"Where's the I-O?" Artificial Intelligence and Machine Learning in Talent Management Systems

Manuel F. Gonzalez
Baruch College and The Graduate Center, City University of New York; Aon Assessment Solutions

John F. Capman
Aon Assessment Solutions

Frederick L. Oswald
Rice University

Evan R. Theys
Aon Assessment Solutions

David L. Tomczak
The George Washington University; Aon Assessment Solutions

Follow this and additional works at: https://scholarworks.bgsu.edu/pad

Part of the Artificial Intelligence and Robotics Commons, Human Resources Management Commons, Industrial and Organizational Psychology Commons, and the Other Psychology Commons

How does access to this work benefit you? Let us know!

"WHERE'S THE I-O?" ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN TALENT MANAGEMENT SYSTEMS

Manuel F. Gonzalez[1,2], John F. Capman[2], Frederick L. Oswald[3], Evan R. Theys[2], and David L. Tomczak[2,4]

1. Baruch College and The Graduate Center, City University of New York
2. Aon Assessment Solutions
3. Rice University
4. The George Washington University

ABSTRACT

Artificial intelligence (AI) and machine learning (ML) have seen widespread adoption by organizations seeking to identify and hire high-quality job applicants. Yet the volume, variety, and velocity of professional involvement among I-O psychologists remains relatively limited when it comes to developing and evaluating AI/ML applications for talent assessment and selection. Furthermore, there is a paucity of empirical research that investigates the reliability, validity, and fairness of AI/ML tools in organizational contexts. To stimulate future involvement and research, we share our review and perspective on the current state of AI/ML in talent assessment as well as its benefits and potential pitfalls; and in addressing the issue of fairness, we present experimental evidence regarding the potential for AI/ML to evoke adverse reactions from job applicants during selection procedures. We close by emphasizing increased collaboration among I-O psychologists, computer scientists, legal scholars, and members of other professional disciplines in developing, implementing, and evaluating AI/ML applications in organizational contexts.

KEYWORDS

artificial intelligence, big data, machine learning, personnel selection, talent analytics

In recent years, organizational researchers and practitioners have witnessed a surge of artificial intelligence (AI) and machine learning (ML) products offered by start-up companies and test vendors. These products promise improved personnel assessment and HR efficiency gains through tailored recruiting methods, large-scale applicant tracking, and perhaps most enticingly, faster and more accurate recruitment and hiring. Even a cursory Internet search for "artificial intelligence in recruiting" yields excerpts such as "faster time to interview," "start making recruitment easy," and "goodbye, manual recruiting."

These benefits obviously appeal to organizations. Globalization, low unemployment rates, automation, rapid technological advances in equipment, and other forces have sparked competition for highly skilled job applicants (Chambers, Handfield-Jones, & Michaels, 1998). By not adopting AI/ML, organizations might fear that they will fall behind their competitors. However, "quicker" talent identification is not equivalent to "better" talent identification. How do we know whether, and to what extent, AI/ML tools are identifying and securing talent more effectively than traditional selection methods?

To address this question, organizations that develop AI/ML-based selection tools have begun seeking guidance from I-O psychologists (a fairly recent development, from the authors' experiences). We welcome these partnerships: Greater input and guidance from I-O psychologists—whose expertise lies in the science of work and the workplace—can help prevent or mitigate legal challenges and other negative reactions to AI/ML from the public, media, and governmental stakeholders. In parallel, AI/ML methods are only recently gaining attention within I-O professional training and practice (Putka, Beatty, & Reeder, 2018). To

the extent such training is neglecting technical AI/ML expertise and applications, I-O psychologists may miss out on powerful predictive modeling tools that uniquely or more powerfully address complex organizational problems (Putka et al., 2018). In short, I-O psychologists can inform other disciplines—and themselves—in the AI/ML talent assessment arena.

Thus, our goals in this paper are threefold: (a) review some key potential benefits and limitations of AI/ML talent assessment applications, (b) highlight areas where I-O psychologists can help improve the development and use of AI/ML, and (c) offer some current empirical evidence regarding applicant reactions to AI/ML technologies.

AI/ML Applications in Talent Assessment: The State of the Market

AI/ML tools have gained great traction within the talent acquisition arena. An estimated 33% of organizations have adopted such tools to identify, recruit, and select job applicants with greater speed and efficiency (Stephan, Brown, & Erickson, 2017). Numerous existing and forthcoming technology companies offer various AI/ML-based applications, such as those claiming to remove bias from job descriptions to identify more diverse candidates (e.g., TalentSonar), score asynchronous video interviews that aid with hiring (e.g., HireVue, Montage), leverage neuroscience and the engaging nature of games (e.g., Pymetrics, Knack), and evaluate person–job fit by scraping social media profiles (e.g., Entelo). Of course, these tools demonstrate their value when accompanied by professionally developed and standards-based research evidence that supports their reliability, validity, and fairness. Unfortunately, from our viewpoint, such evidence *generally* seems to be lacking.

