

# Algorithmic Equity in the Hiring of Underrepresented IT Job Candidates

## Authors

Lynette Yarger, Penn State  
Fay Cobb Payton, NC State  
Bikalpa Neupane, Penn State

## Abstract

**Purpose:** This conceptual paper offers a critical analysis of talent acquisition software and its potential for fostering equity in the hiring process for underrepresented IT professionals. The under-representation of women, African-American, and Latinx professionals in the IT workforce is a longstanding issue that contributes to and is impacted by algorithmic bias.

**Design/methodology/approach:** Sources of algorithmic bias in talent acquisition software are presented. Feminist design thinking is presented as a theoretical lens for mitigating algorithmic bias.

**Findings:** Data is just one tool for recruiters to use; human expertise is still necessary. Even well-intentioned algorithms are not neutral and should be audited for morally and legally unacceptable decisions. Feminist design thinking provides a theoretical framework for considering equity in the hiring decisions made by talent acquisition systems and their users.

**Social Implications:** This research implies that algorithms may serve to codify deep-seated biases, making IT work environments just as homogeneous as they are currently. If bias exists in talent acquisition software, the potential for propagating inequity and harm is far more significant and widespread due to the homogeneity of the specialists creating AI systems.

**Originality/value:** This work uses equity as a central concept for considering algorithmic bias in talent acquisition. Feminist design thinking provides a framework for fostering a richer understanding of what fairness means and evaluating how AI software might impact marginalized populations.

## Classification

Conceptual paper

## Keywords

algorithmic bias, applicant tracking systems, diversity, equity, feminist theory, hiring, IT workforce, recruiting, talent acquisition

## 1. Introduction

Artificial Intelligence (AI) systems are increasingly used to support workforce development, education, criminal justice, financial lending, and other essential domains that shape the life chances of historically underserved populations. The algorithms that power AI systems are described as mirrors that reflect the unconscious biases that inform our research questions and our data (O'Neil, 2016). Sometimes, biases become embedded in the computer code when the algorithm is trained on data that does not represent a population well enough, or when the algorithm is irresponsibly designed to optimize a single type of decision (Edionwe, 2017). Bias also occurs when data scientists and software engineers fail to adequately consider the consequences of their design choices and the broader societal context in which

the results of the algorithmic decision making are going to be used (Edionwe, 2017). Thus, biases can become embedded into algorithms through design principles, feature selection, and training data.

As opaque mathematical models and data are applied increasingly to personal aspects of our lives, there is a concern that autonomous decision making may be reinforcing discrimination and widening inequality. For example, biased algorithms have been reported in convictions for crimes and parole decisions (Angwin, Larson, Mattu, and Kirchner, 2017), hiring decisions (Danieli, Hillis, and Lucas, 2016) and financial lending (Lohr, 2015). With the intentional hiring of underrepresented technology workers and the promotion of equitable workplace policies, a more diverse set of data and computer professionals would be included to think through the societal risks and implications of AI systems.

In this concept paper, we frame a discussion regarding equity, or the freedom from bias, in the talent acquisition software. Talent acquisition is the long-term planning approach for finding and acquiring skilled specialists to meet organizational labor requirements. Equity is giving all skilled specialists access to technology careers in inclusive work environments that enable them to be successful. According to data feminists (D'Ignazio and Klein, 2019), equity is an approach that locates the source of problems like bias in technical systems in historical and political systems that lead to unjust outcomes ("oppression"). Certain groups in society are historically disadvantaged in distinct and quantifiable ways. Racial and ethnic minorities in the US, for instance, are more likely to face socio-economic challenges, complete fewer years of schooling, and receive fewer employment opportunities. Algorithmic approaches, based on these traditional systems of oppression, may perpetuate or even magnify underlying disadvantage when equity is not considered (Kleinburg and Mullainathan, 2019).

The purpose of this conceptual paper is to discuss AI systems and their potential for fostering equity in the hiring of underrepresented IT professionals. In the following sections, we describe the context of under-representation in the IT workforce and how this intersects with equity and algorithmic bias. Next, we present a framework for ideation, design, and implementation of talent acquisition software with equity as a guiding principle. We conclude by using the framework to examine bias issues in AI talent acquisition software.

## 2. Underrepresentation in the IT Workforce

Women, African American, and Latinx computing professionals have long suffered discriminatory barriers that have contributed to their under-representation in the IT workforce. In the US, women held 24 percent of technical jobs, while African Americans and Latinxs made up 5% of the workforce (Atlassian, 2018). While leading technology companies have implemented incentives and targeted internal recruiting strategies designed to bring in female, African American, and Latinx software engineers, progress has been limited. For example, at Facebook, the number of women in IT roles grew from 16 percent to 17 percent, and its proportion of African American and Latinx workers stayed flat at 1 and 3 percent, respectively (Huet, 2017). At Apple, 54 percent of IT employees are White, 21 percent are Asian, 9 percent are African American, and 13 percent are Latinx. Women make up 23% of workers in IT roles at Apple (Rayome, 2018a). Lam (2015) notes that technology work remains one of the least diverse jobs in America.

The Tech Leavers study (Scott, Klein, and Onovakpuri, 2017) reports that diversity in technology remains stagnant because of employee turnover. Using a nationally representative sample of 2,006 U.S. adults who have left a job in a technology-related industry or function within the last three years, the study examines the reason why women and under-represented minorities leave technology careers. They found that a complex set of biases occur throughout the educational and employment pipeline, and these biases encourage people to leave. This unfair treatment of women and people of color creates a revolving door and costs the industry over 16 billion US dollars a year in employee replacement costs (Scott, Klein and Onovakpuri, 2017).

