RESEARCH

THE FUTURE IS FAIR: HOW AI IS ELIMINATING BIAS

HireVue

In the age of artificial intelligence (AI), bias has become a major concern.

Many examples of “intelligent” machines behaving in biased ways have appeared in the press. These biased bots are born when programmers inadvertently train them to replicate biased sets of data.

INTRODUCTION

One notorious example is Microsoft's doomed chatbot Tay, who mimicked racist attitudes in her conversation that she quickly learned from her interactions with human users.[1]

In the hiring world, bias has been a potent concern for many decades. When making employee selection decisions, Title VII of the Civil Rights Act of 1964 prohibits discrimination on the basis of race, creed, religion, sex, national origin, and more recently, physical/mental disabilities and sexual orientation. **The law dictates that selection, promotion, and termination decisions be made on the basis of job performance, not class status.**

> *"Companies in the top quartile for gender diversity are 15% more likely to have financial returns above their respective national industry medians."*
>
> *HTTPS://WWW.MCKINSEY.COM/BUSINESS-FUNCTIONS/ORGANIZATION/OUR-INSIGHTS/WHY-DIVERSITY-MATTERS*

1. *https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter?CMP=twt_a-technology_b-gdntech*

THE STUBBORN ISSUE OF UNCONSCIOUS BIAS

Discrimination is difficult to eradicate because humans often struggle to rely solely on objective criteria when making decisions.

We sometimes act on hunches or take pride in our instincts about what makes a good choice. Or we are vulnerable to prejudices we might not know we even have (or worse, are fully aware we do). Consequently, since Title VII came into being in 1964, job applicants and employees claiming to be victims of unfair and illegal discrimination have pursued the remedies allowed them by law.

Companies using hiring tools must be scrupulous about ensuring their products do not result in differences in test scores on the basis of protected group classes, and hiring personnel must document and support all decisions to prevent and defend against bias claims.

> *"A Yale University study found that male and female scientists, both trained to be objective, were more likely to hire men, and consider them more competent than women, and pay them $4,000 more per year than women."*
>
> *HTTPS://WWW.FORBES.COM/SITES/PRAGYAAGARWALEUROPE/2018/12/03/UNCONSCIOUS-BIAS-HOW-IT-AFFECTS-USMORE-THAN-WE-KNOW/#584C17666E13*

LEARN MORE AT HIREVUE.COM

EARLY ADOPTION, PROMISING RESULTS

Those in the industry producing AI based assessments, or assessments of any kind, must be mindful of the insidious role of implicit bias and actively work to promote diversity and fairness.

A legally defensible assessment should not only show significant ability to predict job performance, it must also show low or nonexistent gender or racial group differences. That is, members of Title VII protected classes must score similarly with respect to non-protected or majority group members. It is vitally important to both document a lack of group differences at the time of validation and also to continuously monitor scores to ensure that group differences do not unknowingly drift into trouble over time.

> *"Caucasians received 50% more callbacks than those from African American descent."*
>
> *HTTPS://WWW.POVERTYACTIONLAB.ORG/EVALUATION/DISCRIMINATION-JOBMARKET-UNITED-STATES*

LEARN MORE AT HIREVUE.COM

WHERE BIAS LURKS

While the selection world is familiar with investigating and documenting Title VII-type discrimination, bias is a relatively new concept in the AI world. And while, as cited above, AI can create bias when applied imprudently, it also has the potential to make great strides toward its elimination.

First, let's take a step back and clarify some definitions. At its most basic, bias is a prejudice for or against something not based on reason or actual experience—frequently resulting in unfair treatment. In the hiring world, Title VII defines bias in a very specific way for very specific categories of people, but it is essential to realize that bias is certainly not limited to distinct categories like race, sex, and religion.

Psychological and brain research tells us that the human brain uses many emotional and intuitive mechanisms to make decisions, and these very routines of thinking are often the foundation of our bias.[2] As noted by Nobel Prize-winning psychologist, Daniel Kahneman, "The confidence people have in their beliefs is not a measure of the quality of evidence but of the coherence of the story the mind has managed to construct."[3]

2. *https://pbs.dartmouth.edu/new-study-biases-during-learning-and-decision-making*
3. *https://www.brainpickings.org/2013/10/30/daniel-kahneman-intuition/*

BIASES AND DEFINITIONS

Humans have many known biases in decision making. Some of the common ones that influence hiring decisions are:

Confirmation bias – When we seek out information to confirm our pre-existing beliefs

Halo/Horns effect – When we assume one piece of information (i.e., a candidate worked for a top competitor) generalizes to the rest of the candidate's skills

Similarity bias – When we are drawn to people similar to ourselves (this bias can lead to a lack of diversity)

While AI can unintentionally replicate existing biases, when trained properly it can also powerfully predict sources of bias, allowing for their control and elimination. The industry should have a singular goal: **The total elimination of all forms of bias.**

The definition of bias needs to be much broader than that of Title VII to include any basis for bias, not just the traditional protected classes, and also recognize that membership in a certain class may not be a binary affair.[4, 5] For example, an individual might have parents of different races or not identify as their biological sex.

Multiple states including Washington, D.C. and Pennsylvania offer a third gender option on driver's licenses and the trend is growing at a rapid pace.[6]

As part of the hiring industry, we—and we're hoping not to be the only ones—are excited to embrace and bring about innovative forms of psychometric approaches commensurate with a more refined understanding of identity.

The modern definition of bias

A prejudice based on irrelevant or unfair factors, including not only the Title VII protected classes, but any characteristic unrelated to outcomes of interest. Further, membership in any class is considered on a continuum rather than as a dichotomy.

