should extend beyond just financial loss. However, variations exist among national laws regarding

[29]The PETL are essentially a compilation of fundamental principles that, despite variations in specific details, are universally shared across the tort laws of Member-States. These principles serve as the foundational framework for the respective domestic laws governing non-contractual liability (Magnus, 2012).

which non-pecuniary losses are eligible for compensation and to what degree (Wurmnest, 2012). This focus particularly examines the possibility of obtaining compensation for moral damages based solely on emotional distress, public embarrassment, and humiliation resulting from a misidentification by an FR system.

In essence, legal systems influenced by the French Code civil tend to be more liberal in granting compensation for non-pecuniary loss. In contrast, other countries only allow such compensation when explicitly outlined by special laws. This caution stems from concerns about excessive liability and the challenges of assessing this kind of loss. This restrictive approach is present in Austrian, Dutch, German, and Italian laws, though with differing criteria for plaintiff success (Wurmnest, 2012).

To illustrate this point further, let's consider a few examples from different Member-States. In Germany, the Civil Code restricts compensation for non-pecuniary loss mainly to cases involving bodily injuries and false imprisonment. Specifically, the Bürgerliches Gesetzbuch (BGB) § 253[30] outlines that monetary compensation for non-pecuniary damage is only permissible in situations explicitly defined by the law. Additionally, if damages are applicable due to harm to the body, health, freedom, or sexual self-determination, equitable monetary compensation might also be sought for damages that aren't related to financial loss.

Similarly, *as per* the Italian Civil Code, Article 2.059 stipulates that non-pecuniary damages (*danno morale)* should only be compensated in cases explicitly provided for by the law, *in verbis:* "*Il danno non patrimoniale deve essere risarcito solo nei casi determinati dalla legge*".

Moreover, article 185 of the Italian Criminal Code[31] states that every crime obliges restitution, in accordance with civil laws. Every crime, which has caused pecuniary or non-pecuniary damage, shall oblige the offender and the persons who, according to civil laws, shall be liable for it, to make restitution. Nevertheless, in some

[30] § 253 Bürgerliches Gesetzbuch (BGB): "Immaterieller Schaden (1) Wegen eines Schadens, der nicht Vermögensschaden ist, kann Entschädigung in Geld nur in den durch das Gesetz bestimmten Fällen gefordert werden. (2) Ist wegen einer Verletzung des Körpers, der Gesundheit, der Freiheit oder der sexuellen Selbstbestimmung Schadensersatz zu leisten, kann auch wegen des Schadens, der nicht Vermögensschaden ist, eine billige Entschädigung in Geld gefordert werden".

[31] Article 185 Italian Criminal Code: (Restituzioni e risarcimento del danno) Ogni reato obbliga alle restituzioni, a norma delle leggi civili. Ogni reato, che abbia cagionato un danno patrimoniale o non patrimoniale, obbliga al risarcimento il colpevole e le persone che, a norma delle leggi civili, debbono rispondere per il fatto di lui".

cases, this legal protection was not enough. Italian courts introduced the concept of "*danno biologico*" or "*danno alla salute*," which means harm to health or well-being. This was done to ensure that people can get compensation for harm that affects their well-being even if no criminal law was broken. According to this rule, if someone's health or well-being is affected by someone's actions, they can seek compensation under the general law for torts, which is a law that deals with civil wrongs, regardless of whether a criminal law was violated (Wurmnest, 2012).

In situations where mistaken identifications happen due to FRT and cause harm, it's challenging to imagine a scenario where criminal liability would apply. This is especially true when the harm is mainly emotional and doesn't involve physical or pecuniary (financial) harm. Criminal liability usually relates to more serious offenses, and in this context, it might only come into play in cases involving crimes against honor (Raposo, 2023, p. 5). In such cases, the concept of "*danno biologico*" or "*danno alla salute*" discussed earlier might not apply directly, as it's more focused on physical and health-related harm.

In France, the initial version of the Civil Code did not address the concept of moral damages (“*dommage moral”)*. However, the legal landscape has since evolved, creating a well-defined avenue for pursuing compensation for moral damages (Palmer, 2015, pp. 58-76). French courts now widely acknowledge and award compensation for moral damages, even in cases where the harm is purely emotional[32]. Moreover, this possibility extends to cases involving breaches of contract (Borghetti, 2015, pp. 268-288)

The pursuit of compensation for moral damages arising from emotional distress, public embarrassment, and humiliation due to misidentification by an FRT faces a complex and often uncertain legal landscape in Europe. While the legal principles within both common law and civil law jurisdictions provide potential avenues for seeking such compensation, the variability in laws and interpretations across different European countries adds an additional layer of complexity. As a result, the determination of whether to award moral damages in such cases often hinges on the discretion of the courts. This uncertainty underscores the need for

[32] Article 1382 and 1383 of the French Civil Code.

ongoing discussions and potential legal reforms to provide clearer guidelines on the eligibility and extent of compensation for these unique and evolving forms of harm.

4.3.2. Liability of Stakeholders

In the broader context of exploring the liability landscape surrounding biased FRT, it's imperative to bridge the gap between the issue itself and the responsibilities shared by stakeholders. The examination of biased FRT and its associated liability considerations spans across a spectrum of individuals and entities involved.

Specifically, when it comes to the liability of users, be they individuals or organizations employing FRT systems, it is crucial to take into consideration that FRT's probabilistic nature introduces an element of decision-making responsibility. This interplay between biased FRT and user liability underscores the complex ethical and legal implications that arise in this technological landscape.

Misidentification in FRT sometimes can be attributed to human errors during oversight due to the inherent probabilistic nature of this technology. Users of FRT must be aware of this unique characteristic and remain attentive to the potential automatic biases that may lead them to overlook external evidence, consequently resulting in inaccurate judgments (Fernandez *et al.*, 2020, p 39). A possible approach to mitigate this challenge involves setting a higher threshold. For instance, Amazon's FRT, Rekognition, advises utilizing a threshold of 99% or higher in scenarios where the precision of classification could negatively impact the individuals[33].

