CREST Centre for Research and Evidence on Security Threats

Bias in Emerging Biometric Systems: A Scoping Review

FULL REPORT

DECEMBER 2022

KAT GIBBS

SOPHIE NIGHTINGALE

DECEMBER 2022

Bias in Emerging Biometric Systems: A Scoping Review

FULL REPORT

Kat Gibbs | Lancaster University
Sophie Nightingale | Lancaster University

This scoping review was produced as part of the Digital Emerging Biometrics project. The project provides summaries of cutting-edge research and knowledge with the aim of offering user-friendly guidance to help the NCA to put evidence into practice as well as guide future planning and policy. You can find all the outputs from this project at: *www.crestresearch.ac.uk/projects/digital-emerging-biometrics/*

ABOUT CREST

The Centre for Research and Evidence on Security Threats (CREST) is a national hub for understanding, countering, and mitigating security threats. It is an independent centre, commissioned by the Economic and Social Research Council (ESRC) and funded in part by the UK security and intelligence agencies (ESRC Award: ES/N009614/1).

www.crestresearch.ac.uk

TABLE OF CONTENTS

ABSTRACT

In light of the rapid development and implementation of systems such as automatic face recognition (AFR) technology (Furnell & Clarke, 2014), speech recognition technology (Savchenko & Savchenko, 2021), and behavioural biometric identification such as how people use a computer mouse (Siddiqui et al., 2022), research must come to understand the flaws and biases in these systems.

This scoping review aimed to identify the types of bias in AI-biometric systems, steps being taken to mitigate these biases, how effective these steps are, and to identify any gaps in the literature.

Database searches were conducted on WebofScience and PsychInfo. In total, the searches identified 80 papers and a further 10 found through scanning the selected articles for relevant references. After title/abstract review 28 papers were read in full and 23 identified as fitting the criteria for this review. From the 23 selected papers four main themes emerged: racial bias, age bias, gender bias, and solutions.

Despite some searches including the terms “disability” and “sexuality” no papers were found to fit the inclusion criteria. The implications of systems having demographic biases include risk of discrimination that potentially breaks equality laws across the world (Wang & Deng, 2019). Research proposed solutions for mitigating bias yet there did not appear to be a cohesive or interdisciplinary approach, meaning that even solutions effective in one context might not generalise more widely, thus potentially limiting their usefulness. More cross-discipline research is needed to assess and mitigate the biases within AI-biometric systems; ideally before these systems are applied throughout society.

INTRODUCTION

Biometrics concern the use of biological or behavioural characteristics of humans primarily for identification purposes, e.g., fingerprints, DNA. New technologies, especially those drawing on the incredible capabilities of Artificial Intelligence (AI), allow for a wide range of new and hopeful ways to identify an individual—these newer types of biometric are referred to as "emerging biometrics" and include, for example, the use of face and voice recognition software. A biometric trait is a measurable physiological or behavioural characteristic of a person that can be used to determine and verify the identity of a person (Jasserand, 2015). The earliest example of wide use of biometrics in security would be photo identification (Bellamy et al., 1999) such as on passports, which still relied on a human to verify the photo identification is that of the holder.

Now with AI and machine learning, biometric identification has been streamlined to reduce human involvement in the process; for example, with automatic facial recognition technology to identify suspects of crime from anywhere such as shops to the streets, and now even in officer's body-worn cameras (Bromberg et al., 2020). Biometric technology is also used to secure information and places, from anything as simple as unlocking a phone, or unlocking rooms in government buildings (Galterio et al., 2018).

The newly developing field of biometric (AI) systems used in security is helpful for identity verification, and also, potentially, provides more security for system access than using a PIN which can easily be forgotten or stolen. Fingerprint biometrics technology was the first mainstream security biometric, but technologies are now developing beyond Touch ID fingerprint scanners on phones (Al-Daraiseh et al., 2015) that was pushed into the mainstream tech industry by Apple in 2013 (Goode, 2014). Advances in technology, including use of machine learning, has allowed facial recognition to develop at pace, and over the past decade have emerged into mainstream security use (Furnell & Clarke, 2014).

Emerging biometric technologies are providing novel ways to identify people based on their biological traits such as new types of feature recognition, for example iris recognition (Alwawi & Althabhawee, 2022), contactless fingerprint recognition (Yin et al., 2020), and speech identification (Savchenko & Savchenko, 2021). Behavioural biometric traits are also being researched and used as a means to identify people, for example analysing people's gait (Tahir Sabir, 2021), typing pattern and keystroke dynamics (Saini et al., 2018) and how people move a mouse (Siddiqui et al., 2022).

Using technology to improve security is beneficial as it saves time and is designed to have a higher accuracy in recognition and intended to reduce human error (Jain & Kumar, 2012). It has become common to develop and use systems that incorporate various biometric-based algorithms (e.g., face/voice recognition software), to support, or replace, human decision making in complex tasks, such as establishing or verifying the identify of individuals. Such automated decision systems can be highly accurate, however, there has been much debate concerning the existence of algorithmic bias and fairness in these systems (Drozdowski et al., 2020). The wide use of these systems alongside concerns about the fairness of such approaches makes it crucial to gain a better appreciation of how these are being used, what biases have been documented, how have these issues been evaluated and what has/should be done to mitigate bias to make automated decision systems fair for all.

