What does it mean to ‘solve’ the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems

Javier Sánchez-Monedero
Sanchez-monederoJ@cardiff.ac.uk
Cardiff University
Cardiff, Wales, United Kingdom

Lina Dencik
DencikL@cardiff.ac.uk
Cardiff University
Cardiff, Wales, United Kingdom

Lilian Edwards
lilian.edwards@newcastle.ac.uk
University of Newcastle
Newcastle upon Tyne, England
United Kingdom

ABSTRACT

Discriminatory practices in recruitment and hiring are an ongoing issue that is a concern not just for workplace relations, but also for wider understandings of economic justice and inequality. The ability to get and keep a job is a key aspect of participating in society and sustaining livelihoods. Yet the way decisions are made on who is eligible for jobs, and why, are rapidly changing with the advent and growth in uptake of automated hiring systems (AHSs) powered by data-driven tools. Evidence of the extent of this uptake around the globe is scarce, but a recent report estimated that 98% of Fortune 500 companies use Applicant Tracking Systems of some kind in their hiring process, a trend driven by perceived efficiency measures and cost-savings. Key concerns about such AHSs include the lack of transparency and potential limitation of access to jobs for specific profiles. In relation to the latter, however, several of these AHSs claim to detect and mitigate discriminatory practices against protected groups and promote diversity and inclusion at work. Yet whilst these tools have a growing user-base around the world, such claims of ‘bias mitigation’ are rarely scrutinised and evaluated, and when done so, have almost exclusively been from a US socio-legal perspective.

In this paper, we introduce a perspective outside the US by critically examining how three prominent automated hiring systems (AHSs) in regular use in the UK, HireVue, Pymetrics and Applied, understand and attempt to mitigate bias and discrimination. These systems have been chosen as they explicitly claim to address issues of discrimination in hiring and, unlike many of their competitors, provide some information about how their systems work that can inform an analysis. Using publicly available documents, we describe how their tools are designed, validated and audited for bias, highlighting assumptions and limitations, before situating these in the socio-legal context of the UK. The UK has a very different legal background to the US in terms not only of hiring and equality law, but also in terms of data protection (DP) law. We argue that this might be important for addressing concerns about transparency and could mean a challenge to building bias mitigation into AHSs definitively capable of meeting EU legal standards. This is significant as these AHSs, especially those developed in the US, may obscure rather than improve systemic discrimination in the workplace.

CCS CONCEPTS

• **Social and professional topics → Socio-technical systems**; *Systems analysis and design*; • **Applied computing → *Law***; *Sociology*.

KEYWORDS

Socio-technical systems, automated hiring, algorithmic decision-making, fairness, discrimination, GDPR, social justice

ACM Reference Format:
Javier Sánchez-Monedero, Lina Dencik, and Lilian Edwards. 2020. What does it mean to ‘solve’ the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In *Conference on Fairness, Accountability, and Transparency (FAT* ’20), January 27–30, 2020, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3351095.3372849

1 INTRODUCTION

The use of data systems and automated decision-making as a way to monitor, allocate, assess and manage labour is a growing feature of the contemporary workplace. Of increasing significance is the way Human Resources, and hiring practices in particular, are being transformed through various forms of automation and the use of data-driven technologies which we will collectively term Automated Hiring Systems (AHSs) [15, 41, 48]. Whilst there is a lack of data on the global uptake of such technologies, a recent report estimated that 98% of Fortune 500 companies use Applicant Tracking Systems of some kind in their hiring process [45]. The so-called ‘hiring funnel’ [15] consists of sourcing, screening, interviewing, and selection/rejection as a set of progressive filtering stages to identify and recruit the most suitable candidates. Each of these stages are undergoing forms of automation, as part of not only perceived efficiency measures and cost-savings that data-driven technologies afford, but also as a means to detect and mitigate discriminatory practices against protected groups and promoting diversity and inclusion at work [15]. Hiring platforms such as PeopleStrong or TribePad implement basic measures to mitigate human unconscious biases, such as anonymisation of candidates, while other platforms such as HireVue, Pymetrics and Applied[1] claim to specifically tackle

FAT ’20, January 27–30, 2020, Barcelona, Spain*

ACM ISBN 978-1-4503-6936-7/20/01.
https://doi.org/10.1145/3351095.3372849

[1] Note that Applied does not automate the evaluation of candidates but it assists in the process of bias discovery and mitigation.

the problem of discrimination in hiring. Yet the basis upon which such claims are made is rarely scrutinised and evaluated. While algorithmic bias generally and in employment law specifically [14] has had extensive investigation in the FAT* community, literature on bias mitigation in AHSs is at an early stage and so far primarily focused on the US context of employment both socially and legally [7, 15, 40]. We know much less about how this phenomena is developing in Europe [8]. This is the first scholarly attempt to consider the question of how satisfactorily bias and discrimination might be mitigated in AHSs within the UK context.[2]

We start by outlining how data-driven technologies are transforming hiring practices, before turning to focus on three AHSs widely in use in the UK which make claims to deal with bias and whose claims could be evaluated using publicly available materials such as company white papers and reports, patents, marketing resources, seminars and, in one case, source code. Access to further information relating to code, data sets, features design, trained models, or even the application user interface was not possible, and will often vary depending on client. Based on this publicly available material that outline their technological and procedural frameworks, we examine how these products implement bias discovery and mitigation. In doing so, and in building on recent work in this area [5, 15, 40, 48] we explore the assumptions made about the meaning of bias and discrimination in hiring practices embedded within these tools. The three prominent systems we examined were Pymetrics, HireVue and Applied. For each of these, we outline the design of the tool, focusing on how bias in hiring is addressed. The first two of these come from the US, and the third, Applied, from the UK, for comparative purposes. Other UK systems were considered but either made no claims as to debiasing, or did not make available sufficient public material to analyse. Although not rigorously addressed, this gap, either in implementation of debiasing, or its disclosure, is in itself, we feel, a significant finding deserving further research.

Next we briefly discuss the background to UK and EU equality law in the context of employment. Importantly, we raise the point that in the EU, all natural persons whose personal data is processed are given data rights, including rights to transparency and protection against power imbalance, which may be as useful in combating bias in AHSs as employment equality rights; these data protection (DP) rights, ensconced in the EU General Data Protection Regulation (GDPR), are not found in omnibus fashion in the US and thus have largely not been analysed in the US hiring algorithm literature.

In analysing these systems, we started from the premise that there is no singular or unified way of interpreting the meaning of discrimination, or how it might feature in hiring practices, nor is there consensus on any computational criteria for how "bias" should be defined, made explicit, or mitigated [22, 37, 40, 47, 54]. Ongoing analysis point out issues such as the unsuitability of maths to capture the full meaning of social concepts such as fairness, especially in a general sense, or the risk of technological solutionism [43]. Even when considering one statistical definition for bias such as the error rate balance amongst groups, the understanding and implication of that approach radically varies with the context and the consequential decisions that are driven by the algorithmic output [26]. Also, all sociotechnical systems, even when designed to mitigate biases, are designed with use cases in mind that may not hold in all scenarios [43]. Moreover, fairness can be procedural, as in the equal treatment of individuals, or substantive, as in the equal impact on groups [14], what is also referred to as opportunity-based vs. outcome-based notions of bias. These do not necessarily align and may actually be contradictory [33].

