Harvard Business Review
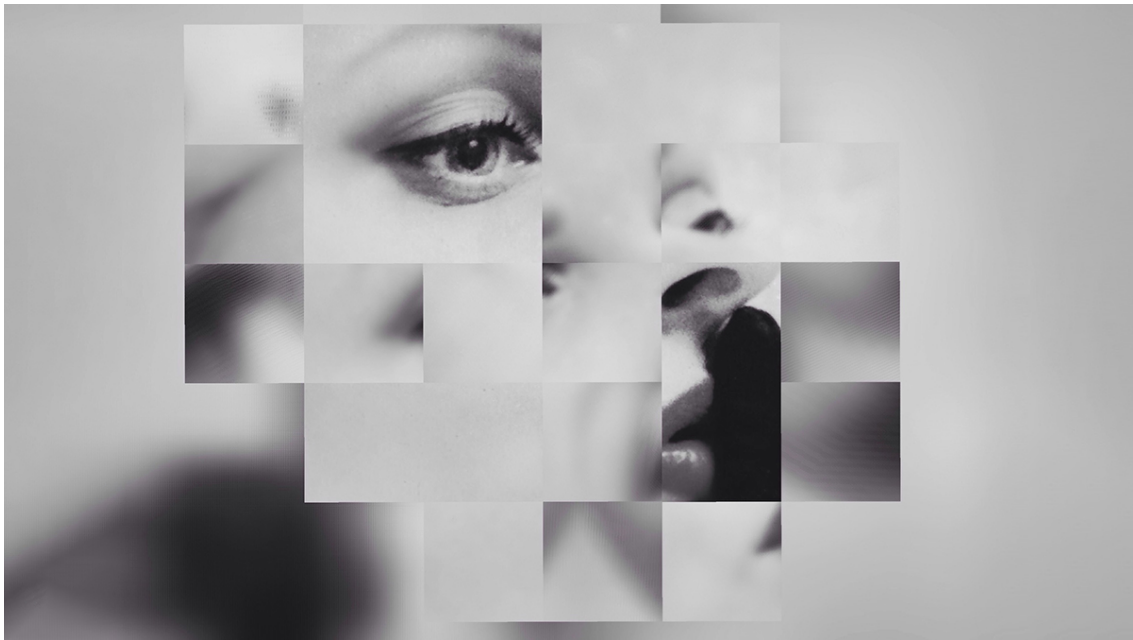
Hiring

Will AI Reduce Gender Bias in Hiring?

by Tomas Chamorro-Premuzic

June 10, 2019

Vlaaka Kocvarová/EyeEm/Getty Images

Summary.   Would women be better off if AI and algorithms were in charge of hiring? There are clearly reasons to be hopeful. One of the big advantages of AI is that, aside from being better at spotting things (i.e., millions of data points), it is also superior at ignoring things. AI can... **more**

AI is disrupting every area of life, including how organizations find talent. Companies are generally aware of the ROI that comes from finding the right person for the right job. McKinsey estimated that,

for highly complex roles, stars can be expected to produce 800% more than average performers. And a recent Harvard Business School study showed that there are even bigger benefits to avoiding toxic workers.

Despite this crucial role of talent, organizations are still unable to attract the right talent, relying more on intuitive rather than data-driven talent identification practices — especially at the top, where the stake are actually highest. Indeed, too many leaders are hired on the basis of their technical expertise, political influence, or interview performance. As I illustrate in my latest book, *Why Do So Many Incompetent Men Become Leaders?: (And How to Fix It),* most companies focus on the wrong traits, hiring on confidence rather than competence, charisma rather than humility, and narcissistic tendencies rather than integrity, which explains the surplus of incompetent and male leaders. The result is a pathological disconnect between the qualities that seduce us in a leader, and those that are needed to be an effective leader.

An interesting question that arises is to what degree new technologies within the brave new world of AI-based hiring tools could help us reduce error, noise, and bias in our talent identification processes. For example, would women be better off if AI and algorithms were in charge of hiring? Previous research has highlighted a clear inconsistency around gender and leadership. On the one hand, women are often evaluated more negatively by others – even when there are few granular behavioral differences between women and men. On the other hand, large scale meta-analyses suggest that women have a slight advantage when it comes to the soft skills that predispose individuals to be more effective leaders, and that they generally adopt more effective leadership styles than men do. For instance, if leaders were selected on the basis of their emotional intelligence, self-awareness, humility, integrity, and coachability, the majority of leaders would be female rather than male.

And yet there have been salient news stories recently indicating that AI may actually contribute to even more bias and adverse impact against women–and that when algorithms are trained to emulate human recruiters, they may not just reproduce human biases, but also exacerbate them, engaging in a much more efficient form of discrimination.