In order for an effective AI biometric system the underlying neural network has to be trained and programmed on what to recognise. To train the neural network, it's fed data from a dataset and learns from the inputs and the outputs it has to achieve, creating a 'black box' in which the AI learns and makes its own assumptions on how to get to the desired outcomes (Buhrmester et al., 2021). Depending on the datasets fed to the neural networks and the coding used to

classify and categorise the data, bias may develop (Vincent & Hecht, 2021). Bias within these systems makes them less effective for any groups they are biased towards, and could present a weakness that hackers and fraudsters can exploit (Vincent & Hecht, 2021). Risk assessments carried out on defendants using AI technology to assess risk of reoffending, frequently predicting white defendants as less risky than they were and black defendants as more risky than they were (Angwin et al., 2016). The commercial AI system used to generate these risk scores; COMPAS has been found to be no more accurate or fair than the estimations of people with no expertise in the criminal justice system, showing how the bias in this system causes it to be detrimental to both people and the systems it tries to support (Dressel & Farid, 2018).

Although it is claimed that COMPAS doesn't directly collect or use any racial demographic data, the bias can arise through other correlated factors that can confound the statistics, for example, including employment or socioeconomic status. In addition, COMPAS considers scores of intelligence, extroversion and introversion (Angwin at al., 2016; Dressel & Farid, 2018); yet intelligence and personality tests have been previously shown to be biased against black people and other minority groups (Reynolds et al., 2021). Therefore, even AI that does not directly collect or use data about a person's race can become racially biased indirectly. The implications of racial biases are that people of different ethnicities may face more unfairness when it comes to the criminal justice system, and that the bias these systems hold can have negative impacts on people's lives. These issues highlight that before the systems can be used responsibly and ethically, work must be done to understand and eliminate the bias within them.

While all of these new technologies are racing ahead with their production and implementation, measures of fairness and reducing bias can be overlooked (Röösli et al., 2020). One concern that has been noted is that there are very few people belonging to minority groups involved in the creation of these systems (Martinez-Martin et al., 2021) which is likely to result in the creation of less diverse systems that are less effective for certain groups of people.

Failing AI-based biometric systems have already been reported in sectors such as hiring, with the failings of Amazon and HireVue becoming publicly known, with lawsuits filed against them for bias and data mismanagement. HireVue have since dropped their facial monitoring biometrics AI after a lawsuit following the discovery of racial and gender bias (Kahn, 2021). Amazon, Microsoft and Google's variants of biometric recognition technology, particularly facial recognition technology, also show evidence of gender and race bias (Wiggers, 2021). Mitigating bias within these systems is a challenging problem because it's hard to predict where bias will arise, whether it's from the planning stage, coding stage, or the bias the machines learn from the datasets (Martinez-Martin et al., 2021). Yet to gain a system that works well for everyone in all contexts, whether it's recruitment, law enforcement and surveillance, or healthcare, AI biometric systems must be free of bias in order to be used ethically and responsibly.

This project has produced a scoping review of bias in emerging biometrics: Specifically, the aim is to learn about bias in emerging biometrics, who these biases primarily impact, where the bias in the systems is most prevalent, as well as any steps forward to mitigate such biases. The review identifies where there are gaps in the research literature in terms of biases that exist yet the risk these pose has not been sufficiently researched. The report provides an overview of the current state of bias in biometrics. As such, this review aims to address four main research questions:

- To identify the different types of bias in emerging biometrics systems
- To look at the effectiveness of actions to mitigate bias in emerging biometrics
- To suggest next steps in mitigating bias in emerging biometrics based on current research
- To identify gaps in current research looking at bias in emerging biometrics

METHOD

This scoping review was informed by PRISMA scoping review guidelines (Tricco et al., 2018).

To answer the research questions, inclusion criteria used in this review stated any article must be:

1. Published in the last 10 years
2. A scientific paper published and peer reviewed in a journal or published online
3. Include information about bias concerning demographic attributes such as race, gender, sexuality, disability or other personal characteristics used in emerging biometric systems such as behavioural identifiers, facial recognition and wireless fingerprint scanners.

Exclusion criteria for this scoping review include:

- Biometric identifiers already established and widely used such as traditional fingerprint scanning will be excluded in the search as technology already implemented and well developed 10 or more years ago does not qualify as an emerging biometric for the purposes of this project.
- Articles that do not talk about bias in biometrics systems will also not be included in this review as they would hold no relevance to answering the main research questions.
- Articles discussing non-human biometrics will also be excluded from the review as this review is focusing on bias in human biometrics.
- Furthermore, the review will exclude articles that are not written in English. This is because the researchers do not have the tools to make accurate translations.
- Case studies and ethnographical studies which only contain one person will be excluded as findings may not generalise beyond the given individual.
- All grey literature will be excluded as it may not be peer reviewed.

DATABASES TO SEARCH

Databases searched covered a Psychology database PsycInfo with the aim of identifying literature about bias, and a multi-disciplinary citation database Web of Science, with the search set to scan the full papers for the keywords.

