INTRODUCTION

Today’s dynamic work environment places a premium on developing a robust and diverse talent pipeline to ensure sustainable, competitive positioning in the marketplace.

Fortunately, rapid advances in technology and the explosive growth of the Internet over the past decade have led to a dramatic shift in how we acquire the best and brightest talent. From sophisticated recruiting and applicant tracking systems to high-fidelity, online assessments capable of global reach, this technology-fueled transformation has delivered clear value as a strategic investment.

As the next generation of these technology-driven applications comes of age, a host of new issues and opportunities have emerged which require careful attention to ensure program integrity and racial and gender fairness associated with these applications. Some well known organizations such as IBM, Amazon, and Microsoft who have been pioneers in the use of technology to assist in talent decisions have recently paused or otherwise discontinued their artificial intelligence (AI) enabled efforts due to emerging concerns regarding the racial bias that such approaches can create[1,2]. While the deployment of facial recognition, machine learning (ML), and variations of AI are attractive options for some organizations, many companies are resisting the attraction of such “bright, shiny objects” and instead prioritizing sound selection practices, legal considerations, and their support for pressing socio-political issues to ensure that critical talent acquisition decisions are made efficiently, accurately, and free of bias.

Working to increase diversity is not only good for business[3], it is a timely and necessary response to correct for long-standing inequities and biases. Unfortunately, there appears to be growing evidence that evolving technologies in the talent acquisition space may be exacerbating rather than helping solve this intransigent challenge. In this article we explore how AI and other technologies can sometimes veer off track without organizations even realizing it, and we offer a set of recommendations for properly evaluating and implementing these increasingly popular talent acquisition technologies.
Artificial intelligence and advances in technology have altered what is possible in talent acquisition. Over the past decade, human resources (HR) systems have been configured and connected for large-scale applications where candidates are recruited online, drawn into an internet-based application process, automatically screened and prioritized, invited to complete automated tests, and presented with online interview questions based on the results of their assessments. Human contact in these scenarios generally does not occur until hundreds–or even thousands–of data points have been collected and evaluated on candidates.

The development of computer systems which can think and learn is known as artificial intelligence. In talent acquisition applications, the computer has learned to apply decision rules and evaluate massive amounts of applicant data more quickly and more consistently than the HR representatives who previously screened applications.

**Artificial Intelligence & Predictive Algorithms**

So, how does artificial intelligence identify the “best” candidates for us without human intervention? A wide range of data is available today on candidates ranging from Web-based applications, resumes, and external social media sites to internal talent management programs. These data can now be aggregated into large, multifaceted data sets (i.e., “big data”) and analyzed through machine learning algorithms to inform talent acquisition decisions. Through successive approximations, the “machine” learns data patterns from diverse and seemingly unrelated variables (such as social-media behavior, specific combinations of words on resumes, facial recognition software, email content, and available data within the employees’ organization). These data are analyzed, combined, and weighted to best predict the characteristics of the group of interest.

The use of predictive analytics for mining the wide range of available data sets has become a much-welcomed development for organizations. The reasons for this enthusiasm are clear. AI-powered software applications provide significant practical benefits and immediate payoff for organizations by reducing painstaking hours of administratively complex work required to move candidates through the application and selection processes. Organizations can leverage artificial-intelligence platforms to automate initial interactions with applicants, create up-to-date job postings, discover and target passive candidates with personalized messages, and generate engaging candidate experiences[4].

Once in the system, candidates can be screened across thousands of different dimensions, allowing organizations to automatically prioritize their talent and make both internal movement and external talent decisions more efficiently[5]. Analytics can also be aimed at identifying flight risks among high-performing women and minorities to help improve and sustain diversity representation within an organization. In fact, big data accumulated through employees’ online activity, job performance, promotional history, payroll data, and other sources have been used by several large organizations to “red flag” valued employees as flight risks and to increase retention by up to 20%[6].
According to a recent global benchmarking study\[7\], 88% of companies worldwide are already using artificial intelligence in some way for HR. Among the many technology-driven approaches used to acquire talent, two of the most prominent techniques are:

- **Al-powered video interviews** which have been employed to replace the phone screen and present evaluation data on thousands of dimensions, including facial cues, body language, voice intonation, and speech cadence. These artificial intelligence solutions leverage natural-language processing and machine learning to evaluate speech patterns that purport to measure a variety of personality characteristics\[8\].