4. *Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American psychologist, 50(9), 741.*
5. *Messick, S. (1995). Standards of validity and the validity of standards in performance asessment. Educational measurement: Issues and practice, 14(4), 5-8.*
6. *https://www.cnn.com/2019/08/01/health/washington-pennsylvania-gender-x-id/index.html*

CONTROLLING FOR ADVERSE IMPACT

How is the elimination of all forms of bias in the hiring process possible?

To answer this question, we need to consider how we detect bias in the first place. When we couple assessment data with protected class information, we can examine whether members of one group score significantly higher or lower than members of another group. If the protected group scores lower by a certain margin, this gap could be evidence that the assessment is having an adverse impact on a particular class of candidates.

Adverse impact is not necessarily illegal, if the selection process itself predicts performance, is well-designed, and meets professional standards. But the goal is to eliminate adverse impact whenever possible to promote fair and equitable treatment and to support diversity, which has been shown in many studies to contribute positively to organizational success.

> *"Companies with a more culturally and ethnically diverse executive team were 33% more likely to see better-than-average profits."*
>
> *HTTPS://WWW.MCKINSEY.COM/BUSINESS-FUNCTIONS/ORGANIZATION/OUR-INSIGHTS/DELIVERING-THROUGHDIVERSITY*
>
> *HTTPS://WWW.FORBES.COM/SITES/KARSTENSTRAUSS/2018/01/25/MORE-EVIDENCE-THAT-COMPANYDIVERSITY-LEADS-TO-BETTERPROFITS/#5705E64B1BC7*

FREQUENT MONITORING IS ESSENTIAL

Group differences analyses, of the sort described above, are typically done at defined intervals, such as on an annual basis and often in keeping with the classic strategies such as concurrent validation (i.e., with incumbents) or predictive design validation (i.e., with newly hired applicants).

But as the applicant population evolves, the nature of predictor-criterion relationships can change, and so organizations and their hiring vendors must monitor these statistics as frequently as possible. In fact, with enough available data paired with realtime multimedia predictions of construct irrelevant variance, algorithms can be developed that monitor the size of group differences continuously—the ultimate way to combat adverse impact. The future of fairness is an extension of the automated scoring approach such that we **predict, measure,** and **programmatically control** for bias.[7]

> *"A full 66% of companies have strategies for diversity hiring, but only 25% set gender diversity targets."*
>
> *HTTPS://BUSINESS.LINKEDIN.COM/TALENTSOLUTIONS/BLOG/DIVERSITY/2017/11-STATS-FROM-SHERYL-SANDBERGSGENDER-DIVERSITY-REPO*

7. *Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. Journal of Applied Psychology, 101(7), 958-975.*

LEARN MORE AT HIREVUE.COM

THE FUTURE IS FAIR

Bias, meaning prejudice against certain characteristics that are unrelated to job performance, hampers the ability of an organization to bring in the best and brightest talent possible.

At HireVue, we aim to not only minimize protected class differences but eliminate them along with other forms of bias. Bias, meaning prejudice against certain characteristics that are unrelated to job performance, hampers the ability of an organization to bring in the best and brightest talent possible. It is unfortunately pervasive in human decisions, and, as described above, often operates at an unconscious level.

With enough of the right type of data paired with state-of-the-art AI models analyzing text, audio, and video, we can identify and eliminate biases today in ways that were not conceivable just a few years ago.

LEARN MORE AT HIREVUE.COM

EXAMPLE BIASES

Let's take a look at some example biases and how emerging AI technology can thwart them.

Protected classes: As described above, AI algorithms can continuously monitor group differences on collected data, generating alerts when sizable differences are detected. This requires asking candidates questions to identify protected group membership, and since these questions are voluntary, that enough respondents answer them to generate reliable findings.

Dissimilar people: As humans, we tend to be drawn toward those who are most similar to us. But by using intelligent hiring tools to select people scientifically based on characteristics that are proven to be job-related, you can eliminate this bias and help increase the diversity and performance of your team.

Social awkwardness: Some people interview better than others. People who are naturally reserved or introverted may have more difficulty connecting during a brief interview, but they are often just as qualified and capable of performing the job. AI can help to root out these types of biases to ensure you hire the best regardless of interpersonal style or social anxiety.

Other characteristics: Humans have innate biases toward all manner of characteristics not related to job performance, including hair styles, tattoos, height, clothing choices, smells, voice styles, and on and on. You may prefer one over the other unconsciously or for completely idiosyncratic reasons. Scientific selection helps prevent us from making decisions based on such potentially harmful prejudices.

People who are naturally reserved or introverted may have more difficulty connecting during a brief interview, but they are often just as qualified and capable of performing the job.

LEADING-EDGE RESEARCH AND PRACTICE

While today's leading assessments are designed to control bias, other data in the hiring process—especially more passive data not necessarily collected within the bounds of an assessment and characterized by missing or inconsistent data—may show evidence of group differences. To counteract this, HireVue is researching and deploying several exciting new AI capabilities, including:

Using natural language processing to anonymize text or other information that might identify group membership so human decision makers are not influenced by it.

Predicting group membership status if it is not known—to the tune of around 99% accuracy—so it can be programmatically controlled for presentation to hiring stakeholders.

Developing algorithms that can spot and control for various biases in large databases and also alert system users to any issues in real time.

> The field of bias detection and prevention is advancing quickly thanks to large data sets and advanced machine learning techniques.

The field of bias detection and prevention is advancing quickly thanks to large data sets and advanced machine learning techniques. However, much work needs to be done in the talent acquisition arena to make sure bias is eventually eliminated from the hiring process. **A commitment and meaningful effort from everyone in the industry is the only way to create a future of fair and transparent hiring practices.**

LEARN MORE AT HIREVUE.COM