Drawing a parallel to the previously discussed case of the ride-sharing driver, a similar situation could unfold where the company — user — could potentially face liability in relation to biased FRT. A driver was removed from the company's platform after being suspended due to alleged account sharing. He claimed that the company's facial recognition system failed to recognize him when he attempted to

[33] “We recommend using a threshold of 99% or more for use cases where the accuracy of classification could have any negative impact on the subjects of the images” (Amazon, 2023, p. 155).

sign in to work. Imagine a scenario where the driver did not share his account, and the system failed to recognize his face due to bias in the algorithm.

Despite undergoing a human review by the company, the definitive decision to expel the driver from the platform was upheld. In such a case, the company's reliance on the outcome of the biased FRT system and the subsequent human review could lead to questions of liability. Despite the FRT not being directly developed by the company in question, the fact that a human review took place and the driver's expulsion was upheld raises the possibility of holding the user (the company) accountable for potential liability.

This scenario gives rise to two additional crucial points that need to be considered. Firstly, there is the potential bias of human operators towards placing trust in algorithmic findings, particularly when certain matches are indicated. This bias might inadvertently influence the outcome of human reviews, impacting the overall fairness and objectivity of the decision-making process[34]. Secondly, users often lack comprehensive information about the inner workings of the facial recognition system they are using, including details about the data quality and training databases employed. This lack of transparency can be attributed to concerns related to copyright issues and the protection of trade secrets, making it challenging for users to fully understand how the system operates and assess its reliability (European Union Agency for Fundamental Rights, 2019, p.10).

According to Raposo (2020, p. 5), the liability of the user becomes quite evident in certain cases involving the misuse of FRT. For instance, consider a scenario where a user tampers with the high-resolution camera included in the FRT system and replaces it with a lower-quality one. This alteration leads to erroneous identifications. In this situation, attributing liability to the manufacturer becomes difficult, as the issue arises solely from the user's misuse.

Another example to consider is when the intended usage of the FRT system is clearly stipulated to be in well-lit environments, as specified in usage guidelines. However, the user decides to employ it in a poorly lit area. This decision could result in images of lower quality, leading to inaccurate identifications. Consequently, the

[34] See II.3.

responsibility for such occurrences falls on the user, as they have deviated from the recommended usage conditions (Raposo, 2020, p. 5).

Furthermore, examining the responsibility of both developers and manufacturers necessitates a consideration of the Product Liability Directive (PLD) and the AI Liability Directive, which has been recently proposed by the European Commission.

As mentioned before, the PLD, which has been in place for several years, lays down the groundwork for assigning liability to manufacturers when their products are defective. Under Article 6(1) “A product is defective when it does not provide the safety which a person is entitled to expect” (EU, 1985).

However, it's important to note the PLD introduces a stringent definition of what constitutes a defect, tightly linked to safety concerns. Consequently, a scenario involving misidentification by a FRT system would likely fall outside the scope of this framework. This is because a misidentification issue, while significant in its impact, may not be classified as a safety-related malfunction, thus potentially limiting the applicability of the PLD in addressing liability matters arising from non-safety-related defects (Raposo, 2020, p. 6).

For instance, consider a scenario where an individual is wrongfully arrested due to a misidentification facilitated by a flawed FRT. Despite the emotional distress and harm caused by the wrongful arrest, this situation might not fall within the ambit of the PLD. This is because the PLD, *as per* in sector IV.3 notes, is primarily oriented towards protecting the physical well-being and property of consumers. While the emotional harm suffered by the misidentified person is undoubtedly substantial, it doesn't inherently correspond to the concept of a lack of safety, as envisaged by the PLD. Therefore, even in cases where serious moral harm results from misidentification – which is particularly likely in identification activities carried out by law enforcement agencies – the PLD's scope may not extend to cover such non-safety-related defects (Raposo, 2020, p. 6).

Furthermore, it has been firmly established that the PLD is applicable exclusively to products and does not encompass services. The significance of this distinction was underscored in the ruling of Case C-65/20 VI v. KRONE in 2021,

where the court's interpretation clarified this matter. The pertinent excerpt from the case reads as follows:

> “ By its question, the referring court asks, in essence, whether Article 2 of Directive 85/374, read in the light of Articles 1 and 6 thereof, must be interpreted as meaning that a copy of a printed newspaper that, concerning paramedical matters, gives inaccurate health advice relating to the use of a plant which, when followed, has proved injurious to the health of a reader of that newspaper, constitutes a ‘defective product’ within the meaning of those provisions [...] **It is apparent from the wording of that article that services do not come within the scope of that directive** [...] **Were such advice to come within the scope of Directive 85/374, this would not only result in the negation of the distinction drawn by the EU legislature between goods and services and the exclusion of the latter from the scope of that directive, but would also make newspaper publishers strictly liable without it being possible for them – or with a limited possibility for them – to avoid that liability.** However, such a consequence would be detrimental to the objective of ensuring that risk is fairly apportioned between the injured person and the producer, as recalled in the seventh recital of that directive” (VI v. KRONE-Verlag Gesellschaft mbH & Co KG, 2021).

In this ruling, the Advocate General's opinion further clarifies that the directive solely applies to the physical components of a product, *“In my view, it is perfectly clear from the language, objectives and context of that directive that it applies to the physical properties of products only, so that it is not applicable in a case of this kind*” (Advocate General, 2021).

Indeed, it is challenging to envision a scenario where a case of misidentification stemming from FRT would fall under the purview of the PLD. A revised version has been proposed, with the aim of modernizing the current no-fault-based (strict) product liability framework within the European Union. This proposed revision is designed to address the evolving landscape of technological advancements and their associated risks (Madiega, 2023, p. 5).