In job recruitment for technology workers, screening biases have been well documented in creating an unfair advantage for applicants based on specific physical appearance, gender, and ethnicity (Beattie and Johnson, 2012). Women and people of color get worse treatment

and ethnicity (Beattie and Johnson, 2012). Women and people of color get worse treatment than their white male peers in the technology job market (Wachter-Boettcher, 2017). In a meta-analysis of every available field experiment of hiring discrimination against African American or Latin's applicants (n = 28), Quillian and colleagues (2017) found a striking persistence of racial discrimination in US labor markets. The researchers analyzed callback rates for fictionalized matched candidates from different racial or ethnic groups applying for jobs. Since 1989, Whites receive, on average, 36% more callbacks than African Americans and 24% more callbacks than Latinx applicants. Moreover, they found no change in the levels of discrimination against African Americans since 1989, even after controlling for applicant education, gender, study method, occupational groups, and local labor market conditions. In 2016, with machine learning algorithms that screen resumes and perform personality tests, nearly 72% of resumes were weeded out before a human ever saw them (Mann and O'Neil, 2016). However, there is no well-established way to know whether the algorithms are systematically screening out African American and Latinx applicants (Mann and O'Neil, 2016).

One of the most troubling aspects of this longstanding pattern of under-representation is that it persists concurrent with increases in minority worker protections, anti-discrimination law, and interventions aimed at diversification of the field (Quillian et al., 2017). Under-representation has typically been investigated through demographic characteristics, such as race and gender, as well as structural factors, such as sexism and racism, perceptions of discrimination, and justice concerns (Dovidio and Gaertner, 1996). These approaches are now beginning to inform studies on algorithmic bias in the employment selection process (Kim, 2017; McCarthy et al., 2017).

For instance, in a study that used scenarios to measure how people from marginalized racial and socioeconomic groups perceived algorithmic fairness, Woodruff and colleagues (2018) report that almost half of the respondents viewed the hiring scenario as a moderate or severe problem. Moreover, Black respondents rated hiring scenarios at the highest severity because they elicited negative feelings that connect to national concerns about racial injustice and economic inequality.

Diversity in the IT workforce remains important as machine learning, predictive analytics, and other AI technologies increasingly influencing facets of our daily lives. "The people creating this technology have the power to influence how it works, and that's too big a responsibility for any single demographic to have full control. A lack of diverse ideas and representation could lead to further disparities between gender, race, and class" (Johansson, 2017). Not only diversity a moral imperative, it is a business imperative. Researchers report that teams of diverse people creating AI talent acquisition technologies can help companies to be more innovative and profitable (Winning, 2018). A study by Hunt, Layton and Prince (2015) found that, of the 366 public companies surveyed, those in the top quintile for ethnic and racial diversity were 35 percent more likely to have financial returns above their industry mean. The top quartile of companies for gender diversity are 15 percent more likely to have financial returns above their respective industry medians.

AI is one of the most critical places in technology to attack the diversity problem. Byrne (2018) asks: "If diverse people across racial, ethnic, gender, sexual identities, and socioeconomic backgrounds are absent from the IT workforce that is designing and developing AI systems, how well will the software foster inclusion and equity? How will these systems engage with and support these populations if their voices are not present to raise questions, illuminate blind spots, and check implicit assumptions?" To frame a discussion about algorithmic bias and equity in the IT workforce, we present a feminist design justice framework in the following section.

### 3. An Equity-based Framework for Talent Acquisition Software

Bardzell (2010) argues that feminist theories and methods can advance our understanding of cyber-human systems by critiquing core operational concepts, assumptions, and epistemologies. Feminist theories are useful for exploring the intersections of equity and fairness, bias and oppression, majority and minority, algorithmic and human decision-making because these theories pose critical questions about power relations. "Which groups are benefiting from AI?" "Which groups are harmed?" "How do we design a more equitable IT workforce informed by data and algorithms?" Moreover, "By whose values do we re-make the IT workforce?" Feminist approaches make apparent the ways that design configures users as gendered and raced social subjects—and what implications these configurations bear for the

life chances of historically marginalized groups (Bardzell, 2010; D'Ignazio and Klein, 2019; Stone et al., 2015). For feminist researchers, IT workforce diversity is fundamental to establishing equitable AI software. Design justice, a feminist framework devised by Costanza-Chock (2018), provides prompts that commit the architects of AI systems to engage with the design process in ways that support an ethic of equity.

- Equity: Who gets to do design?
- Beneficiaries: Who do we design for, or with?
- Values: What values do we encode and reproduce in the objects and systems that we design?
- Scope: How do we scope and frame design problems?
- Sites: Where do we do design, what design sites are privileged, and what sites are ignored or marginalized, and how do we make design sites accessible to those who will be most impacted?
- Ownership, accountability, and political economy: Who owns and profits from design outcomes, what social relationships are reproduced by design, and how do we move towards community control of design processes?
- Discourse: What stories do we tell about how things are designed?

In the following section, use the prompts from the design justice approach to interrogate approaches for addressing bias in AI hiring tools with diversity and equity as guiding principles, and how to identify discrimination as an outcome of poor design or unrepresentative data. First, we discuss AI talent acquisition software intentionally designed to mitigate known sources of human bias in the talent acquisition process. By design, these tools seek to encode equity into applicant sourcing and lead generation, resume review, interviewing, and skills assessment. Second, we consider more conventional means of auditing the results of algorithmic decision-making for bias. Here the focus is on dataset biases and statistical modeling based on biased cultural assumptions and associations.