To this latter point, I-O psychologists will continue to learn more as a community, as AI/ML tools are more frequently evaluated for their reliability, validity, and fairness across various contexts (i.e., organizational, consulting, legal, and policy making). However, the development and implementation rates of AI/ML applications are already outpacing relevant scientific research and legal guidelines. Companies' proprietary concerns restrict sharing; limited access by researchers to AI/ML tools prevents publications from emerging (Dastin, 2018); and relevant case law that would circumscribe the nature and application of AI/ML tools is scant. Therefore, at least currently, organizations seem to be placing great trust and acceptance in AI/ML talent acquisition tools, despite potential lingering limitations and ethical questions.

Organizational Benefits of AI/ML

Most appealingly, AI/ML applications offer efficiency by helping organizations sift through massive volumes of applicant data and consider larger applicant pools in a shorter time frame (e.g., Das, Pandey, & Rautaray, 2018). Even

if AI/ML algorithms did not produce better employees, they lend potentially greater decision-making speed and efficiency than traditional assessments, saving organizations time and money. These savings may be the primary bottom-line improvement, such that AI/ML tools are actually changing or replacing the work of HR as much or more than their intended effect of changing the quality of hired applicants (Scholz, 2017). Regarding this latter effect, as we noted earlier, any improvements in selection accuracy from AI/ML over traditional methods remain a largely open question, and more evidence seems necessary in this space.

AI/ML methods *could* yield higher validity for organizations by offering access to new sources of data and applying sophisticated algorithms to them (e.g., Sajjadani, Sojourner, Kammeyer-Mueller, & Mykerezi, 2019). AI/ML tools have made various types of data more accessible for organizations to analyze (Bâra, Șimonca, Belciu, & Nedelcu, 2015), such as resumé content (e.g., Indira & Kumar, 2016, online social media activity (e.g., Park et al., 2015; Zang & Ye, 2015), and open-ended applicant responses (e.g., Campion, Campion, Campion, & Reider, 2016). Techniques such as Natural Language Processing (NLP) can collect and analyze these new data sources in a rapid, automated manner. One might argue that NLP, then, *might* improve reliability because, even if the text itself is unstandardized, the NLP analysis is standardized and faster to train than humans (given that the accuracy of human judgments fluctuates unreliably within and across people). Keeping the context and ethics of AI/ML applications in mind, organizations might use such text data to assign scores on psychological variables (e.g., verbal fluency, honesty, emotionality, aggression; Tausczik & Pennebaker, 2010) to predict important employee and organizational outcomes.

Within limited construct domains and settings, researchers have found comparable quality measurement between human raters and algorithmic techniques such as NLP (Campion et al., 2016; Park et al., 2015). For example, social media data have been shown to produce reliable personality measurement, at a level comparable to self-report measures and informant ratings (Park et al., 2015; Youyou, Kosinski, & Stillwell, 2015), although additional convergent and discriminant validity evidence is needed. Therefore, if used judiciously – and with the critical guidance of job analysis information – AI/ML tools can *potentially* yield more effective recruitment and hiring decisions over traditional approaches by measuring job-relevant attributes (e.g., using natural language that describes and predicts leadership potential; Campion et al., 2016).

Potential Limitations of AI/ML

As I-O psychologists collaborate on AI/ML tools with HR specialists, data scientists, organizational decision makers, lawyers, and policy makers, they should (strategically, over the course of their collaboration) soberly balance the

actual and potential strengths and implications of AI/ML with its potential limitations (see Table 1).

Potential limitation #1: Data quality and decision making. To be clear, many of the primary concerns that arise with AI/ML tools and techniques pertain equally to traditional approaches: for example, (a) assessing the appropriateness of the nature and quality of the data captured, and the inferences and decisions derived from them (e.g., Sessions & Valtorta, 2006); and (b) assessing and addressing the potential problems (selection irrelevancies, biases) imbued within AI/ML datasets that impact reliability, validity, and fairness (see Dastin, 2018, for an example from Amazon). AI/ML applications must satisfy the same core psychometric, professional, and legal standards as traditional selection systems, the latter being outlined in Title VII as supported by the Uniform Guidelines (EEOC, 1978) and as addressed in the recently updated *SIOP Principles for the Validation and Use of Personnel Selection Procedures* (5th ed., 2018). In short, companies must provide evidence of minimal risk of adverse impact against members of protected groups (EEOC, 1978) and protect such findings through thorough job analytic and validity-based evidence that supports their employment systems. Ideally, AI/ML tools and algorithms will someday help to fulfill these obligations—not obscure or threaten them.