Search Terms:

- Emerg* (Emerge, emerging, emergent)
- Biometric* (Biometric, biometrics, biometrical)
- Bias* (Bias, biased)
- Fair* (Fairness)

As our initial searches conducted in April 2022 with these terms did not yield many results on either database (PsychInfo N=0, WebOfScience N=1). We reran the literature search excluding one of the four key terms per search, and then with the inclusion of each of the following more specific terms:

- Rac* (race, racism, racist)
- Sexuality
- Gender
- Disabilit* (disability, disabilities)

In total this resulted in 14 searches, each combination of search terms is shown in *Table 1*.

Search	Key Terms used	Language	Dates	Document types	Results	Selected
Webofscience Search 1	TS=(Emerg*, Biometric*, Bias*, fair)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	1	1
Webofscience Search 2	TS=(Emerg*, Biometric*, Bias*)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	5	1
Webofscience Search 3	TS=(Biometric*, Bias*, fair)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	22	5
Webofscience Search 4	TS=(Biometric*, Bias*, Rac*)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	17	5
Webofscience Search 5	TS-(Biometric*, Bias*, Gender)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	33	6
Webofscience Search 6	TS-(Biometric*, Bias*, Sexuality)	English	2012-2022	All document types	0	0
Webofscience Search 7	TS=(Biometric*, Bias*, Disabilit*)	English	2012-2022	All document types	2	0
PsychInfo Search 1	TS=(Emerg*, Biometric*, Bias*, fair)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	0	0
PsychInfo Search 2	TS=(Emerg*, Biometric*, Bias*)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	0	0
PsychInfo Search 3	TS=(Biometric*, Bias*, Fair*)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	0	0

Table 1. Number Of Searches In Databases With Each Search Term

PsychInfo Search 4	TS=(Biometric*, Bias*, Rac*)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	0	0
PsychInfo Search 5	TS-(Biometric*, Bias*, Gender)	English	2012-2022	Article, Book, Book Chapter, Database Review and Software Review	0	0
PsychInfo Search 6	TS-(Biometric*, Bias*, Sexuality)	English	2012-2022	All document types	0	0
PsychInfo Search 7	TS=(Biometric*, Bias*, Disabilit*)	English	2012-2022	All document types	0	0

Table 1. (continued) Number Of Searches In Databases With Each Search Term

PROCEDURE

We used the extended PRISMA protocol for scoping reviews as a guide when conducting this research (Page et al., 2021). Researchers initially ran searches across two databases; PsychInfo to assess the information available on psychology databases about bias within systems, and WebOfScience to broaden the search across a larger multidisciplinary database. The articles deemed relevant based on the title/abstract were then assessed on a full text basis to see if they meet the inclusion or exclusion criteria. Any discrepancies were discussed by the research team. A final list of papers to be included in the review was created. Finally, these papers underwent data charting to extract information to help answer the proposed research questions.

A spreadsheet was used to record progress—this recorded the search term, the database, the date of the search, the number of hits, the number of articles screened, and the number of relevant results found. Papers deemed relevant were also downloaded and kept in organised folders. The reasons for paper exclusion after reading the full text was documented.

Any articles read on a full text basis were also scanned for any potentially relevant articles in the reference lists as a snowball sampling technique. A separate page on the spreadsheet was created to keep note of these articles. After the database searching, these articles were then reviewed on a full text basis to see if they meet the inclusion or exclusion criteria. Through this snowballing, any articles missed via the formal database searching may be found through scanning citations.

DATA CHARTING

The researchers used data charting to extract and organise key items of information from the included sources. The data was charted using a ‘data charting form’ in Excel. The extracted content included a mixture of specific and more general information. We included article characteristics (e.g., author(s), year of publication, country of origin, funder), content focus (e.g., type of biometric, type of bias, benefits, unintended consequences, ethical issues, technical issues, bias evaluation/mitigation, future directions), and for research studies also more general information (e.g., aims of the study, method, important findings).

RESULTS

The search on the WebOfScience database identified 80 papers. The researchers then screened the abstracts for relevance against the inclusion and exclusion criteria, selecting 18 of these to be read and screened in full. The search on PsychInfo yielded no results for any of the combinations of search terms.

Using a snowball approach and scanning the reference lists of the 18 selected papers lead to the review an additional 10 papers, resulting in a total of 28 papers to be reviewed in full. Five papers were excluded after reading in full as they were deemed not to fit the research criteria—namely they could not contribute to answering any of the research questions posed in this scoping review. Therefore, the total number of papers included in this review is 23.

The primary researcher (KG) extracted data from the papers after reading them fully, and coded the papers fully. Then as themes emerged another researcher (SN) on the team coded 15% of these papers. Researchers agreed on the data extraction and the themes the papers fell under, and Cohen's Kappa was run to establish intra-rater reliability to assess the agreement of two researchers as a measure of quality control before (Cohen's $\kappa = 0.67$) and after discussion of the coding (Cohen's $\kappa = 0.89$) to ensure a strong agreement.