#### 4. Design Justice at the Intersection of Algorithmic Bias and Equity

##### 4.1 Talent Acquisition Software with Equity as a Design Intention

A growing number of vendors are using AI to remove gender, ethnic, and racial biases from the hiring process. We present four companies (Blendoor, GasJumpers, Interviewing.io, and Textio) as examples of intentional design justice approaches that apply established human resource practices for mitigating bias in the hiring process. Rather than auditing systems for bias after the fact, these tools are conceived with a design justice intention of removing known sources of human bias from the hiring process.

###### 4.1.1. Blendoor (blendoor.com): Remove Identifying Information from the Resume

Resumes reveal information that indicates membership in a protected class based on gender, race, ethnicity, religion, sexual orientation, age, veteran status, and disability. An address is a possible proxy for race or income. Hobbies and interests could reveal religion, age, political affiliation, or whether a person has children. The name of a college could be linked to race or institutions that are rivals of the hiring manager. Names, however, are one of the most pervasive examples of how unconscious biases affect job applicants. Prior research on racial discrimination, for example, has reported significant discrimination against African-American names across industries and occupations (O'Connor, 2016; Bertrand and Mullainatha, 2016; Quillian, Pager, Hexel and Midtboen, 2017). Kang and colleagues (2016) examined how African American job seekers use "resume whitening" to adapt to anticipated discrimination by concealing or downplaying racial cues in job applications. These findings suggest that job applicants from underrepresented groups may engage in self-presentation strategies to improve their chances of being hired.

The design intention of Blendoor is that removing personally-identifying information on resumes helps to circumvent unconscious bias and improve equity. Many companies are turning to blind hiring, which enables employers to consider candidates based primarily on their skills. Blendoor matches diverse job seekers with employment opportunities. To mitigate unconscious bias, Blendoor captures applicants' profiles from existing online job boards and applicant tracking systems, and "blendorizes" the candidate profiles to remove names, photos,

and dates. In this way, Blendoor promotes design justice by assisting its beneficiaries (under-represented job seekers) and encoding the values of equity of opportunity in its algorithms.

#### 4.1.2. Gas-Jumper ([www.gapjumpers.me](http://www.gapjumpers.me)): Offer a Blind, Skills-Based Test

For software engineering jobs, potential employees are often asked to complete a coding challenge to test their skill level. Typically, a skills assessment is performed along with the interview process. However, in blind hiring, skills testing is done much sooner. A skill test can help companies get a feel for how the candidate thinks and if they possess the talents that the position requires. GasJumpers uses blind performance auditions to evaluate talent on their performance rather than keywords on a resume. Unlike job boards, GasJumpers offers a platform where companies can post challenges tailored to their work environment. Applicants are anonymized as they perform the challenge. GapJumpers technology grades the exams and creates a scorecard for each candidate. The client company then receives a ranking of applicants by their performance—without any personally identifying information. Hiring managers select which job seekers to bring in for interviews based squarely on performance metrics. This blind review process helps companies to diversify the workforce by screening applicants based on skills and avoids missing desirable candidates based on preconceived notions of fit (Bortz, 2018). The blind auditions reflect a design justice method of framing and scoping Hiring algorithms that may help women, older adults, people with disabilities, and applicants from diverse educational backgrounds to make the shortlist of top talent.

#### 4.1.3. Interviewing.io ([www.interviewing.io](http://www.interviewing.io)): Conduct Anonymous Technical Interviews

Interviewing.io is a hiring platform that provides university students with free opportunities to practice technical interviewing. During these interviews, the company collects audio transcripts, data and metadata describing the code written by the interviewee, and detailed feedback from both the interviewer and interviewee. Communication and coding skills demonstrated during the interview, rather than resumes, are used to identify top performers. Top performers then get to interview with companies on the platform. This alternative to campus career fairs provide students with pathways into tech that do not involve attending an elite school or extensive professional social networks. The company reports that 38 percent of their university candidates come from backgrounds that are under-represented in the IT workforce, and 82 percent are from schools that are not in the "Elite Tier." Interviewing.io adopts a design justice approach that broadens the scope and sites of participation, which extends the value of equity in sourcing and lead generation for employers.

#### 4.1.4. Textio ([www.textio.com](http://www.textio.com)): Write job advertisements that attract a diverse pool of applicants

Textio analyzes the hiring outcomes of more than 10 million job posts a month and predicts the competitive performance of clients' job listings. The software scans the job ad for key phrases that will statistically impact the real-world performance of job posts and calculates a Textio Score for each job ad. For example, a job advertisement with a Textio Score of 75 is predicted to perform better than 75 percent of comparable posts. To increase the diversity of the applicant pool, Textio includes a Tone Meter that determines whether the overall tone of the writing is likely to attract more men or more women. Gender-neutral language typically attracts the most diverse applicant pool. The platform also guides how to improve the inclusiveness of the language in the clients' job listings and raise the Textio Score. For example, according to Textio's analyst, "Rock star" attracts more male job seekers. The software suggests using a "high performer" instead. Textio, therefore, encodes the design justice principles of discourse and scope by helping companies to present jobs using more inclusive language that may attract a more diverse talent pool.

#### 4.1.6 Applicant Reaction to AI Hiring

Applicants' perceptions of fairness are important considerations from a design justice perspective that centers equitable employment outcomes for job applicants from under-represented populations as a fundamental design rule. Applicant reactions research could benefit under-represented job seekers by offering practical advice for system designers and administrators of talent acquisition systems (Ryan and Huth, 2008). Moreover, issues of ownership, accountability, and political economy that are raised by the feminist design justice prompts are essential for employers who may benefit from equitable hiring processes. For instance, adverse applicant perceptions affect negative outcomes, such as public complaints and litigation types of behaviors. However, if correctly implemented, diversity efforts could net

and mitigation types of behaviors. However, if correctly implemented, diversity efforts could net the IT industry an extra \$400 billion in revenue each year, according to CompTIA CEO Todd Thibodeaux (Rayome, 2018b).