We conclude by arguing that, while common practices may be emerging to *mitigate* bias in AHSs built in the US, these are inevitably likely to reflect US legal and societal conceptions of bias and discrimination, and yet are exported wholesale as products to UK and other markets beyond the US. If this mismatch is not explicitly disclosed and analysed, there is a patent danger that inappropriate US-centric values and laws relating to bias in hiring (and more generally in society) may be exported to UK workplaces. Data protection rights, such as the alleged "right to an explanation" may by contrast not be implemented, potentially rendering useless the rights of candidates. What is more, while biased workplace hiring practices in traditional modes of hiring may at least to some extent be evident, and combatted in traditional ways by unions, strategic litigation and regulators, there are severe dangers that such biases buried within "black box" AHSs may not be manifestly obvious and so remain unchallenged. As AHSs become ever more popular, especially in sectors which are precarious, such as retail and the gig economy where prospective employees (or contractors) may have little economic power or knowledge of rights [7, 15], this is, we argue, a serious problem.

2 THE DATA-DRIVEN HIRING FUNNEL

The automation of hiring is an important part of the broader discussion on the future of work, subject to a range of ethical concerns and issues of fairness across the different stages of the hiring funnel, from sourcing to screening to interviewing to selection [15]. At each of these stages, the increasing use of data to automate or partially automate the process is significantly changing the way decisions are made on who is eligible for jobs, and why. Recommender engines based on hybrid collaborative filtering methods are used to capture user (both job seekers and recruiters) preferences; tools are used to filter candidates and identify the most promising candidates; automated skills tests and video interviews evaluate candidates; and analytics dashboards are used to select candidates and generate ad hoc job offers [48]. Data is collected from a range of sources and can be self-reported by the candidate in the form of unstructured documents such as resumés or as structured professional network profiles or online application forms. Often this information will be extended and/or scored through additional sources of information and assessment tests.

Central to the hiring process, only made more salient with automation, is the goal of hiring for 'fit' with an organisation, a criterion that leverages the employer significant discretion and may

[2] In this paper we are not looking at the general scientific validation of the system design as we do not have access to any independent studies of that kind. Instead, we are focusing on how 'bias' and discrimination is understood in the design. Our assessment does therefore not directly engage with the effectiveness of using AHSs to evaluate candidates, even though issues of discrimination would be very pertinent in such an assessment.

only be useful if an employer can make an accurate determination of such a fit [5]. The algorithmic specification of such a fit often relies on abstracting candidate profiles in relation to historical data of the company and/or top performers for specific roles. That is, at each stage of the hiring process, tools are designed around a set of variables, optimized for a particular criteria as to who might constitute an appropriate and 'good' employee. Deciding on such a criteria has long been imbued with the potential for discriminatory practices, and the aim here is not to suggest that those practices have only emerged with the introduction of data-driven tools nor is it to further the construct of a (false) binary between human vs. automated decision-making. Rather, in line with Ajunwa [5], we see value in addressing issues that such tools have made salient with the understanding that technologies are the product of human action and that, as Wajcman puts it, "histories of discrimination live on in digital platforms and become part of the logic of everyday algorithmic systems." [57].

At the same time, the turn to AHSs is significant also for the potential scale of impact, the difficulty with interrogating decision-making, increasing information asymmetry between labour and management, the standardisation of techniques, the obfuscation of accountability, and the veneer of objectivity that such technologies often afford employers, despite evidence of discrimination at each stage of the data-driven hiring funnel. For instance, when sourcing, specific groups have been found excluded from viewing job ads based on age [10] or gender [12]. Masculine gendered language in job descriptions can discourage women to apply for certain type of jobs [24] whilst language used by candidates has been demonstrated to be a proxy for social class that impact on the chances of being selected for an interview [50]. As such, screening of candidates based on collaborative filtering can perpetuate existing discriminatory practices [15]. Even when a recommendation engine is generally not discriminating against women, some job titles and rank results can place men in better positions [18]. Automated interviews, including the use of speech and facial recognition software, has been shown to perform poorly, particularly with regards to women of colour [15, 16]. Moreover, using 'cultural fit' as part of the selection criterion has been shown to lead to exclusive hiring practices, and is now outlawed by some companies as part of an effort to decrease unconscious bias [5]. These findings are part of a broader debate that engages with the way data-driven technologies systematically introduce and entrench forms of discrimination and social and economic inequality [9, 35, 36, 38, 47].

The role that AHSs come to play in entrenching and furthering the information asymmetry between labour and management inherent in the hiring process, the increased lack of accountability in decision-making, and the advancement of a global standard of management techniques changes the terms of control in the workplace; what Ajunwa [4] refers to as 'platform authoritarianism' in which employers gain penetrating new insights into current or potential employees, but the latter have no room to negotiate. What is more, the reliance on what are often candidate profiles based on incomplete data, proxies and inferences, may further this sense of authoritarianism on the basis of incomplete or inaccurate profiles. Similarly, Moore [34] argues that the on-going quantification of the workplace comes to discipline workers as they continuously seek to adapt to the needs of the technologies in place to assess them in a process of 'self-quantification'. Importantly, who might be best positioned to adapt to such measures and who is likely to be excluded rarely forms part of approaches to bias mitigation. Instead, the focus of mitigation tends to be on the technicalities of the model, at the point of the interface, situating the relationship between discrimination and inequality within the confines of 'unconscious bias'.

Of course, asking providers of AHSs to attend to the dynamics of power in labour relations and society more broadly might seem unnecessarily burdensome, but by not recognising the broader functions of automation in shaping those dynamics when considering forms of bias mitigation, we run the risk of neutralising challenges in a way that actively facilitates discrimination and inequality under a banner of fairness [20, 23, 28]. This is particularly important as these companies are part of standardising not only managerial techniques [7], but also how we should both understand and 'solve' the problem of discrimination in hiring, within a potentially global market.

3 TECHNOLOGICAL BIAS AUDITING AND MITIGATION IN HIRING

The data-driven hiring funnel therefore demands attention and scrutiny, particularly in relation to issues of discrimination and rights. This is especially pertinent as a number of AHSs specifically claim to address bias and discrimination in hiring, and seek to do so across organisational and national contexts. The claim is that automation reduces hirer bias, replacing messy human decisions with a neutral technical process [7]. In this section, we introduce three such software systems that specifically address the issue of bias in hiring. These AHSs were selected as they are known to be used in the UK, and, unlike many of their competitors, there is publicly available information to inform an evaluation of their approach. We base this evaluation on what documents are available through their websites and registered patents. The complete models or design of the tools, or outlines of specific data sources, have not been available to us for auditing and will vary depending on the client.

3.1 Pymetrics

Pymetrics is a vendor of hiring technology that performs a pre-employment assessment of candidates with games tests that are based on neuroscience research[3]. By analysing how participants behave with these games, the software generates metrics of cognitive, social and emotional traits. The profile is evaluated with a statistical model trained on the game results of top performers in each role so that the model can calculate a score and categorize the candidate as out-of-group and in-group [39]. This fit-to-role score is an aggregation of the scores of the individual tests. To create a description of optimal traits values for a professional role, the system uses an unsupervised learning clustering algorithm to identify representative scores of traits for the reference workers. For instance, according to Pymetrics [58], one of the desired characteristics of the best performing software developers are 'delayed gratification' and 'learning', that are two of the characteristics of a person that the software can measure (see Figure 1). A person

[3] https://www.pymetrics.com/

with a good score in these characteristics is more likely to fit in the position.

