Harvard Business Review

REPRINT H05AD9
PUBLISHED ON HBR.ORG
NOVEMBER 21, 2019

ARTICLE
HIRING

Building Ethical AI for Talent Management

by Tomas Chamorro-Premuzic, Frida Polli and Ben Dattner

Harvard Business Review

HIRING

Building Ethical AI for Talent Management

by Tomas Chamorro-Premuzic, Frida Polli and Ben Dattner

NOVEMBER 21, 2019

HBR STAFF/LISHAKOV/UNSPLASH

Artificial intelligence has disrupted every area of our lives — from the curated shopping experiences we've come to expect from companies like Amazon and Alibaba to the personalized recommendations that channels like YouTube and Netflix use to market their latest content. But, when it comes to the workplace, in many ways, AI is still in its infancy. This is particularly true when we consider the ways it is beginning to change talent management. To use a familiar analogy: AI at work is in the dial-up mode. The 5G WiFi phase has yet to arrive, but we have no doubt that it will.

COPYRIGHT © 2019 HARVARD BUSINESS SCHOOL PUBLISHING CORPORATION. ALL RIGHTS RESERVED.

To be sure, there is much confusion around what AI can and cannot do, as well as different perspectives on how to define it. In the war for talent, however, AI plays a very specific role: to give organizations more accurate and more efficient predictions of a candidate's work-related behaviors and performance potential. Unlike traditional recruitment methods, such as employee referrals, CV screening, and face-to-face interviews, AI is able to find patterns unseen by the human eye.

Many AI systems use real people as models for what success looks like in certain roles. This group of individuals is referred to as a "training data set" and often includes managers or staff who have been defined as "high performers." AI systems process and compare the profiles of various job applicants to the "model" employee it has created based on the training set. Then, it gives the company a probabilistic estimate of how closely a candidate's attributes match those of the ideal employee.

Theoretically, this method could be used to find the right person for the right role faster and more efficiently than ever before. But, as you may have realized, it has become a source of both promise and peril. If the training set is diverse, if demographically unbiased data is used to measure the people in it, and if the algorithms are also debiased, this technique can actually mitigate human prejudice and expand diversity and socioeconomic inclusion better than humans ever could. However, if the training set, the data, or both are biased, and algorithms are not sufficiently audited, AI will only exacerbate the problem of bias in hiring and homogeneity in organizations.

In order to rapidly improve talent management and take full advantage of the power and potential AI offers, then, we need to shift our focus from developing more ethical HR systems to developing more ethical AI. Of course, removing bias from AI is not easy. In fact, it is very hard. But our argument is based on our belief that it is far more feasible than removing it from humans themselves.

When it comes to identifying talent or potential, most organizations still play it by ear. Recruiters spend just a few seconds looking at a resume before deciding who to "weed out." Hiring managers make quick judgments and call them "intuition" or overlook hard data and hire based on cultural fit — a problem made worse by the general absence of objective and rigorous performance measures. Further, the unconscious bias training implemented by a growing number of companies has often been found to be ineffective, and at times, can even make things worse. Often, training focuses too much on individual bias and too little on the structural biases narrowing the pipeline of underrepresented groups.

Though critics argue that AI is not much better, they often forget that these systems are mirroring our own behavior. We are quick to blame AI for predicting that white men will receive higher performance ratings from their (probably also white male) managers. But this is happening because we have failed to fix bias in the performance ratings that are often used in training data sets. We are shocked that AI can makes biased hiring decisions but fine living in a world where human biases dominate them. Just take a look at Amazon. The outcry of criticism about their biased recruiting algorithm ignored the overwhelming evidence that current human-driven hiring in most organizations is ineradicably worse. It's akin to expressing more concern over a very small number of

COPYRIGHT © 2019 HARVARD BUSINESS SCHOOL PUBLISHING CORPORATION. ALL RIGHTS RESERVED.

driverless car deaths than the 1.2 million traffic deaths a year caused by flawed and possibly also distracted or intoxicated humans.

Realistically, we have a greater ability to ensure both accuracy and fairness in AI systems than we do to influence or enlighten recruiters and hiring managers. Humans are very good at learning but very bad at unlearning. The cognitive mechanisms that make us biased are often the same tools we use to survive in our day-to-day lives. The world is far too complex for us to process logically and deliberately all the time; if we did, we would be overwhelmed by information overload and unable to make simple decisions, such as buying a cup of coffee (after all, why should we trust the barista if we don't know him?). That's why it's easier to ensure that our data and training sets are unbiased than it is to change the behaviors of Sam or Sally, from whom we can neither remove bias nor extract a printout of the variables that influence their decisions. Essentially, it easier to unpack AI algorithms than to understand and change the human mind.