However, even in the revised version of the PLD, it remains challenging to envision a scenario where a case of misidentification resulting from this technology would fall under its purview. The updated definition within this new version specifically includes software as a product in Article 3, stating: “‘*product’ means all movables, even if integrated into another movable or into an immovable. ‘Product’ includes electricity, digital manufacturing files and software*” (European Commission, 2022b). Consequently, FRT software would be encompassed within the scope of the PLD. Therefore, those adversely affected can pursue legal action against the manufacturer under the established strict liability regime, while not as foreseeable within the context of lack of safety (Raposo, 2023, p. 6).

In an effort to modernize the EU's liability framework, the European Commission proposed the 'AI liability directive' in September 2022, with the aim of adapting non-contractual fault-base civil liability rules to claim for damages of AI (Madiega, 2023, p. 1). The concept of an AI system in the AI Act is consistent with that of the AI liability directive as well as high-risk AI systems. According to the latest version of the AI Act, an AI system is a machine-based system that has differents levels of autonomy and employs techniques to generate outputs with implicit or explicit objectives[35] and FRT categorizes as a high-risk system[36] (European Parliament, 2023).

The AI Liability Directive introduces a presumption of causality, simplifying the process for individuals seeking compensation to establish liability claims successfully (Madiega, 2023, p. 6)[37]. Moreover, the court is empowered to grant a disclosure of evidence (Article 3) in a manner that is both reasonable and proportionate to the plaintiff's liability claim request. This requirement for evidence disclosure extends to all parties involved, including third parties protecting trade secrets (Article 3(4)). In cases where a defendant fails to comply with a court order, it will be presumed that they have not fulfilled their duty of care (Article 3(5)) (European Commission, 2022d). In accordance with Raposo's analysis (2022, p. 6) this disclosure of evidence might pertain to the quality and quantity of images utilized for training and testing the FR system, as well as details about the system's code.

In cases where the court orders the disclosure of evidence that includes trade secrets or alleged trade secrets considered confidential, national courts have the authority to take steps to protect the confidentiality of that information during legal proceedings. This ensures that while evidence is being used or discussed in court, any trade secrets or confidential information will be kept confidential as required by Directive (EU) 2016/943 (European Commission, 2022c).

[35] AI Act “Article 3(1): ‘‘artificial intelligence system’ (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments;.”
[36] Annex III AI Act: “The AI systems specifically refered to in under points 1 to 8a stand for critical use cases and are each considered to be high-risk AI systems pursuant to Article 6(2), provided that they fulfil the criteria set out in that Article. 1. Biometric and biometrics-based systems.”.
[37] See article Article 4, Recital 25 and 26 of the draft of AI Liability Directive.

The path to obtaining compensation in liability lawsuits is marked by vagueness, complexity, and various legal gaps. Despite Europe's endeavor to modernize its laws through the draft of the AI Liability Directive, seeking compensation in cases involving FRT faces several challenges due to the inherent characteristics of the system. Opacity within the production chain makes it challenging to attribute responsibility. Emerging biases may only become evident once the technology is already in use. Moreover, trade secret considerations further complicate the landscape. Additionally, as a probabilistic technology, FRT provides identifications based on probabilities, such as surpassing a certain threshold, raising issues of human oversight. In light of these complexities, the path to securing compensation remains intricate and multifaceted.

4.3.4. Trade Secrets and Transparency

Within the domain of FRT, the intricate interplay between trade secrets and AI transparency becomes a pivotal subject of examination. The concept of AI transparency can be defined as articulated by Rita Matulionyte (2023), “a requirement to provide information about the AI model, its algorithm and data [...] could require disclosing very general information, such as ‘when AI is being used’, or more specific information about the AI module, e.g. its algorithmic parameters, training, validation and testing information” (Matulionyte, 2023, p. 3). This discussion converges with the provisions laid out in the draft of the AI Liability Directive, which empowers courts to order the disclosure of evidence (Article 3) in a manner that aligns with the reasonableness and proportionality of the plaintiff's liability claim request (European Commission, 2022d).

According to Article 2(1) of Directive (EU) 2016/943, trade secrets are defined as information that (i) is known only among key individuals, (ii) holds commercial value, and (iii) has been kept confidential through appropriate measures by the lawful possessor of the information. Trade secrets are a form of protecting intellectual property (IP) rights in AI systems, such as FRT (Matulionyte, 2023, p. 9).

Another layer of complexity arises from trade secrets in the path to seek compensation in liability lawsuits in FRT. Companies might resist disclosing the inner workings of their technology due to concerns about protecting their trade secrets information.

To comprehend which information within FRT could potentially be safeguarded as trade secrets, Rita Matulionyte (2023) outlines which could potentially fall under trade secrets protection, considering the prerequisites of confidentiality and commercial significance. These encompass: (i) the "algorithm's architecture"; (ii) "details regarding training"; (iii) "validation and verification of the algorithm, encompassing training and validation/verification data, methods, and procedures"; and (iv) "information derived from real-life testing"(Matulionyte, 2023, p. 9).

Even if stakeholders assert that such information about FRT is safeguarded by trade secrets, in the context of liability claims, court-ordered evidence disclosure might be sought to gain insights into data quality or training methodologies. This effort could involve checking for compliance with legal requisites during FRT system development and assessing the technology's accuracy as well as determining where liability lies. The Article 3 of Directive (EU) 2016/943 safeguards the confidentiality of trade secrets during legal proceedings and the issue can be addressed through established confidentiality and trade secret rules in certification/auditing processes or court procedures (Matulionyte, 2023, p. 12).

4.6. Mitigating Bias and Ethical considerations

In an evolving landscape where FRT is permeating diverse sectors such as law enforcement and commercial applications, the potential for biased outcomes and ethical dilemmas becomes increasingly pronounced. This heightened risk raises concerns about subjecting individuals to potential discrimination, public embarrassment, or infringements on their rights. The urgency of addressing bias inherent in FRT algorithms and ensuring their ethical deployment goes beyond safeguarding societal well-being — it extends directly to liability considerations.