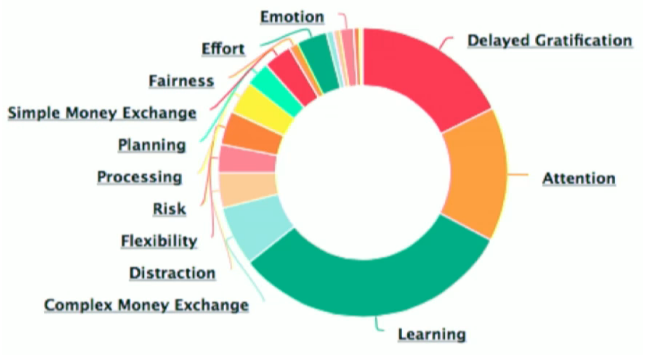


Figure 1: Pymetrics software engineer profile of traits. Screenshot from Pymetrics seminar recording available in [58].

Pymetrics features bias mitigation in candidate assessment based on the game design that is intended to avoid score correlation with protected groups and be agnostic with respect to non-verbal communication and culture [58]. To evaluate the fairness of each game score and of the fit-to-role score, Pymetrics performs statistical tests to compare both the individual and aggregated scores with respect to candidates grouped by age, gender and ethnicity. Since each game produces one to ten metrics, each one of those metrics is analysed to check for negative impact on protected groups scores.

The Pymetrics's US patent [39] reports detailed statistical test results of multiple group comparison. The worker and candidate data used for the test corresponds to several Pymetrics customers. The impact of age was analysed by grouping candidates in four age groups ($\leq$ 29; 30-34; 35-39; and $\geq$ 40), the impact on gender was evaluated considering binary gender classes (male; female) and the impact of race considered eight categories (Asian, Black, Hispanic, Middle Eastern, Native American, White, other, and mixed race). The multivariate ANOVA and Hotelling's T-squared tests concluded that for each game none of the tasks showed significant differences by ethnic group and a subset of the tasks showed different scores based on age and gender. When comparing the final score that assesses the suitability of a person for a position, the multivariate statistical analyses concluded that there were no differences, statistical bias, between samples groups by age ($p>0.05$), gender ($p>0.05$) and race ($p>>0.1$).

Pymetrics developed a specific tool, audit-AI, to perform this auditing and released it as open source software[4]. The software also performs US regulatory compliance checks to comply to fair standard treatment of all protected groups indicated in the Uniform Guidelines on Employee Selection Procedures by the Equal Employment Opportunity Commission (EEOC) [52]. The EEOC requires that the ratio of the proportion of pass rates, selection of candidates in this context, of the highest-passing and lowest-passing demographic categories has to follow the 4/5ths rule, meaning the ratio comparing the two extreme cases cannot be smaller than 0.80. For example, if there are 1,000 candidates who were hired and they belong to three groups, A, B and C, with passing frequency of 350, 320 and 330 respectively, the highest and lowest passing groups are A and B and, so the bias ratio is 320/350, or 0.91. Since this ratio is greater than 0.80 the selection procedure meets the legal requirements of the EEOC.

3.2 HireVue

HireVue[5] is a product to automate the pre-interview assessment of candidates from a pool. It performs automated video interview and games to profile candidates. The games and questions are designed based on Industrial Organization psychology research. The tool extracts three types of *indicator* data from applicants: categorical, audio and video [49]. HireVue promises to eliminate human biases in the assessment of candidates whilst simultaneously finding the subset of candidates that are most likely to be successful in a job by comparing them to employees already performing that job. Therefore, the software automates the screening of candidates by directly selecting them for a later human interview. In contrast to information provided by Pymetrics, HireVue provides more general information about what precise features are extracted from candidates and the specific statistical definition of bias.

Bias detection is performed by measuring demographic parity as defined by the US EEOC. To mitigate bias, HireVue has two strategies:

The first strategy consists of the removal of indicators that have an adverse impact on protected groups based on previous knowledge. As described in the website documentation [27] and patent [49], the abstract process first defines performance indicators and questions to elicit responses that can be measured and related to job performance. Indicators can include not only what the candidate says but also how they say it by extracting audio features such as pitch or duration. Then a model is trained to learn how to predict the suitability of the candidate from all these indicators. The bias mitigation consists of evaluating the adverse impact on protected groups by detecting violation of the 4/5ths rule. Features that cause biased results are removed and the models are re-trained and re-evaluated until bias is not detected. As an illustrative example, HireVue presents the case of speaking slowly as a characteristic of the top performers in a technical support role that also is more common in men[6]. Testing the tool should reveal this correlation so that the feature will be suppressed from the model input[7]. Alternatively, according to the patent [49], the bias discovering process can consist of applying clustering methods to detect protected groups in the feature space[8]. Clustering methods are unsupervised learning algorithms that try to structure unlabelled data points into different groups based on their arrangement in the input space. In this case, data points are composed of indicators excluding those ones related to group and performance. The proposal consists of running these methods to try to find groups based on the features used to evaluate candidates. If the method is capable of discovering protected groups

[4] https://github.com/pymetrics/audit-ai

[5] https://www.hirevue.com/

[6] https://www.hirevue.com/blog/hirevue-assessments-and-preventing-algorithmic-bias

[7] HireVue does not specify how this correlation can be discovered.

[8] Here we refer to as *feature space* to input data that consist of features extracted from candidates.

in this unsupervised manner, this means that one or more features, such as weight or hair colour, are correlated with the categorical variable so that learning algorithms could potentially use these features to learn to discriminate. If this is the case, the input data will be examined to identify and remove such correlated features.

The second strategy consists of modifying the learning algorithm to account for fairness. In machine learning, the objective function is a mathematical expression of how well the model is fitted to the data. It guides the learning algorithms in the process of learning from data and creating data transformations that contribute to improving accuracy. The patent [31] proposes to replace the objective function, typically a global sum of squared errors, with a corrected function that sums the separate error of the model for each protected group. By doing so, the objective function incorporates a fairness constraint that will indirectly introduce pressure on the learning algorithm to build a model that considers that the accuracy of the model with respect to all the protected groups (race, gender, age, etc.) must be equal. To account for the equal influence of underrepresented or minority classes, each group error term is normalized to ensure that the majority class does not influence the model more than the rest of the classes. The general expression for corrected error can be written as:

$$E_{\text{corrected}} = E_A + E_B + E_C + \cdots$$

where E_A, E_B, etc. are the errors for each protected group.

Additionally, the patent proposes to sum a penalty term to the corrected error to account for the regulations such as the EEOC. An example term of the 4/5ths rule can be represented as follows:

$$P(X) = \begin{cases} p_m & \text{if } f(X) \text{ violates 4/5ths rule} \\ 0 & \text{otherwise} \end{cases}$$

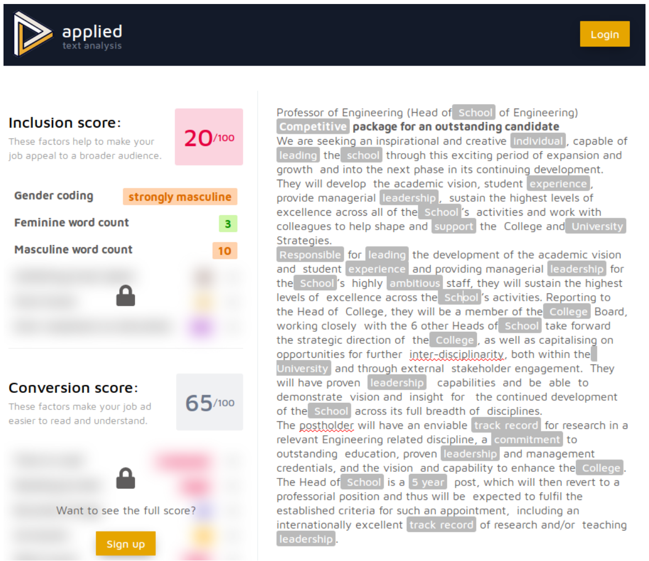
where p_m is the cost the user wants to associate with the rule violation and $f(X)$ refers to the candidate evaluation model whose output will be checked for demographic parity. Therefore, the objective error function becomes:

$$E_{\text{with\_ penalty}} = E_{\text{corrected}} + P(X)$$

3.3 Applied

Applied[9] is a hiring platform specialised in promoting diversity and inclusion in recruitment. The system includes a numerical, analytical and problem-solving testing platform called Mapped[10] that designs the tests by excluding patterns that are found to negatively impact on different demographic groups and improve pass rates of candidates of different groups[11].