To do this, organizations using AI for talent management, at any stage, should start by taking the following steps.

1) Educate candidates and obtain their consent. Ask prospective employees to opt in or to provide their personal data to the company, knowing that it will be analyzed, stored, and used by AI systems for making HR-related decisions. Be ready to explain the *what*, *who*, *how*, and *why*. It's not ethical for AI systems to rely on black-box models. If a candidate has an attribute that is associated with success in a role, the organization needs to not only understand why that is the case but also be able to explain the causal links. In short, AI systems should be designed to predict and explain "causation," not just find "correlation." You should also be sure to preserve candidate anonymity to protect personal data and comply with GDPR, California privacy laws, and similar regulations.

2) Invest in systems that optimize for fairness *and* accuracy. Historically, organizational psychologists have pointed to a drop in accuracy when candidate assessments are optimized for fairness. For example, much academic research indicates that while cognitive ability tests are a consistent predictor of job performance, particularly in high-complexity jobs, their deployment has adverse impact on underrepresented groups, particularly individuals with a lower socioeconomic status. This means that companies interested in boosting diversity and creating an inclusive culture often de-emphasize traditional cognitive tests when hiring new workers so that diverse candidates are not disadvantaged in the process. This is known as the fairness/accuracy trade-off.

However, this trade-off is based on techniques from half a century ago, prior to the advent of AI models that can treat the data very differently than traditional models. There is increasing evidence that AI could overcome this trade-off by deploying more dynamic and personalized scoring algorithms that are sensitive as much to accuracy as to fairness, optimizing for a mix of both. Therefore, developers of AI have no excuse for not doing so. Further, because these new systems now exist, we should question whether the widespread use of traditional cognitive assessments,

COPYRIGHT © 2019 HARVARD BUSINESS SCHOOL PUBLISHING CORPORATION. ALL RIGHTS RESERVED.

which are known to have an adverse impact on minorities, should continue without some form of bias-mitigation.

3) Develop open-source systems and third-party audits. Hold companies and developers accountable by allowing others to audit the tools being used to analyze their applications. One solution is open-sourcing non-proprietary yet critical aspects of the AI technology the organization uses. For proprietary components, third-party audits conducted by credible experts in the field are a tool companies can use to show the public how they are mitigating bias.

4) Follow the same laws — as well as data collection and usage practices — used in traditional hiring. Any data that shouldn't be collected or included in a traditional hiring process for legal or ethical reasons should not be used by AI systems. Private information about physical, mental or emotional conditions, genetic information, and substance use or abuse should never be entered.

If organizations address these issues, we believe that ethical AI could vastly improve organizations by not only reducing bias in hiring but also by enhancing meritocracy and making the association between talent, effort, and employee success far greater than it has been in the past. Further, it will be good for the global economy. Once we mitigate bias, our candidate pools will grow beyond employee referrals and Ivy League graduates. People from a wider range of socioeconomic backgrounds will have more access to better jobs — which can help create balance and begin to remedy class divides.

To make the above happen, however, businesses need to make the right investments, not just in cutting-edge AI technologies, but also (and especially) in human expertise — people who understand how to leverage the advantages that these new technologies offer while minimizing potential risks and drawbacks. In any area of performance, a combination of artificial and human intelligence is likely to produce a better result than one without the other. Ethical AI should be viewed as one of the tools we can use to counter our own biases, not as a final panacea.

Tomas Chamorro-Premuzic is the Chief Talent Scientist at ManpowerGroup, a professor of business psychology at University College London and at Columbia University, and an associate at Harvard's Entrepreneurial Finance Lab. He is the author of *Why Do So Many Incompetent Men Become Leaders? (and How to Fix It)*, upon which his TEDx talk was based. Find him on Twitter: @drtcp or at www.drtomas.com.

Frida Polli, a cognitive neuroscientist, is the co-founder and CEO of pymetrics, a platform that uses behavioral science and AI to match people to jobs accurately and fairly.

Ben Dattner is an executive coach and organizational development consultant, and the founder of New York City–based Dattner Consulting, LLC. You can follow him on Twitter at @bendattner.

COPYRIGHT © 2019 HARVARD BUSINESS SCHOOL PUBLISHING CORPORATION. ALL RIGHTS RESERVED.