Analytics tools in the talent acquisition context are on the rise. They are marketed as the most objective and efficient means of sourcing and selecting the perfect candidate for a particular job. However, government enforcement agencies such as the EEOC have expressed concern that big data algorithms and machine learning may result in discriminatory hiring practices in violation of Title VII, the ADEA, the ADA, and other laws against workplace discrimination. The purpose of this paper is to explore how companies are using big data and other analytics tools in the employment context, to examine the legal considerations and risks that should be considered when doing so, and to offer practical guidance to employers who use data analytics to source, recruit, and hire candidates for employment.

A. Big Data Tools To Source, Recruit, and Hire

Big data tools use algorithms to find and select applicants. Although such screening devices have the appearance of objectivity, government enforcement agencies have raised concerns that they are not neutral insofar as they either replicate the prejudices of the employer or perpetuate the selection of employees who share the same characteristics of the existing workforce.¹ This section addresses some of the big data tools that are being marketed and how they may violate non-discrimination statutes.

1. Search Algorithms

Recruiting algorithms scan the internet to find qualified candidates for a particular job and then encourage them to apply. The vendors who sell these search tools claim to include variables that can predict who will be a top performer if hired. Some algorithms are known as “smart algorithms” that constantly update based upon an employer’s past hiring to target candidates with similar characteristics to the employer’s top performers. For example, the company Entelo offers a service that scans over 275 million job candidates and identifies which candidates are likely to change jobs based on its collection of publicly-available information about applicants from social media websites including LinkedIn. Other companies create algorithms to narrow the applicant pool. For instance, some companies are marketing algorithms created based upon words or phrases from employment applications or resumes of high-performing incumbents. The algorithms are then used to narrow applicant pools to those candidates with similar words or phrases on their applications or resumes.

Recruiting algorithms that serve as an employer’s primary tool, or exclude candidates from the pool, may be in operation excluding members of protected groups.

As an example, the Communications Workers of America recently filed a putative age discrimination class action against several major corporations that use Facebook to source and recruit candidates through sponsored advertisements for job openings. According to the lawsuit, Facebook tells users who click on a link that they are receiving the sponsored advertisement due to their age: “want[ ] to reach people ages 21 to 55.” Unlike the targeted recruiting of minority candidates that the EEOC and courts have blessed, the targeted recruiting of younger workers here may have a significant disparate impact on the protected class of older workers. Of course, it remains to be seen whether the plaintiffs can prove actionable

---


3 See Solon Barocas & Andrew D. Selbst, Big Data’s Disparate Impact, 104 CAL. L. REV. 671, 680-81 (2016) (“[I]f data mining treats cases in which prejudice has played some role as valid examples to learn from, that rule may simply reproduce the prejudice involved in these earlier cases; or [] if data mining draws inferences from a biased sample of the population, any decision that rests on these inferences may systematically disadvantage those who are under- or overrepresented in the dataset.”).

4 See Complaint [Dkt. 1], Comm’n Workers of Am. v. T-Mobile US, Inc., et. al., No. 17-7232 (N.D. Cal.).

5 Id. ¶ 3.

6 See Duffy v. Wolle, 123 F.3d 1026, 1038-39 (8th Cir. 1997) (“[A]n employer’s affirmative efforts to recruit minority and female applicants does not constitute discrimination.”); see also EEOC, EEOC COMPL. MAN., SECTION 15: RACE & COLOR DISCRIMINATION (Apr. 19, 2006), http://www.eeoc.gov/policy/docs/race-color.html (“[I]f an employer notices that African Americans are not applying for jobs in the numbers that would be expected given their availability in the labor force, the employer could adopt strategies to expand the applicant pool of
discrimination based upon the practice of targeted recruiting based on age. However, even if the practice itself is not found to be unlawful, plaintiffs may use such targeting as evidence of discrimination with respect to age and potentially other aspects of employment.

2. Applicant Sourcing Systems

Applicant tracking systems are also being marketed for sourcing applicants at the initial stages of employee hiring.7 These systems incorporate big data algorithms by mining the available public data on both passive and active candidates, “looking for statistical correlations that connect seemingly unrelated variables, such as patterns of social media behavior, with workplace performance.”8

One example is a chatbot that communicates with applicants by conducting the initial stages of a hiring process such as sourcing, screening resumes, and scheduling interviews.9 At least 75 providers are competing to sell these services to recruiters and employers.10 Some chatbots incorporate machine learning, which is “the science of getting computers to act without being explicitly programmed.”11 An example of the legal risks of chatbots arose in 2016 when Microsoft created a Twitter account to “respond[] to users’ queries and emulate[] the casual, jokey speech patterns of a stereotypical millennial.”12 The chatbot used machine learning insofar as Microsoft let the chatbot search for general structures in the dataset of tweets it received.13 The result was unfortunate: the chatbot quickly received many racist, misogynistic tweets and qualified African Americans such as recruiting at schools with high African American enrollment.”

--

13 Id.; see also Barocas & Selbst, 104 CAL. L. REV. at 678 n.24 (explaining the difference between this sort of “unsupervised” machine learning and “supervised” machine learning like Internet search algorithms where an employer “must actively specify a target variable of interest”).
mimicked them such that it “blithely use[d] racial slurs, defend[ed] white-supremacist propaganda, and even outright call[ed] for genocide.”