Researchers have studied job applicants' reactions to technology-enabled employment selection procedures, and have built a robust body of theories and measurement tools (Chambers, 2003; Wiechmann and Ryan, 2003; Bauer et al., 2004; Truxillo, Steiner, and Gilliland, 2004; Jansen, Jansen and Spink, 2005; Strohmeier, 2007; Bauer et al., 2011; Thielsch, Träumer, and Pytlik, 2012; Konradt, Warszta, and Ellwart, 2013; Chowdhury, and Srimannarayana, 2014; Stone et al., 2015; Walker et al., 2015; Langer, König, and Fitili, 2018). The key findings of over 145 studies demonstrate that applicant reactions have significant and meaningful effects on attitudes, intentions, and behaviors (McCarthy et al., 2017). Moreover, applicant reactions are informative for identifying procedures in the talent acquisition process where perceptions of fairness may be formulated (Strohmeier, 2007).

However, the use of AI technology by recruiters, most notably those in large multinational organizations, has, in some cases, forged well ahead of the research base (Anderson, 2003). Highhouse (2008) posits that two implicit beliefs may inhibit the widespread acceptance of AI talent acquisition technologies – the belief that humans can achieve near-perfect precision in predicting performance on the job and the belief that human intuition and prediction can be improved by experience. However, the ways that people perceive decisions made by algorithms as compared to decisions made by humans is still not well understood (Lee, 2018). Decision agents can capture the potential tradeoffs between human and AI agents that afford various justice characteristics (e.g., the consistency of administration offered by an algorithmic agent versus missed opportunity to perform before a live person). Researchers (Dineen, Noe, and Wang, 2004) argue that these affordances may represent a hierarchy of justice characteristics, and that understanding this hierarchy may inform the design of HR practices and software that advance equity.

#### 4.1.5 Summary

These AI solutions are innovative in their use of design justice approaches to frame and scope inequities in talent acquisition, extend fairness to under-represented job seekers by blinding personally identifiable information, and evaluate applicants based on their skills and promote inclusive language in job advertisements. However, we do not yet have empirical evidence that measures how this talent acquisition software is helping or hindering efforts to diversify the technology field. We also have a limited understanding of how job seekers, especially those from under-represented groups, perceive this software. In a study of people's perceptions of fairness and trust in algorithmic decision-making, Lee (2018) reports that participants thought that human decisions were fairer than algorithmic decisions in the hiring and evaluation tasks. Participants believed that the algorithms would not have the ability to discern good candidates because they lack human judgment, make judgments based on keywords, or ignore qualities that are hard to quantify. Participants expected the human manager would be able to identify top candidates based on "skills and experiences" and "merit," and would have the qualifications and authority to do so. Participants further believed that the algorithm would lack human intuition.

#### 4.2 Software Audits and Representative Data to Alleviate Discrimination

Because the financial, legal, and reputations cost of discriminatory behavior is exceptionally high, Thryft (2019) argues that data scientists need additional tools to detect and correct underlying bias in algorithms or datasets that lead to discrimination. Current approaches for detecting algorithmic bias focus on auditing AI systems for intentional discrimination, statistical and classification bias, as well as data errors and absences that may perpetuate structural disadvantage (Williams, Brooks, and Schmargad, 2018; Sandvig et al., 2014; Kim, 2017). Corporations are also creating tools for evaluating AI models for bias and explainability such as Google's What-If Tool, IBM's AI Fairness 360 and FactSheets for AI, and Pymetrics' Audit-AI algorithm bias detection tool.

A design justice approach is informative when designing methods and tools for auditing systems. Chou and colleagues (2017) categorize sources of data and algorithmic biases that may result in discriminatory decision-making.

- Dataset bias: data used to train machine learning models does not represent the diversity of the customer base (e.g., Voice recognition technologies that only work well for male users because the initial training data excluded women.)
- Association bias: data used to train a model reinforces and multiplies a cultural bias (e.g., Language translation tools that make gender assumptions.)
- Automation bias: automated decisions override social and cultural considerations (e.g., Beautification photo filters reinforcing a European notion of beauty on facial images.)
- Interaction bias: humans interact with AI and create biased results (e.g., Humans deliberately input sexist language into a chatbot to train it to say offensive things.)
- Confirmation bias: overly simplified personalization makes biased assumptions for a group or an individual (e.g., Job advertisements for executive positions are displayed only to male users.)

The categories provided by Chou and colleagues (2017) parallel feminist design justice prompts of scoping and framing design problems and datasets, cultural and social sites that are ignored or marginalized, who gets to design systems, and social relationships that are reproduced by the design. In the following sections, we discuss each of these issues.

#### 4.2.1 Dataset Bias

In their paper on the future of work, Rosenblat, Kneese and Boyd (2014) note that "although most companies do not intentionally engage in discriminatory hiring practices, their reliance on automated systems, algorithms, and existing networks systematically benefits some at the expense of others, often without employers even recognizing the biases of such mechanisms." For data scientists, discrimination consists of conduct, practices, and laws that impose some disadvantage or harm on the persons based on their membership in a marginalized social group (d'Alessandro, O'Neil and LaGatta, 2017). Discrimination may occur when personal data available, such as social media profiles, is used in applicant screening algorithms even though there is no apparent reason why these data are relevant to employment (Rosenblat, Kneese and Boyd, 2014). From a design justice perspective, the use of personal data is harmful in exerting and extending power over job applicants.