In contrast to Pymetrics and HireVue, it performs bias discovery and mitigation by providing a de-biasing guide and demographic analytics reports for the hiring pipeline [11]. Note this tool does not perform automatic candidate assessmet but it semi-automates the task of discrimination monitoring. For example, regarding advertisement, the platform can analyse gendered language use and inclusiveness of position descriptions (see Figure 2). To ‘remove bias’ in the rest of the steps, the platform collects demographic information and then performs candidate anonymization and removes direct and indirect group information. Applied recommends to chunk assessment tests and compare results across candidates rather than performing full reviews of applications. In addition, it suggests to randomise the order of the chunks and get more than one person to score each chunk. Rather than performing formal statistical tests, the platform provides aggregated analytics to evaluate the whole process to visually detect biases at different stages (see Figure 3). Other comparison analytics are chunk scoring for each group or the degree of scoring agreement between multiple reviewers.



Inclusion score: 20/100
These factors help to make your job appeal to a broader audience.

Gender coding: strongly masculine
Feminine word count: 3
Masculine word count: 10

Conversion score: 65/100
These factors make your job ad easier to read and understand.

Want to see the full score?

Professor of Engineering (Head of School of Engineering)
Competitive package for an outstanding candidate
We are seeking an inspirational and creative individual, capable of leading the school through this exciting period of expansion and growth and into the next phase in its continuing development. They will develop the academic vision, student experience, provide managerial leadership, sustain the highest levels of excellence across all of the School’s activities and work with colleagues to help shape and support the College and University Strategies.
Responsible for leading the development of the academic vision and student experience and providing managerial leadership for the School’s highly ambitious staff, they will sustain the highest levels of excellence across the School’s activities. Reporting to the Head of College, they will be a member of the College Board, working closely with the 6 other Heads of School take forward the strategic direction of the College, as well as capitalising on opportunities for further inter-disciplinarity, both within the University and through external stakeholder engagement. They will have proven leadership capabilities and be able to demonstrate vision and insight for the continued development of the School across its full breadth of disciplines.
The postholder will have an enviable track record for research in a relevant Engineering related discipline, a commitment to outstanding education, proven leadership and management credentials, and the vision and capability to enhance the College. The Head of School is a 5 year post, which will then revert to a professorial position and thus will be expected to fulfil the established criteria for such an appointment, including an internationally excellent track record of research and/or teaching leadership.

Figure 2: Example of gendered language and inclusion analysis of Applied. Generated with trial version available at https://textanalysis.beapplied.com/ and an academic job ad of a British university.

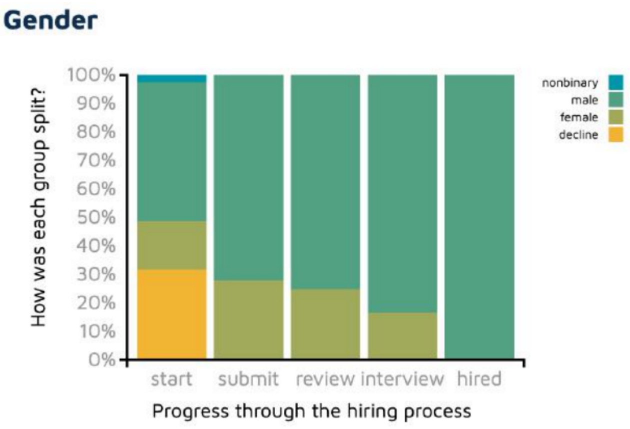


Figure 3: Visual auditing of bias implemented by Applied. Source [11].

[9] https://www.beapplied.com
[10] http://www.get-mapped.com/
[11] http://www.get-mapped.com/#about

4 EVALUATING CLAIMS OF BIAS MITIGATION

Having outlined the different claims and approaches that our three AHSs under study provide for mitigating bias in hiring we now draw on FAT-related debates to briefly sketch some of the limitations of such approaches.[12]

With different levels of formalization, the three AHSs understand fairness in the broader categories identified in recent literature [19]: (1) *anti-classification*, the omission of group variables and proxies in the decision making model, also known as debiasing; (2) *classification parity* in terms of equal passing rates; and (3) *calibration*, the requirement that outcomes are independent of the group variables. Notice that all the categories depend on clear definitions of groups. Examples of group definitions are binary gender, ethnicity (using definitions informed specifically by US demographics) and age interval. Social class or disabilities are not found in either the examples nor in the available validation tests. (2) and (3) depend on the comparison of an outcome that is ultimately the selection of candidates, i.e. passing rates, but also candidate scoring. HireVue and Pymetrics audit bias by checking the 4/5ths rule (classification parity). Additionally, Pymetrics uses statistical tests to compare group scores and passing rates which are a common way of comparing groups represented by samples.

In line with on-going discussions on criteria for fairness in data systems, notions of bias mitigation in AHSs present some inherent limitations [15, 40]. One fundamental limitation is that reference data is, by definition, extracted from current and past employees, and in many cases from the 'best performing' ones. Regardless of attention to 'fairness', the algorithmic specification of 'best performing' or its close association best 'fit', can itself become a vehicle for bias [7]. Within this setting, it is not clear that fair evaluation methods can be built on past data that reflect historical injustices [15, 40]. In other words, the issue with bias in AHSs may be in the very logic of prediction itself. Related to this, if the data used to build and validate the models only incorporates data from employees in the company and not the rejected candidates, this limits the use of fairness metrics that consider that discrimination can be better reflected when accounting for disparate mistreatment [59], for example by comparing false negative cases, i.e., candidates that would be suitable employees and were wrongly rejected. Validating fairness with respect to passing rates (disparate impact), as in true positives, has been discussed as problematic since it represents the degree of belief we have in the model prediction with respect to an individual or group rather than a measure of discrimination [26]. Hellman [26] proposes to compare error ratios instead as a better measure to compare different treatment of groups, which opens up a complex discussion about the validity of metrics to compare groups.

Discrimination with respect to different attributes is partially covered in the frameworks outlined by the three AHSs under study. Yet, as is a common feature of bias mitigation, they appear limited to single-axis understandings of identification that neglects substantial engagement with intersectional forms of discrimination [28]. Pymetrics validates fairness of each assessment test by performing multiple group comparisons across each attribute, e.g. comparing its determined racial groups scorings. Multiple axes of group identification are not directly compared, e.g. hispanic women with white men. HireVue, meanwhile, proposes in its patent to create a model that mitigates discrimination by modifying a standard objective function to equally account for the prediction errors of all the groups and also to penalize solutions that violate the classification parity constraint. However, this proposal is abstract and with no validation reports making it difficult to analyse. Moreover, the idea of adding a set of penalty terms to the objective function does not necessarily generate useful models. The problem arises in the fact that adding many terms to these functions will decrease the influence of each term, causing convergence problems in the learning algorithm. Indeed, the study of intersectionality and rich subgroups fairness in ML remains limited and in an early stage because of the problem of multiple objectives for optimization and the complexity of expressing the concept of intersectionality in a mathematical way [13, 16, 29, 59]. In this, understandings of bias mitigation in AHSs tend to overlook established limits of dominant antidiscrimination discourses that have also featured in legal debates [28].