Papers in this literature review overwhelmingly focused on bias within facial recognition technology. The main themes extracted were racial bias, gender bias, age bias, and potential solutions to mitigate any bias. Papers focus on facial recognition looking at mapping features on the entire face (Khiyari & Wechsler, 2016) and also iris recognition (Alshareef et al., 2021) and explore the bias prevalent within technologies used in a security setting such as gaining access to systems and data (Liang et al., 2019) or looking at bias in systems such as those designed to recognise people in public places (Wang et al., 2019). Research examined in this review highlights in particular, racial bias (Buolamwini & Gebru, 2018; Robinson et al., 2020), age bias (Terhörst et al., 2019), and gender bias (Costa-Pazo et al., 2021) within biometric systems and how this may impact accuracy. Much of the literature involves papers with research into novel ways of reducing bias (Gong et al., 2020; Kloppenburg & van der Ploeg, 2018), with only a minority of the literature suggesting that the problems of systematic bias in biometric systems lie more broadly in society which influences the datasets used for training and how the programs are coded.

RACIAL BIAS

Literature looking at racial bias focuses on highlighting how AI-biometric systems perform on people of different ethnic backgrounds. Most commonly, research suggests that often non-white ethnic groups are disadvantaged by AI-biometric systems (Bacchini & Lorusso, 2019; Buolamwini & Gebru, 2018; Gong et al., 2020; Robinson et al., 2020). Bias in training datasets has been detected in multiple instances which suggests that how the AI programs are trained in the beginning influences their fairness (Ortega et al., 2021). Research suggests that unbalanced datasets consisting of more biometric data from white individuals than other racial groups could be a large contributor to racial bias within biometrics, which in turn could have a large impact on how effective the systems are in practice (Buolamwini & Gebru, 2018; Robinson et al., 2020). One proposed method for engaging with this problem is not to ignore or discount demographic features from datasets, but rather to code them differently and separately, and recognise demographic features to also improve demographic estimation (Gong et al., 2020). A downfall to this would be the current capabilities of these systems to be able to code multiple different demographics;

research shows that facial recognition software can effectively identify one demographic group but struggle to identify different features in a group made up of different demographics (Khiyari & Wechsler, 2016). This result suggests that systems can cope with processing one or two different demographics, but they struggle when more variation is added.

There is strong evidence showing there is bias within biometric systems which suggests that they will have a measurable impact on people's lives (Khiyari & Wechsler, 2016), with one individual claiming that biometric technology should not exist at all, as it will only have discriminatory effects on minority groups (Williams, 2020). While this is perhaps an impractical stance as the technology will likely be continually developed regardless of the ethics surrounding it, it raises issues that are crucial to solve as these technologies are developed; to benefit society as a whole technology should not be implicitly biased in any way towards any demographics.

In contrast some of the research included in this review suggests technology is not the issue, but rather humanity's relationship with the way people are grouped into each of the demographic categories and sub-groups of humans and labels that are coded into the systems (Kloppenburg & van der Ploeg, 2018). It has been suggested that recognising sub-groups based on pre-defined categories of human that societies have decided upon may factor into the biometric systems' poor performance and that producing and enacting new ethnic and gender categorisations from the datasets available may be a way of mitigating bias (Kloppenburg & van der Ploeg, 2018).

Using AI-based biometric systems could also create problems for any companies or organisations using them, as any discrimination found within such systems has the potential to raise lawsuits. Algorithmic bias and racial discrimination within biometric algorithms fall foul of many different countries anti-discrimination and equality laws (Wang & Deng, 2019). Terhörst et al. (2020b) suggest that racial bias within facial recognition technology comes from bias in current face quality assessment technology, which decides whether an image can be used for recognition—for example high quality images of white people might be more accessible than for some other racial groups. They suggest that trade-offs have been made to gain high performance of these systems at the cost of strong demographic bias against sub-groups. Such a trade-off will not be beneficial when it comes to fairness.

Furthermore, research suggests that demographic bias in biometric systems rarely just affects one type of demographic group, and that the way that machine learning works means that a range of biases can be created and that attempts to eliminate one bias could introduce other types of bias (Serna et al., 2021). Most of the papers included in this review did not focus on a single demographic bias but instead covered several different types, with racial bias featured most commonly. Racial bias in particular is well documented within state-of-the-art facial recognition systems (Muhammad et al., 2021), however beyond facial recognition systems and iris recognition systems there appears to be a paucity of literature studying racial bias in other biometric systems such as voice recognition and gait recognition systems, as no literature for these systems showed up in our literature search.

AGE BIAS

Automated estimation of biometric attributes such as age are becoming increasingly important in technology ranging from forensics to use in social media as it is acknowledged that these systems will have to recognise people over time as they age. As such, recent research stresses the importance of having systems that are not biased towards different age demographics (Terhörst et al., 2019). Literature examined within this review also highlighted age bias as a problem when it comes to creating a fair biometric system, framing it as a problem that needs to be solved by new technology and ways of coding data. Current algorithms tend to mis-predict age with high confidence scores which makes them less successful than human predictors

who consider surrounding conditions and factor in their lack of experience (Terhörst et al., 2019). As with race bias, much of the problem with age discrimination and worse performance of biometric systems when it comes to age results from the datasets the systems are trained on: If a particular demographic is under represented when it comes to biometric identification, the systems will have worse performance compared to the largest demographics in the dataset (Buolamwini & Gebru, 2018; Terhorst et al., 2019).