To be sure, we are much more easily shocked and scandalized by hiring mistakes done by AI, than by human errors or biases. It's a bit like with self-driving cars: it takes one autonomous car crash to convince us that the technology is flawed, but we are OK with having 1.2 million fatal accidents and 50 million driving injuries per year, courtesy of humans. So, let us start with the important realization that most hiring practices are (a) intuitive and (b) ineffective. For every company that appoints most of its leaders based on objective and meritocratic criteria, there are many more where such appointments are a true rarity — something that may be happening by accident, occasionally, and independently of their intentions. It is also clear that AI cannot be biased in the way humans are: that would require AI to have emotions, feelings, or opinions. AI does not need to engage in unconscious biases to penalize women or other underprivileged groups in order to get a self-esteem boost. Of course, if AI is trained with biased data — for instance, if we teach it to predict which candidates will be rated positively by *human* interviewers — it will not just emulate, but also exacerbate, human bias: augmenting it and making it far more efficient. But this can be addressed by teaching AI to predict relevant and objective outcomes, rather than mimic human intuition.

In addition, there *are* reasons to expect AI-talent tools to be more accurate and predictive than humans (not just because humans are generally bad at this):

- Our favorite method for screening and vetting candidates — including leaders — is the interview, and large-scale scientific studies have shown that interviews are most predictive when they are highly structured. Whereas in-person/analogue interviews are

hard to standardize, video interviews allow us to put people through exactly the same experience, capture millions of data points on their behaviors (e.g., what they say, how they say it, language use, body language, and micro-expressions), and remove prejudiced human observers from the process. It is safe to assume that automating *all* unstructured and humanly-rated interviews would reduce bias and nepotism while increasing meritocracy and predictive accuracy. This should be good for women (and bad for men).

- Of course there are some incredibly smart human interviewers who may generally outperform the algorithms (though watch out for the next Netflix documentary on how AI beats the best human interviewers, much like they beat the greatest chess or AlphaGo players). The main problem, however, is that most people are not as intuitive as they think. And for every brilliant interviewer, there are hundreds or thousands who think they are brilliant, but in reality, are not. We all think highly of our own intuition, especially when we are not intuitive. As one of the founders of the behavioral economics movement — and Nobel laureate — Daniel Kahneman noted: "We're generally overconfident in our opinions and our impressions and judgments." Regardless of AI's *ability* to detect talent, we can expect it to be much more *aware* of its ability than humans are of their own ability. This will also allow AI to improve (more than humans can be expected to do). Consider that the average human interviewer will never even admit to making a hiring mistake, for they will indulge in confirmation bias to see the candidates they personally hired in a positive vein once they are tasked with rating their performance. Humans have skin in the game: accepting mistakes makes them look stupid — AI does not care about looking stupid.

- One of the big advantages of AI is that, aside from being better at spotting things (i.e., millions of data points), it is also superior at *ignoring things*. Imagine an ethical, well-meaning, and open-minded human who has every intention of being fair in their hiring

practices and is therefore determined to avoid gender bias in his —
let's assume he is male — hiring process. Regardless of how hard he
tries, it will be very hard for him to ignore candidates' gender.
Imagine him sitting in front of a female candidate, repeating to
himself: "I must not think about the fact that this person is a
woman," or "I must not let this person's gender interfere with my
evaluation." In fact, the more he tries to suppress this thought, the
more prominent it will be in his mind. This will also lead to
distraction or over-compensation. In contrast, AI can be trained to
ignore people's gender and focus only on the relevant signals of
talent or potential. For example, algorithms can be trained to pick
up relevant signals of EQ, competence, or communication skills,
while being truly blind to gender. This would definitely favor
women.

- The critical factor for this to work is that organizations identify *real*
  performance data to train the algorithms. If AI is taught to predict
  or anticipate human preferences — like whether a candidate will be
  liked by their (human) boss once they are hired — we can expect
  bias to remain... and be augmented. However, if AI is trained to
  identify the actual drivers of performance — defined broadly as an
  individual's contribution to the organization — then we can expect
  a much fairer, more accurate, and replicable assessment of people's
  potential than what humans can do. This, again, should be good for
  women.

In sum, for those who are interested in not just helping women to be
more represented in the leadership ranks, but also improving the
quality of our leaders, there are clearly reasons to be hopeful about
AI. However, many of the emerging innovations in this brave new
world of technologically-enhanced and data-driven talent
identification are still a "work in progress," and we need to ensure
that they are not only accurate, but that they are also ethical and legal
alternatives to existing methods. Above all, it is time to admit that
most of the practices that are in place are far from effective, and that

they have contributed to much of the unfairness and nepotism that governs the average workplace. So, here's to finding the necessary self-awareness to begin to improve.

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.