AI and Ethics (2023) 3:1407–1422
https://doi.org/10.1007/s43681-022-00249-2

ORIGINAL RESEARCH

There is an elephant in the room: towards a critique on the use of fairness in biometrics

Ana Valdivia[1] · **Júlia Corbera Serrajòrdia**[2] · **Aneta Swianiewicz**[2]

Received: 5 September 2022 / Accepted: 5 December 2022 / Published online: 22 December 2022

Abstract
The proliferation of biometric systems in our societies is shaping public debates around its political, social and ethical implications. Yet, whilst concerns towards the racialised use of this technology have been on the rise, the field of biometrics remains unperturbed by these debates. Despite the lack of critical analysis, algorithmic fairness has recently been adopted by biometrics. Different studies have been published to understand and mitigate demographic bias in biometric systems, without analysing the political consequences. In this paper, we offer a critical reading of recent debates about biometric fairness and show its detachment from political debates. Building on previous fairness demonstrations, we prove that biometrics will be always biased. Yet, we claim algorithmic fairness cannot distribute justice in scenarios which are broken or whose intended purpose is to discriminate. By focusing on demographic biases rather than examine how these systems reproduce historical and political injustices, fairness has overshadowed the elephant in the room of biometrics.

Keywords Biometric systems · Fairness · Racialised borders · Migration

1 Introduction

Biometric systems are being designed and implemented by public and private organisations for law enforcement, migration control and security purposes. This technology is used to identify or verify the unique identity of a person through physical, physiological or behavioural characteristics, such as fingerprint, face, voice, gait or finger veins of human bodies. Individual bodily characteristics are transformed into biometric data and processed by algorithmic systems which output is considered the *truthful* identity of an individual. Biometric systems are part of our daily lives, we use them to unlock our mobile phones or to 'efficiently' cross borders at airports.[1] However, this technology has different social and political consequences depending on a subject's migration status [59, 71, 73]. In Europe, the Dublin Regulation is the

[1] The concept of technological efficiency can be questioned at the border. In many cases, the technology fails, the automatic gate does not recognise the passport and efficiency may not be achieved by digitisation, but by the border police, i.e. humans.

✉ Ana Valdivia
ana.valdivia@oii.ox.ac.uk

Júlia Corbera Serrajòrdia
julia.corbera_serrajordia@kcl.ac.uk

Aneta Swianiewicz
aneta.swianiewicz@kcl.ac.uk

[1] Oxford Internet Institute (University of Oxford), 34 St Giles, Oxford OX1 3LD, UK

[2] King's College London, Strand, London WC2R 2LS, UK

legal framework that establishes which Member State within the European Union (EU) is responsible of the application of asylum seekers. This law provides that the asylum seeker should seek asylum in the first country of arrival.[2] Under this law, asylum seekers are forced to remain in this first country where their asylum procedure will be examined, without the possibility of travelling or moving to another Member State in general. In order to enforce this law, the EU implemented a biometric database (EURODAC) to make it easier for Member States to determine the country responsible for the asylum application [17, 77, 92]. This large-scale database stores fingerprints of asylum seekers to identify the first country of arrival where they should have sought asylum. Then, in the case of a border crossing, border authorities can determine the country responsible of the asylum case by extracting the fingerprint of the asylum seeker and compare it to the fingerprints stored in EURODAC. In the case of a biometric match, the database will show the Member State responsible of the asylum case and the asylum seeker will be push-back. In other words, biometrics are implemented to immobilise, control and obstruct migrants' movements within Member States through fingerprint identification [82, 87]. Meanwhile, European citizens do have the freedom of free movement across Member States. Moreover, the World Food Programme (WFP) in partnership with the United Nations High Commissioner for Refugees (UNHCR) implemented iris recognition to register migrants and provide cash assistance in Jordan's Za'atari refugee camp [5]. More recently, Privacy International has denounced the plans of the UK Home Office to use facial recognition smartwatches to monitor migrants with criminal records [76]. The use of biometrics in these scenarios has been largely criticised by academics, activists and human and digital rights organisations who have argued that 'undermines democracy, freedom and justice' [30, 77]. In the specific context of migration, biometric systems are used to criminalise migrants and illegalise border crossing between Member States through the rule of law such as Dublin Regulation, infringing a fundamental right: the freedom of movement. Yet, algorithmic fairness has plunged into biometrics as a new research area which aims at debiasing these systems, without considering the historical, political, legal and social context in which this technology is embedded.

This paper starts by questioning the 'emergent challenge' of fairness in biometrics. Since the publication of *Gender Shades* by Buolamwini and Gebru in 2018, the field of biometrics has experienced a surge in studies on bias and disparate impact in algorithmic systems, such as facial recognition, fingerprints, finger veins and iris recognition, among others [26]. By examining the recent literature on fairness in biometrics, we observe a general lack of engagement with the political and social aspects in which these systems are implemented. The literature examined also neglect current debates and critiques towards algorithmic fairness, such as the impossibility of fairness and the need to analyse the politics of algorithms. Notably, a recent work in finger vein recognition systems has suggested a lack of bias for different biometric algorithms [27]. However, critical scholars find limitations to reproduce biometric systems and compare results. Data and code are not usually made open and publicly available. The sensitivity of biometric and demographic data stems from the fact that subjects identity can be put at stake if the anonymisation process is not carried out properly. Moreover, obtaining access to biometric code is also a constraint as engineers and even academics are sometimes reluctant to share it. Thus, challenging the biometric research field in fairness from a quantitative and technical perspective becomes a very arduous task for critical scholars and digital rights organisations.

In this paper, we propose to critically analyse the field of fairness in biometrics. Whilst the impossibility of fairness in machine learning has been widely investigated [22, 40, 45, 58], this result has not been proved in the field of biometrics yet. Building on these previous works in machine learning, the first contribution of this work is to show the impossibility of fair biometric systems from a mathematical perspective. The second contribution is to empirically demonstrate this theoretical framework by implementing three fairness metrics in four different biometric systems. These biometric systems are reproduced in [27] where authors claimed that these systems are not biased by analysing statistical differences. Although this might seem pragmatic to evaluate fairness, we outline some technical limitations. First, the decision threshold[3] which sets the cut-off between a biometric match or non-match plays a key role in the assessment of fairness in biometric systems [36]. Second, intersectional demographic evaluation of gender and age must be assessed. Third, holding on our theoretical framework that demonstrates the impossibility of fairness in biometrics, the lack of bias is almost technically impossible in any biometric system. Moreover, we highlight that the dataset used to train these systems and which we had access to reproduce the racialisation of subjects, proposing race and gender categories that are colonial, archaic and offensive.

[2] Council of the European Union, Regulation (EU) No 604/2013 of the European Parliament and of the Council of 26 June 2013 establishing the criteria and mechanisms for determining the Member State responsible for examining an application for international protection lodged in one of the Member States by a third-country national or a stateless person (recast), 29 June 2013, OJ L. 180/31-180/59; 29.6.2013, (EU)No 604/2013, available at: https://www.refworld.org/docid/51d298f04.html (Accessed 01 Nov 2022).

[3] The decision threshold in biometrics plays a different role than in machine learning. In the biometrics field, the performance of a system is analysed setting different decision thresholds.

Yet, we move the argument forward by contending that a critical questioning of the use of fairness in biometric systems should consider the historical, political and social contexts in which biometrics are deployed, rather than narrow to error differences on demographic groups. The third contribution of this paper is to expose how biometric systems are unfair considering the political context in which these algorithms are deployed. Moving our discourse towards a 'critical biometric consciousness' [16], we analyse the case of an asylum appeal where the migrant's credibility was challenged in part due to an inconsistency between his testimony and a biometric trace stored in a database and matched by a biometric system. Given this fact and other inconsistencies, the UK Deputy Upper Tribunal Judge denied his asylum claim, dismissing any other evidence provided by the asylum seeker. This case, we argue, unveils *the elephant in the room* of biometrics: the fact that is being intentionally ignored or left unaddressed, showing how biometrics cannot be fair if they are deployed in scenarios that are intended to discriminate, even in the hypothetical case of lack of bias. Whilst academics, private companies and biometric engineers are focusing their efforts in building more accurate, explainable and fairer biometrics, little attention is paid to the political implications of biometrics and how these systems are used to jeopardise rights that have been recognised by institutions such as United Nations (UN) and require high degree of protection, such as the freedom of movement [48]. Importantly, whilst the European Commission has proposed the first legal framework for artificial intelligence and biometrics, the Artificial Intelligence Act (AI Act hereinafter), to protect European citizens' digital and fundamental rights [33], biometric systems used at the border for migration control will be explicitly exempt from such regulation. Therefore, migrants' digital and fundamental rights will not obtain the same lawful protection. Since fairness in biometrics has the risk of becoming more prominent in the incoming years, we urge for a critical, political and radical examination of this field. We suggest that we must also situate current debates in biometrics within a broader historical context of struggles against discrimination, moving beyond the technological dilemmas about ethics and demographic bias. To the best of our knowledge, this is the first academic work that investigates fairness in biometrics from a critical and political perspective, and pushes the argument further showing how debates about the core function of borders and the use of biometrics to criminalise migration are undermined by the focus on ethics and more equitable biometric systems.