In a recent landmark case, the Third Section of the European Court of Human Rights (ECtHR, 2023) delved into the compatibility of FRT with human rights in the Glukhin v. Russia case. While the primary focus of the case revolved around the law enforcement application of FRT, an area beyond the scope of this research, several noteworthy observations emerged, shedding light on the importance and ongoing dialogues surrounding this technology. The court highlighted that " the use of facial recognition has direct or indirect impact on a number of fundamental rights and freedoms enshrined in the EU Charter of Fundamental Rights that may go beyond privacy and data protection, such as human dignity, freedom of movement, freedom of assembly, and others". It also noted the potential for widespread discrimination due to this technology's large-scale implementation (Glukhin v. Russia, 2023).

Nevertheless, as previously mentioned, the technologically advanced landscape has already led to the widespread use of FRT. Consequently, considering an outright ban or complete avoidance of its usage is both unrealistic and impractical. However, FRT does exhibit biases, capable of causing harm to specific demographic groups.

While the European legal framework may still present gaps that pose challenges for obtaining compensation in cases of misidentification, efforts are underway to bridge this gap. Legislative measures are being actively pursued to provide individuals with avenues for seeking damages resulting from emerging technologies, just as they would for conventional products or technologies. Notable developments in this regard include the AI Act and the AI Liability Directive, which signify Europe's commitment to modernizing its legal framework to align with the complexities of contemporary technological advancements. These initiatives are trying to establish a path that holds organizations/companies accountable for potential harms stemming from their technology, thereby ensuring that individuals can seek compensation for damages caused by FRT misidentification and similar emergent technologies.

In this context, there is a clear need for the exploration of strategies aimed at mitigating bias, upholding ethical principles and promoting innovation. According to Fernandez *et al.* (2020, p. 80) significant consideration lies in the dominance of the U.S. National Institute for Standards and Technology (NIST) in global facial

recognition technology standards. NIST's evaluation criteria hold sway globally, even in European tenders. Facial Recognition Vendor Test (FRVT) is present in most reports, academic and government texts as the main reference, according to Fernandez *et al.* (2020) “this predominance is also made possible by the absence of a European equivalent to the NIST”.

However, these standards primarily focus on technical aspects. To protect fundamental rights and freedoms while countering biases and ethical challenges, the EU needs a robust standardization system for facial recognition technologies. Ongoing assessment of compliance with evolving European standards is crucial to keep pace with the technology's changes and implications (Fernandez *et al.*, 2020, p. 7).

Thus, maintaining continuous monitoring policies, including impact assessments, is essential. Currently, there is no mandatory mechanism to assess the impact on fundamental rights before deploying FRT, except for regulations related to personal data processing. Auditing forms the bedrock of any standardization system; without auditable standards, compliance tracking is challenging. To ensure accountability, establishing such standards could lead to a common certification framework across EU member states. It's important to highlight that certifications for facial recognition technology should be constant due to the ever-evolving nature of this technology, requiring ongoing monitoring by certifying bodies (Fernandez *et al.*, 2020, p. 95).

As well mentioned by Fernandez *et al.* (2020, p. 6) “any decision made in which facial recognition technology is involved is the result of a chain of events. Therefore, it is essential to ensure that the decisions at each step in the chain are explainable, right up to the human decision”.

Drawing from Coeckelbergh (2020) and Friedman and Nissenbaum's (1996) classification of bias, this study identifies three primary sources of bias in AI systems: programmers, data, and human interactions. To effectively mitigate these biases, several strategic measures need to be implemented.

Addressing the issue of biased programming requires tackling the lack of diversity within development teams, which often consist predominantly of white

males. Introducing a more diverse team composition is essential to counteract this bias at the source. Concerning biased data, a main step involves creating a more inclusive dataset that accurately represents a broader range of demographics. Lastly, bias stemming from human interactions necessitates heightened awareness about the potential biases embedded within AI systems. Additionally, understanding the probabilistic nature of this technology is crucial, as it prompts individuals to acknowledge external factors rather than relying solely on algorithmic outcomes. This collective approach stands as a comprehensive strategy to mitigate bias effectively (Coeckelbergh, 2020; Friedman and Nissenbaum, 1996).

V. CONCLUSION

Throughout this dissertation, it was analysed algorithmic bias in FRT within the context of liability lawsuits in the current legal framework of the EU. In this context, it has become clear that the rapid advancement of FRT has shed the light of legal and ethical concerns related to bias inherent in these systems.

Bias is an ethical and societal issue and it is perpetuating societal prejudices within algorithms. Several cases have been reported in which FRT does not work properly in certain demographic groups, harming individuals in various spheres of everyday life. As examined in this research, academia is already addressing this issue. It is noteworthy to highlight the predominance of NIST as the primary institute addressing reference standards for FRT algorithms.

The widespread use of these technologies is unquestionable, as is the damage they can do to certain individuals when they don't work to the same accuracy. Even so, banning shouldn't be discussed, every AI system has its advantages and they should be used by everyone. Instead it should be discussed effective mechanisms to enforce accountability, address bias, and provide avenues for compensation to the affected individuals.

The study identified three primary sources of bias in FRT: programmers, data, and human interactions. Addressing these sources are necessary to ensure fairness

and non-discriminatory outcomes, such as diversity within development teams, a more inclusive dataset and awareness about the potential biases. While it has been discussed methods to prevent and detect bias, there remains a significant challenge in seeking compensation through liability lawsuits.

Addressing liability issues in AI systems is difficult, mostly due to certain features, such as opacity, transparency, autonomous behavior, and explainability. Addressing liability issues in cases of misidentification caused by biased FRT is challenging, mainly because current law lacks enforcement in AI systems. As illustrated, there is no legal personhood in AI system. Therefore, addressing the liability resulting from bias in FRT is related to the humans who own and control such systems.

The current legal framework in the European Union regarding liability is partially harmonized. Cases of misidentification by FRT are unlikely to fall within the scope of PLD. This is because the concept of ‘defect’ under the PLD is closely tied to safety considerations, which may not apply directly to cases of misidentification.