Age bias in biometric systems is acknowledged across both hard biometrics (technology to measure these was built specifically for identification and with a substantial evidentiary basis, such as DNA, fingerprints) and soft biometrics (those we use more to naturally identify each other, e.g., skin and hair colour, weight, accessories) (Khiyari & Wechsler, 2016; Marsico et al., 2017; Muhammad et al., 2021; Terhörst et al., 2019). Therefore, although age bias in facial recognition was most commonly the focus in the papers in this review, this issue is not unique to facial recognition (Marsico et al., 2016). The suggestion to mitigate age bias is that datasets, especially face recognition datasets should not have explicitly pre-determined demographic groupings within the data for age (or any other demographic). Instead, the system has a number of classifiers each trained on a specific demographic class with the most accurate classifier's result being used for a given input (Marsico et al., 2016).

A study by Gong et al. (2020) proposes that to mitigate age bias from biometric facial recognition systems the data must be disentangled, along with race and gender data, saying it will also lead to more accurate age estimation. The solution Terhörst et al. (2019) propose is a multi-algorithmic fusion approach for age and gender estimation that is able to state the reliability of the model's prediction. This reliability measure should give an indication to the people using the model as to how accurate its output may be, and also mitigate the issue of users of biometric technology placing too much trust in the accuracy of the algorithms. Research also highlights security risks in which the age bias in AI algorithms can be exploited in ways such as physical presentation attacks, disguise/makeup, digital adversarial attacks, and tampering to alter the perceived age of a person in an image. This possible risk shows how not only different age demographics can be discriminated against by biometric face recognition systems, it can also pose a weakness which aggressive actors can exploit (Singh et al., 2020). Algorithms have a lower performance on younger age demographics between the ages of 18-30 (Khiyari & Wechsler, 2016) even with the proposed solution of convoluted neural networks for feature extraction which is supposed to filter age out from the other demographic attributes. Child face recognition in relation to aging is still a major issue within biometrics, with a large portion of biometric technology aimed at recognising the biometrics of people above 18 years of age (Srinivas et al., 2019). Much of the reviewed literature only considered age biases for people over 18 years of age.

GENDER BIAS

As with both race and age, studies examined in this review point out the problems of gender bias within existing biometric systems and how this can lead not only to unfair systems but weaknesses in security that people could potentially exploit. Discrimination and bias in biometric security systems like face attack presentation detection (which identifies whether the image is genuine or fake) looking at irises can result in women having a lower measure of protection from these systems than men because any flaws in the recognition system could be a weakness that hackers could exploit (Costa-Pazo et al., 2021; Fang et al., 2020). An indirect gender bias could come from that women are more likely to wear makeup, and biometric facial recognition systems have worse performances on faces altered by makeup (Fang et al., 2020). Such weaknesses could pose as a security risk if attackers realise they can hoodwink large systems and databases using the biases in the systems.

In line with the problems concerning data and racial and age biases, datasets with lower gender variation and less diverse gender representations will perform significantly worse than datasets trained on faces with larger amounts of gender diversity and gender expression. Similarly, larger datasets will typically yield better performance than smaller datasets due to the greater diversity within them, therefore datasets used to train biometric AI algorithms need more gender diversity and representation to increase better gender performance (Buolamwini & Gebru, 2018; Terhorst et al., 2019).

Another proposed solution to gender bias in AI systems is a 'normalisation approach' in which all individuals are treated similarly so as to reduce the gender bias in face recognition, and a solution in which gender and age are sorted under a multi-algorithmic fusion approach so that it reduces risk of one bias in one demographic affecting another, for example a system that performs worse on younger people, and one system that performs worse on women, and the two systems being used together giving poor performance for young women (Terhörst et al., 2020). Though it's suggested that this could still produce lower performance for certain groups if the multi-algorithm fusion itself is biased.

SOLUTIONS PROPOSED BY THE RESEARCH

Research included this review not only acknowledged and investigated bias in biometric systems, but also looked at how to reduce it, and explored ways to create fairer systems. One solution proposed is for the better sorting and coding of demographic features and to combine demographics with identity features (Gong et al., 2020) whilst another suggests the best way is to disentangle demographic information from the domain-differences (the differences in how each feature is coded e.g., facial mapping) (Liang et al., 2019). Another paper suggests that lower quality images of faces can lead to more demographic bias in systems, so the aim should always be to use the highest quality face images in order to avoid demographic bias, as face quality assessments will assign lower values to subgroups affected by recognition bias if the systems are not properly trained (Terhorst et al., 2019). In a novel approach, some researchers have begun trying to decipher the 'black box' of machine learning to examine how the AIs learn bias and how bias becomes coded into the systems, with the proposal that deciphering the black box is the key to being able to create fairness (Ortega et al., 2021). All of research implementing these solutions suggested the results were bias free, or that the bias was reduced significantly (Gong et al., 2020; Ortega et al., 2021). However, these solutions may not stand up in other contexts and while all these methods may reduce bias in specific scenarios, to an extent, it remains to be seen how well the results generalise (Gong et al., 2020; Ortega et al., 2021). More varied solutions might be required as those covered in the review largely involved training AI systems on different datasets or on more demographically mixed datasets with the idea of producing adaptive margins (margins to allow greater or lesser error) to reduce biases (Wang & Deng 2019). There's a gap in the papers in this review looking at any solutions including the human in the loop; the human that helps program the AI or monitor decisions or the machine learning it undertakes.