2 Fairness in biometrics: An emergent challenge?

In 2018, Buolamwini and Gebru published *Gender Shades*, an academic work that assessed bias in gender classification algorithms through facial recognition. They analysed several commercial gender classification models and found significant disparities based on individuals' characteristics: white skins had better results than dark skins whilst males obtained better results than females. This intersectional benchmark opened up a new avenue of critical analysis towards biases in racialised technologies. It influenced public and academic debates towards the use of facial recognition [13, 54, 69], creating awareness about the risks of algorithms that encode and propagate historical, political and social disparities. Consequently, it also has disrupted the field of biometrics. Since the publication of *Gender Shades* [18], fairness has emerged as a major challenge within biometrics [26].

The main goal of fairness in the field of biometrics is to identify and understand bias in systems that determine or validate the identity or other characteristics like the gender of individuals [80]. In some other cases, bias mitigation can be also part of this goal. Researchers have analysed demographic biases in facial recognition, fingerprints, palm prints, iris and even in finger veins. Face recognition has been notably the most scrutinised system in the last decade, where gender and race are the demographic features scrutinised. In general, males and white skins obtain higher biometric performance [3, 36, 60, 85]. The annual reports published by National Institute of Standards and Technology (NIST) [46] also found error disparities based on gender and race on more than 189 commercial facial recognition systems that are used for border control. Similarly, other biometric systems such as iris recognition find noteworthy differences between females and males, with the former having higher error rates [34]. A recent study has analysed demographic bias in fingerprint recognition [44], concluding that bias depends on external factors such as image quality rather than gender or race of individuals. Yet, age is the demographic feature that has the greatest impact on the performance of fingerprint systems. For instance, the conclusion achieved in recent works is that fingerprint verification systems get higher error rates on minors [63, 75]. Indeed, most studies about fairness in biometrics focus on analysing bias in race and gender rather than age. Given the lack of biometric samples of children, these biometric systems perform worse on minors when compare to adult samples. Despite the complexity of mitigating demographic bias in biometric systems, researchers have suggested that 'statistically significant bias' on age and gender 'have not been detected on five finger vein recognition algorithms tested on four datasets' [27].

In spite of the numerous biometric articles in fairness that have been recently published, there is a lack of critical analysis on how these systems are implemented in real scenarios and jeopardise fundamental rights, such as freedom of movement. According to Guild and Groenendijk: '[F]reedom of movement did not only amount to the right to travel freely, to take up residence and to work, but also involved the enjoyment of a legal status characterised by security of residence, the right to family reunification and the right to be treated equally with nationals' [48, pp. 206]. What is the purpose of developing biometric systems that perform equally well across different demographic groups if they are implemented to push-back migrants at the border, limit border crossings and deny asylum applications? On the other hand, part of these studies also neglect the current debates in the field of fairness and critical artificial intelligence studies. First, these works analyse algorithmic bias without considering the more than 20 definitions in fairness in machine learning that have been proposed in the last decade [12, 28, 95]. For instance, in [27] demographic bias is calculated by analysing differences of score distributions by groups (males vs. females or children vs. adults). However, differences on error rates which is an standard approach to evaluate fairness are not considered. Second, despite the popularity of *Gender Shades* [18] and its proposed benchmark, intersectionality is not considered in most of the studies analysed. In general, error disparity is analysed taking only a single demographic feature into account (gender, race or age). Third, some studies also suggest a lack of error disparities of biometric systems such as in [27]. However, different fairness works have demonstrated the impossibility of fairness, proving mathematically and empirically that fairness definitions are mutually exclusive [22, 40, 45, 58]. As a result, it has been proved that despite the efforts on mitigating disparities, error differences among demographics still persists [46].

As we have previously stated, part of the literature of fairness in biometrics that we have examined do not take into account the politics of these systems and how they are impacting on fundamental rights [20]. In contrast, Buolamwini and Gebru clearly exposed that 'all evaluated companies provide a "gender classification" feature that uses the binary sex labels of female and male. However, this view on gender does not adequately capture its complexities or address transgender identities' [18, p. 6]. In fact, the mere idea of gender classification algorithms infringes fundamental rights by designing and targeting trans people which are more likely to be 'miss-classified'. Examining the performance of classification models with respect to gender will likely improve these models. Yet, we ought to consider if it is even a desirable goal. What is the politics of a gender classification system? What is the benefit of gender classification algorithms bring to our societies? How could communities such as LGBTQIA+ benefit from a fair and transparent gender facial recognition?

In this paper, we focus on the critique towards the use of biometrics for migration control which impacts on asylum seekers' rights. Since 2000, the EU has implemented biometric systems to identify travellers for migration control [10, 64, 82]. Judges, officials in migration administration and border guards among others are currently making use of biometric technology in asylum cases or visa applications [43, 74] to make 'unequivocal' and 'efficient' decisions. The output of the biometric systems are considered more reliable than migrants or asylum seekers, which are perceived as deceptive subjects [9].

Fairness and the study of bias has become the *elephant in the room* of biometrics. Whilst the effort is focused on solving the challenge of addressing demographic biases and design 'fair' biometric systems, less attention is paid on the context of how it is used by the biometric community. These systems are nowadays reproducing political injustices which are linked to a colonial legacy which is ignored in contemporary biometrics [16]. As Maguire argued [62], fingerprints were used by a British officer at the Indian Civil Service to avoid fraud on colonial subjects. At the same time, Galton investigated 'the heritable characteristics of race' on racialised individuals in British prisons. In fact, biometrics 'offered 19th century innovators more than the prospect of identifying criminals: early biometrics promised a utopia of bio-governmentality in which individual identity verification was at the heart of population control' [62]. In the 21st century, this approach to biometrics has evolved with the use of algorithms that automatically find patterns and calculate matches. Yet, its colonial legacy remains still visible. At the border, these systems are used to identify racialised subjects. There, fairness has emerged to make the identification of asylum seekers 'fairer'.

3 A critique of fairness in biometrics from a technological perspective

Fairness is a contested concept that has been historically discussed by social scientists, legal experts and philosophers [41, 79]. Yet, there is no consensus on what this concept means or implies. The surge of socio-technical systems deployed in our society has revealed algorithmic disparate impacts [8, 32, 68]. As a response, the concept of fairness has been translated into mathematical definitions which quantify the fairness of algorithmic-based models [52]. Different formulations [12, 28, 95] have been proposed to calculate and mitigate algorithmic biases. In recent years, the research field in biometrics has also evaluated the fairness of these systems as well as developing novel methods for bias mitigation. Most of the studies do not take into

account the most relevant state-of-the-art definitions of fairness in machine learning and do not engage with critical debates [27, 34, 63, 84, 85]. Moreover, one of these studies analysed suggests lack of bias in score distributions [27] which is technically impossible as studies in the field of algorithmic fairness in machine learning have shown [22, 40, 45, 58]. As there are multiple definitions to evaluate demographic bias, the field of fairness has proved that it is generally impossible to satisfy several fairness criteria simultaneously. Therefore, algorithmic systems, including biometrics as we will prove, will be generally biased.

3.1 Formulation of a biometric verification system

In the discussion, herein, we focus on biometric verification systems. These systems confirm (or reject) whether a biometric sample belongs to an specific individual based on similarity to their learned biometric representation. Formally, we use the following notation:

- $x^{(i)}$: learned representation of the biometric ith-sample within the dataset $\mathcal{D}$,
- $y^{(i,j)}$: element (i, j) of the binary outcome vector y,
- θ: set of parameters of the biometric system,
- $\tilde{y}^{(i,j)} = f((x^{(i)}, x^{(j)})|\theta)$: element (i, j) of the binary prediction vector $\tilde{y}$,
- τ: decision threshold,
- $s^{(i,j)}$: similarity score between samples i and j.

Usually, these systems employ learning algorithms which transform images into features and numerical representations. They are driven by an optimisation function to minimise errors between two samples ($x^{(i)}$ and $x^{(j)}$) and the outcome ($y^{(i,j)}$) which is whether they belong to the same individual or not:

$$\theta^* = \arg\min_{\theta} \sum_{x^{(i)}, x^{(j)} \in \mathcal{D}} \mathcal{L}\Big(f\big((x^{(i)}, x^{(j)})|\theta\big), y^{(i,j)}\Big) \tag{1}$$

The biometric model compares two learned representation samples ($x^{(i)}$ and $x^{(j)}$) and obtains a similarity score ($s^{(i,j)}$). This score is then translated into a binary prediction ($\tilde{y}^{(i,j)}$) given a threshold τ:

$$\tilde{y}^{(i,j)} = \begin{cases} 0 \text{ if } s \leq \tau \\ 1 \text{ if } s > \tau \end{cases} \tag{2}$$

The model learns the best representation of parameters (θ) that achieves the minimum number of 'miss-identifications'. This means that the binary prediction ($\tilde{y}^{(i,j)}$) should be as similar as the binary outcome ($y^{(i,j)}$). In the case of biometric verification systems, two pairs of biometric samples are labelled as *genuine* ($y^{(i,j)} = 1$) if they correspond to the same individual and as *impostor* ($y^{(i,j)} = 0$) otherwise.[4] Therefore, a *true genuine* ($TG = Pr(\tilde{y}^{(i,j)} = 1|y^{(i,j)} = 1)$) refers to those samples that correspond to the same individual and are correctly matched by the biometric system; the *true impostor* ($TN = Pr(\tilde{y}^{(i,j)} = 0|y^{(i,j)} = 0)$) refers to samples that correspond to different individuals and they are rejected; *false genuine* ($FG = Pr(\tilde{y}^{(i,j)} = 1|y^{(i,j)} = 0)$) refers to different individuals whose samples are matched; and *false impostor* ($FI = Pr(\tilde{y}^{(i,j)} = 0|y^{(i,j)} = 1)$) refers to samples of the same individual that are miss-matched.[5] Thus, the errors that biometric systems commit are: (1) incorrect association of two subjects or (2) failed association of one subject.

