

Hiring

All the Ways Hiring Algorithms Can Introduce Bias

by Miranda Bogen

May 06, 2019



Eric Raptosh Photography/Getty Images

Summary. Understanding bias in hiring algorithms and ways to mitigate it requires us to explore how predictive technologies work at each step of the hiring process. Though they commonly share a backbone of machine learning, tools used earlier in the process can... [**more**](#)

Do hiring algorithms prevent bias, or amplify it? This fundamental question has emerged as a point of tension between the technology's proponents and its skeptics, but arriving at the answer is more

complicated than it appears.

Hiring is rarely a single decision, but rather the culmination of a series of smaller, sequential decisions. Algorithms play different roles throughout this process: Some steer job ads toward certain candidates, while others flag passive candidates for recruitment. Predictive tools parse and score resumes, and help hiring managers assess candidate competencies in new ways, using both traditional and novel data.

Many hope that algorithms will help human decision-makers avoid their own prejudices by adding consistency to the hiring process. But algorithms introduce new risks of their own. They can replicate institutional and historical biases, amplifying disadvantages lurking in data points like university attendance or performance evaluation scores. Even if algorithms remove some subjectivity from the hiring process, humans are still very much involved in final hiring decisions. Arguments that cast “objective” algorithms as fairer and more accurate than fallible humans fail to fully recognize that in most cases, both play a role.

Understanding bias in hiring algorithms and ways to mitigate it requires us to explore how predictive technologies work at *each step* of the hiring process. Though they commonly share a backbone of machine learning, tools used earlier in the process can be fundamentally different than those used later on. Even tools that appear to perform the same task may rely on completely different types of data, or present predictions in substantially different ways.

Our analysis of predictive tools across the hiring process helps to clarify just what “hiring algorithms” do, and where and how bias can enter into the process. Unfortunately, we found that most hiring algorithms will drift toward bias by default. While their potential to help reduce interpersonal bias shouldn’t be discounted, only tools that proactively tackle deeper disparities will offer any hope that predictive technology can help promote equity, rather than erode it.

Shaping the candidate pool

The hiring process starts well before a jobseeker submits an application. During the “sourcing” or recruiting stage, predictive technologies help to advertise job openings, notify jobseekers about potentially appealing positions, and surface prospective candidates to recruiters for proactive outreach.

To attract applicants, many employers use algorithmic ad platforms and job boards to reach the most “relevant” job seekers. These systems, which promise employers more efficient use of recruitment budgets, are often making highly superficial predictions: they predict not who will be successful in the role, but who is most likely to click on that job ad.

These predictions can lead jobs ads to be delivered in a way that reinforces gender and racial stereotypes, even when employers have no such intent. In a recent study we conducted together with colleagues from Northeastern University and USC, we found, among other things, that broadly targeted ads on Facebook for supermarket cashier positions were shown to an audience of 85% women, while jobs with taxi companies went to an audience that was approximately 75% black. This is a quintessential case of an algorithm reproducing bias from the real world, without human intervention.

Meanwhile, personalized job boards like ZipRecruiter aim to automatically learn recruiters’ preferences and use those predictions to solicit similar applicants. Like Facebook, such recommendation systems are purpose-built to find and replicate patterns in user behavior, updating predictions dynamically as employers and jobseekers interact. If the system notices that recruiters happen to interact more frequently with white men, it may well find proxies for those characteristics (like being named Jared or playing high school lacrosse) and replicate that pattern. This sort of adverse impact can happen without explicit instruction, and worse, without anyone realizing.

Sourcing algorithms are not likely top of mind for most people when they think “hiring algorithm.” But automated decisions at this early stage of the hiring funnel are widespread. For example, the tool Amazon scrapped for disadvantaging women was not a selection tool to assess actual applicants, but a tool to help uncover passive candidates for recruiters to solicit.

Sourcing algorithms may not overtly reject applicants, but as legal scholar Pauline Kim has argued, “not informing people of a job opportunity is a highly effective barrier” to people seeking jobs. These tools may not always make for dystopian headlines, but they play a critical role in determining who has access to the hiring process at all.

Narrowing the funnel

Once applications start flowing in, employers seek to focus on the strongest candidates. While algorithms used at this stage are often framed as decision aids for hiring managers, in reality, they can automatically reject a significant proportion of candidates.

Some of these screening algorithms are simply old techniques dressed up in new technology. Employers have long asked “knockout questions” to establish whether candidates are minimally qualified; now, chatbots and resume parsing tools perform this task. Other tools go further, using machine learning to make predictions based on past screening decisions, saving employers time and, purportedly, minimizing the effect of human prejudice. At first glance, it might seem natural for screening tools to model past hiring decisions. But those decisions often reflect the very patterns many employers are actively trying to change through diversity and inclusion initiatives.

Other selection tools incorporate machine learning to predict which applicants will be “successful” on the job, often measured by signals related to tenure, productivity, or performance (or by the absence of signals like tardiness or disciplinary action). Newer tools in this space claim to help employers use subtler signals to make their predictions, like game play or video analysis.

Notably, in the United States, these kinds of selection procedures fall under traditional regulations. Employers are obligated to inspect their assessment instruments for adverse impact against demographic subgroups, and can be held liable for using procedures that overly favor a certain group of applicants. Several assessment vendors describe in detail the steps they take to “de-bias” their algorithms — steps that also happen to ensure their clients are compliant with the law.

But the very act of differentiating high performers from low performers often reflects subjective evaluations, which is a notorious source of discrimination within workplaces. If the underlying performance data is polluted by lingering effects of sexism, racism, or other forms of structural bias, de-biasing a hiring algorithm built from that data is merely a band-aid on a festering wound. And if an employer can prove that its selection tool serves a concrete business interest — a relatively low bar — they can easily justify using a selection algorithm that leads to inequitable outcomes. Some industrial-organizational psychologists, who are often involved in the development of hiring procedures, are skeptical of relying solely on atheoretical correlations as a basis for new selection tools, but nothing in current regulatory guidelines requires employers do much more.

Finally, once an employer selects a candidate to hire, other predictive tools seek to help the employer make an offer that the candidate is likely to accept. Such tools could subvert laws banning employers from asking about salary history directly, locking in — or at least making it more difficult to correct — longstanding patterns of pay disparity.

Bending hiring algorithms toward equity

While existing U.S. law places some constraints on employers using predictive hiring tools, it is ill-equipped to address the evolving risks presented by machine learning-enhanced hiring tools.

So, how can we ensure hiring algorithms actually promote equity? Regulation (which is sluggish) and industry-wide best practices (which are nascent) certainly have roles to play. In the meantime, vendors building predictive hiring tools and employers using them must think *beyond* minimum compliance requirements. They must clearly consider whether their algorithms actually produce more equitable hiring outcomes. Before deploying any predictive tool, they should evaluate how subjective measures of success might adversely shape a tool's predictions over time. And beyond simply checking for adverse impact at the selection phase, employers should monitor their pipeline from start to finish to detect places where latent bias lurks or emerges anew.

Unless those touting the potential for algorithms to reduce bias in hiring choose to proactively build and test their tools with that goal in mind, the technology will at best struggle to fulfill that promise — and at worst, undermine it.

MB

Miranda Bogen is a Senior Policy Analyst at Upturn, a nonprofit research and advocacy group that promotes equity and justice in the design, governance, and use of digital technology.