The path to seek competition regarding moral damages will vary between jurisdictions, with potential avenues for seeking such compensation differing across European countries and depending on the courts. When considering stakeholder liability, a few factors must be examined, especially the probabilistic nature of this technology. As discussed, clear liability arises in cases of misuse, typically attributed to the user or when there is human oversight. In cases of liability claims, data can be requested by the court, and trade secrets may become an issue as companies might resist disclosing the data of their technology. Nevertheless, it is enforced through legal means the confidentiality of trade secrets throughout judicial processes.

In the end, despite efforts to adapt to new technologies, the current legal framework in the EU struggles to adequately address compensation for biased FRT. As FRT continues to expand, it is imperative that the legal framework and standards evolve to foster innovation while maintaining accountability and robust enforcement of liability rules.

REFERENCES

Advocate General. (2021). VI v. KRONE-Verlag Gesellschaft mbH & Co KG, 10 June 2021 (Case C-65/20), ECLI:EU:C:2021:471.

ALTURIS AI. (2022). Ethics of AI - Opacity of AI Systems. https://www.alturis.ai/post/ethics-of-ai-opacity-of-ai-systems.

Amazon. (2023). Amazon Rekognition Developer Guide. https://docs.aws.amazon.com/pdfs/rekognition/latest/dg/rekognition-dg.pdf.

Apple. (2023). About Face ID Advanced Technology. https://support.apple.com/en-us/HT208108.

Article 29 Data Protection Working Party. (2012). Opinion 02/2012 on facial recognition in online and mobile services (00727/12/EN, WP 192). https://www.pdpjournals.com/docs/87997.pdf

Bansal, V. (2022). Uber's Facial Recognition Is Locking Indian Drivers Out of Their Accounts. MIT Technology Review. https://www.technologyreview.com/2022/12/06/1064287/ubers-facial-recognition-is-locking-indian-drivers-out-of-their-accounts/.

Barocas, S., Rosenblat, A., Boyd, D., Gangadharan, S. P., and Yu, C. (2017). Data & Civil Rights: Technology Primer. Data & Society Research Institute. https://www.datacivilrights.org/pubs/2014-1030/Technology.pdf

Barry, E. (2021). Uber Drivers Say a 'Racist' Algorithm Is Putting Them Out of Work. Time. https://time.com/6104844/uber-facial-recognition-racist/.

Basenko, R., Avanesian, H., & Strilko, D. (2022). Institute of compensation for moral damage: international legal experience and legislative innovations. Entrepreneurship, Economy and Law, 1, 5–10.https://doi.org/10.32849/2663-5313/2022.1.01

BBC. (2020). IBM Abandons 'Biased' Facial Recognition Tech. https://www.bbc.com/news/technology-52978191.

Ben-Shahar, O., & Porat, A. (2018). The restoration remedy in private law. Columbia Law Review, 118(6), 1901-1952. https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=13819&context=journal_articles

Benjamin, R. (2019). Race after Technology: Abolitionist Tools for the New Jim Code. Polity Press.

Bibe, M. (2021). DABUS: The 'Natural Person' Problem. Inventa. https://www.inventa.com/pt/noticias/artigo/681/dabus-the-natural-person-problem.

Bledsoe, W.W. (1963). A Facial Recognition Project Report. Public Domain. https://archive.org/details/firstfacialrecognitionresearch/FirstReport/.

Boden, M. A. (2016). AI: Its Nature and Future. Oxford University Press UK.

Borghetti, J-S. (2015). Non-Pecuniary Damages in France. The Chinese Journal of Comparative Law, 3(2).https://doi.org/10.1093/cjcl/cxv012

Buolamwini, J. (2016). InCoding — In The Beginning Was The Coded Gaze. MIT Media Lab.https://medium.com/mit-media-lab/incoding-in-the-beginning-4e2a5c51a45d.

Buolamwini, J. (2017). Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers. Massachusetts Institute of Technology, 3-116. https://dspace.mit.edu/handle/1721.1/114068

Buolamwini, J. (2018). Fighting the "Coded Gaze": How We Make Artificial Intelligence Benefit All. Public Interest Tech, Ford Foundation. https://www.fordfoundation.org/news-and-stories/videos/joy-buolamwini-fighting-the-coded-gaze-how-we-make-artificial-intelligence-benefit-all-public-interest-tech/.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research, 81, 1–15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Chellappa, R., Wilson, C.L., and Sirohey, S. (1995). Human and machine recognition of faces: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(5), 705-741. 10.1109/5.381842.

Chen, J. (2022). Common Law: What It Is, How It's Used, and How It Differs From Civil Law. Investopedia.https://www.investopedia.com/terms/c/common-law.asp

Clark, M. (2021). Students of color are getting flagged to their teachers because testing software can't see them. The Verge. https://www.theverge.com/2021/4/8/22374386/proctorio-racial-bias-issues-opencv-facial-detection-schools-tests-remote-learning;

Clayton, J., & Derico, B. (2023). Clearview AI used nearly 1m times by US police, it tells the BBC. BBC. https://www.bbc.com/news/technology-65057011.

Coeckelbergh, M. (2020). AI Ethics. The MIT Press.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). Introduction to Algorithms (3rd ed.). The MIT Press. https://pd.daffodilvarsity.edu.bd/course/material/book-430/pdf_content

Crumpler, W. (2020). How Accurate are Facial Recognition Systems – and Why Does It Matter? Center for Strategic & International Studies. https://www.csis.org/blogs/strategic-technologies-blog/how-accurate-are-facial-recognition-systems-and-why-does-it.

Danks, David and London, Alex. (2017). Algorithmic Bias in Autonomous Systems. 4691-4697. 10.24963/ijcai.2017/654.