These computational limitations of bias mitigation in AHSs point to several directions that such efforts might take. However, none of these discussions take account of the different national contexts in which these systems are being deployed and how different legal frameworks might apply, despite the prominence of legal definitions in the outline of the system design. In order to illustrate the significance of this, we now turn to discuss the relevant legal framework for the deployment of such systems in a UK context where we know they are used, and how the approaches to bias mitigation they propose relate.

5 LEGAL FRAMEWORK

Raghavan et al's [40] comprehensive study of automated hiring systems (AHSs) makes it clear that the majority of such systems on the market are developed in the US. Exceptions include Applied and Thrivemap (both UK), Teamscope (Estonia) and ActiView (Israel). No rigorous evidence seems to be available as to the market share of systems in actual use across the globe, yet in terms of bias mitigation, where the system is developed is crucial. While hiring goals and values may be universal (itself questionable), legal regulation of data-driven automated systems is decidedly not. As outlined above, considerations of bias and its mitigation in US-built systems – such as Pymetrics and HireView - have clearly aimed at meeting the constraints of US equality law [14], notably the 4/5ths rule set by the Uniform Guidelines on Employee Selection Procedures by the Equal Employment Opportunity Commission (EEOC) [52]. Yet UK law (itself heavily permeated by EU law) on equality and discrimination in employment matters is considerably different; most obviously in the UK there is no 4/5ths rule and hard statistical goals to (dis)prove bias do not seem formally to exist, in either statute or case law. It is noticeable that the two US-originated systems make use of the 4/5ths rule while the UK system discussed (Applied) does not. Interestingly, also, another UK system used widely, Thrive Map makes no claims as to bias mitigation at all and is therefore not included in this paper.

[12] This is not intended as an exhaustive list, but merely to highlight key concerns and debates.

Furthermore, EU DP law, now codified and reformed in the GDPR, provides a suite of rights to “data subjects” (natural persons whose personal data is processed) which are unknown to Federal US law (although sectoral data rights do exist in the US in health, finance and some other domains, and significant *state* laws have been enacted eg the Californian Consumer Privacy Act 2018, which replicates many features of EU DP law). DP rights may prove in the UK/EU context to be as or more important in uncovering or combating bias than sectoral employment or equality rights; yet US systems may (unsurprisingly) not be optimised to meet these rights.

Turning first to UK discrimination law, similarly to US law, it is based initially on the idea of a closed list of “protected characteristics”, here laid out in the Equality Act 2010, s 4. UK law also however (currently - Brexit may change this) has to respect supranational EU law, which includes a number of Directives relevant to equality and bias, notably the Equal Opportunities and Equal Treatment Directive 2006/54/EC, the Gender Equality Directive 2004/113/EC and the Framework Directive for Equal Treatment in Employment and Occupation 2000/78/EC. Human rights standards under the European Convention of Human Rights (ECHR) are also relevant and refers to broader protection against discrimination on *any* ground (art 14) in relation to the rights and freedoms guaranteed by the ECHR. The EU itself now has as a binding source of law its own human rights instrument, the Charter of Fundamental Rights of the EU [3]. This complex legislative patchwork, divided across common law and civilian approaches to law-making and interpretation [30], possibly contributes to fewer “hard and fast” standards for measuring unlawful bias in UK employment law than in the US, which itself arguably makes the task of proving debiasing harder for builders of AHSs for the UK market and may have contributed to abandoning the effort altogether for some companies.

The UK’s Equality Act 2010 attempts to replace a number of piecemeal prior laws relating to different types of inequality with a coherent statute covering inter alia sex, race and disability discrimination. Rather as with US law, discrimination can be direct or indirect. Direct discrimination in employment is nowadays regarded as rare [42]. Indirect discrimination is defined in s 19(1) as “a provision, criterion or practice which is discriminatory in relation to a protected characteristic”. Effectively it occurs when a policy that applies in the same way to everybody has an effect that particularly disadvantages people with a protected characteristic. This is similar to the US idea of disparate impact. However the Code of Practice which accompanies the Equality Act [1] does not lay down a statistical rule of thumb akin to the 4/5ths rule to prove bias. Instead, reflecting s 19(2), the Code provides only that a comparison must be made between workers with the protected characteristic and those without it. The circumstances of the two groups must be sufficiently similar for a comparison to be made and there must be no material differences in circumstances (para 4.15). This “pool for comparison” consists of the group which the provision, criterion or practice affects (or would affect) either positively or negatively, while excluding workers who are not affected by it, either positively or negatively. Importantly, the guidance does not always require a formal comparative exercise using statistical evidence. Such an approach was however adopted by the Court of Justice of the European Union (CJEU) to prove indirect sex discrimination in *R v Secretary for State for Employment es parte Seymour-Smith* [2000] IRLR 263 and is endorsed in some cases by the Equality Act Code for the UK as below (para 4.21):

> “• What proportion of the pool has the particular protected characteristic?
> • Within the pool, does the provision, criterion or practice affect workers
> without the protected characteristic?
> • How many of these workers are (or would be) disadvantaged by it?
> How is this expressed as a proportion (‘x’)?
> • Within the pool, how does the provision, criterion or practice affect
> people who share the protected characteristic?
> • How many of these workers are (or would be) put at a disadvantage by it?
> How is this expressed as a proportion (‘y’)?
> Using this approach, the Employment Tribunal will then compare (x) with (y).”

However there is no particular prescribed ratio of outcomes that proves bias. “Whether a difference is significant will depend on the context, such as the size of the pool and the numbers behind the proportions. It is not necessary to show that the majority of those within the pool who share the protected characteristic are placed at a “disadvantage” (para 4.22). Furthermore according to s 19(2), bias can be justified if it is shown to be a “proportionate” means of achieving a “legitimate” aim. Legitimacy is not even defined in the 2010 Act though guidance can be drawn from the CJEU and the Code of Practice, which states that the aim of the discriminatory provision, criterion or practice “should be legal, should not be discriminatory in itself, and must represent a real, objective consideration”. It adds: “Although reasonable business needs and economic efficiency may be legitimate aims, an employer solely aiming to reduce costs cannot expect to satisfy the test” (para 4.29). Given this kind of language, it is hard to imagine how either proportionality or legitimacy could be coded into a hiring tool. In legal discourse, proportionality is what is known as an “open textured” concept: it is impossible to predict what factors will come relevantly into play in advance, lacking a sufficiently large dataset of case law, nor how factors would be ranked. Even if sufficient data was available to mine using ML techniques, we would argue that it would not contain the individualised policy factors which drive courts or tribunals to make decisions around proportionality and legitimacy. by contrast a hard edged heuristic like the 4/5ths rule is simplistic to implement.