Talent acquisition software enables companies to mesh traditional data sources of employment data like resumes and candidates' performance on computer programming tasks, along with audio, video, text, and social data posts to construct psychological profiles (Rosenblat, Kneese and Boyd, 2014). When algorithms use these profiles as a proxy for measuring organizational fit and predicting the ability of a candidate to perform the job, people from significantly different cultural backgrounds may be systematically disadvantaged. The algorithms serve to codify deep-seated biases and surface applicants who have specific attributes, making workplaces just as homogeneous as they were before (Miller, 2015). The exclusion of under-represented IT professionals is fundamentally an issue of equity and determines who gets to design.

#### 4.2.2 Automation and Interaction Bias

Another issue is that algorithms can obscure discrimination in ways that are unfair and unfamiliar (Barocas and Selbst, 2014). The AI software becomes more intelligent and exhibits more agency, but the predictive and decision-making processes used by algorithms are often opaque—it is difficult to explain why a particular decision was made (Dineen, Noe and Wang, 2004; Konrandt et al., 2016; Vasconcelos, Cardonha and Goncalves, 2017; McCarthy et al., 2017). Also, when algorithms use non-work related data to make inferences about applicants' age, race, religion, and sex, it makes it difficult to determine if firms are adhering to federal laws that protect job applicants against discrimination (Vasconcelos, Cardonha, and Goncalves, 2017; Valentino-DeVries, 2013). Legally, social media content is public data, but is it ethical to use it for employment decisions? The scrutiny of applicants' private lives on the Internet also raises the question of whether social media posts are a reliable and accurate indicator of personality (Rosenblat, Kneese and Boyd, 2014).

Once the system deployed in a social context, it is difficult to retroactively identify the source of the bias and develop a solution to remove it. Selbst and colleagues (2019) argue that by abstracting away the social context in which AI systems will be deployed, machine learning developers miss the information necessary to understand fairness and create equitable outcomes. However, some researchers have begun to uncover inequities that manifest in social contexts as a first step in the data generation process and before building machine learning models to correct for historical bias within the system (Suresh et al., 2019).

Social context is also abstracted away when diverse populations are excluded from the IT workforce. "There is a bias to what kinds of problems we think are important, what kinds of research we think are important, and where we think AI should go. If we don't have diversity in our set of researchers, we are not going to address problems that are faced by the majority of people in the world. When problems don't affect us, we don't think they're that important, and we might not even know what these problems are, because we're not interacting with the people who are experiencing them... The reason diversity is really important in AI, not just in data sets but also in researchers, is that you need people who just have this social sense of how things are. We are in a diversity crisis for AI" (Snow, 2018.)

#### 4.2.3 Confirmation and Association Bias

Historical data used to train machine learning models may reflect longstanding societal biases, which may unintentionally harm marginalized populations. If the historical data has an implicit bias that favors white men over Latinos, for example, then the measure of a strong candidate may have a strong correlation to race, ethnicity and gender, even if the algorithm-designer has no intention of replicating past hiring decisions that marginalize groups of people based on these categories (Barocas and Selbst, 2014). IBM researchers (Vasconcelos, Cardonha, and Goncalves, 2017) found that, in creating these algorithms, companies model historical patterns of hiring from data describing "high-performance" employees as a basis for selecting candidates with similar profiles. Consequently, biases rooted in the traditional hiring process are reproduced and encoded in the data used to train the new systems and can have drastic impacts on human lives (O'Neil, 2016; Kilpatrick, 2016, Giang 2018).

In October 2018, for example, Amazon scrapped an AI recruiting tool after realizing that this new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way (Dastin, 2018; Higginbottom, 2018; Miller, 2015). Because the computer models were trained to rate applicants by observing patterns in submitted resumes and most of those resumes came from men, the algorithm was systemically excluding female candidates. What is novel, in this case, is that the discriminatory effects are data-driven (Kim, 2017; Mann and O'Neil, 2016). Perhaps more importantly, the Amazon case serves as an example of how historically systems of exclusion may be perpetuated unwittingly by algorithms. To mitigate racial and gender bias, researchers have constructed preprocessing methods to maintain the accuracy of the data set. These methods include assigning more weight to underrepresented populations within the data set and duplicating data points in order to make up for underrepresentation (Edionwe, 2017).

#### 4.2.4 Summary

In the past, laws were used to "debias" the hiring process. Today, these controls are being embedded in software (Jolls and Sundtein, 2016). Many hoped that algorithms would help human decision-makers avoid their prejudices by adding consistency to the hiring process, but algorithms can reproduce historical biases (Bogen 2019). Hao (2019) provides several technological and human reasons for why discrimination by algorithms is so hard to control. The introduction of bias is not always apparent during a model's construction because computer scientists are not generally trained to contextualize social problems. Computer scientists may not realize the impacts of their choices about the model, the sources and types of training data, and the selection of attributes that the algorithms should consider. Design justice provides a useful set of questions that can help to remedy this oversight.

### 5. Conclusion

Companies are increasingly adopting AI tools to minimize human bias, reduce costs, and



streamline the recruitment process (Florentine, 2016). Big data and AI technologies support HR decision-making in a variety of talent acquisition tasks such as sorting through resumes, making predictive matches between job seekers and positions using data, correcting biases in the language used in job descriptions, and using bots to schedule candidate interviews (Danieli, Hillis, and Lucas, 2016). While these AI tools can help organizations navigate the vast pool of potential applicants efficiently, researchers caution that these algorithms are ultimately human decision-making processes embedded in code (Tiku, 2018). In other words, even well-intentioned algorithms are not neutral and should be audited for morally and legally unacceptable decision-making (Mann and O'Neil, 2016; Miller, 2015).