Data quality and the appropriateness of the algorithm can jointly facilitate (or hinder) the effectiveness of personnel selection decisions. In general, algorithms seek out relationships or regularities in datasets for the purposes of clustering/classification or prediction. AI/ML algorithms can more flexibly adjust to the data to seek robust prediction than traditional modeling approaches, especially with larger quantities of data (e.g., via *k*-fold cross-validation or bootstrapping, or by averaging predictions across trees in a random forest or other model ensembles; see Putka et al., 2018). Specifically, algorithms seek out regularities in the data that are unconstrained by traditional model specifications found in *t*-tests, ANOVA, or linear regression models. To improve prediction (i.e., reduce mean squared error), predictive models are developed within a subset of the data (training sample) and subsequently tested on the remainder of the data that were held out (test sample). When based on sophisticated AI/ML algorithms, cross-validated predictions are *potentially* based on more complex relationships than can possibly be modeled in traditional approaches. AI/ML algorithms can be applied to any data set, yet the aforementioned complexity will generally not be detected when the data set is not sufficiently large enough and/or there is no complex relationship to detect in the first place.

Like more traditional modeling approaches (e.g., regression, ANOVA), AI/ML algorithms risk "garbage in, garbage out" when provided with low-quality data (Redman, 2018). For example, despite their sophistication, AI/ML algorithms often face the same impediment as traditional selection tools/methods: the criterion problem (e.g., Austin & Villanova, 1992; Campbell, McCloy, Oppler, & Sager, 1993). Limited appreciation for measuring the multidimensionality of criteria (e.g., performance, satisfaction) can conspire with having low-quality measures and a limited window of measurement to capture the performance processes of interest. A measure might be conceptually multidimensional, but empirically unidimensional, for instance, when supervisors hurriedly rate employees or have insufficient opportunity to observe certain behaviors, among other reasons (e.g., Murphy, 2008), or because the instrument does not measure the intended constructs reliably or representatively. AI/ML algorithms cannot correct for such real-world problems, just like psychometric corrections to

TABLE 1.
Potential Organizational Benefits and Limitations of Artificial Intelligence/Machine Learning

Potential Benefits of AI/ML	Potential Limitations of AI/ML
Reduced costs in terms of time, effort, money, and human resources	Quality of prediction is bound by the quality of the data obtained and the appropriateness of the algorithm used
Increased power to handle large quantities of data	Algorithms are to some extent a "black box," such that it is often unclear how or why AI/ML arrived at a given prediction
Greater accessibility to previously burdensome forms of data (e.g., large volumes of resumes, social media content, interview responses and other forms of open-ended assessment data)	Potential for future legal issues around data privacy and automated handling of personal data
Potential for increased predictive accuracy for important individual and organizational outcomes of interest	Unfavorable reactions toward AI/ML from job applicants, media, and the general public

Note. AI/ML refers to "artificial intelligence/machine learning".

correlations for measurement error cannot improve actual real-world predictions at the individual level (Oswald, Ercan, McAbee, & Shaw, 2015). Perhaps future AI/ML-enhanced technologies will aid in gathering better criterion data that, in turn, improve the performance of AI/ML-based tests and algorithms. But history could disappointingly repeat itself here should AI/ML technologies and techniques remain more focused on the predictor than on the criterion.

Potential limitation #2: "Black box" predictions. As a second and related limitation, the application of AI/ML algorithms to big data (e.g., random forests, deep learning) can yield "black box" predictive results. Here, "black box" means that even when the algorithm itself is well understood, and even when prediction is impressive, the predictive relationships are too complex to interpret (Adadi & Berrada, 2018). For example, random forests average across hundreds or thousands of different predictive trees that each reflect interactions (Breiman, 2001). In this case, end-users of AI/ML technologies may face challenges in justifying *how* specific predictors and outcomes in big data led to a set of hiring decisions (Tonidandel, King, & Cortina, 2018).

Even when professionals obtain personnel selection data using sound practices—and even when those data, in tandem with AI/ML algorithms, demonstrate robust incremental validity over traditional statistical approaches—it is *still not guaranteed* that the detected patterns of prediction or clustering predominantly reflect important and interpretable constructs (for a similar argument regarding big data and adverse impact, see Jacobs, Murphy, & Silva, 2013). For example, research on educational testing has revealed that automated and human scores can sometimes diverge for certain linguistic subgroups (Bridgeman, Trapani, & Attali, 2012), and both sets of scores can share common sources of contamination, such as essay length (Attali, 2007). Research of this nature is critical for exploring the benefits and limitations of AI/ML applications, and how job applicants can potentially "game" AI/ML talent assessment tools (e.g., receiving higher scores simply by typing longer responses).