- **Chatbots** are also increasingly used by organizations to connect with and evaluate candidates. As used in hiring, the chatbot can serve as an initial screening tool to ask “interview questions,” gather candidate responses, and evaluate the answers to identify viable candidates. It is typically designed to improve in efficiency and predictive accuracy over time through feedback loops with the company’s applicant tracking system.

## The Viability and Defensibility of AI Products

Artificial intelligence has been touted by HR-technology vendors and organizational users as a means to conquer a variety of intransient talent-acquisition and retention issues including the: 1) reduction of “unconscious” bias in the hiring process, 2) retention of top diverse talent through programs that single out flight risks, and 3) increased diversity in slating and hiring processes through algorithms that mandate pre-defined representation on slates and in hiring decisions.

While these methods present the possibility of identifying previously unrecognized talent and reducing some aspects of subjective decision making, their use can be compromised by errors inherent in the data and the decisions that underlie the design of the algorithms themselves. Moreover, a number of vendors claim that the variables being measured in their AI-driven products are proprietary and operate as a “black box” – the user must accept the output without understanding the specific bases on which the candidates are screened.

- **Data Integrity Issues.** A key question is whether the data that are aggregated and analyzed are accurate and complete enough to provide valid and reliable assessments of important characteristics necessary for successful job performance. Algorithms used for recruiting often include data obtained by searching publicly available databases where the accuracy or completeness of the data may be questionable, leading to missing and incorrect data in the selection process. In situations where the scoring algorithms contain hundreds or even thousands of data points, not all those data points will necessarily be populated for all candidates including highly qualified diverse candidates. Lundquist\[9\] described a case in which an organization developed an algorithm with over 100,000 potentially-scoreable individual data points. However, due to missing data and the nature of the characteristics being measured, only about 500 data points were scored—and in fact, those 500 data points were not necessarily the same from person to person. In that situation, candidates were being evaluated on different pieces of data, i.e., arguably different bases were used for the selection decisions.

- **Social-Media Data Scraping.** Big data analytics that rely on social-media data scraping are typically suspect—for many reasons—not the least of which is that not all prospective candidates will have a comparable social-media presence. Furthermore, certain protected classes may have differential access to the social-media activities that drive these big data models. Practically meaningful
and statistically significant differences exist in social-media usage based on race, age and gender across multiple platforms\textsuperscript{[10,11]}. Adverse impact becomes a real concern when big data analytics leverage social media as a key source of information for selection purposes.

- **Comparing Apples to Oranges.** Consider also the situation where the algorithm only draws information from resumes provided by applicants. The algorithm may give credit (positive or negative) for the inclusion of specific words in the resume (e.g., “technical” or “football”) or the absolute number of words in the resume. Two equally qualified candidates might describe their experience and backgrounds using different words or different styles of expression and be scored quite differently by the algorithm.

- **Exacerbating Bias.** Finally, algorithms may be trained to predict outcomes which are themselves the result of previous discrimination. For example, an algorithm that is developed based upon a non-diverse, high-performing group may simply be reflecting this group’s demographics more than the skills or abilities required for job success. The algorithm is matching people characteristics rather than defensible factors which predict job performance. As a result, the organization would not expect to increase demographic diversity among its hires by using such an algorithm and might be unintentionally institutionalizing a bias in its screening of applicants.

- **The Black Box.** The process of developing an algorithm by machine learning is often atheoretical—an “black box,” driven by correlations in the data rather than by any understanding of the causal relationship between the variables and the behavior it seeks to predict. Finding a correlation does not mean that one variable caused the other (i.e., correlation is not causation), but rather that the relationship could be incidental or due to another factor altogether. For example, if there is a correlation between having red hair and being hired, it is highly unlikely that the hair color caused the hiring decision.

- **Large-Sample Artifacts.** The ease with which statistically significant correlations with job performance can be achieved in large data sets challenges the interpretation and stability of such evidence. Statistically significant correlations can easily be found in situations with extremely large sample sizes. It is necessary to do more than show a statistically significant prediction; it is necessary to understand why that correlation exists and what it tells us about a person’s ability to perform the job. Despite the ease with which these data are collected, there is surprisingly little real validation evidence being collected to substantiate the job relatedness of the variables measured by the algorithms.