In this paper, we presented several areas where bias in the talent acquisition process intersect with socio-cultural notions of equity. Longstanding societal systems that perpetuate inequities need to be evaluated as metrics or benchmarks in the construction of these systems. We also need more diversity in the computer and data specialists who build AI systems. Feminist thought and methods can aid in providing an ethical and moral compass for designing and auditing AI systems.

## 7. References

- Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2017), "Machine bias", available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (accessed December 3, 2017).
- Atlassian (2018), "State of diversity and inclusion in US tech: Stats summary", available at: <https://www.atlassian.com/diversity/survey/2018> (accessed October 30, 2018).
- Bardzell, S. (2011), "Towards a Feminist HCI Methodology: Social Science, Feminism, and HCI", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, May 7-12, pp. 675-684.
- Bardzell, S. (2010), "Feminist HCI: Taking Stock and Outlining an Agenda for Design", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA, April 10-15, pp. 1301-1310.
- Bauer, T., Truxillo, D., Mack, K. and Costa, A. (2011), "Applicant reactions to technology-based selection: What we know so far", in Tippins, N., Adler, S. and Kraut, A. (Eds.), *Technology-Enhanced Assessment of Talent*, Josey-Bass, San Francisco, CA, pp. 120-223.
- Bauer, T. N., Truxillo, D. M., Paronto, M. E., Weekley, J. A. and Campion, M. A. (2004), "Applicant reactions to different selection technology: Face-to-Face, interactive voice response, and computer-assisted telephone screening interviews", *International Journal of Selection and Assessment*, Vol. 12, No. 1-2, pp. 135-148.
- Beattie, G. and Johnson, P. (2012), "Possible unconscious bias in recruitment and promotion and the need to promote equality", *Perspectives: Policy and Practice in Higher Education*, Vol. 16, No. 1, p. 7.
- Bertrand, M. and Mullainathan, S. (2004), "Are Emily And Greg more employable than Lakisha And Jamal? A field experiment on labor market discrimination", *American Economic Review*, Vol. 94, pp. 991-1013.
- Bogen, M. (2019). "All the ways hiring algorithms can introduce bias", *Harvard Business Review*, May 6, available at: <https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias> (accessed May 14, 2019).
- Bortz, D. (2018), "Can blind hiring improve workplace diversity?", *Society for Human Resource Management*, available at: <https://www.shrm.org/hr-today/news/hr-magazine/0418/pages/can-blind-hiring-improve-workplace-diversity.aspx> (accessed October 30, 2018).
- Byrne, W. (2018), "Now Is the time to act to end bias in AI", *Fast Company*, February 2, available at: <https://www.fastcompany.com/40536485/now-is-the-time-to-act-to-stop-bias-in-ai>

ai (accessed March 14, 2019).

Chambers, B. (2003), "Applicant reactions and their consequences: Review, advice, and recommendations for future research", *International Journal of Management Review*, Vol. 4, No. 4, pp. 317-333.

Chou, J., Murillo, O. and Ibars, R. (2017), "In pursuit of inclusive AI", Microsoft Research, available at: [https://msdesignstorage.blob.core.windows.net/microsoftdesign/inclusive/InclusiveDesign\\_InclusiveAI.pdf](https://msdesignstorage.blob.core.windows.net/microsoftdesign/inclusive/InclusiveDesign_InclusiveAI.pdf) (accessed October 20, 2019).

Chowdhury, T. and Srimannarayana, M. (2014), "Applicants' perceptions on online recruitment", *Management and Labour Studies*, Vol. 38, No. 3, pp.185-199.

Costanza-Chock, S. (2018, June 3). "Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice", *Proceedings of the Design Research Society*, available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3189696](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3189696) (accessed March 13, 2019).

d'Alessandro, B., O'Neil, C., and LaGatta, T. (2017), "Conscientious classification: A data scientist's guide to discrimination-aware classification", *Big Data*, Vol. 38, No. 2, pp. 120-134.

Danieli, O., Hillis, A. and Lucas, M. (2016), "How to hire with algorithms", *Harvard Business Review*, October 17, available at: <https://hbr.org/2016/10/how-to-hire-with-algorithms> (accessed December 3, 2017).

Dastin, J. (2018), "Amazon scraps secret AI recruiting tool that showed bias against women", *Reuters*, available at: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> (accessed October 20, 2018).

Dineen, B. R., Noe, R. A., and Wang, C. (2004), "Perceived fairness of web-based applicant screening procedures: Weighing the rules of justice and the role of individual differences", *Human Resource Management*, Vol. 43, No. 2-3, pp. 127-145.

D'Ignazio, C. and Klein, L. (2019), "Data feminism", MIT University Press, Cambridge, MA.

Dovidio, J. F. and Gaertner, S. L. (1996), "Affirmative Action, unintentional racial biases, and intergroup relations", *Journal of Social Issues*, Vol. 52, pp. 51-75.

Eder, S. (2018). "How can we eliminate bias in our algorithms?", *Forbes*, June 27, available at: <https://www.forbes.com/sites/theyec/2018/06/27/how-can-we-eliminate-bias-in-our-algorithms/#781c4d99337e> (accessed May 14, 2019).

Edionwe, T. (2017), "The fight against racist algorithms: Can we teach our machines to unlearn racism?", *The Outline*, May 22, available at: <https://theoutline.com/post/1571/the-fight-against-racist-algorithms?zd=1&zi=uahpmzdb> (accessed May 14, 2019).

Florentine, S. (2016), "How artificial intelligence can eliminate bias in hiring", *CIO Magazine*, December 22, available at: <https://www.cio.com/article/3152798/artificial-intelligence/how-artificial-intelligence-can-eliminate-bias-in-hiring.html> (accessed October 27, 2018).