Turning to DP law, all data subjects have a number of rights in relation to the processing of personal data (itself defined in GDPR, art 4(1)), including rights to transparency and access to data held about them (“subject access rights” or SARs)(GDPR, arts 12-15) and to object to decisions which have legal or similarly significant effects being made about them using their personal data and by solely automated means (GDPR, art 22). The latter provision caused great academic stir in 2016 when it was claimed somewhat controversially it could be interpreted to provide data subjects, not just with a right to a “human in the loop” as had been known to be the case (albeit with little publicity) since 1995, but also with a “right to an explanation” of how the decision was made [21, 25, 55]. Given the importance of access to employment, automated hiring systems almost certainly make a decision which has legal or significant effect. Indeed, the Information Commissioner's Office (ICO), which regulates DP in the UK, gives "e-recruiting practices without human intervention" as a canonical example for the application of art 22[13].

Thus it could be argued that any use of a fully automated AHS, regardless of whether bias can be proved or is, indeed, said to have been mitigated, can be refused by the prospective employee on the ground that it is a solely automated decision under art 22, requiring explicit consent (art 22(2)(c)) and instead, the candidate could ask for a human to make that decision instead, or reconsider it. This rule is augmented further by the fact that it may be impossible to give valid explicit consent in the context of employment anyhow. Consent under the GDPR art 4(11) must be "freely given" and the Art 29 Working Party (A29 WP - now replaced by the European Data Protection Board or EDPB) who provide persuasive guidance on the GDPR, have indicated that truly free, and therefore valid, consent can probably never be given in the context of employment relations [2]. Of course it could be argued that a hiring (though not a firing or promotion/demotion) system *precedes* employment, and consent can therefore be valid; it seems unlikely the CJEU would take kindly to this, especially in times of austerity and precarity. Thus by chain of deduction it seems plausible that it is in fact illegal to use a solely automated AHS in the EU.

Issues arise with this analysis however. First, anecdotally, very few hiring (or even firing, promotion, demotion, allocation of hours, etc.) decisions seem to be taken without any human intervention at all. To take a recent US example, Amazon came under fire in April 2019 for apparently automatedly sacking up to 10% a year of their employees whose productivity fell below certain measured efficiency levels in environments regarded as highly datafied [32]. However later evidence emerged that no "automatic" sackings in fact occurred and a human supervisor was always there to reverse the sacking. Thus art 22 would arguably not have applied. Uber's global driver terms and conditions state that automated firings can take place but then add that in the EU a right to object to a human is available [51]. This suggests however that mere rubber stamping by a manager with no real intervention into decision making might be sufficient to render art 22 nugatory. What constitutes "enough" interaction by a human such that art 22 is not triggered remains an unclear issue in the GDPR, and may vary from member state to member state (see [2, 53]). Secondly, art 22 does not require consent from the data subject if the decision is "necessary for entering into or performance of a contract between an organisation and the individual" (art 22 (1)(a))[14]. Could submission to an AHS ever be "necessary" for entering the contract of employment? It seems prima facie unlikely but an employer might argue eg triage when many 1000s of applications are received does not just benefit from but actually *requires* solely automated systems.

If the right to object under art 22 is excluded by the lack of a "solely automated" decision, then any "right to an explanation" read from art 22 may also fall. However it is possible, though also controversial, that such a right may then be derived from art 15(h) which provides that users have a right to information about "the existence of automated decision-making, including profiling" (so the use of an AHS in hiring has to be notified to candidates) and "at least in those cases, meaningful information about the logic involved" [53]. It can be argued that the use of the phrase "at least" means that semi-automated decision making may not exclude the right to "meaningful information" [53]. Again such a right need not be chained to proof of bias, or failure to mitigate bias, and yet could prove highly effective in exposing discriminatory or even simply arbitrary or erroneous practices, at an individual and possibly even at a group level, given the possibility for collective redress actions within the GDPR (see arts 80 and 82). Thus at this point an easier route to disincentivising bias in AHSs might, in the EU context, be seen as coming via DP rather than equality rights - especially given the probable difficulty of building in bias mitigation into AHSs definitively capable of meeting EU legal standards. It is also worth noting here that any machine learning system is likely to be regarded as "high risk" processing requiring a prior Data Protection Impact Assessment (DPIA) (see art 35(3)(a), which should show inter alia that potential for unfairness and bias had been considered and steps taken to avert. This might arguably be seen as implying a *requirement* for debiasing in AHS tools deployed in EU (see also [5] for a similar point in a US context).

However, even if a right to algorithmic transparency does exist in the solely-automated hiring context, what does it practically *mean*? This has again been the subject of much academic debate. Selbst and Powles argue, for example, that for a right to "meaningful information about the logic involved" to be (sic) meaningful, it must be more than a simple regurgitation of source code [44]. The A29 Working Party recommend that the data subject should be provided with "general information (notably, on factors taken into account for the decision-making process, and on their respective 'weight' on an aggregate level) which is also useful for him or her to challenge the decision" [17]. To date there has been no relevant case law in the UK and the provision is not expanded on in the Data Protection Act (DPA) 2018 which implements the GDPR in the UK (in comparison to some other member states such as France).

6 DISCUSSION

The computational and legal challenges of AHSs that we have outlined here raise significant concerns for tackling discrimination and providing transparency to enable challenges to AHS hiring decisions in the context of the UK. This is particularly pertinent as bias mitigation in hiring is one of the key selling points for several of these tools. They are part of a growing 'diversity, equity, and inclusion' (D.E.I) industry that has boomed in the last couple of years [60]. Whilst a few of the companies developing AHSs provide some documentation to evidence such claims, access to relevant information remains a key problem for conducting any thorough analysis. Claims and validation are often vague and abstract, if they are provided at all. Moreover, it is not clear how relevant stakeholders, not least job seekers, are able to access and understand information about how decisions about their eligibility might have been reached through AHSs. This makes it difficult to assess if and

[13] See ICO guidance on the GDPR and DPA 2018 at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/,

[14] A second exemption relates to where the decision is authorised by member state or EU law. This refers to governmental "public tasks" and it seems unlikely it could ever apply to an AHS.

how discriminatory practices might have been part of the hiring process, and leaves little room for anyone to challenge the decision made. Given the transparency rights attributed to data subjects by the GDPR, this haziness as to transparency is unacceptable in the EU and UK. On the other hand, what approach to transparency is required by EU law, remains itself vague. It would be good to see AHSs built in, or sold into the EU market meeting this challenge of “meaningful information” explicitly. Might it take the form of the counterfactual explanations which have become fashionable but actually offer little by way of practical remedies? [56] What if vendors say that greater transparency is simply not possible? Again we may be back at a conclusion that such systems simply cannot be lawful in the EU.

The AHSs we have looked at in this paper are relatively unique in providing some information about their workings, even if the exact data sources and model remain obscure and will vary according to different clients. In particular, in seeking to explicitly tackle issues of discrimination in hiring practices, these systems provide some insights into how such issues are understood and approached by AHSs. This is significant, and welcomed, as it provides an opportunity to engage with what could be considered as emerging standards in managerial techniques that are being exported to a global marketplace.

Whilst a desire to address the prevalence of bias and discrimination in hiring is a significant pursuit (not least in the context of the UK where levels of discrimination against ethnic minorities in accessing jobs have remained relatively unchanged since the 1960s [46]), the approaches to bias mitigation provided by the three AHSs we have looked at come with important **limitations.** Here, we are not attempting to provide a comprehensive list of the issues that might come with the use of AHSs in general, but want to summarise a few of the key points that emerge from the evidence base we have provided in this paper. First, attempts at mitigation within AHSs run into on-going concerns with computational *fairness*. These relate not only to the inherent problems with data-driven predictions and with relying on quantification for determining criteria of ‘good’ or appropriate ‘fit’, but also to the necessary reductionist nature of group identification in computational systems and the neglect of intersectionality that have also been the subject of significant criticism in legal understandings of discrimination.