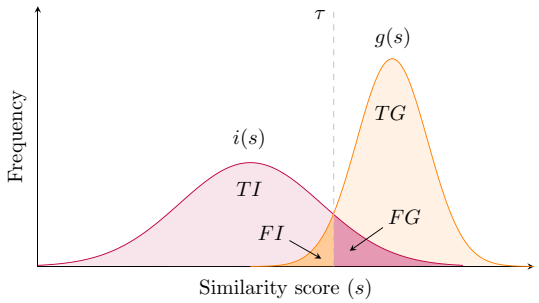


Fig. 1 Relationship between the similarity score (s), decision threshold (τ) and the genuine ($g(s)$) and impostor ($i(s)$) score distributions in a biometric system. The figure visually shows the dependence with the confusion matrix elements (true genuine (*TG*), true impostor (*TI*), false genuine (*FG*), and false impostor (*FI*))

In contrast to machine learning, biometric systems are evaluated setting different decision thresholds (τ) which clearly affects on the distribution of errors (see Fig. 1). Equal Error Rate (*EER*) is the value where the false genuine and impostor rates curves intersect ($FGR = FIR$). False genuine rate at 0.001 (FGR_{1000}) is the value of false impostor rate when false genuine rate is 0.001 ($FGR = 0.001$), and vice versa. Systems are also evaluated when one of these rates is 0 (*ZFGR* when $FIR = 0$ or *ZFIR* when $FGR = 0$). Thus, the decision threshold is set targeting different error values.

3.2 Formulation of three fairness definitions for biometrics

A fairness measure is a mathematical function that quantifies and assesses biased systems and algorithmic discrimination. These measures aim at evaluating model performance across

[4] Note the language used in biometrics: a miss-match is defined as an *impostor* which denotes that identity's deception is taken for granted. In the machine learning literature, there is an absence of critical, transdisciplinary and genealogical examination of these terms as pointed out in [15].

[5] In the biometric literature, false genuine and impostor rates are also known as false match (FMR) and non-match (FNMR) rates respectively.

demographics groups ($\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$) and ensure that there is no disparate impact among them. In this setting, typically a demographic feature such as gender, age, class or race is proposed to define the 'advantaged' and 'disadvantaged' group.

We identify that part of the literature on fairness in biometrics disengages from the established fairness definitions in machine learning. In some of the works analysed [27, 34, 88], authors assessed biases in biometric systems by proposing their own fairness definitions without considering previous works on algorithmic discrimination. However, recent studies have adopted these definitions into biometrics [36, 44]. Therefore, we propose three well-known definitions of fairness translated into biometrics notation to demonstrate the impossibility of fairness in these systems.

3.2.1 Equalised odds (also *disparate mistreatment* or *error rate balance*) [50]

A biometric system satisfies *equalised odds* if TGR and FGR are similar across demographics groups:

$$|TGR_{C_a} - TGR_{C_b}| < \epsilon \tag{3}$$

and,

$$|FGR_{C_a} - FGR_{C_b}| < \epsilon, \forall\, C_a, C_b \subset \mathcal{C} \tag{4}$$

3.2.2 Statistical parity (also *group fairness* or *demographic parity*) [29]

A biometric system satisfies *statistical parity* if the probability of predicted genuine is similar across demographics groups. This definition is based on the predicted outcome. Mathematically, this definition is expressed as follows:

$$|Pr(\tilde{y} = 1|C_a) - Pr(\tilde{y} = 1|C_b)| < \epsilon, \forall\, C_a, C_b \subset \mathcal{C} \tag{5}$$

3.2.3 Predictive parity (also *outcome test*) [22]

A biometric system satisfies *predictive parity* if the probability of being predicted genuine of actual genuine is similar across demographic groups. More formally:

$$\begin{aligned} &|Pr(y = 1|\tilde{y} = 1, C_a) \\ &\quad - Pr(y = 1|\tilde{y} = 1, C_b)| < \epsilon, \forall\, C_a, C_b \subset \mathcal{C} \end{aligned} \tag{6}$$

3.3 The impossibility of unbiased biometric systems

In this section, we provide a theoretical framework to demonstrate that the previous fairness criteria are mutually exclusive. The impossibility of fairness in machine learning has been widely studied by [22, 40, 45, 58]. However, there is a lack in the literature about the impossibility of fairness in biometrics that this paper attempts to address. We mathematically show that only under very unrealistic conditions (equal ratios among demographics groups, trivial or perfect biometric system),[6] these three definitions can be simultaneously satisfied. Thus, we prove the impossibility of any unbiased biometric system.

To simplify the notation, we assume that the demographic features is categorised into two groups: $\mathcal{C} = \{C_1, C_2\}$.

Proposition 3.3.1 *Given a biometric system which is non-trivial with unequal ratios among groups that satisfies equalised odds and statistical parity, then predictive parity cannot hold.*

Proof Suppose that predictive parity is held, then equalised odds or statistical parity is not held.

Given Bayes' theorem, we obtain that:

$$Pr(y = 1|\tilde{y} = 1, \mathcal{C}) = \frac{Pr(\tilde{y} = 1|y = 1, \mathcal{C}) Pr(y = 1|\mathcal{C})}{Pr(\tilde{y} = 1|\mathcal{C})} \tag{7}$$

If predictive parity is satisfied, then:

$$\begin{aligned} &|Pr(y = 1|\tilde{y} = 1, C_1) - Pr(y = 1|\tilde{y} = 1, C_2)| < \epsilon \overset{\text{Equation 7}}{\Longrightarrow} \\ &\left| \frac{Pr(\tilde{y} = 1|y = 1, C_1) Pr(y = 1|C_1)}{Pr(\tilde{y} = 1|C_1)} \right. \\ &\left. - \frac{Pr(\tilde{y} = 1|y = 1, C_2) Pr(y = 1|C_2)}{Pr(\tilde{y} = 1|C_2)} \right| < \epsilon \end{aligned}$$

Assuming that equalised odds on the TGR ($|TGR_{C_1} - TGR_{C_2}| < \epsilon$) and statistical parity ($|Pr(\tilde{y} = 1|C_a) - Pr(\tilde{y} = 1|C_b)| < \epsilon$) are satisfied in the previous expression:

$$\left| \frac{Pr(\tilde{y} = 1|y = 1)[Pr(y = 1|C_1) - Pr(y = 1|C_2)]}{Pr(\tilde{y} = 1)} \right| < \epsilon$$

Given that $Pr(\tilde{y}|y), Pr(y|\mathcal{C}), Pr(\tilde{y}) \in [0, 1]$, predictive parity is only satisfied when:

$$|Pr(\tilde{y} = 1|y = 1)[Pr(y = 1|C_1) - Pr(y = 1|C_2)]| < \epsilon$$

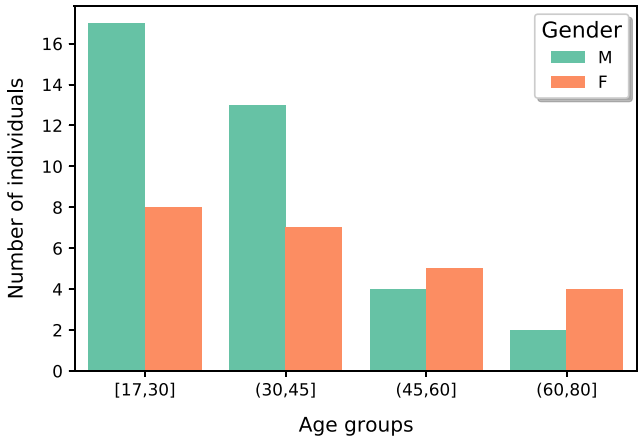
which implies:

$$Pr(\tilde{y} = 1|y = 1) < \epsilon$$

or

$$|Pr(y = 1|C_1) - Pr(y = 1|C_2)| < \epsilon.$$

[6] The trivial biometric system is the system that only classifies or outputs one class (genuine or impostor). The perfect biometric system is an Utopian system that achieves zero classification error, i.e. $FGR = 0$ and $FIR = 0$.