Considering this formidable "black box" issue, recent interest in the topic of "explainable AI" has surged. Explainable AI refers to a set of techniques applied to AI/ML results to enhance interpretation (Adadi & Berrada, 2018). For example, regarding our earlier example of random forests, researchers can use techniques such as local interpretable model-agnostic explanations (LIME) to understand which features most strongly drive predictions within specific predictive regions (Ribeiro, Singh, & Guestrin, 2016). However, this research area is relatively young, and challenges remain. Several explainable AI techniques involve decrements to flexibility and accuracy, such that more easily understood models tend to be less flexible

and less predictive (Adadi & Berrada, 2018), which can diminish the original promise of incremental validity of AI/ML tools over traditional statistical methods. Furthermore, explainable AI methods may indicate *which* features drive prediction but not *why* they drive prediction. Traditional job analysis should therefore drive the selection of measures as well as the understanding and interpretation of data.

Importantly, we note how AI/ML algorithms are often applied to very large datasets with many possible predictors (features) that may yield predictive benefit. In these cases, drilling large datasets down to a small handful of predictors for the sake of interpretability may be difficult or even counterproductive. Yet ultimately, the set of predictors used should remain defensible as job relevant. Here, the expertise of I-O psychologists seems critical, then, in partnering with relevant technical disciplines (e.g., computer science, applied statistics) to inform relevant data and the AI/ML tools applied to them. In this context, the traditional emphasis and practice of job analysis by I-O psychologists also remain highly relevant.

Potential limitation #3: Ethical and legal issues. Third, organizations should identify and appropriately address relevant ethical, professional, and legal issues when accessing and analyzing applicant data, such as by obtaining applicants' and employees' consent to collect and use their data and by using such data appropriately. To provide some ethical context: The media has celebrated AI/ML algorithms' successes at solving problems with clear goals and criteria that are almost uniformly valued in their own right: for example, winning at games of chess, Go, or Jeopardy!; or successfully identifying zip codes on envelopes (Markoff, 2011; Schaeffer & van den Herik, 2002; Thórisson, Bieger, Thorarensen, Sigurðardóttir, & Steunebrink, 2016). By contrast, employee selection involves multiple and complex goals and criteria that, as such, are informed by ethical, professional, and legal contexts (e.g., capturing selection-relevant characteristics of jobs and employees, avoiding adverse impact, accommodating disability). Thus, the regularities found in historical employment data by algorithms (AI/ML, regression, or otherwise) should not *necessarily* guide tomorrow's decisions. Sometimes organizations should pursue something very different from or even *opposite* of their empirical findings, such as when an algorithm captures and relies upon irrelevant or legally inappropriate data (e.g., race/ethnicity, gender, age; Tonidandel et al., 2018).

Vendors who develop AI/ML talent assessment products and the companies that use them should strive to maximize benefits while mitigating risks to both job applicants (e.g., data privacy) and the organization (e.g., legal repercussions). Regarding data privacy, the Global Data Protection Regulations (GDPR, 2016) in the European Union (EU) and European Economic Area (EEA) present challenges by

requiring companies to disclose the use of AI/ML to applicants and remain transparent about the data used to make selection decisions (Liem et al., 2018). Similar legislation is emerging in the United States, such as in Illinois, where employers must obtain applicants' consent to use AI/ML in their hiring processes (Bologna, 2019). Furthermore, HIPAA requires AI/ML users to specify which types of data are needed to fulfill a specific objective, given the risk that AI/ML techniques could potentially identify sensitive health information from nonclinical sources such as social media (Weintraub, 2017). Research regarding the implications of such regulations on employers and applicants is needed, as AI/ML and the interpretation of these regulations co-evolve.

Potential limitation #4: Applicant reactions. A final factor that could impede the progress of AI/ML-based talent assessment products is negative reactions, both from job applicants and from the broader population. Recent events like the Cambridge Analytica scandal, in which big data were used to influence national elections (Granville, 2018), and the Chinese government's use of AI/ML to establish a "social credit score system" for its citizens (Cao, 2019), have generated data privacy concerns and high levels of distrust of AI/ML systems (e.g., Meos & Roosipuu, 2018; Rao & Cameron, 2018). It would be unfortunate if litigation and negative perceptions of AI/ML forestalled significant developments that would provide high benefit to both employees and organizations.