Secondly, attempts at bias mitigation in AHSs within a UK context also show problems with *accountability*. When such technologies are used for decision-making, to whom are the companies making AHSs responsible? The employer or the candidate? The employer might argue that some types of transparency are sometimes undesirable, as indeed might the software company defending its intellectual property. Thus, the question is whose job it is to fulfil the obligation of transparency alongside the obligation of bias mitigation; the system builder, the employer who utilises it, or another actor altogether.

Connectedly , the transfer of AHSs developed within a US socio-legal context to a UK (and arguably EU) context introduces a number of fundamental legal problems of fit, not just with regards to discrimination and equality law, but as we have argued, perhaps more significantly in relation to DP law. GDPR transparency rights in arts 15 and 22 may provide avenues to overturn aspects of the candidate-employer information asymmetry and might even outright prohibit the use of AHSs for wholly automated decision-making in hiring. Yet these rights may be ignored or ill-implemented in systems not built within the EU.

A number of other points might be addressed in **future work**. US literature on bias, especially racial bias, in the algorithmic workplace focuses on hiring because the US in general has “at will” firing with few legal constraints [6]. In Europe and in the UK specifically, things are very different and a quick survey of Employment Appeal tribunal cases shows most revolve around firing or issues of in-work conditions. There is a real need in extending FAT work on algorithmic bias in the workplace to Europe to consider these other loci for datafication, bias and opacity. Furthermore, more work is needed looking specifically at systems developed in the UK and the EU that also connects these to the actual practices and experiences of employers and candidates to get a sense of how AHSs shape those interactions. Such work requires more qualitative research that we seek to pursue in future project work.

In conclusion, the lack of information about how AHSs work, the approach they take to tackling discriminatory hiring practices, and crucially, where and how they are used around the world is therefore a significant problem. Given it is not even clear that AHSs provide significant benefits to employers [17], it could even be asked if their use should actually be restricted or discouraged by regulators from the EU data protection and equality sectors. As this trend is set to become more pervasive, there is an urgent need to assess if and how AHSs should be used so as to uphold fundamental rights and protect the interests of candidates and employees.

ACKNOWLEDGMENTS

The research of Lina Dencik and Javier Sánchez-Monedero has been funded by the ERC Starting Grant DATAJUSTICE (grant no. 759903).

REFERENCES

[1] 2011. *Equality Act 2010 Statutory Code of Practice Employment*. Equality and Human Rights Commission. https://www.equalityhumanrights.com/en/publication-download/employment-statutory-code-practice

[2] 2016. Article 29 Working Party Guidelines on consent under Regulation 2016/679 WP259 rev.01.

[3] 2016. Charter of Fundamental Rights of the European Union. [2016] OJ C202/1. , 389–405 pages. https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3AC%3A2016%3A202%3ATOC&uri=uriserv%3AOJ.C_.2016.202.01.0389.01.ENG

[4] Ifeoma Ajunwa. 2018. The Rise of Platform Authoritarianism. https://www.aclu.org/issues/privacy-technology/surveillance-technologies/rise-platform-authoritarianism

[5] Ifeoma Ajunwa. 2020. The Paradox of Automation as Anti-Bias Intervention. *41 Cardozo, L. Rev.* Forthcoming (2020). https://ssrn.com/abstract=2746078

[6] Ifeoma Ajunwa, Kate Crawford, and Jason Schultz. 2016. Limitless Worker Surveillance. *105 Calif. L. Rev. 735 (2017)* Rev. 735, 2017 (March 2016). https://doi.org/10.15779/Z38BR8MF94

[7] Ajunwa Ifeoma. 2019. Platforms at Work: Automated Hiring Platforms and Other New Intermediaries in the Organization of Work. In *Work and Labor in the Digital Age*, Greene Daniel, Steve P. Vallas, and Anne Kovalainen (Eds.). Research in the Sociology of Work, Vol. 33. Emerald Publishing Limited, 61–91. https://doi.org/10.1108/S0277-283320190000033005

[8] AlgorithmWatch. 2019. *Automating Society. Taking Stock of Automated Decision-Making in the EU.* Technical Report. AlgorithmWatch in cooperation with Bertelsmann Stiftung. https://www.algorithmwatch.org/automating-society

[9] Julia Angwin and Jeff Larson. 2016. Machine Bias. *ProPublica* (May 2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[10] Julia Angwin and Noam Scheiber. 2017. Dozens of Companies Are Using Facebook to Exclude Older Workers From Job Ads. *ProPublica* (Dec. 2017). https://www.propublica.org/article/facebook-ads-age-discrimination-targeting

[11] Applied. 2019. *Scaling fast: how to get hiring right.* Technical Report. https://www.beapplied.com/whitepaper-signup

[12] Jeremy B. Merrill Ariana Tobin. 2018. Facebook Is Letting Job Advertisers Target Only Men. *ProPublica* (Sept. 2018). https://www.propublica.org/article/facebook-is-letting-job-advertisers-target-only-men

[13] Ananth Balashankar, Alyssa Lees, Chris Welty, and Lakshminarayanan Subramanian. 2019. Pareto-Efficient Fairness for Skewed Subgroup Data. In *International Conference on Machine Learning AI for Social Good Workshop.* Long Beach, United States, 8.

[14] Barocas, Solon and Andrew D. Selbst. 2016. Big Data's Disparate Impact. *104 Calif. L. Rev. 671* (2016), 671–732. https://doi.org/10.15779/Z38BG31

[15] Miranda Bogen and Rieke Aaron. 2018. *Help Wanted: An Exploration of Hiring Algorithms, Equity and Bias.* Technical Report. Upturn. https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20--%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf

[16] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research)*, Sorelle A. Friedler and Christo Wilson (Eds.), Vol. 81. PMLR, New York, NY, USA, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html

[17] Peter Cappelli. 2019. Data Science Can't Fix Hiring (Yet). *Harvard Business Review* (May 2019). https://hbr.org/2019/05/recruiting

[18] Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. 2018. Investigating the Impact of Gender on Rank in Resume Search Engines. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18.* ACM Press, Montreal QC, Canada, 1–14. https://doi.org/10.1145/3173574.3174225

[19] Sam Corbett-Davies and Sharad Goel. 2018. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. (July 2018). https://arxiv.org/abs/1808.00023

[20] Lina Dencik, Fieke Jansen, and Philippa Metcalfe. 2018. A conceptual framework for approaching social justice in an age of datafication. https://datajusticeproject.net/2018/08/30/a-conceptual-framework-for-approaching-social-justice-in-an-age-of-datafication/

[21] Lilian Edwards and Michael Veale. 2017. Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. *16 Duke Law & Technology Review* (May 2017), 18–84. https://scholarship.law.duke.edu/dltr/vol16/iss1/2

[22] Sorelle A. Friedler, Carlos Scheidegger, Suresh Venkatasubramanian, Sonam Choudhary, Evan P. Hamilton, and Derek Roth. 2019. A Comparative Study of Fairness-enhancing Interventions in Machine Learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19).* ACM, New York, NY, USA, 329–338. https://doi.org/10.1145/3287560.3287589 event-place: Atlanta, GA, USA.