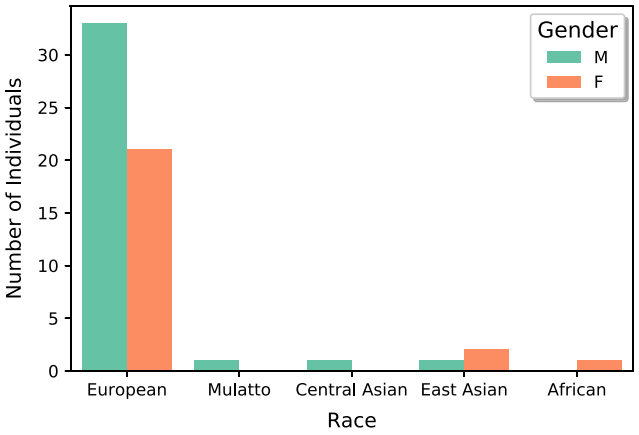


Fig. 2 Intersectional ratios on age, race and gender of individuals in PLUSVein-FV3 [56]. There are large disparities among groups: more males (M) than females (F), young than old adults, and majority of Europeans. The proposed race labels lack of diversity, considering one of them ('Mulatto') rather archaic and offensive

On the one hand, if $Pr(\tilde{y} = 1|y = 1) < \epsilon$, we obtain that Proposition 3.3.1 is only satisfied when $TGR = 0$ which means that the system fails to correctly classify any genuine instance or the system is a trivial classifier, e.g. there are only impostor instances. On the other hand, $|Pr(y = 1|C_1) - Pr(y = 1|C_2) < \epsilon| \implies Pr(y = 1|C_1) \sim Pr(y = 1|C_2)$ which implies that Proposition 3.3.1 is satisfied only under equal ratios. □

Thus, we prove that the impossibility of fairness also holds for biometric systems.

4 A biometric experiment: Why this system is biased?

Getting access to off-the-shelf biometric algorithms together with biometric and demographic data is challenging. Reasonably, biometric data is not available in the public domain due to privacy and consent policies. However, we got access to a finger vein dataset together with its demographic data developed by scholars at the University of Salzburg (Austria) that allow us to empirically prove the impossibility of fairness in biometrics. We examine four biometric systems (finger veins) that are publicly available to empirically demonstrate the impossibility of fairness in biometric systems (see Proposition 3.3.1).

These biometric algorithms aimed at identifying subjects through vascular patterns on the human body, i.e. finger, palm or human eye veins [90]. These four systems are evaluated in [27] to assess demographic bias by differences on statistics (mean and standard deviation) of genuine and impostor score distributions. The conclusion achieved is that statistically significant biases in score distributions do not exist and the authors proposed to evaluate this framework in the future with more individuals, given that the number of subjects in each of the databases is very low. Rather than reproduce their experiments with larger databases, we empirically demonstrate that this framework is biased from three different perspectives: (1) ratios, (2) fairness criteria, and (3) intersectionality.

4.1 Ratios

Broken links and APIs hampered the access to three of four publicly available datasets [61, 89, 93]. Through an online petition, we had access to PLUSVein-FV3 [56].[7], [8] This database contains 1440 finger vein images from hands and fingers of 60 individuals. Figure 2 shows the number of individuals based on age, race[9] and gender[10]. We observe that the mean of age within this dataset is 37.9 years, whilst Q_3[11] is 46.5 years, which implies that the database is not

[7] See: https://wavelab.at/sources/PLUSVein-FV3 (Accessed 06 Dec 2021).

[8] The dataset provided is completely anonymised. There is no possibility that the direct linking of this information to an individual could lead to their identification. Our sole research intention in processing this data is to demonstrate the impossibility of fairness in biometric systems, which is in the public interest.

[9] In the original work, authors proposed to use 'ethnicity'. However, we propose to rather use 'race' which is related to the colonial and legal construction of human categories.

[10] Note that in the original work, authors used 'sex' instead of 'gender'. They argued in [27] that: '[T]erms "gender" and "sex" are often used in a binary and conflated manner'. Moreover, they proposed to use ISO/IEC's definitions that distinguish biological sex from cultural gender. Building on Butler's claim that there are no distinctions between sex and gender [19], we propose to use the 'gender' concept.

[11] Third quartile.

Table 1 Error rates of biometric methods

		Performance metric				
		EER	FGR_{1000}	FGR_{100}	FGR_{10}	*ZFIR*
Method	LBP	0.16	0.89	0.78	0.20	0.99
	MC	0.02	0.03	0.02	0.01	0.99
	PC	0.02	0.05	0.03	0.02	0.98
	SIFT	0.02	0.04	0.02	0.01	1.00

LBP is the method with the poorest performance. MC, PC, and SIFT obtain low and similar error rates

representative for elders. Analysing the gender feature, we observe that rates are unbalanced: 60% males and 40% females. Yet, this feature does not consider other gender expressions rather than binary ones. Europeans are the most represented race in PLUS-VeinFV3: 90% European, 5% East Asian, 1.6% Central Asian, 1.6% 'Mulatto', 1.6% African. The proportion of non-European individuals is significantly low (see Fig. 2).

Remarkably, 'Mulatto' is a label proposed for this category which is not linked to any continent. Whilst European, East Asian and Central Asian correspond to race labels related with geographical expressions, 'Mulatto' was a label used during the Spanish colonial period to mark the slave status of children born to Spaniards and African women slaves. This category exposes the colonial and racial legacy of biometrics through a conceptualisation of racialised and colonial bodies. 'Mulatto' is a clear expression of the making of bodies through their qualities of 'colour' and colonial slavery [16, 42]. In this paper, we contend that the literature in biometrics creates race categories to calculate demographic biases without critically analysing how these systems are embedded in structrual mechanisms such as borders, the intention of which is to discriminate [2]. As other critical scholars pointed out: '[t]reating race as an attribute, rather than a structural, institutional, and relational phenomenon, can serve to minimise the structural aspects of algorithmic unfairness' [49].

Bias in demographic groups in PLUSVein-FV3 becomes also evident (see Fig. 2). Any biometric system trained in this dataset will perform better on individuals whose demographic characteristics are widely presented in the database, i.e. young male Europeans [18]. The performance of the biometric system will be poor on individuals who are underrepresented: non-Europeans, elderly and females. Consequently, given that the majority of samples are taken from European males, rates of genuine and impostor outcomes will be significantly unequal across different demographic groups. Thus, as previously demonstrated (Proposition 3.3.1) equalised odds, statistical parity and predictive parity cannot hold together.

4.2 Fairness criteria

The experiments are run using the four biometric systems in [27] and the PLUSVein-FV3 dataset. These systems are designed using different types of vein recognition schemes: LBP [35], MC [67], PC [21], and SIFT [55, 96]. These methods are designed following different mechanisms of finger vein recognition. Whilst LBP is based on a pattern-based method that identifies veins textures using filters, MC and PC extract patterns analysing curvatures. SIFT is designed to identify keypoints on the image that guide the algorithm to recognise veins.

Table 1 shows the results of these four different biometric systems on the PLUSVein-FV3 dataset. In general, the error rates of these systems are very low (see MC, PC and SIFT). However, we observe that LBP has the worst performance, obtaining 0.89 and 0.78 of *FIR* at FGR_{1000} and FGR_{100} respectively. In this case, if the rate of false match is very low ($\downarrow$ *FGR*), the false non-match rate is extremely high ($\uparrow$ *FIR*). Nevertheless, the aim of this section is to empirically demonstrate the impossibility of fairness in biometric systems (see Proposition 3.3.1). To do so, we calculate three fairness criteria (equalised odds, statistical parity and predictive parity) on three different demographics (age, gender, and race). Demographics groups are categorised as: young (≤ 45) and old (> 45), male and female, and European and non-European.

As de Freitas Pereira and Marcel observed, several works of fairness in biometrics set a single decision threshold τ for every demographic group. However, they argue that this is a 'serious flaw' and 'can give a false impression that a biometric verification system is fair' [36, p. 3]. They conclude that '[f]air biometric recognition systems are fair if a decision threshold τ is "fair" for all demographic groups with respect to $FGR(\tau)$ and $FIR(\tau)$' [36, p. 10]. Following this suggestion, fairness metrics are calculated after setting the same decision thresholds for the three demographic groups: FGR_{1000} and near *ZFIR*.