Dignum, V. (2018). On Bias, Black-Boxes, and the Quest for Transparency in Artificial Intelligence. Medium. https://medium.com/@virginiadignum/on-bias-black-boxes-and-the-quest-for-transparency-in-artificial-intelligence-bcde64f59f5b.

Drozdowski, P., Rathge, C., Dantcheva, A., Damer, N., & Busch, C. (2020). Demographic Bias in Biometrics: A Survey on an Emerging Challenge. IEEE Transactions on Technology and Society, 1(2), 98-99. 10.1109/TTS.2020.2992344

Dutch Human Rights Court. Case No. 2022-146 (2022). Tussenoordeel. De Stichting Vrije Universiteit krijgt de gelegenheid om te bewijzen dat de door haar ingezette antispieksoftware een studente met een donkere huidskleur niet heeft gediscrimineerd. https://oordelen.mensenrechten.nl/oordeel/2022-146

DW. (2021). Privacy activists challenge Clearview AI in EU. https://learngerman.dw.com/en/privacy-activists-challenge-clearview-ai-in-eu/a-57691756;

Dwoskin, E., and Harwell, D. (2021). Facebook Is Ending Use of Facial Recognition Software, Deleting Data on More Than a Billion People. The Washington Post.https://www.washingtonpost.com/technology/2021/11/02/facebook-ends-facial-recognition/.

ECHR. (1950). European Convention on Human Rights. https://www.echr.coe.int/documents/d/echr/convention_ENG

Eubanks, V. (2018). Automating Inequality. New York: St. Martin's Press. ISBN: 978-1-250-07431-7

Eubanks, V. (2018). Automating Inequality. St Martin’s Press.

European Commission (2016a). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 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). https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Commission. (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 products. Official Journal of the European Communities, L

210, 7.8.1985, 29. As amended by Directive 1999/34/EC of the European Parliament and of the Council, Official Journal of the European Communities, L 141, 20.4.1999.

European Commission. (2014). Directive 2014/104/EU of the European Parliament and of the Council of 26 November 2014 on certain rules governing actions for damages under national law for infringements of the competition law provisions of the Member States and of the European Union. Official Journal of the European Union, L 349, 5.12.2014. Text with EEA relevance.

European Commission. (2016b). Defective products: liability. https://eur-lex.europa.eu/EN/legal-content/summary/defective-products-liability.html.

European Commission. (2022a). Impact Assessment Report Accompanying the Document Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence. Commission Staff Working Document, 319. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022SC0319

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

European Commission. (2022c). Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.

European Commission. (2022d). 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).

European Court of Human Rights. (2023). Third Section. Case of Glukhin v. Russia (Application no. 11519/20).

European Data Protection Board. (2023). Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement. https://edpb.europa.eu/system/files/2023-05/edpb_guidelines_202304_frtlawenforcement_v2_en.pdf.

European Group on Tort Law. (2023). Principles of European Tort Law. http://www.egtl.org/docs/PETL.pdf

European Parliament. (2020). Civil Liability Regime for Artificial Intelligence. P9_TA(2020)0276. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.pdf.

European Patent Office, J 0008/20 (Designation of inventor/DABUS). (2021). https://new.epo.org/en/boards-of-appeal/decisions/j200008eu1.html.

European Union Agency for Fundamental Rights. (2019). Facial recognition technology: fundamental rights considerations in the context of law enforcement. https://fra.europa.eu/en/publication/2019/facial-recognition-technology-fundamental-rights-considerations-context-law.

Expert Group on Liability and New Technologies. (2019). Liability for Artificial Intelligence and Other Emerging Digital Technologies. European Commission, 3-65https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2020/01-09/AI-report_EN.pdf.

Federation of Law Societies of Canada National Committee on Accreditation. (2021). https://nca.legal/wp-content/uploads/2021/10/NCA-Jurisdictions-Policies-Oct-2021.pdf

Fernandez, V., Galissaire, J., Laugier, L., Morat, G., Pouyat, M., Richard, A. (2020). Facial recognition: Embodying European values. Renaissance Numérique. https://www.renaissancenumerique.org/en/publications/facial-recognition-embodying-european-values/

Friedman, B., & Nissenbaum, H. (1996). Bias in Computer Systems. ACM Transactions on Information Systems, 14(3), 330–347. https://nissenbaum.tech.cornell.edu/papers/Bias%20in%20Computer%20Systems.pdf.

Girish, D. (2020). Coded Bias' Review: When the Bots Are Racist. The New York Times. https://www.nytimes.com/2020/11/11/movies/coded-bias-review.html.

Global Times. (2015). AI technology used to find missing child after 19 years. URL. https://www.globaltimes.cn/content/1165558.shtml.

Grossman, L. (2001). Welcome to the Snooper Bowl. Time. https://content.time.com/time/magazine/article/0,9171,999210,00.html.

Grother, P., Ngan, M., and Hanaoka, K. (2019a). Face Recognition Vendor Test (FRVT) Part 2: Identification. National Institute of Standards and Technology, 2-183. https://doi.org/10.6028/NIST.IR.8271.

Grother, P., Ngan, M., Hanaoka, K. (2019b). Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/ir/2019/nist.ir.8280.pdf

Hacker, P. (2018). Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law. Common Market Law Review, 55, 1143-1186. https://ssrn.com/abstract=3164973.

Hern, A. (2022). Microsoft Limits Access to Facial Recognition Tool in AI Ethics Overhaul. The Guardian. https://www.theguardian.com/technology/2022/jun/22/microsoft-limits-access-to-facial-recognition-tool-in-ai-ethics-overhaul.

Hill, K. (2020). The Secretive Company That Might End Privacy as We Know It. The New York Times. https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

Hill, K., and Mac, R. (2021). Facebook, Citing Societal Concerns, Plans to Shut Down Facial Recognition System. The New York Times. https://www.nytimes.com/2021/11/02/technology/facebook-facial-recognition.html

IANS. (2021). Meta to Continue Use of Facial Recognition Feature. Economic Times.https://economictimes.indiatimes.com/tech/technology/meta-to-continue-use-of-facial-recognition-feature/articleshow/87535165.cms.