Bias in computational biometric systems may be reflecting the biases already present and systematic in society. A solution needs to not only be at the computational level, as research has shown that bias may be coded into the systems by the humans who create them (Terhörst et al., 2019). Humans have been shown to have an equivalent bias when it comes to measuring and categorising humans, which in turn may impact the kinds of bias that are unintentionally being coded into AI biometric systems, including age, race, and gender bias (Robinson et al., 2020). This therefore could indicate that whatever methods of removing bias from the computational aspects of biometric AI, bias will not be mitigated given that the humans creating them influence systems through existence of their own biases.

ABLEISM IN BIOMETRIC SYSTEMS

The search terms included in this literature review found very little research highlighting disability bias as a problem area that needs addressing and solving. The review did reveal research indicating that people with dexterity issues may struggle with fingerprint scanners or holding a face or iris scanning device, and someone with a voice tremor or non-typical speech may find that voice recognition does not work for them (Young-Powell, 2021). These results can be considered as a double-edged sword; firstly, disabled people are less able to engage in a digital lifestyle which might in turn disable them further through their exclusion from datasets used to train AI software. Although going beyond the results of this literature search, it is important to draw attention to reporting of disability bias in biometric systems in the media (Engler, 2022; Young-Powell, 2021). Specifically, news reports describe how existing technologies, such as AI hiring systems making it more difficult for those with social and communication disabilities to gain employment. Ableism bias is largely overlooked and we highlight that this bias requires much greater attention in the research literature to ensure fairness in future technology.[1]

1 References used in this section were not found as a result of the systematic review to identify bias, and instead are media sources which suggests a lack of research examining disability bias within biometrics.

CONCLUSION

Current literature on emerging biometric systems shows that bias is prevalent throughout the use of this technology in different sectors such as healthcare, employment sectors, and security. Overwhelmingly a large portion of the literature focuses on the biometrics surrounding facial features such as facial identification (Bacchini & Lorusso, 2019) and iris identification (Alshareef et al., 2021; Alwawi et al., 2022) and highlights how flaws in these systems can weaken the safety when it comes to their real-world application.

This focus on facial recognition technology suggests a need for research into bias in other types of biometrics including, but not limited to; voice recognition technology, speech pattern recognition, gait recognition, typing pattern recognition and recognition of other features such as ear shape. Racial biases were most prominently noted throughout the reviewed literature, although it was common that papers discuss multiple types of bias (Bacchini & Lorusso, 2019; Buolamwini & Gebru, 2018; Gong et al., 2020; Robinson et al., 2020). Age bias is also prevalent with one system described as being less effective when it came to identifying younger people (Khiyari & Wechsler, 2016). This bias against younger people is somewhat contrary to prior research expectations suggesting that a deeper understanding of the types of age bias in biometric systems is needed before attempts can be made to tackle them. Gender bias shows that biometric systems are often less effective for women (Costa-Pazo et al., 2021; Fang et al., 2020).

There is a lack of literature on disability bias within biometric systems for the search terms used in this scoping review, which also included disability as one of its search terms, thus highlighting disability bias as an underdeveloped field of research and as an area the field needs to develop in. Furthermore, the inclusion of "sexuality" as a search term revealed no results indicating that sexual orientation might be another area that has been overlooked to date. The bias found within these AI-biometric systems suggests that more research needs to be done to allow development of fair systems—ideally, this should happen before such systems are used in a mainstream setting, otherwise the risk of discrimination, and discrimination charges for those who use will likely become a more prevalent issue. As things stand, AI-biometric technologies are biased in such a way that could be detrimental to several demographics and revealing a lack of fairness. As the systems can only be used for limited demographics effectively, this renders them ineffective in a diverse society.

Only 23 papers were identified as relevant to review for this paper which highlights the need for more research within this area, and in particular interdisciplinary research. Future research should consider bias in biometrics outside of facial recognition and iris recognition technology, and to also consider a wider range of biases (Young-Powell, 2021). More research needs to be done in this area as there is a clear lack of evidence showing whether multimodal biometrics systems would be any less biased, or whether other search terms could be included to gain more papers for a review. Future research in this area may also want to include the search terms sex* to encapture papers that use the term sexual orientation as well as sexuality, and disab* and neurodiver* to capture more disability inclusive terms such as disabled, disability, neurodivergent and neurodiverse.

From the research examined in this scoping review, the field paints a picture of flawed systems with no cohesive strategy to trying to mitigate the bias within them. There appears to be no agreed upon approach or methodology to mitigate bias in biometric systems and each solution might only work on a specific type of bias or in a specific system. By nature, this is a multidisciplinary problem and yet all solutions

proposed by papers in this scoping review tended to each only focus on a single discipline rather than multiple disciplines that blend both computational science and psychology. The area of looking at solutions to mitigate bias currently appears incohesive and discipline-specific, and future research should look on a more universal solution that factors in multiple methods of bias mitigation.