For several reasons, applicants may perceive selection procedures and decisions as unfair when such processes are left to AI/ML alone. First, many people are currently unfamiliar with how AI/ML works (Johnson & Verdicchio, 2017), and therefore they may perceive AI/ML-based decision processes as too narrow, impersonal, inaccurate, or opaque. Second, AI/ML can operate more or less automatically on a massive number of applicants using a uniform process, which might deprive applicants of their ability (or perceived ability) to appeal decisions, provide voice, exercise other forms of process control (Leventhal, 1980) or otherwise erode the reality and perceptions of procedural fairness (Thibaut & Walker, 1975). Conversely, with human decision makers such as recruiters or hiring managers, applicants might feel that they have more control in an interview, where they can provide unique information, ask questions, and engage in interpersonal conversation (or impression management; Gilmore & Ferris, 1989). However, human decision makers clearly have their own idiosyncratic and fluctuating biases that may be no better or more manageable by applicants than those found in AI/ML applications (Judge, Cable, & Higgins, 2000). Third, organizational practices signal the organization's values to applicants (e.g., Bangerter, Roulin, & König, 2012), and applicants could perceive the organization, rightly or wrongly,

as not caring about applicants (or human capital in general) by essentially outsourcing the selection process to AI/ML technologies.

Empirical research: Reactions to AI/ML. A subset of the authors (Gonzalez, Capman, Martin, Justenhoven, & Preuss, 2019) conducted research on reactions toward AI/ML in employee selection. We hypothesized that participants in an applicant simulation would generally react negatively during selection procedures automated by current AI/ML tools rather than managed by a human (e.g., recruiter, hiring manager) and that these negative reactions would spill over toward the organizations that adopt these procedures. Furthermore, these differences in applicant reactions would likely be more apparent in hired applicants, given how past applicant reactions research shows that applicants who are not hired display uniformly negative reactions (Bauer, Maertz, Dolen, & Campion, 1998).

The study followed a 2 (decision-maker: human, AI) x 2 (outcome favorability: hired, not hired) between-subjects experimental design. Data were collected from 192 Amazon Mechanical Turk participants (59.4% male; 79.2% Caucasian; age $M = 35.49$ years, $SD = 10.90$; work experience $M = 13.68$ years, $SD = 10.70$). Participants imagined they were applying for a job in which either a hiring manager or an AI/ML tool would make the hiring decision. For the sake of brevity, we note they completed several reaction measures, mostly adapted from existing organizational justice measures (Bauer et al., 2001; Colquitt, 2001). Participants were next told to imagine being either hired or not hired for the job, and reported additional reactions toward the organization. Last, participants reported their knowledge, experience, and general attitudes involving AI/ML. All multi-item measures demonstrated acceptable reliability ($\alpha s \geq .80$).[1]

First, we found that participants generally reacted less favorably to AI/ML decision makers, compared to human decision makers (Table 2)—particularly expressing stronger interpersonal concerns (e.g., dignified treatment, communication) than procedural concerns (e.g., consistency, accuracy). Furthermore, participants trusted the organization less and they were less likely to promote the organization if they either had an AI/ML decision maker or were not hired (Table 3). Participants' open-ended comments revealed concerns regarding the impersonal nature and inaccuracy of AI/ML. The latter concern regarding inaccuracy is striking because meta-analytic evidence supports the validity of selection decisions that use standardized decision rules, versus subjective or holistic judgments that tend to base themselves on more inconsistent and unreliable information (Kuncel, Klieger, Connelly, & Ones, 2013; see also, Highhouse, 1997). Perhaps inaccuracy concerns arose because appli-

---

1 More detailed information about the study procedures and measures can be obtained by contacting Manuel F. Gonzalez.

TABLE 2.

Main Effects of Decision Maker on Procedural Reactions Using Independent Samples t-Tests

Dependent variable	Human decision maker (N = 97) M (SD)	AI/ML decision maker (N = 95) M (SD)	t-value	Cohen's d	p-value
Procedural justice[a]	3.62 (0.79)	3.53 (0.82)	.83	.11	.409
Interactional justice[a]	3.95 (0.84)	3.40 (1.00)	4.13	.60	< .001
Trust in decision factors[a]	3.56 (1.00)	3.15 (1.27)	2.48	.36	.014
Communication concerns[b]	3.22 (1.35)	3.82 (1.18)	3.31	-.47	.001
Privacy concerns[b]	2.39 (1.20)	2.83 (1.22)	2.53	-.36	.012

Note. N = 192. Scores on all dependent variables ranged from 1 to 5. [a] Higher scores reflect more organizationally favorable reactions. [b] Higher scores reflect more organizationally unfavorable reactions.

TABLE 3.