[23] Seeta Peña Gangadharan and Jędrzej Niklas. 2019. Decentering technology in discourse on discrimination. *Information, Communication & Society* 22, 7 (June 2019), 882–899. https://doi.org/10.1080/1369118X.2019.1593484

[24] Danielle Gaucher, Justin Friesen, and Aaron C. Kay. 2011. Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology* 101, 1 (July 2011), 109–128. https://doi.org/10.1037/a0022530

[25] Bryce Goodman and Seth Flaxman. 2017. European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine* 38, 3 (Oct. 2017), 50. https://doi.org/10.1609/aimag.v38i3.2741

[26] Deborah Hellman. 2019. *Measuring Algorithmic Fairness.* Technical Report. https://papers.ssrn.com/abstract=3418528

[27] HireVue. 2019. Bias, AI Ethics, and the HireVue Approach. https://www.hirevue.com/why-hirevue/ethical-ai

[28] Anna Lauren Hoffmann. 2019. Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society* 22, 7 (June 2019), 900–915. https://doi.org/10.1080/1369118X.2019.1573912

[29] Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. 2018. Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Jennifer Dy and Andreas Krause (Eds.), Vol. 80. PMLR, Stockholmsmässan, Stockholm Sweden, 2564–2572. http://proceedings.mlr.press/v80/kearns18a.html

[30] Jackie A. Lane and Rachel Ingleby. 2018. Indirect Discrimination, Justification and Proportionality: Are UK Claimants at a Disadvantage? *Industrial Law Journal* 47, 4 (Dec. 2018), 531–552. https://doi.org/10.1093/indlaw/dwx009

[31] Loren Larsen and Benjamin Taylor. 2017. Performance model adverse impact correction. https://patents.google.com/patent/US20170293858A1/en Patent No. US20170293858A1, Filed Sep. 27 , 2016, Issued Oct . 12 , 2017.

[32] Colin Lecher. 2019. How Amazon automatically tracks and fires warehouse workers for 'productivity'. *The Verge* (April 2019). https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations

[33] Zachary Lipton, Julian McAuley, and Alexandra Chouldechova. 2018. Does mitigating ML's impact disparity require treatment disparity? In *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.). Curran Associates, Inc., 8125–8135. http://papers.nips.cc/paper/8035-does-mitigating-mls-impact-disparity-require-treatment-disparity.pdf

[34] Phoebe Moore and Andrew Robinson. 2016. The quantified self: What counts in the neoliberal workplace. *New Media & Society* 18, 11 (Dec. 2016), 2774–2792. https://doi.org/10.1177/1461444815604328

[35] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York University Press.

[36] Cathy O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown Publishing Group, New York, NY, USA.

[37] Rebekah Overdorf, Bogdan Kulynych, Ero Balsa, Carmela Troncoso, and Seda Gürses. 2018. Questioning the assumptions behind fairness solutions. In *Critiquing and Correcting Trends in Machine Learning.* Montréal, Canada. http://arxiv.org/abs/1811.11293 arXiv: 1811.11293.

[38] Seeta Peña Gangadharan, Virginia Eubanks, and Solon Barocas (Eds.). 2014. *Data and Discrimination: Collected Essays.* Open Technology Institute, New America. https://newamerica.org/documents/945/data-and-discrimination.pdf

[39] Frida Polli and Julie Yoo. 2019. Systems and methods for data-driven identification of talent. https://patents.google.com/patent/US20190026681A1/en Patent No. US20190026681A1, Filed Jun. 20, 2018, Issued Jan. 24, 2019.

[40] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating Bias in Algorithmic Employment Screening: Evaluating Claims and Practices. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Vol. Accepted Papers. ACM. http://arxiv.org/abs/1906.09208 arXiv: 1906.09208.

[41] Adler-Bell Sam and Miller Michelle. 2018. *The Datafication of Employment.* Technical Report. The Century Foundation. https://tcf.org/content/report/datafication-employment-surveillance-capitalism-shaping-workers-futures-without-knowledge/

[42] Malcolm Sargeant. 2017. *Discrimination and the Law* (2nd edition ed.). Routledge, London.

[43] Andrew D. Selbst, Danah Boyd, Sorelle Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, Vol. Accepted Papers. ACM, Atlanta, GA, USA, 59–68. https://doi.org/10.1145/3287560.3287598

[44] Andrew D. Selbst and Julia Powles. 2017. Meaningful information and the right to explanation. *International Data Privacy Law* 7, 4 (Dec. 2017), 233–242. https://doi.org/10.1093/idpl/ipx022

[45] Jon Shields. 2018. Over 98% of Fortune 500 Companies Use Applicant Tracking Systems (ATS). https://www.jobscan.co/blog/fortune-500-use-applicant-tracking-systems/

[46] Haroon Siddique. 2019. Minority ethnic Britons face 'shocking' job discrimination. *The Guardian* (Jan. 2019). https://www.theguardian.com/world/2019/jan/17/minority-ethnic-britons-face-shocking-job-discrimination

[47] Javier Sánchez-Monedero and Lina Dencik. 2018. *How to (partially) evaluate automated decision systems.* Technical Report. Cardiff University. 15 pages. https://datajusticeproject.net/wp-content/uploads/sites/30/2018/12/WP-How-to-evaluate-automated-decision-systems.pdf

[48] Javier Sánchez-Monedero and Lina Dencik. 2019. *The datafication of the workplace.* Technical Report. Cardiff University. 48 pages. https://datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-The-datafication-of-the-workplace.pdf

[49] Benjamin Taylor and Loren Larsen. 2017. Model-driven evaluator bias detection. https://patents.google.com/patent/US9652745B2/en Patent No. US9652745B2, Filed Nov. 17, 2014, Issued May 16, 2017.

[50] Kyla Thomas. 2018. The Labor Market Value of Taste: An Experimental Study of Class Bias in U.S. Employment. *Sociological Science* 5 (Sept. 2018), 562–595. https://doi.org/10.15195/v5.a24

[51] Uber Technologies Inc. 2019. Uber Privacy. https://privacy.uber.com/policy/

[52] US EEOC. 1979. *Adoption of Questions and Answers To Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures.* Text VOL. 44, NO. 43. The U.S. Equal Employment Opportunity Commission. https://www.eeoc.gov/policy/docs/qanda_clarify_procedures.html

[53] Michael Veale and Lilian Edwards. 2018. Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. *Computer Law & Security Review* 34, 2 (April 2018), 398–404. https://doi.org/10.1016/j.clsr.2017.12.002

[54] Sahil Verma and Julia Rubin. 2018. Fairness Definitions Explained. In *Proceedings of the International Workshop on Software Fairness (FairWare '18).* ACM, New York, NY, USA, 1–7. https://doi.org/10.1145/3194770.3194776 event-place: Gothenburg, Sweden.

[55] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 2017. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law* 7, 2 (June 2017), 76–99.

https://doi.org/10.1093/idpl/ipx005

[56] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2018. Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology* 31, 2 (2018).

[57] Judy Wajcman. 2017. Automation: is it really different this time? *The British Journal of Sociology* 68, 1 (2017), 119–127. https://doi.org/10.1111/1468-4446.12239

[58] Julie Yoo. 2017. Pymetrics with Dr. Julie Yoo. https://www.youtube.com/watch?v=9fF1FDLyEmM

[59] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez-Rodriguez, and Krishna P. Gummadi. 2019. Fairness Constraints: A Flexible Approach for Fair Classification. *Journal of Machine Learning Research* 20, 75 (2019), 1–42. http://jmlr.org/papers/v20/18-262.html

[60] Nora Zelevansky. 2019. The Big Business of Unconscious Bias. *The New York Times* (Nov. 2019). https://www.nytimes.com/2019/11/20/style/diversity-consultants.html