IPWatchdog. (2021). DABUS Gets Its First Patent in South Africa Under Formalities Examination. https://ipwatchdog.com/2021/07/29/dabus-gets-first-patent-south-africa-formalities-examination/id=136116/.

Jain, G., & Parsheera, S. (2021). Cinderella's Shoe Won't Fit Soundarya: An Audit of Facial Processing Tools on Indian Faces. https://doi.org/10.48550/arXiv.2112.09326

Jasserand, C. (2022, May 5). Clearview AI: illegally collecting and selling our faces in total impunity? (Part II). CitTip. https://www.law.kuleuven.be/citip/blog/clearview-ai-illegally-collecting-and-selling-our-faces-in-total-impunity-part-ii/;

Johnson, A. (2023). Racism And AI: Here's How It's Been Criticized For Amplifying Bias. Forbes. https://www.forbes.com/sites/ariannajohnson/2023/05/25/racism-and-ai-heres-how-its-been-criticized-for-amplifying-bias/?sh=3673acc3269d;
Kantayya, S. (2020). Coded Bias. Brooklyn, New York City, New York, USA.

Kaur, P., Krishan, K., and Kanchan, T. (2020). Facial-recognition algorithms: A literature review. Medicine, Science and the Law, 60(2), 131-139. 10.1177/0025802419893168

Kelion, L. (2017). Apple iPhone X adopts facial recognition and OLED screen. BBC. https://www.bbc.com/news/technology-41228126.

Klare, B., Burge, M., Klontz, J., Vorder Bruegge, R., & Jain, A. (2012). Face Recognition Performance: Role of Demographic Information. IEEE Transactions on Information Forensics and Security, 7, 1789-1801. 10.1109/TIFS.2012.2214212.

Klosowski, T. (2020). Facial Recognition Is Everywhere: Here's What We Can Do About It. The New York Times. https://www.nytimes.com/wirecutter/blog/how-facial-recognition-works/

KPMG (2021). Facial recognition: Privacy considerations in access control, 1-6. https://assets.kpmg.com/content/dam/kpmg/nl/pdf/2021/services/facial-recognition-privacy-considerations-in-access-control.pdf.

Lai, Alicia. (2020). Artificial Intelligence, LLC: Corporate Personhood as Tort Reform. 2021 Mich, 2021, 597.http://dx.doi.org/10.2139/ssrn.3677360

Lee, N. T., Resnick, P., & Barton, G. (2019). Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms. Brookings. http://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/.

Lohr, S. (2018, February 9). Facial Recognition Is Accurate, if You're a White Guy. The New York Times. https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html;

López, A. (2001). Biometrics and Law Enforcement: Staring Privacy in the Face. Information Studies 246: Social and Cultural Impact of New Information, UCLA. https://besser.tsoa.nyu.edu/impact/w01/Papers/Lopez.htm#_ftn2.

Lowry, S., and Macpherson, G. (1988). A Blot on the Profession. British Medical Journal (Clinical Research Ed.), 296(6623), 657–658. https://doi.org/10.1136/bmj.296.6623.657.

Loyola-González, O. (2019). Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View. IEEE Access, 4, 1-19.10.1109/ACCESS.2019.2949286.

Lui, A. (2023). Snapchat Filters Don't Work On My Face: Here's Why. Nudest. https://www.nudest.co/naked-truths/snapchat-filters.

Mac, R. (2021). Facebook Apologizes After A.I. Puts 'Primates' Label on Video of Black Men. The New York Times. https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html.

Mac, R., Haskins, C., and McDonald, L. (2020). Clearview's Facial Recognition App Has Been Used By The Justice Department, ICE, Macy's, Walmart, And The NBA. BuzzFeed News. https://www.buzzfeednews.com/article/ryanmac/clearview-ai-fbi-ice-global-law-enforcement.

Madiega, T. (2023). Artificial Intelligence Liability Directive. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf

Magnus, U. (2012). Principles of European Tort Law (PETL). https://max-eup2012.mpipriv.de/index.php/Principles_of_European_Tort_Law_(PETL)#4._Guiding_principles

Mallon, B. (2003). Every Breath You Take, Every Move You Make, I'll Be Watching You: The Use of Face Recognition Technology. Villanova Law Review, 48 (3). https://digitalcommons.law.villanova.edu/vlr/vol48/iss3/6

Meaker, M. (2023). This Student Is Taking On 'Biased' Exam Software. Wired. https://www.wired.co.uk/article/student-exam-software-bias-proctorio.

Meaker, M. (2023). This Student Is Taking On 'Biased' Exam Software. Wired. https://www.wired.co.uk/article/student-exam-software-bias-proctorio;

Metz, C., and Singer, N. (2019). A.I. Experts Question Amazon's Facial-Recognition Technology. The New York Times. https://www.nytimes.com/2019/04/03/technology/amazon-facial-recognition-technology.html.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2). https://doi.org/10.1177/2053951716679679

Moreau, S. (2020). Equality and Discrimination. In J. Tasioulas (Ed.), The Cambridge Companion to the Philosophy of Law, Cambridge Companions to Law. Cambridge University Press.

Mr P E Manjang v. Uber Eats UK Ltd and others, Case No. 3206212/202 (2021). Employment Tribunals. https://assets.publishing.service.gov.uk/media/62dab66b8fa8f5649dbef494/Mr_P_E_Manjang_-v-_Uber_Eats_UK_Ltd___Others_-_3206212_2021_-_Preliminary_Judgment.pdf.

NIST (2017). Face Recognition Technology (FERET). U.S. Department of Commerce. https://www.nist.gov/programs-projects/face-recognition-technology-feret

NIST. (2019). NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software.https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software.

Nkonde, M. (2020). Automated Anti-Blackness: Facial Recognition in Brooklyn, New York. Harvard Kennedy School Journal of African American Policy, 30-36. https://pacscenter.stanford.edu/wp-content/uploads/2020/12/mutalenkonde.pdf.