Analysis of Variance Main Effects and Interactions of Decision Maker and Outcome Favorability on Reactions to Selection Decisions

Decision maker	Means (SD) Outcome favorability			Outcome favorability			Decision maker			2-way interaction		
				F	η_p^2	p	F	η_p^2	p	F	η_p^2	p
Organizational distrust												
	Not hired	Hired	Overall[a]									
Human	3.10 (1.07)	1.38 (0.81)	2.25 (1.28)	72.16	.28	< .001	6.68	.03	.011	7.14	.04	.008
AI	3.08 (1.20)	2.19 (1.13)	2.62 (1.25)									
Overall[b]	3.09 (1.13)	1.78 (1.05)	2.45 (1.27)[c]									
Organizational promotion												
	Not hired	Hired	Overall[a]									
Human	2.65 (1.13)	4.08 (0.74)	3.36 (1.19)	58.62	.24	< .001	10.20	.05	.002	4.86	.03	.029
AI	2.51 (1.08)	3.30 (0.99)	2.89 (1.11)									
Overall[b]	2.58 (1.10)	3.70 (0.95)	3.13 (1.17)[c]									

Note. N = 192. Scores on all dependent variables ranged from 1 to 5. For all F-tests, $df_1 = 1$ and $df_2 = 188$. [a] To aid with interpreting outcome favorability main effects, we present the M (SD) when collapsing across the two outcome favorability conditions. [b] To aid with interpreting decision-maker main effects, we present the M (SD) when collapsing across the two decision-maker conditions. [c] We present the grand mean and SD, collapsing across all four experimental conditions.

cants do not consider AI/ML to be standardized in the same way as job knowledge tests or structured application forms.

Second, *hired* participants were significantly more distrustful and less likely to promote the organization if the decision maker was an AI/ML rather than a human, whereas *nonhired* participants exhibited uniformly less positive reactions, regardless of the type of decision maker (Figures 1 and 2). Thus, perhaps the use of AI/ML in employee selec-

tion sets a negative tone even for hired applicants entering the organization. Last, some of these less positive reactions toward AI/ML (i.e., lower trust in the decision criteria, lower likelihood of promoting the organization) were reduced when participants had higher familiarity toward AI/ML (Table 4; Figures 3 and 4). Thus, less trust toward AI/ML may partly stem from unfamiliarity with how AI/ML functions (Johnson & Verdicchio, 2017). Certainly these findings can

FIGURE 1.

Significant two-way interaction between decision-maker and outcome favorability on organizational distrust, $F(1,188) = 7.14$, $p < .01$, $\eta_p^2 = .04$, using two-way analysis of variance (ANOVA). With the exception of the human/rejected versus AI/rejected conditions, all conditions differed significantly from one another, $ps < .001$.
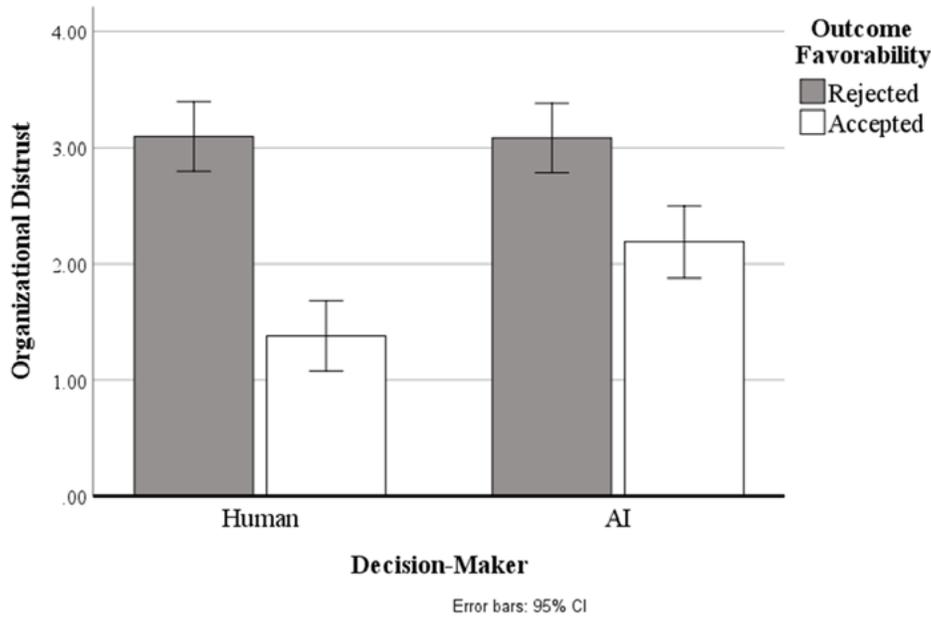

FIGURE 2.