REFERENCES:

Al-Daraiseh, A., Al Omari, D., Al Hamid, H., Hamad, N., & Althemali, R. (2015). Effectiveness of Iphone's Touch ID: KSA Case Study. *International Journal Of Advanced Computer Science And Applications, 6*(1). *https://doi.org/10.14569/ijacsa.2015.060122*

Alshareef, N., Yuan, X., Roy, K., & Atay, M. (2021). A Study of Gender Bias in Face Presentation Attack and Its Mitigation. *Future Internet, 13*(9), 234. *https://doi.org/10.3390/fi13090234*

Alwawi, B., & Althabhawee, A. (2022). Towards more accurate and efficient human iris recognition model using deep learning technology. *TELKOMNIKA (Telecommunication Computing Electronics And Control), 20*(4), 817. *https://doi.org/10.12928/telkomnika.v20i4.23759*

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. ProPublica. Retrieved 12 September 2022, from *https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing*

Bacchini, F., & Lorusso, L. (2019). Race, again: How face recognition technology reinforces racial discrimination. *Journal of Information, Communication and Ethics in Society, 17*(3), 321–335. *https://doi.org/10.1108/jices-05-2018-0050*

Bellamy, B., Mason, J., & Ellis, M. (1999). Photograph Signatures for the Protection of Identification Documents. *Cryptography And Coding*, 119-128. *https://doi.org/10.1007/3-540-46665-7_13*

Bromberg, D., Charbonneau, É., & Smith, A. (2020). Public support for facial recognition via police body-worn cameras: Findings from a list experiment. Government Information *Quarterly, 37*(1), 101415. *https://doi.org/10.1016/j.giq.2019.101415*

Buhrmester, V., Münch, D., & Arens, M. (2021). Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. *Machine Learning And Knowledge Extraction, 3*(4), 966-989. *https://doi.org/10.3390/make3040048*

Buolamwini, J,; Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proc. Mach. Learn. Conf. Fairness Account. Trans. 81*, 1–15.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*(1). *https://doi.org/10.1126/sciadv.aao5580*

Drozdowski, P., Rathgeb, 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), 89-103. *https://doi.org/10.1109/tts.2020.2992344*

Engler, A. (2022). The EEOC wants to make AI hiring fairer for people with disabilities. Brookings. Retrieved 12 October 2022, from *https://www.brookings.edu/blog/techtank/2022/05/26/the-eeoc-wants-to-make-ai-hiring-fairer-for-people-with-disabilities/*

Galterio, M., Shavit, S., & Hayajneh, T. (2018). A Review of Facial Biometrics Security for Smart Devices. *Computers, 7*(3), 37. *https://doi.org/10.3390/computers7030037*

Gong, S., Liu, X., & Jain, A. K. (2020). Jointly de-biasing face recognition and demographic attribute estimation. In European conference on computer vision, 330-347. Springer, Cham.

Goode, A. (2014). Bring your own finger – how mobile is bringing biometrics to consumers. *Biometric Technology Today, 2014*(5), 5-9. *https://doi.org/10.1016/s0969-4765(14)70088-8*

Jain, A., & Kumar, A. (2012). Biometric Recognition: An Overview. The International Library Of Ethics, *Law And Technology, 11*, 49-79. *https://doi.org/10.1007/978-94-007-3892-8_3*

Jasserand, C. (2015). Avoiding terminological confusion between the notions of 'biometrics' and 'biometric data': an investigation into the meanings of the terms from a European data protection and a scientific perspective. *International Data Privacy Law, 6*(1), 63-76. *https://doi.org/10.1093/idpl/ipv020*

Kahn, J. (2021). Why HireVue will no longer assess job seekers' facial expressions. Fortune. Retrieved 12 September 2022, from *https://fortune.com/2021/01/19/hirevue-drops-facial-monitoring-amid-a-i-algorithm-audit/*

Khiyari, H., & Wechsler, H. (2016). Face Verification Subject to Varying (Age, Ethnicity, and Gender) Demographics Using Deep Learning. J*ournal Of Biometrics & Biostatistics, 07*(04). *https://doi.org/10.4172/2155-6180.1000323*

Kloppenburg, S., & van der Ploeg, I. (2018). Securing Identities: Biometric Technologies and the Enactment of Human Bodily Differences. *Science As Culture, 29*(1), 57-76. *https://doi.org/10.1080/09505431.2018.1519534*

Liang, J., Cao, Y., Zhang, C., Chang, S., Bai, K., & Xu, Z. (2019). Additive adversarial learning for unbiased authentication. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11428-11437).

Marsico, M., Nappi, M., Riccio, D., & Wechsler, H. (2016). Leveraging implicit demographic information for face recognition using a multi-expert system. *Multimedia Tools And Applications, 76*(22), 23383-23411. *https://doi.org/10.1007/s11042-016-4085-8*

Martinez-Martin, N., Greely, H. T., & Cho, M. K. (2021). Ethical development of digital phenotyping tools for mental health applications: Delphi study. *JMIR mHealth and uHealth, 9*(7), e27343.

Muhammad, J., Wang, Y., Wang, C., Zhang, K., & Sun, Z. (2021). CASIA-Face-Africa: A Large-Scale African Face Image Database. *IEEE Transactions On Information Forensics And Security, 16*, 3634-3646. *https://doi.org/10.1109/tifs.2021.3080496*

Ortega, A., Fierrez, J., Morales, A., Wang, Z., de la Cruz, M., Alonso, C., & Ribeiro, T. (2021). Symbolic AI for XAI: Evaluating LFIT Inductive Programming for Explaining Biases in Machine Learning. *Computers, 10*(11), 154. *https://doi.org/10.3390/computers10110154*

Page, M. J., McKenzie, J.E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. T*he British Medical Journal, 372, 71*. *https://doi.org/10.1136/bmj.n71*

Reynolds, C., Altmann, R., & Allen, D. (2021). The Problem of Bias in Psychological Assessment. Mastering Modern Psychological Testing, 573-613. *https://doi.org/10.1007/978-3-030-59455-8_15*

Robinson, J. P., Livitz, G., Henon, Y., Qin, C., Fu, Y., & Timoner, S. (2020). Face recognition: too bias, or not too bias?. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-1).