Giang, C. (2018), "The Potential Hidden Bias In Automated Hiring Systems", *Fast Company*, May 8, available at: <https://www.fastcompany.com/40566971/the-potential-hidden-bias-in-automated-hiring-systems>

Hao, K. (2019), "This is how AI bias really happens—and why it's so hard to fix", *Technology Review*, February 9, available at: <https://www.technologyreview.com/s/612876/this-is-how-ai->

bias-really-happensand-why-its-so-hard-to-fix (accessed October 27, 2018).

Higginbottom, K. (2018), "The pros and cons of algorithms in recruitment", *Forbes*, October 10, available at: <https://www.forbes.com/sites/karenhigginbottom/2018/10/19/the-pros-and-cons-of-algorithms-in-recruitment/#35464e677340> (accessed October 27, 2018).

Highhouse, S. (2008), "Stubborn reliance on intuition and subjectivity in employee selection", *Industrial and Organizational Psychology*, Vol. 1, pp. 333–342.

Huet, E. (2017), "Facebook's hiring process hinders its effort to create a diverse workforce", *Bloomberg Technology*, January 9, available at: <https://www.bloomberg.com/news/articles/2017-01-09/facebook-s-hiring-process-hinders-its-effort-to-create-a-diverse-workforce> (accessed January 25, 2018).

Hunt, V., Layton, D. And Prince, S. (2015), "Why diversity matters", *Mc Kinsey and Company*, available at: <https://www.mckinsey.com/business-functions/organization/our-insights/why-diversity-matters> (accessed October 30, 2018).

Jansen, B. J., Jansen, K. J. and Spink, A. (2005), "Using the web to look for work: Implications for online job seeking and recruiting", *Internet Research*, Vol. 15, No. 1, available at: <http://dx.doi.org.ezaccess.libraries.psu.edu/10.1108/10662240510577068> (accessed October 27, 2018).

Jolls, C. and Sunstein, C.R. (2016), "The law of implicit bias", *California Law Review*, Vol. 969, available at: <http://scholarship.law.berkeley.edu/californialawreview/vol94/iss4/2> (accessed October 30, 2018).

Johansson, A. (2017), "Why millennials are demanding even more diversity in tech", *Forbes*, December 19, available at: <https://www.forbes.com/sites/annajohansson/2017/12/19/why-millennials-are-demanding-even-more-diversity-in-tech/#66bb36d4386b> (accessed October 30, 2018).

Kang, S., DeCelles, K., Tilcsik, A. and Jun, S. (2016), "Whitened Résumés: Race and Self-Presentation in the Labor Market", *Administrative Science Quarterly*, Vol. 61, No. 3, pp. 469–502.

Kilpatrick, K. (2016), "Battling algorithmic bias: How do we ensure algorithms treat us fairly?", *Communications of the ACM*, Vol. 59, No. 10, pp. 16-17.

Kim, P. (2017), "Data-driven discrimination at work", *William and Mary Law School Scholarship*, Rev. Vol. 857, No. 3-4, available at: <http://scholarship.law.wm.edu/wmlr/vol58/iss3/4> (accessed January 26, 2018).

Kleinberg, J. and S. Mullainathan. (2019), "Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability", *Proc. 20th ACM Conference on Economics and Computation*.

Konradt, U., Warszta, T. and Ellwart, T. (2013), "Fairness perceptions in Web-based selection: Impact on applicants' pursuit intentions, recommendation intentions, and intentions to reapply", *International Journal of Selection and Assessment*, Vol. 21, No. 2, pp. 155-169.

Konradt, U., Garbers, Y., Erdogan, B. and Bauer, T. (2016), "Patterns of change in fairness perceptions during the hiring process", *International Journal of Selection and Assessment*, Vol. 24, No. 3, pp. 246-259.

Lam, B. (2015), "The least diverse jobs in America", *The Atlantic*, June 29, available at: <https://www.theatlantic.com/business/archive/2015/06/diversity-jobs-professions-america/396632> (accessed October 27, 2018).

Langer, M., König, C. and Fitili, A. (2018), "Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection", *Computers in Human Behavior*, Vol. 18(April), pp. 19-30.

Lee, M.K. (2018), "Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management", *Big Data and Society*, Vol. 5, No. 1, pp. 1-16.

Lohr, S. (2015), "Banking start-ups adopt new tool for lending", *The New York Times*, January 18, available at: <https://www.nytimes.com/2015/01/19/technology/banking-start-ups-adopt-new-tools-for-lending.html> (accessed December 3, 2017).

Mann, G. and O'Neil, C. (2016), "Hiring algorithms are not neutral", *Harvard Business Review*, December 9, available at: <https://hbr.org/2016/12/hiring-algorithms-are-not-neutral> (accessed December 3, 2017).

McCarthy, J., Bauer, T., Truxillo, D., Anderson, N., Costa, A. and Ahmed, S. (2017), "Applicant perspectives during selection: A review addressing 'So what?', 'What's new?', and 'Where to next?'" , *Journal of Management*, Vol. 46, No. 3, pp. 1693 - 1725.

Miller, C. (2015), "Can an algorithm hire better than a human?", *The New York Times*, June 25, available at: <https://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html> (accessed December 3, 2017).

O'Connor, C. (2016), "Black woman engineer launches 'Blind' job match app to take bias out of tech hiring", *Forbes*, March 3, available at: <https://www.forbes.com/sites/clareoconnor/2016/03/03/black-woman-engineer-launches-blind-job-match-app-to-take-bias-out-of-tech-hiring/#228f757d2394> (accessed January 14, 2018).

O'Neil, C. (2016), "Weapons of math destruction: How big data increases inequality and threatens democracy", Crown Publishing Group, New York.