- **Unstable Algorithms.** Also, when the group on which the algorithm is built is not representative of a wide range of candidates, the model will not have a basis to make an evaluation for the unusual candidate. Hence, the algorithm may produce unstable models which do not predict equally well for different populations (e.g., applicants vs. incumbents, more-diverse applicant pools, applicants with different levels of experience, disabled applicants). As just one example, recent data has shown that some facial recognition algorithms have had difficulty recognizing people with dark skin tone and have falsely assessed Asians’ face shots as being people with their eyes closed\textsuperscript{[12,2]}. Awareness of these limitations and concerns has led data scientists to: 1) develop debiasing algorithms\textsuperscript{[13]}, 2) search for variables that unexpectedly serve as proxies for race or gender\textsuperscript{[12]}, and 3) carefully vet the training data sets and variables that constitute the model. In some cases, a second layer of artificial intelligence is implemented to monitor the model and flag anomalies for further inspection\textsuperscript{[14]}. 
Recent legal reviews regarding the use of artificial intelligence and machine learning in employee hiring, as well as expert testimony provided to government reviewing agencies, highlight legal and professional considerations when implementing AI-driven technology for talent acquisition. This section introduces a four-point checklist designed to help organizations minimize risk and successfully implement the next generation of assessment technologies.

1. **Selection Procedures Must be Job Related**

   The algorithms associated with artificial intelligence, machine learning, and big data are considered selection procedures under the federal *Uniform Guidelines for Employee Selection Procedures*, which govern the defensibility of selection procedures. As with any selection procedure which has adverse impact against protected groups, the selection procedure must be shown to be job related.

   Many organizations unwittingly attempt to demonstrate the job relatedness of their algorithms by simply analyzing and databasing the personal characteristics associated with their best-performing employees. To many users, this process of selecting and weighting variables in an algorithm to best predict and correlate with successful job performance appears to be evidence of criterion-related validity. An example of potential bias associated with approaches like this was recently reported regarding a technology company's multi-year effort to develop a machine learning approach for reviewing job applicants' resumes with the aim of mechanizing the search for top talent. The organization decided to abandon its efforts when it was discovered that the new recruiting process was not rating candidates for software developer and other technical jobs in a gender-neutral way, apparently due to penalizing certain resumes that included such words as “women's,” as in “women's chess club captain.”

   Without a critical analysis of how the characteristics being predicted are required for success on the job (i.e., through a structured job analysis), the validity of these algorithms cannot be supported. The atheoretical basis of many predictive algorithms is also inconsistent with professional standards for validation, which specify the “variables chosen as predictors should have an empirical, logical, or theoretical foundation. The rationale for a choice of predictor(s) should be specified.”

---

**02 Recommendations to Minimize Bias and Implement AI Successfully**

From a user’s perspective, it is essential to understand what content is being measured by the algorithm as well as the rationale for measuring it. Organizations should also plan to monitor the continued effectiveness of the algorithm by cross-validating it on new data to ensure it is predicting accurately for a full range of candidates. As Baer and Kamalnath point out, users have a responsibility to guide the use of the AI tools they implement. “Using a machine-learning model is more like driving a car than riding an elevator. To get from point A to point B, users cannot simply push a button; they must first learn operating procedures, rules of the road, and safety practices.”
Our specific guidance in addressing the job relatedness challenge:

- Conduct a job analysis to ensure the algorithm is measuring the knowledge, skills and abilities related to job performance, rather than reflecting the demographic characteristics of current employees.
- Conduct a content review of all the variables included in the algorithm to identify variables which may be biased on their face.
- Use a debiasing algorithm to discover and minimize the impact of surrogate variables which are proxies for demographics (e.g., race, age, gender) which may have unintentionally crept into the procedure.
- Validate the predictive model’s accuracy over time and with different employee segments and monitor the results of the algorithm on a regular basis to detect potential adverse impact.