Röösli, E., Rice, B., & Hernandez-Boussard, T. (2020). Bias at warp speed: how AI may contribute to the disparities gap in the time of COVID-19. *Journal Of The American Medical Informatics Association, 28*(1), 190-192. *https://doi.org/10.1093/jamia/ocaa210*

Saini, B., Kaur, N., & Bhatia, K. (2018). Authenticating Mobile Phone User using Keystroke Dynamics. *International Journal Of Computer Sciences And Engineering, 6*(12), 372-377. *https://doi.org/10.26438/ijcse/v6i12.372377*

Savchenko, V., & Savchenko, A. (2021). Method for Measuring Distortions in Speech Signals during Transmission over a Communication Channel to

a Biometric Identification System. *Measurement Techniques, 63*(11), 917-925. *https://doi.org/10.1007/s11018-021-01864-x*

Serna, I., Pena, A., Morales, A., & Fierrez, J. (2021). InsideBias: Measuring bias in deep networks and application to face gender biometrics. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 3720-3727). IEEE.

Siddiqui, N., Dave, R., Vanamala, M., & Seliya, N. (2022). Machine and Deep Learning Applications to Mouse Dynamics for Continuous User Authentication. *Machine Learning And Knowledge Extraction, 4*(2), 502-518. *https://doi.org/10.3390/make4020023*

Singh, R., Agarwal, A., Singh, M., Nagpal, S., & Vatsa, M. (2020, April). On the robustness of face recognition algorithms against attacks and bias. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 09, pp. 13583-13589).

Srinivas, N., Hivner, M., Gay, K., Atwal, H., King, M., & Ricanek, K. (2019, January). Exploring automatic face recognition on match performance and gender bias for children. In 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW) (pp. 107-115). IEEE.

Tahir Sabir, A. (2021). Gait-based Biometric Identification System using Triangulated Skeletal Models (TSM). *Academic Journal Of Nawroz University, 10*(3), 202-208. *https://doi.org/10.25007/ajnu.v10n3a1223*

Terhörst, P., Huber, M., Kolf, J.N., Damer, N., Kirchbuchner, F., & Kuijper, A. (2019). Multi-algorithmic Fusion for Reliable Age and Gender Estimation from Face Images. 2019 22th International Conference on Information Fusion (FUSION), 1-8.

Terhörst, P., Kolf, J., Damer, N., Kirchbuchner, F., & Kuijper, A. (2020). Post-comparison mitigation of demographic bias in face recognition using fair score normalization. *Pattern Recognition Letters, 140*, 332-338. *https://doi.org/10.1016/j.patrec.2020.11.007*

Terhörst, P., Kolf, J. N., Damer, N., Kirchbuchner, F., & Kuijper, A. (2020) b. Face quality estimation and its correlation to demographic and non-demographic bias in face recognition. In 2020 IEEE International Joint Conference on Biometrics (IJCB) (pp. 1-11). IEEE.

Tricco, A., Lillie, E., Zarin, W., O'Brien, K., Colquhoun, H., & Levac, D. et al. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals Of Internal Medicine, 169*(7), 467-473. *https://doi.org/10.7326/m18-0850*

Vincent, N., & Hecht, B. (2021). Preview of "Data and its (dis)contents: A survey of dataset development and use in machine learning research". *Patterns, 2*(11), 100388. *https://doi.org/10.1016/j.patter.2021.100388*

Wang, M., & Deng, W. (2019). Mitigate bias in face recognition using skewness-aware reinforcement learning. *arXiv preprint arXiv:1911.10692*

Wiggers, K. (2021). Bias persists in face detection systems from Amazon, Microsoft, and Google. VentureBeat. Retrieved 12 September 2022, from *https://venturebeat.com/ai/bias-persists-in-face-detection-systems-from-amazon-microsoft-and-google/*

Williams, D. (2020). Fitting the description: historical and sociotechnical elements of facial recognition and anti-black surveillance. *Journal Of Responsible Innovation, 7*(sup1), 74-83. *https://doi.org/10.1080/23299460.2020.1831365*

Yin, X., Zhu, Y., & Hu, J. (2020). Contactless Fingerprint Recognition Based on Global Minutia Topology and Loose Genetic Algorithm. I*EEE Transactions On Information Forensics And Security, 15*, 28-41. h*ttps://doi.org/10.1109/tifs.2019.2918083*

Young-Powell, A. (2021). Ensuring biometrics work for everyone - Raconteur. Retrieved 12 September 2022, from *https://www.raconteur.net/hr/diversity-inclusion/ensuring-biometrics-work-for-everyone/*

For more information on CREST
and other CREST resources, visit
www.crestresearch.ac.uk

CREST

CENTRE FOR RESEARCH AND
EVIDENCE ON SECURITY THREATS

22-064-01