Quillian, L., Pager, D., Hexel, O. and Midtbøen, A. H. (2017), "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 114, No. 41, pp. 10870–10875.

Rayome, A. (2018a), "Five eye-opening statistics about minorities in tech", *TechRepublic*, February 7, available at: <https://www.techrepublic.com/article/5-eye-opening-statistics-about-minorities-in-tech> (accessed March 15, 2018).

Rayome, A. (2018b), "Why AI could be the tool your HR team needs to hire and retain the best talent", *Tech Republic*, January 2, available at: <https://www.techrepublic.com/article/why-ai-could-be-the-tool-your-hr-team-needs-to-hire-and-retain-the-best-talent> (accessed January 27, 2018).

Rosenblat, A., Kneese, T. and Boyd, D. (2014) , "Networked Employment Discrimination", *Data and Society Working Paper*, October 8, available at: <https://www.datasociety.net/pubs/fow/EmploymentDiscrimination.pdf> (accessed October 13, 2018).

Ryan, A. M. and Huth, M. (2008), "Not much more than platitudes? A critical look at the utility of applicant reactions research", *Human Resource Management Review*, Vol. 18, No. 3, pp.119-132.

Sandvig, C., Hamilton, K., Karahalios, K. and Langbort, C. (2014), "Auditing algorithms: Research methods for detecting discrimination on Internet platforms", Paper presented to *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, 64th Annual Meeting of the American Psychological Association, Washington, DC, 2014. <https://www.apa.org/pubs/journals/00000000>

of the International Communication Association. May 22, 2014, Seattle, WA.

Scott, A., Klein, F. and Onovakpuri, U. (2017), "Tech Leavers, Kapor Center for Social Impact", available at: <https://www.kaporcenter.org/wp-content/uploads/2017/08/TechLeavers2017.pdf> (accessed March 13, 2019).

Selbst, A., Boyd, D., Friedler, S., Venkatasubramanian, S. and Verweis, J. (2019). "Fairness and Abstraction in Sociotechnical Systems", Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 59-68.

Snow, J. (2018). "We're in a diversity crisis: cofounder of Black in AI on what's poisoning algorithms in our lives", MIT Technology Review, Feb 14, available at: <https://www.technologyreview.com/s/610192/were-in-a-diversity-crisis-black-in-ais-founder-on-whats-poisoning-the-algorithms-in-our> (accessed March 13, 2019).

Stone, D., Deadrick, D., Lukaszewski, K., and Johnson, R. (2015), "The influence of technology on the future of human resource management", Human Resource Management Review, Vol. 25, pp. 216-231.

Strohmeier, S. (2007), "Research in e-HRM: Review and implications", Human Resource Management Review, Vol. 17, No. 1, pp. 19-37.

Thielsch, M. T., Träumer, L. and Pytlik, L. (2012), "E-recruiting and fairness: The applicant's point of view", Information Technology and Management, Vol. 13, No. 2, pp. 59-67.

Thryft, A. (2019), "Not a lot of debiasing, auditing tools yet", Electrical Engineering Times, May 5, available at: [https://www.eetimes.com/document.asp?doc\\_id=1334650](https://www.eetimes.com/document.asp?doc_id=1334650) (accessed May 13, 2019).

Tiku, N. (2018), "The dirty war over diversity inside Google", Wired Magazine, January 26, available at: <https://www.wired.com/story/the-dirty-war-over-diversity-inside-google> (accessed January 27, 2018).

Truxillo, D., Steiner, D. and Gilliland, S. (2004), "The importance of organizational justice in personnel selection: Defining when selection fairness really matters", International Journal of Selection and Assessment, Vol.12, No. 2, pp. 39-53.

Valentino-DeVries, J. (2013), "Bosses May Use Social Media to Discriminate Against Job Seekers", Wall Street Journal, November 20, available at: <https://www.wsj.com/articles/bosses-may-use-social-media-to-discriminate-against-job-seekers-1384979412> (accessed January 27, 2018).

Vasconcelos, M., Cardonha, C. and Goncalves, B. (2017), "Modeling epistemological principles for bias mitigation in AI systems: An illustration in hiring decisions", arXiv preprint arXiv:1711.07111, available at: <https://arxiv.org/pdf/1711.07111> (accessed January 27, 2018).

Wachter-Boettcher, S. (2017), "AI recruiting tools do not eliminate bias", Time Magazine, October 25, available at: <http://time.com/4993431/ai-recruiting-tools-do-not-eliminate-bias> (accessed March 13, 2019).

Walker, H. J., Helmuth, C. A., Feild, H. S. and Bauer, T. N. (2015), "Watch what you say: Job applicants' justice perceptions from initial organizational correspondence", Human Resource Management, Vol. 54, No. 6, pp. 999-1011.

Wiechmann, D. and Ryan, A. M. (2003), "Reactions to computerized testing in selection contexts", International Journal of Selection and Assessment, Vol. 11, No. 2-3, pp. 215-229.

Williams, B., Brooks, C. and Shmargad, Y. (2018), "How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications", Journal of Information Policy, Vol. 8,

pp. 78-115.

Winning, L. (2018), "It's time to prioritize diversity across tech", Forbes, March 13, available at: <https://www.forbes.com/sites/lisawinning/2018/03/13/its-time-to-prioritize-diversity-across-tech/#642deab616f8> (accessed October 30, 2018).

Woodruff, S., Fox, S., Rousso-Schindler, S. and Warshaw, J. (2018), "A qualitative exploration of perceptions of algorithmic fairness", Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, April 21-26, Montreal QC, Canada, Paper No. 656, pp. 1-14.