The overall recognition performance results show that the four biometric systems are unfair considering three fairness

Table 2 Biometric recognition performance as measured by fairness criteria differences among three demographic groups at FGR_{1000}

	Equalised odds		
	Age	Gender	Race
LBP	[4.26%, 40.19%]	[29.02% , 11.36%]	[44.29% , 19.87%]
MC	[7.91%, 86.06%]	[4.80%, 129.57%]	[3.60%, 0.07%]
PC	[9.47%, 76.76%]	[3.58%, 421.18%]	[3.96%, 198.80%]
SIFT	[6.33%, 7.90%]	[5.32% , 43.35%]	[5.14% , 49.51%]

	Statistical parity			Predictive parity		
	Age	Gender	Race	Age	Gender	Race
LBP	32.16%	47.91%	45.17%	11.21%	4.98%	3.03%
MC	42.03%	16.05%	7.24%	0.46%	1.27%	0.14%
PC	39.53%	19.12%	10.73%	0.40%	2.59%	2.68%
SIFT	42.74%	12.68%	7.91%	0.74%	1.11%	0.99%

All systems are consistently unfair, showing significant equalised odds and statistical parity differences for age, gender and race

Table 3 Biometric recognition performance as measured by fairness criteria differences among three demographic groups at $\sim ZFIR$

	Equalised odds		
	Age	Gender	Race
LBP	[0.36%, 0.35%]	[0.36%, 0.42%]	[1.01%, 0.96%]
MC	[0.08%, 1.25%]	[0.42%, 1.92%]	[0.32%, 1.44%]
PC	[0.62%, 3.55%]	[1.97%, 9.33%]	[0.58%, 7.24%]
SIFT	[3.40%, 3.45%]	[0.36%, 0.42%]	[1.01%, 0.96%]

	Statistical parity			Predictive parity		
	Age	Gender	Race	Age	Gender	Race
LBP	0.36%	0.51%	1.01%	53.22%	19.19%	3.66%
MC	1.19%	1.84%	1.32%	51.58%	18.67%	5.38%
PC	3.38%	8.88%	6.65%	47.46%	12.71%	11.69%
SIFT	2.79%	0.61%	1.01%	52.47%	19.19%	3.66%

All systems are consistently unfair, showing significant predictive parity differences for age, gender and race. These results empirically demonstrate Proposition 3.3.1

definitions on the three demographic groups assessed.[12] These results contradict the result that these systems lack of demographic bias. Moreover, the results on Table 3 hold our theoretical framework on the impossibility of fairness in biometric systems (Proposition 3.3.1). We observe how the decision threshold value (τ) impacts on the fairness criteria that are satisfied. On one hand, setting τ at FGR_{1000} both equalised odds and statistical parity are not satisfied, yet predictive parity is achieved (Table 2). On the other hand, when τ is set at $\sim ZMR$, then predictive party is held but equalised odds and statistical parity are not achieved (Table 3).

The empirical results at FGR_{1000} clearly hold Chouldechova's observation in [22, p. 157] about predictive parity, *FGR*, and *FIR*. When a system satisfies predictive parity but ratios differ across demographic groups, the system cannot achieve equal *FGR* and *FIR* across those groups. Thus, if *FGR* are not similar across categories in the demographic group scrutinised, equalised odds cannot be satisfied. Analysing disparities among groups, old adults obtain worsen results than young, which could be a consequence of low proportion of old adults in the dataset. Gender also has significant unfair results. For instance, MC and PC obtains higher *FGR* on females than on males (421.18% and 198.8% respectively). Fair criteria are also worsen on non-Europeans than Europeans.

On the other hand, setting thresholds near *ZFIR* ($\sim ZFIR$) implies that predictive parity is not satisfied. These results clearly demonstrate Proposition 3.3.1 that states that the three fairness definition cannot hold simultaneously.[13] Rather than race, we observe that age and gender obtain wider differences. In this case, old people and female individuals obtain larger error rates than young and males respectively.

4.3 Intersectionality

Given the intersectional benchmark proposed in *Gender Shades* [18], we evaluate the results of the four biometric recognition systems based on intersectional groups. The results show that disparities are more prominent on age and gender, rather than on race. Unlike facial recognition, finger veins systems do not take into account skin tones and their performance is more affected by other attributes such as fingers size. Moreover, wide differences in ratios among Europeans and non-Europeans (see Fig. 2) impacts also on these results.

Figure 3 shows the distribution of genuine and impostor scores across four groups (young females, old females, young males and old males) at two decision thresholds. On the four biometric systems, young male scores obtain better results than the other groups. The distribution of genuine scores (orange box) is more right-handed, which results in a higher number of correct matches ($\uparrow TGR$). In contrast, impostor scores (purple box) are less left-handed, which implies high rates of false matches ($\uparrow FGR$). Genuine distributions of young males obtained the highest lower quartile (Q_1) on the four systems. Young females obtain higher genuine scores than old females and males. Overall, all biometric systems perform worst on elderly males. For instance, Q_1 of PC's genuine distribution of old males is the only one below τ when FGR_{1000}. This implies that the *FGR* of this group is substantially higher than other groups. The distribution of impostor scores is similar across the four demographic groups and biometric systems. Impostor distributions are narrower than genuine distributions. However, the number of outliers is considerable (see LBP's distributions). Setting τ at $\sim ZFIR$, young males obtain higher *TIR* than females and elderly adults.

5 The politics beyond fairness in biometric systems

Biometric systems will generally be biased as we have previously demonstrated. Yet these systems are getting more accurate and fairer.[14] The percentage of errors and biases are decreasing consistently over the years. Private companies and research centres are training their systems with better image quality and more sophisticated algorithmic architectures which outperforms previous versions. For instance, the latest publication released by NIST [47] reported that the best vendor's facial recognition has negligible errors with VISA border photos ($FGR_{10^6} = 0.0023$). Moreover, the analysis on demographic disparities shows no significant differences. To demystify myths on the technical details of algorithmic systems, we argue that a critical approach to fairness in biometrics, coined as 'critical biometric consciousness' by the scholar Simone Browne [16, p. 116], should also shift the attention towards the political context and racialised mechanisms where biometrics are embedded in [7, 20].

Since 2001, the EU has established a digital border infrastructure for migration control [17, 74]. The European Asylum Dactyloscopy Database (EURODAC), the Visa Information System (VIS), the Schengen Information System (SIS II), and the new Entry–Exit System (EES) are the four main databases that Member States use to determine

[12] We consider that the system is unfair if the difference is greater than 5%.

[13] Note we cannot set the threshold at *ZFIR* because this implies dividing by zero when calculating fairness definitions. Therefore, we propose the evaluation near the *ZFIR*, noted as $\sim ZFIR$.

[14] See: Facial Recognition Verification Performance by NIST: https://pages.nist.gov/frvt/html/frvt11.html (Accessed 16 Nov 2022).

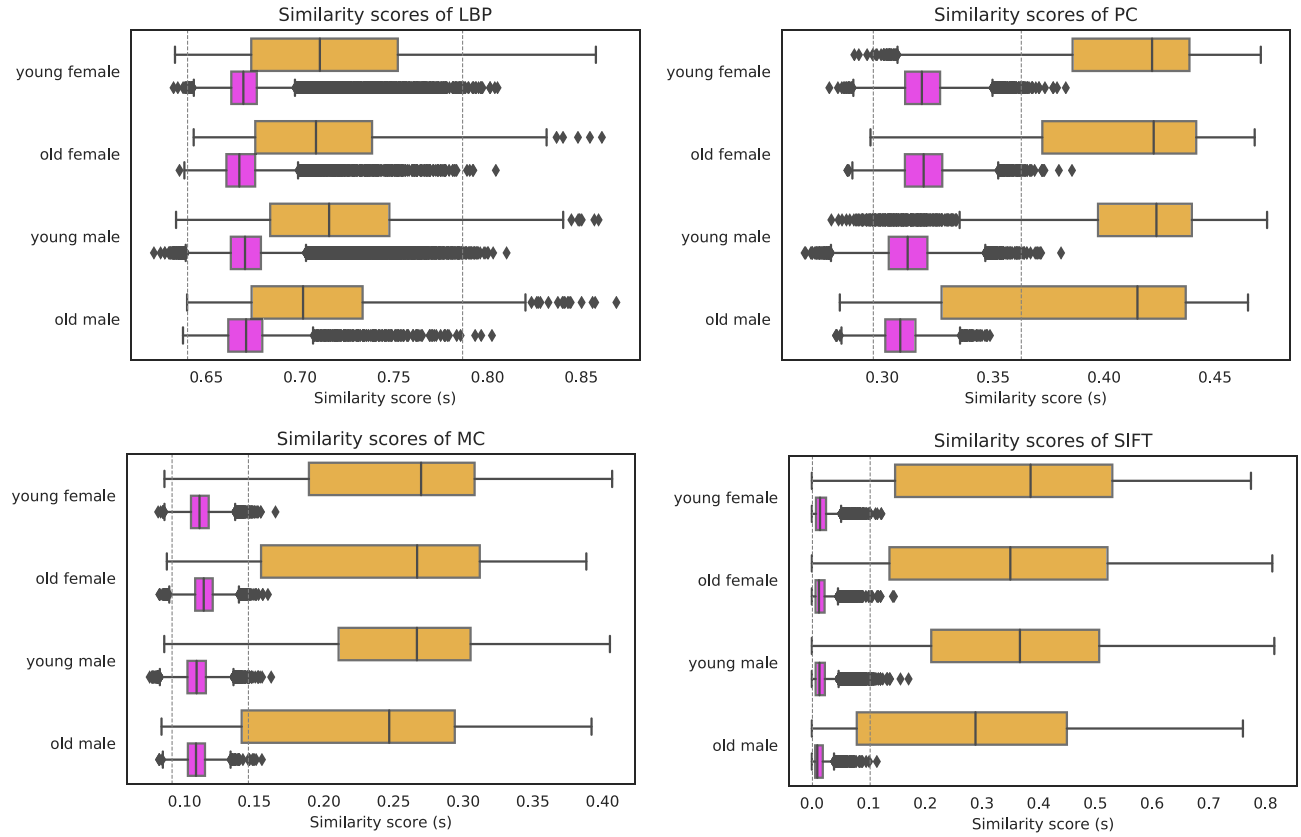


Fig. 3 Intersectional fairness disparities in four biometric recognition systems based on age and gender. Decision threshold (τ) is set at $\sim ZFIR$ (first dashed line) and FGR_{1000} (second dashed line). Distribution of genuine (orange) and impostor (purple) scores differ among groups. Classification disparities across intersectional demographic groups are apparent. Overall, young males obtain better performance than other groups