Significant two-way interaction between decision-maker and outcome favorability on organizational promotion, $F(1,188) = 4.86$, $p < .05$, $\eta_p^2 = .03$, using two-way analysis of variance (ANOVA). With the exception of the human/rejected versus AI/rejected conditions, all conditions differed significantly from one another, $ps < .01$.
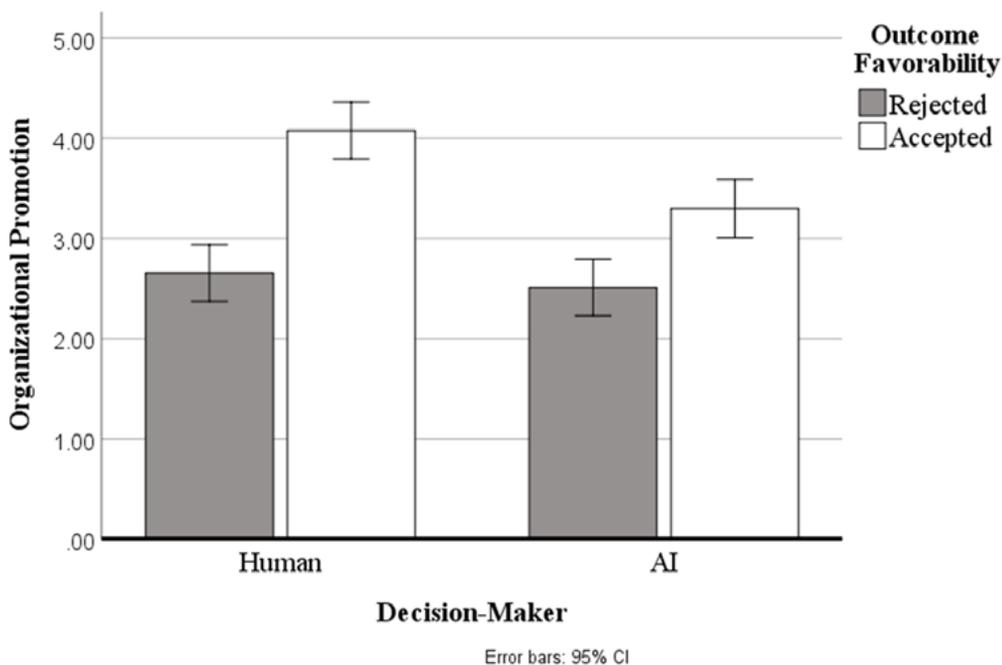
TABLE 4.

Exploratory Ordinary Least Squares Regression Analyses: Moderating Effects of AI/ML Familiarity

Dependent variable: Trust in decision factors			Dependent variable: Organizational promotion		
Predictors[a]	B (SE)	p	Predictors	B (SE)	p
Decision maker (DM)	-1.45 (.50)	.004	Outcome favorability (OF)	3.26 (.65)	< .001
AI/ML familiarity (AF)	.10 (.37)	.465	Decision maker (DM)	.42 (.58)	.473
DM × AF[b]	.38 (.18)	.033	AI/ML familiarity (AF)	.53 (.15)	.001
			OF × DM	-2.43 (.88)	.006
			DM × AF	-.23 (.21)	.262
			OF × AF	-.71 (.24)	.004
			OF × FM × AF[c]	.70 (.32)	.028
R^2	.11	< .001	R^2	.36	< .001

Note. $N = 192$. Results of ordinary least squares (OLS) regression analyses with unstandardized betas presented. [a] Main and interactive effects of outcome favorability were not modeled because trust in decision factors was measured before the outcome favorability manipulation. [b] $\Delta R^2 = .02$, $p < .05$, for the two-way interaction. [c] $\Delta R^2 = .02$, $p < .05$, for the three-way interaction.

FIGURE 3.

Statistically two-way significant interaction between AI/ML familiarity and decision maker on organizational distrust, $B = .38$, $SE = .18$, $p < .05$, using hierarchical linear regression. Based on follow-up simple slopes analyses, slopes for lower and higher AI/ML familiarity are modeled at -1SD below the mean and +1SD above the mean, respectively. Participants with lower AI/ML familiarity reported less trust in the factors used for decision making when there was an AI/ML decision maker, relative to a human decision maker, $B = -.79$, $SE = .23$, $p < .001$. Participants with higher AI/ML familiarity reported similar, higher levels of trust on average, regardless of the decision maker, $B = -.10$, $SE = .27$, $p = .67$. Interactions were plotted and probed using procedures from Aiken and West (1991) and Dawson (2014).
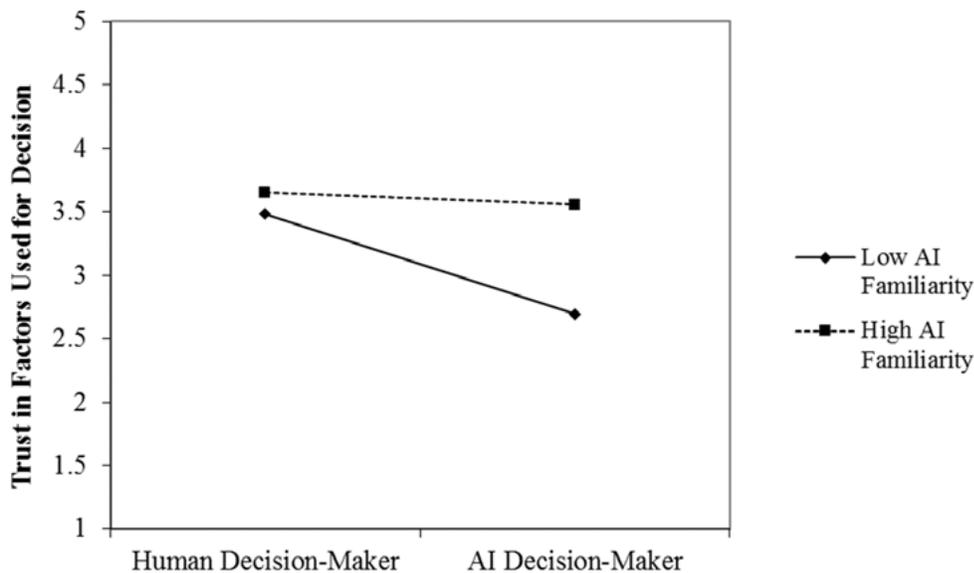
FIGURE 4.