(2) All Candidates Applying for the Same Job Must be Assessed by the Same (or Equivalent) Procedure

Machine learning offers a continual revision of scoring algorithms as new data become available for the same job; however, each variation of the algorithm is a separate selection procedure in need of validation. Where the machine-learning algorithm is not fixed, but rather continues to iterate with additional data, validity must be established for each iteration or successive approximation that is used for decision making. This presents a situation where the “test” is not only variable from person to person, but ever-changing ... and in need of constant re-validation.

While one of the key advantages of machine learning is that algorithms may be trained to constantly learn and improve, the issue in a selection context is that different candidate decisions may be made from Time 1 to Time 2 using the same data. If the decision on a candidate oscillates between qualified and unqualified without the addition of any new data, an organization will need to demonstrate how these “oscillations” are justified, particularly if they result in adverse impact. This requires that organizations retain the history of algorithm changes.

Our specific guidance in addressing the algorithm variability challenge:

- Implement a single, fixed machine-learning algorithm for a defined period of time when used on actual candidates.
- Continue to gather additional data and implement a revised algorithm only after re-validation.
- When changes are made to the algorithm, retain the history of algorithm changes.

(3) Sound Selection Decisions Require Complete and Accurate Candidate Data

A related issue is the concern that candidates are evaluated on different criteria when the data on which a decision is based vary from candidate to candidate. This can result from drawing data into the algorithm which have many missing values or when the underlying data are themselves not accurate. The use of data from social-media profiles, Twitter feeds, and even candidate resumes frequently suffers from these problems. Similarly, the use of algorithms trained on limited and non-diverse data sets runs the risk that the application of the screening criteria to diverse candidates will be flawed and inaccurate. The stability of candidate scores is important for sustaining the reliability and validity of any measure and supporting a common interpretation of the results.
Our specific guidance in addressing the accuracy and reliability challenge:

- Examine the accuracy and fairness of the data inputs on which the algorithm is based to ensure that all relevant data are both correct and inclusive.
- Evaluate the representativeness of the populations included in training the algorithm for use with your applicant population.
- Determine how missing data will be handled and whether the pattern of missing data reveals a possible source of systemic bias which should be controlled.
- Consider limiting the predictors to those with similar response rates across candidate groups.
- Leverage the use of psychometric tests rather than data scraping to create a common playing field for candidates.

(4) Build a Diverse and Inclusive Candidate Pipeline

Efforts to establish valid, fair, and job related technology solutions for talent acquisition will not enable the achievement of organizational goals for diversity and inclusion unless there is a robust candidate pipeline that resembles all members of the communities in which organizations do business. This point may seem obvious; however, countless organizations fail to attract diverse candidates and then are left wondering why their D&I efforts have fallen short of stated expectations.

Our specific guidance for addressing the candidate pipeline challenge:

- Identify sources for diverse candidates and actively reach out (e.g., job fairs, community relations, recruiting fairs, social media, etc.) and build relationships that lead to increased diversity candidate interest in the organization.
- Ensure that minimum qualifications are written in a way that don't needlessly eliminate qualified candidates from diverse backgrounds (i.e., ensure they are job related).
- Use fewer "preferred" (nice to have, but not required) qualifications in job postings as research shows women feel they need to meet a much higher percentage of the qualifications listed in job postings compared to men before they will apply for the job.[20]
- Monitor the health of the candidate pipeline on a regular basis and take steps to improve identified deficiencies (e.g., fill gaps involving experiences, orientations, perspectives, etc.).

Some would argue we are experiencing a great renaissance in the development and application of artificial intelligence[21]. This renaissance certainly translates to its application in talent acquisition. This new technology has the potential to help organizations more accurately and efficiently identify candidates that will get up to speed faster, be more productive, and stay longer on the job. While these benefits are impressive, there appears to be growing evidence that the unchecked use of evolving technologies in the talent acquisition space may be exacerbating rather than helping solve our longstanding fight to combat racial and gender bias in organizational hiring and promotional systems. AI and other technologies can sometimes veer off track without organizations even realizing it, but there are clear actions that can be taken to properly evaluate and implement these increasingly popular talent acquisition technologies. Careful attention must be paid to these issues as organizations continue to invest their energy and resources in the pursuit of fair and unbiased selection and promotion decisions.
Endnotes


Headquarters
One Thorndal Circle
2nd Floor
Darien, CT 06820 USA
+1 203.655.7779
info@aptmetrics.com