responsibility for examining an asylum application, register visa applications or border crossings. These systems are provided with biometric systems in order to register, identify and criminalise migrants [64, 86, 87]. Through these databases, migrants who have entered the Schengen area are registered and identified through their fingerprints. The storage of biometric samples in these databases is used to verify the identity of asylum seekers in countries which 'are not responsible' for their asylum petition (EURODAC) or detect those people whose visa has expired (EES and VIS). However, biometric traces provided by these biometric systems are used in other contexts such as asylum tribunal decisions. On May 2019, the First-tier Tribunal in the UK refused a petition of asylum to an individual national of Iraq of Kurdish origin due to a biometric evidence given by one of these systems [91], alongside other 'discrepancies' on his narrative.[15] The asylum seeker claimed that he was at risk of serious harm, stating that one family member had been killed and his own house was intentionally set on fire. He appealed against the decision, the Deputy Upper Tribunal Judge did not set aside the previous decision:

> A further document was relied upon by the Respondent at that hearing, being a EURODAC search result, demonstrating that a person in the Appellant's identity was fingerprinted in Dresden in Germany on 22 March 2016. The Appellant's account as given in his Statement of Evidence (SEF) interview and confirmed in oral evidence before the judge was that he only left Iraq in December 2017. The Appellant denied before the judge that the person identified in the EURODAC search was him but the judge stated that she was satisfied by the details contained within the document, and looking at the clear photograph on the EURODAC match, that the person fingerprinted in Germany was indeed the Appellant. [...] The evidence provided by the EURODAC document is unequivocal. The photograph contained within the document is clearly the Appellant who appeared before me and I have no reason to doubt that the document relates to him. Therefore, this document upon which I am satisfied I can

[15] Note this date is before Brexit, so the UK had full access to EURODAC.

> place significant weight puts him in Germany on 22 March 2016. Consequently, I find that this evidence undermines the credibility of his entire account and his credibility as a witness in his own cause and that it renders his entire account unreliable. [91, p. 2]

This case unveils how biometric systems are used nowadays as an algorithmic evidence at the border to refuse and fail asylum seekers. As Browne has argued biometrics are a technology 'to make the mute body disclose the truth of its racial identities' and 'that can be employed to do the work of alienating the subject by producing a truth about the racial body and one's identity (or identities) despite the subject's claims' [16, pp. 108–110]. The evidence given by EURODAC's fingerprint system was placed ahead of the person's narrative within the hierarchy of truthfulness [9]. The asylum seeker was portrayed as a deceptive subject given that the biometric system unveiled the 'truthful' of his whereabouts. The inconsistency encountered did not question the credibility of the technology, but rather affected the asylum seeker's credibility. Moreover, this inconsistency became decisive for the UK's judicial power to refuse his asylum petition.

Perhaps less intuitively, this asylum appeal also exposes *the elephant in the room* of biometrics. Whilst engineers are centering their efforts in training better performing, fairer, and more equitable biometrics, the same systems are implemented at the border to deny asylum and push migrants back. As previously shown, we are witnessing a trend on fairness in biometrics [26]. In the previous section, we have demystified the possibility of fairness in biometric systems taken into account different definitions of algorithmic fairness. However, in this paper, we argue that part of the literature of fairness in biometrics neglects how these systems are implemented in real life for migration control. We also observe this trend beyond biometrics, proposing approaches to better distribute asylum seekers within a country [4, 11, 25, 57]. However, algorithmic fairness cannot distribute justice in scenarios which intended purpose is to discriminate and that consistently jeopardise fundamental rights. As Tenday Achiume has argued: '[T]here can be no technological solution to the inequities of digital racial borders' [2, p. 337]. Fairer biometrics and algorithmic solutions implemented at racialised borders conceals the injustices that these infrastructures reproduce. These social, political and historical injustices are *the elephant in the room* of biometrics, the controversial issue that is obvious but remains ignored and unmentioned in debates around borders, biometrics and fairness.

In April 2021, the European Parliament published the AI Act, the first legal framework proposal to regulate artificial intelligence [33]. The scope of the AI Act is to address the risks associated with the use of such a technology and protect safety and fundamental rights. Within this document, biometrics is considered a high-risk system in the following areas: (i) biometrics identification and categorisation of natural persons, (ii) migration, asylum and border control management, (iii) law enforcement and (iv) emotion recognition. Probably, the recent campaigns against mass surveillance using facial recognition organised by several organisations have played a key role for the regulation of biometric systems [14, 30]. Indeed, the proposal opts to ban the use of 'real-time' facial recognition in public spaces [94]. However, certain exceptions are considered regarding the use of biometric systems such as targeted search for specific potential victims of crime, threat to the life or physical safety of natural persons or of a terrorist attack or perpetrator or suspect of a criminal offence. Interestingly, Article 83 exempts large-scale biometric systems used for migration control in the EU from this regulation:

> This Regulation shall not apply to the AI systems which are components of the large-scale IT systems established by the legal acts listed in Annex IX that have been placed on the market or put into service before [12 months after the date of application of this Regulation referred to in Article 85(2)], unless the replacement or amendment of those legal acts leads to a significant change in the design or intended purpose of the AI system or AI systems concerned. [33, p. 88]

Despite of the fact that the legal document categorised as high-risk biometric systems used in the context of migration, border control management and law enforcement, the previous text exposes that this same regulation does not apply on the four biometric databases (EURODAC, VIS, SIS II, and EES), which are used specifically for these purposes. Regulation will only apply when there is a 'significant change' within these systems, but the proposed document does not provide what does a 'significant change' mean. Thus, the AI Act will not entail any substantial legal change for migrants and asylum seekers who are screened by biometric systems upon arrival in Europe.

Whilst facial recognition technologies used by police authorities in public spaces has been banned by the European Parliament [70, 81], the use of fingerprints to immobilise migrants will remain legalised and in the shadows of any public or political debate. As the EU Rapporteur, Petar Vitanov, said after the resolution approved by the European institution: 'This is a huge win for all European citizens'. Yet, fundamental rights for non-Europeans will be put into the background. Asymmetries in legal, political and social rights have been historically confronted. In her classic *Women, Race & Class*, Angela Davis narrates the frictions between the (white) feminist movement and the enslavement of Black people. During the women's rights campaign in the US, she explains the advanced political position of an

American abolitionist and women's rights advocate: 'But Angelina Grimke proposed a principled defence of the unity between Black Liberation and Women's Liberation: "I want to be identified with the Negro", she insisted. "Until he gets his rights, we shall never have ours." ' [24, p. 59]. Bringing Grimke's political consciousness to our discussion about the regulation of biometrics, we suggest that until migrants get their digital and fundamental rights, we shall never have ours.

6 Conclusion

Fairness has emerged in the context of biometrics as an emergent research area. It aims at addressing and mitigating demographic biases in systems designed to identify or authenticate subjects based on body features (eyes, face, finger veins, among others). Yet, fairness has overshadowed *the elephant in the room* of the use of biometrics, the controversial issue which is obviously present but avoided as a subject for discussion. Biometrics has a long-standing colonial and racial legacy which is usually ignored by the biometric industry and research field. This heritage is still latent today with the implementation of biometric systems for the purposes of migration control and law enforcement. Whilst the study of fairness revolves around 'debiasing' biometric systems, migrants' fundamental rights are jeopardised by the use of this technology at the border.

In this paper, we argued that biometrics are and will be always biased. Building on the literature of fairness in machine learning, we demonstrated theoretically that biometric systems cannot mutually satisfy different fairness definitions. Then, we empirically proved the impossibility of fair biometric systems. We also observed that the biometric dataset proposed to train the biometric systems reproduce racialisation of bodies, underrepresenting non-Western subjects and using race categories that are colonial, archaic and offensive. Although our experiment is only tested in one dataset due to limitations on accessing biometric and demographic data, results clearly exposed that biometric systems show differences on three fairness criteria at different decision thresholds. Yet, this paper has pushed this argument further showing how the focus on the fairness of biometrics disregards the political discourse about the use of biometrics at the border. As we have shown, biometric systems are used nowadays by border and judicial authorities for migration control. The algorithmic decision is used to assess and challenge the narrative of the migrant or asylum seeker, and in case of an inconsistency, the biometric output is positioned as the ground truth. Moreover, the proposed AI regulation by the EU that bans and categorises certain biometric systems will not be applied to the large-scale databases that are used to immobilise and criminalise migrants. As J. Khadijah Abdurahman argues: '[I]t is not just that classification systems are inaccurate or biased, it is who has the power to classify, to determine the repercussions / policies associated thereof and their relation to historical and accumulated injustice?'[1].