3. **Analysis Algorithms**

Finally, employers are also using algorithms to analyze video interviews where a candidate prerecords taped answers to interview questions for an employer to later review at its convenience. Goldman Sachs, for example, recently shifted all of its first round interviews to this format. Once videotaped, employers can subject the recordings to an algorithm that spots tens of thousands of hints about intents, habits, personality, and qualities in candidate responses. This assessment of prerecorded video interviews provides employers far more information about candidates than even internet search algorithms. For example, it will check for voice inflections and microexpressions that convey a range of emotion based on psychological research. Employers who consider using an algorithm to analyze prerecorded video interviews in the hiring process could unwittingly exclude candidates with disabilities based on factors completely unrelated to their ability to successfully perform the job.

**B. Practical Guidance For Employers**

In the absence of a sound, professionally developed validation study that demonstrates otherwise, employers should carefully evaluate whether their big data tools will have a disparate impact on some protected group. This is particularly true where the vendor claims that its algorithm screens for characteristics like educational attainment and/or past work experience. Such automated screening lacks the discretion a human reviewer might exercise to mitigate the chance that those characteristics alone disparately impact racial and national origin groups unless job related and consistent with business necessity. Similarly, if job tenure is within the algorithm, women who leave the labor force to have children, or persons with disabilities who have had periods of medical absence from work, could be adversely impacted by the algorithm. And algorithms that incorporate year of high school or college graduation could target workers based on age. Video analyses that pick up on physical traits (expression, movement, voice) which may be impacted by disability are also likely to have an adverse impact. Algorithms — and, even more so, assessments — that attempt to analyze personality characteristics using the five factor model of personality may also screen out individuals with mental health disabilities. Some may even tread on the ADA line barring pre-offer medical evaluations.

---


Since these screening algorithms do not simulate the job or test for human traits, employers will need to conduct a criterion-related validity study in order to document a statistical relationship between the data inputs for the screening algorithm and successful job performance. Most criterion-related validation studies rely on a sample of incumbent employees who are given the screening procedure and whose job performance is rated. Statistical analyses are then performed to see whether success on the screening procedure predicts success on the job. While many vendors claim to have conducted this type of validation study, employers — which should assume Title VII disparate impact liability when using those vendor services — must ensure that the data inputs for the screening algorithm satisfy the rigorous process under the Uniform Guidelines on Employee Selection Procedures (“Uniform Guidelines”). Some industrial-organizational psychologists would not consider algorithms validated if the criteria being screened for is unrelated to the job or job performance. For example, it may be that there is a correlation between the high performers in an organization and liking a social media page on curly fries; that would not mean, however, that all experts would agree that an algorithm can validly include liking curly fries in its criteria.

Another important consideration for big data tools is how they affect the size of an employer’s applicant pool for a particular position. The larger an employer’s applicant pool, the more likely that a policy or practice in selecting applicants from that pool will have a statistically significant disparate impact on a protected class. In addition, any data solution that considers information about candidates to determine who should advance in the hiring process will likely convert all such individuals into “applicants” for purposes of adverse impact analysis. For instance, if an algorithm is used to filter all applicants, a company’s pool of applicants for purposes of assessing adverse impact will arguably include everyone subject to the algorithm, not just those applications or resumes that were reviewed or considered by a person. This remains true regardless of whether the employer is using a data tool to screen the resumes themselves for particular words, or using a data tool to scour the internet for information about all of its applicants to narrow the pool. Therefore, an employer must consider whether to use data screening tools on the entire applicant population or on a smaller randomly selected subset of applicants, taking into account the size of the applicant pool, the strength of validation evidence, and any adverse impact it is likely to produce.

In addition, employers should consider whether the use of these big data tools may heighten the risk of class certification of any substantive claims that may arise. Since the algorithms are centrally controlled and applied to everyone in the potential class, they may be more likely to provide the “glue” to support a finding of commonality than other types of selection practices.

---

Apart from potential discrimination claims, employers should also consider whether the use of vendors to provide them information about applicants would fall under the purview of the Fair Credit Reporting Act (“FCRA”). FCRA applies when an employer obtains a “consumer report” to use in evaluating a candidate for hiring or promotion. A consumer report includes reports on an applicant’s “character, general reputation, personal characteristics, or mode of living . . . .” 15 U.S.C. § 1681a(d)(1). Many algorithms consider information encompassed by those areas in recommending whether someone is suitable for employment or not. A consumer reporting agency includes any company which regularly, for a fee, assembles or evaluates information on consumers to provide consumer reports to a third party. Id. at § 1681a(f). Thus, a vendor providing algorithmic search services could be found to be a consumer reporting agency. If the reports that vendors provide to employers are deemed to be consumer reports, then a host of obligations arise including providing notice to and obtaining consent from the applicant before obtaining the consumer report, providing pre-adverse action notice to the applicant before making a decision not to hire him or her because of the information contained in the report, and providing notice about FCRA rights.

Above all, employers who proceed with purchasing big data hiring solutions should seriously consider retaining an experienced professional who specializes in employment selection procedures to review the validation work performed by the vendor before the product is used on real applicants. Absent such an assessment, any potential benefits of these screening processes are greatly outweighed by the legal risk of a successful disparate impact challenge under Title VII and other employment laws.