O'Neil, C. (2017). Weapons of math destruction. Penguin Books.
Palmer, V. V. (2015). Moral Damages: The French Awakening in the Nineteenth Century. In V. V. Palmer (Ed.), The Recovery of Non-Pecuniary Loss in European Contract Law (pp. 58-76). Cambridge University Press.

Parker, J. (2020). What NIST Data Shows About Facial Recognition and Demographics. Security Industry Association, 1-4. https://www.securityindustry.org/wp-content/uploads/2020/02/SIA-NIST-data-and-facial-recognition.pdf

Pesenti, J. (2021). An Update On Our Use of Face Recognition. Meta.https://about.fb.com/news/2021/11/update-on-use-of-face-recognition/.

Proctor Ninja. (2021). Proctorio's facial recognition is racist.https://web.archive.org/web/20220112203737/https:/proctor.ninja/proctorios-facial-recognition-is-racist;

Ragazzi, F., Kuskonmaz, E. M., Plájás, I., van de Ven, R., & Wagner, B. (2021). Biometrical & Behavioral mass surveillance in the EU member states. The Greens/EFA. https://www.greens-efa.eu/biometricsurveillance/

Raposo, V.L. (2023). When facial recognition does not 'recognise': erroneous identifications and resulting liabilities. AI & Soc. https://doi.org/10.1007/s00146-023-01634-z.

Ruggeri, A. (2023). The problems with TikTok's controversial 'beauty filters'. BBC. https://www.bbc.com/future/article/20230301-the-problems-with-tiktoks-controversial-beauty-filters.

Russell, S. J., & Norvig, P. (2010). Artificial Intelligence: A Modern Approach (3rd ed.). Pearson. https://web.cs.ucla.edu/~srinath/static/pdfs/AIMA.pdf

Rutkin, A. (2015). Facebook Can Recognise You in Photos Even If You're Not Looking. New Scientist.https://www.newscientist.com/article/dn27761-facebook-can-recognise-you-in-photos-even-if-youre-not-looking/

Sharma, S., Bhatt, M., and Sharma, P. (2020). Face Recognition System Using Machine Learning Algorithm. 5th International Conference on Communication and Electronics Systems (ICCES), 1162-1168.10.1109/ICCES48766.2020.9137850

Simonite, T. (2017). Facebook Can Now Find Your Face, Even When It's Not Tagged. Wired. https://www.wired.com/story/facebook-will-find-your-face-even-when-its-not-tagged/.

Singer, N. (2019). Amazon Is Pushing Facial Technology That a Study Says Could Be Biased. The New York Times. https://www.nytimes.com/2019/01/24/technology/amazon-facial-technology-study.html.

Staffer, E. L. (2021). Europe's Next Steps in Regulating Facial Recognition Technology. Columbia Journal of Transnational Law. https://www.jtl.columbia.edu/bulletin-blog/europes-next-steps-in-regulating-facial-recognition-technology

Sullivan, E. (2021). Facial Recognition Technology: Verification vs. Identification. Economic Affairs Interim Committee, Legislative Research Analyst, Montana State Legislature. https://leg.mt.gov/content/Committees/Interim/2021-2022/Economic%20Affairs/Meetings/November%202021/Facial-verification-vs-facial-identification.pdf.

The Cube. (2021). Facial recognition: Clearview AI breaks EU data privacy rules, says French watchdog. Euro News .https://www.euronews.com/my-europe/2021/12/16/facial-recognition-clearview-ai-breaks-eu-data-privacy-rules-says-french-watchdog;

Thorat, S.B., Nayak, S.K., and Dandale, J.P. (2010). Facial Recognition Technology: An analysis with scope in India. International Journal of Computer Science and Information Security, 8(1), 325-330 https://arxiv.org/pdf/1005.4263.pdf

Tolba, A.S., El-Baz, A.H., and El-Harby, A.A. (2014). Face Recognition: A Literature Review. International Journal of Signal Processing, 2(2), 88-103. https://www.researchgate.net/publication/233864740_Face_Recognition_A_Literature_Review

Tomas, T. (2022). Uber Eats treats drivers as 'numbers not humans,' says dismissed UK courier. The Guardian. https://www.theguardian.com/technology/2022/jul/27/uber-eats-treats-drivers-as-numbers-not-humans-says-dismissed-courier.

Tucker, I. (2017). A white mask worked better': why algorithms are not colour blind. The Guardian. https://www.theguardian.com/technology/2017/may/28/joy-buolamwini-when-algorithms-are-racist-facial-recognition-bias;

Turk, M., and Pentland, A. (1991). Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1). https://www.face-rec.org/algorithms/pca/jcn.pdf

Uber India. (2017). Selfie-powered Real-Time ID Check Comes to India. https://www.uber.com/en-IN/blog/chennai/selfie-powered-real-time-id-check-comes-to-india/.

VI v. KRONE-Verlag Gesellschaft mbH & Co KG, Judgment of the Court (First Chamber) of 10 June 2021 (Case C-65/20), ECLI:EU:C:2021:471. https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A62020CJ0065

Vincent, J. (2018). Google 'Fixed' Its Racist Algorithm by Removing Gorillas from Its Image-Labeling Tech. The Verge. https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai.

Wong, K. (2017). Chinese Woman Discovers That Her iPhone X Can Be Unlocked by Her Friend's Face. Mothership.https://mothership.sg/2017/12/iphone-x-asians-look-the-same/.

Worker Info Exchange. (2022). Court Rejects Uber's Attempt to Have Facial Recognition Discrimination Claim Struck Out. https://www.workerinfoexchange.org/post/court-rejects-uber-s-attempt-to-have-facial-recognition-discrimination-claim-struck-out.

Wurmnest, W. (2012). Non-Pecuniary Loss. Max-Eup2012. https://max-eup2012.mpipriv.de/index.php/Non-Pecuniary_Loss