In conclusion, the use of fairness in algorithmic systems installed in social and political contexts which principal and intended function is to discriminate, displaces the breach of fundamental rights because the algorithm is 'fair'. Fairer biometric systems embedded at the border will legitimise denials of asylum, push-backs or secondary movements. Moreover, the current debates around the demographic biases and the ethics of artificial intelligence overshadow political, social and historical discrimination that was there before the technology. As Browne argues: 'a critical biometric consciousness must acknowledge the connexions between contemporary biometric information technologies and their historical antecedents' [16, p. 118]. As this trend will become more prominent in the incoming years, there is an urgent need to shift the debates around the historical and political context in which most of these systems are embedded in the present.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43681-022-00249-2.

Acknowledgements We would like to thank the reviewers of AI and Ethics for their valuable comments. We are deeply grateful to Martina Tazzioli for her detailed and insightful comments on early drafts. We are grateful to Claudia Aradau for her guidance and advice during the course of this research and for sharing the UK Immigration and asylum chamber's digital archive with us. We also thank Lucrezia Canzutti, Sarah Perret and Tobias Blanke for their comments that greatly improved the article; and Lorena Jaume-Palasí and Reuben Binns for their ethical and data protection advice. The work of Ana Valdivia was supported by SECURITY FLOWS (ERC Consolidator Grant, grant number 819213) and the Dieter Schwarz Foundation. The work of Júlia Corbera-Serrajòrdia and Aneta Swianiewicz was supported by the King's Undergraduate Research Fellowship (KURF) scheme.

References

1. Abdurahman, J.K.: FAT* Be Wilin'. Medium blog. https://upfromthecracks.medium.com/fat-be-wilin-deb56bf92539. Accessed 9 Aug 2022 (2019)

2. Achiume, E.: Digital racial borders. Am. J. Int. Law **115**, 333–338 (2021). https://doi.org/10.1017/aju.2021.52
3. Acien, A., Morales, A., Vera-Rodriguez, R., Bartolome, I., Fierrez, J.: Measuring the gender and ethnicity bias in deep models for face recognition. In Iberoamerican Congress On Pattern Recognition, pp. 584–5930
4. Ahmad, N.: Refugees and algorithmic humanitarianism: applying artificial intelligence to RSD procedures and immigration decisions and making global human rights obligations relevant to AI governance. Int. J. Minority Group Rights **1**, 1–69 (2020)
5. Aloudat, A., Michael, K., Abbas, R.: The implications of iris-recognition technologies: will our eyes be our keys? IEEE Consum. Electron. Mag. **5**(3), 95–102 (2016)
6. Amoore, L.: Biometric borders: governing mobilities in the war on terror. Polit. Geogr. **25**, 336–351 (2006)
7. Amoore, L.: The deep border. Polit. Geogr. 102547 (2021)
8. Angwin, J., Larson, J., Mattu, S., Kirchner, L.: Machine bias. Propublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Accessed 8 Aug 2022 (2016)
9. Aradau, C., Perret, S.: The Politics of (Non-)knowledge at Europe's Borders: Errors, Fakes, and Subjectivity Review of International Studies, pp. 1–20 (2022)
10. Amnesty International.: Hotspot Italy: abuses of refugees and migrants. https://www.amnesty.org/en/latest/campaigns/2016/11/hotspot-italy. Accessed 29 Oct 2021 (2016)
11. Bansak, K., Martén, L.: Algorithmic decision-making, fairness, and the distribution of impact: application to refugee matching in Sweden (2021)
12. Barocas, S., Hardt, M., Narayanan, A.: Fairness and machine learning (2019)
13. Benjamin, R.: Race After Technology: Abolitionist Tools for the New Jim Code. Polity Press, Cambridge (2019)
14. Big Brother Watch: Big brother watch briefing on facial recognition surveillance (2020). https://bigbrotherwatch.org.uk/wp-content/uploads/2020/06/Big-Brother-Watch-briefing-on-Facial-recognition-surveillance-June-2020.pdf. Accessed 29 Oct 2021
15. Birhane, A.: The impossibility of automating ambiguity. Artif. Life **27**, 44–61 (2021)
16. Browne, S.: Dark Matters. Duke University Press, Durham (2015)
17. Broeders, D.: The new digital borders of Europe: EU databases and the surveillance of irregular migrants. Int. Sociol. **22**, 71–92 (2007)
18. Buolamwini, J., Gebru, T.: Gender shades: intersectional accuracy disparities in commercial gender classification. In: Conference on fairness, accountability and transparency (FAccT*), pp. 77–91 (2018)
19. Butler, J.: Gender Trouble. Routledge, London (1999)
20. Castelvecchi, D.: Beating biometric bias. Nature **587**, 347–349 (2020)
21. Choi, J., Song, W., Kim, T., Lee, S., Kim, H.: Finger vein extraction using gradient normalization and principal curvature. Image Process. Mach. Vis. Appl. II **7251**, 725111 (2009)
22. Chouldechova, A.: Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big Data **5**, 153–163 (2017)
23. Crenshaw, K.: Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. U. Chi. Legal F., p. 139 (1989)
24. Davis, A.: Women, Race, & Class. Penguin Random House UK, London (2019)
25. Dekker, R., Koot, P., Birbil, S.I., van Embden Andres, M.: Co-designing algorithms for governance: ensuring responsible and accountable algorithmic management of refugee camp supplies. Big Data Soc. **9**(1), 20539517221087856 (2022)
26. Drozdowski, P., Rathgeb, C., Dantcheva, A., Damer, N., Busch, C.: Demographic bias in biometrics: a survey on an emerging challenge. IEEE Trans. Technol. Soc. **1**(2), 89–103 (2020)
27. Drozdowski, P., Prommegger, B., Wimmer, G., Schraml, R., Rathgeb, C., Uhl, A., Busch, C.: Demographic bias: a challenge for fingervein recognition systems? In: 2020 28th European Signal Processing Conference (EUSIPCO), pp. 825–829 (2021)
28. Dunkelau, J., Leuschel, M.: Fairness-Aware Machine Learning An Extensive Overview. Working Paper (2019)
29. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R.: Fairness through awareness. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, pp. 214–226 (2012)
30. EDRi.: Facial Recognition & Biometric Mass Surveillance: Document Pool. EDRi. https://edri.org/our-work/facial-recognition-document-pool/. Accessed 29 Oct 2021 (2020)
31. EDRi.: The rise and rise of biometrics mass surveillance in the EU. https://edri.org/our-work/new-edri-report-reveals-depths-of-biometric-mass-surveillance-in-germany-the-netherlands-and-poland/. Accessed 29 Oct 2021 (2021)
32. Eubanks, V.: Automating inequality: how high-tech tools profile, police, and punish the poor. St. Martin's Press, New York (2018)
33. European Commission: Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021) 206 final) (2021)
34. Fang, M., Damer, N., Kirchbuchner, F., Kuijper, A.: Demographic bias in presentation attack detection of iris recognition systems. In: 2020 28th European Signal Processing Conference (EUSIPCO), pp. 835–839 (2021)
35. Feng, L., Chao, W., Jialiang, P.: Finger vein recognition using log gabor filter and local derivative pattern. Image Process. Pattern Recognit. **9**, 231–242 (2016)
36. de Freitas Pereira, T., Marcel, S.: Fairness in biometrics: a figure of merit to assess biometric verification systems. ArXiv:2011.02395 (2020)
37. Friedler, S., Scheidegger, C., Venkatasubramanian, S.: On the (im) possibility of fairness. ArXiv:1609.07236 (2016)
38. Friedler, S., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E., Roth, D.: A comparative study of fairness-enhancing interventions in machine learning. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 329–338 (2019)
39. Fussey, P., Murray, D.: Independent report on the London Metropolitan Police Service's trial of live facial recognition technology (2019)
40. Garg, P., Villasenor, J., Foggo, V.: Fairness metrics: a comparative analysis. In: 2020 IEEE International Conference on Big Data (BIGDATA), pp. 3662–3666 (2020)
41. Gabriel, I.: Toward a theory of justice for artificial intelligence. Daedalus **151**(2), 218–231 (2022)
42. Gilroy, P.: Against Race: Imagining Political Culture Beyond the Color Line. Harvard University Press, Cambridge (2000)
43. Glouftsios, G., Scheel, S.: An inquiry into the digitisation of border and migration management: performativity, contestation and heterogeneous engineering. Third World Q. **42**, 123–140 (2021)
44. Godbole, A., Grosz, S.A., Nandakumar, K., Jain, A.K.: On demographic bias in fingerprint recognition. arXiv:2205.09318 (2022)
45. Zhao, H., Gordon, G.: Inherent tradeoffs in learning fair representations. In: Advances In Neural Information Processing Systems, Vol. 32, pp. 15675–15685 (2019)
46. Grother, P., Grother, P., Ngan, M., Hanaoka, K.: Face recognition vendor test (FRVT). US Department of Commerce, National Institute of Standards (2019)
47. Grother, P., Ngan, M., Hanaoka, K.: Ongoing Face Recognition Vendor Test (FRVT) Part 1: Verification. NIST. https://www.nist.

gov/programs-projects/face-recognition-vendor-test-frvt-ongoing. Accessed 09 Nov 2021 (2021)
48. Guild, E., Groenendijk, K., Carrera, S.: Illiberal Liberal States: Immigration, Citizenship and Integration in the EU. Ashgate Limited, Surrey (2009). https://doi.org/10.4324/9781315587813
49. Hanna, A., Denton, E., Smart, A., Smith-Loud, J.: Towards a critical race methodology in algorithmic fairness. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 501–512 (2020)
50. Hardt, M., Price, E., Srebro, N.: Equality of opportunity in supervised learning. In: Advances in Neural Information Processing Systems, Vol. 29, pp. 3315–3323 (2016)
51. Hoffmann, A.: Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. Inf. Commun. Soc. **22**, 900–915 (2019)
52. Hutchinson, B., Mitchell, M.: 50 years of test (un)fairness: lessons for machine learning. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 49–58 (2019)
53. Jain, A., Flynn, P., Ross, A.: Handbook of Biometrics. Springer Science & Business Media, New York (2007)
54. Kantayya, S.: Coded bias (Netflix, 2020) (2020)
55. Kauba, C., Reissig, J., Uhl, A.: Pre-processing cascades and fusion in finger vein recognition. In: 2014 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–6 (2014)
56. Kauba, C., Prommegger, B., Uhl, A.: Focusing the beam-a new laser illumination based data set providing insights to finger-vein recognition. In: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–9 (2018)
57. Kinchin, N.: Technology, Displaced? The Risks and Potential of Artificial Intelligence for Fair, Effective, and Efficient Refugee Status Determination. Law Context Socio-legal J. **37** (2021)
58. Kleinberg, J.: Inherent trade-offs in algorithmic fairness. In: ACM International Conference on Measurement and Modeling of Computer Systems, Vol. 46, No. 1, p. 40 (2018)
59. Leese, M.: The new profiling: algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. Secur. Dial. **45**(5), 494–511 (2014)
60. Lohr, S.: Facial recognition is accurate, if you're a white guy. New York Times, Vol. 9, pp. 283
61. Lu, Y., Xie, S., Yoon, S., Wang, Z., Park, D.: An available database for the research of finger vein recognition. In: 2013 6th International Congress on Image and Signal Processing (CISP), Vol. 1, pp. 410–415 (2013)
62. Maguire, M.: The birth of biometric security. Anthropol. Today **25**, 9–14 (2009)
63. Marasco, E.: Biases in fingerprint recognition systems: where are we at? In: 2019 IEEE 10th International Conference On Biometrics Theory, Applications And Systems (BTAS), pp. 1–5 (2019)
64. Metcalfe, P., Dencik, L.: The politics of big borders: data (in) justice and the governance of refugees. First Monday **24** (2019)
65. Miconi, T.: A note on the impossibility of fairness (2017)
66. Mitchell, S., Potash, E., Barocas, S., D'Amour, A., Lum, K.: Algorithmic fairness: choices, assumptions, and definitions. Annu. Rev. Stat. Appl. **8**, 141–163 (2021)
67. Miura, N., Nagasaka, A., Miyatake, T.: Extraction of finger-vein patterns using maximum curvature points in image profiles. IEICE Trans. Inf. Syst. **90**, 1185–1194 (2007)
68. Noble, S.U.: Algorithms of oppression. In: Algorithms of Oppression. New York University Press, New York (2018)
69. O'Flaherty, M.: Facial recognition technology and fundamental rights. Eur. Data Prot. Law Rev. **6**(2), 170–173 (2020)
70. Ojamo, J.: Use of artificial intelligence by the police: MEPs oppose mass surveillance. European Parliament. https://www.europarl.europa.eu/news/en/press-room/20210930IPR13925/use-of-artificial-intelligence-by-the-police-meps-oppose-mass-surveillance. Accessed 12 Nov 2021 (2021)
71. Oliveira Martins, B., Lidén, K., Jumbert, M.G.: Border security and the digitalisation of sovereignty: insights from EU borderwork. Eur. Secur. **31**(3), 475–494 (2022)
72. Paullada, A., Raji, I., Bender, E., Denton, E., Hanna, A.: Data and its (dis) contents: a survey of dataset development and use in machine learning research. Patterns **2**, 100336 (2021)
73. Stanley, E.: Borders, Mobility and Technologies of Control, Sharon Pickering and Leanne Weber (eds). Current Issues in Criminal Justice, Vol. 19, No. 2, pp. 252–253. Springer, Dordrecht. https://doi.org/10.1080/10345329.2007.12036432 (2007)
74. PICUM and Statewatch: Data Protection, Immigration, Enforcement and Fundamental Rights. PICUM and Statewatch. https://picum.org/wp-content/uploads/2019/11/Data-Protection-Immigration-Enforcement-and-Fundamental-Rights-Full-Report-EN.pdf. Accessed 5 Aug 2022 (2019)
75. Preciozzi, J., Garella, G., Camacho, V., Franzoni, F., Di Martino, L., Carbajal, G., Fernandez, A.: Fingerprint biometrics from newborn to adult: a study from a national identity database system. IEEE Trans. Biometr. Behav. Identity Sci. **2**, 68–79 (2020)
76. Privacy International: PI Submission on the GPS Tracking of Migrants in the UK. https://privacyinternational.org/report/4866/pi-submission-gps-tracking-migrants-uk. Accessed 05 Aug 2022 (2022)
77. Queiroz, B.: The impact of EURODAC in EU migration law: the era of crimmigration? Market Compet. Law Rev. **3**, 157–183 (2019)
78. Raji, I., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J., Denton, E. Saving face. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (2020)
79. Rawls, J.: Justice as fairness: political not metaphysical. In: Equality and Liberty, pp. 145–173. Palgrave Macmillan, London (1991)
80. Ross, A., Banerjee, S., Chen, C., Chowdhury, A., Mirjalili, V., Sharma, R., Swearingen, T., Yadav, S.: Some research problems in biometrics: the future beckons. In: 2019 International Conference on Biometrics (ICB), pp. 1–8 (2019)
81. Sánchez Nicolás, E.: MEPs back EU facial-recognition ban for police. EUObserver. https://euobserver.com/democracy/153135. Accessed 12 Nov 2021 (2021)
82. Scheel, S.: Autonomy of migration despite its securitisation? Facing the terms and conditions of biometric rebordering. Millennium **41**, 575–600 (2013)
83. Scheel, S., Squire, V.: Forced migrants as illegal migrants. In: The Oxford Handbook of Refugee and Forced Migration Studies, pp. 188–199 (2014)
84. Serna, I., Morales, A., Fierrez, J., Cebrian, M., Obradovich, N., Rahwan, I.: Algorithmic discrimination: formulation and exploration in deep learning-based face biometrics. ArXiv:1912.01842 (2019)
85. Serna, I., Peña, A., Morales, A., Fierrez, J.: InsideBias: measuring bias in deep networks and application to face gender biometrics. In: 2020 25th International Conference On Pattern Recognition (ICPR), pp. 3720–3727 (2021)
86. Stenum, H.: The body-border. Governing irregular migration through biometric technology. Spheres J. Dig. Cult. **4**, 1–16 (2017)
87. Tazzioli, M.: The making of migration: the biopolitics of mobility at Europe's borders. SAGE, California (2019)
88. Terhörst, P., Kolf, J., Huber, M., Kirchbuchner, F., Damer, N., Morales, A., Fierrez, J., Kuijper, A.: A comprehensive study on face recognition biases beyond demographics. ArXiv:2103.01592 (2021)
89. Ton, B., Veldhuis, R.: A high quality finger vascular pattern dataset collected using a custom designed capturing device. In: 2013 International Conference On Biometrics (ICB), pp. 1–5 (2013)
90. Uhl, A., Busch, C., Marcel, S., Veldhuis, R.: Handbook of Vascular Biometrics. Springer Nature (2020)

91. Upper Tribunal (Immigration and Asylum Chamber): Appeal Number: PA/00240/2019. https://tribunalsdecisions.service.gov.uk/utiac/pa-00240-2019. Accessed 01 Nov 2021 (2019)
92. Van der Ploeg, I.: The illegal body: Eurodac and the politics of biometric identification. Ethics Inf. Technol. **1**(4), 295–302 (1999)
93. Vanoni, M., Tome, P., El Shafey, L., Marcel, S.: Cross-database evaluation using an open finger vein sensor. In: 2014 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, pp. 30–35 (2014)
94. Veale, M., Borgesius, F.: Demystifying the draft EU artificial intelligence act-analysing the good, the bad, and the unclear elements of the proposed approach. Comput. Law Rev. Int. **22**, 97–112 (2021)
95. Verma, S., Rubin, J.: Fairness definitions explained. In: 2018 IEEE ACM International Workshop On Software Fairness (FAIRWARE), pp. 1–7 (2018)
96. Xie, C., Kumar, A.: Finger vein identification using convolutional neural network and supervised discrete hashing. Pattern Recognit. Lett. **119**, 148–156 (2019)