Significant three-way interaction between AI/ML familiarity, outcome favorability, and decision maker on organizational promotion, $B = .70$, $se = .32$, $p < .05$, using hierarchical linear regression analysis. Based on follow-up simple slopes analyses, slopes for lower and higher AI/ML familiarity are modeled at $-1SD$ below the mean and $+1SD$ above the mean, respectively. Among participants with lower AI/ML familiarity, participants with a favorable outcome (i.e., accepted) reported were generally less likely to promote the organization when there was an AI/ML decision maker, relative to a human decision maker, whereas the decision maker did not affect organizational promotion among participants with an unfavorable outcome (i.e., rejected), $B = -1.21$, $se = .40$, $p < .01$. This interaction was not found among participants with higher AI/ML familiarity, $B = .05$, $se = .40$, $p = .91$. Interactions were plotted and probed using procedures from Aiken and West (1991), Dawson (2014), and Dawson and Richter (2006).
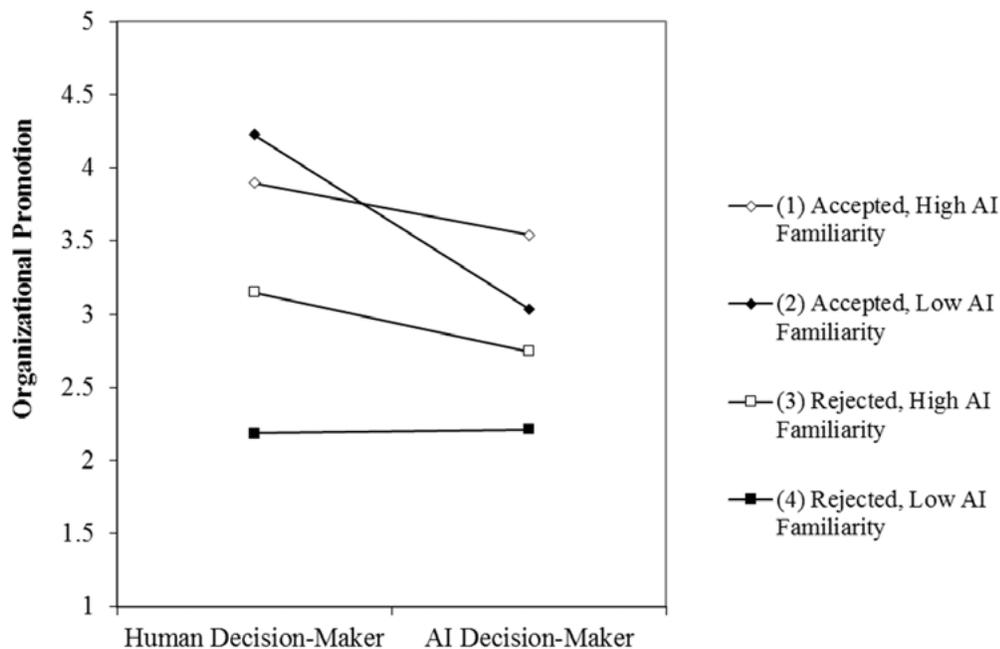
be usefully replicated in actual hiring situations versus this simulation study; yet the findings may still have important implications for organizations that use AI/ML-based selection tools, as we discuss below.

Where Will AI/ML Go Next? A Call for Collaboration

Given the many organizational issues surrounding data privacy, job relevance, legal challenges, and negative applicant reactions, we believe that well-trained I-O psychology researchers and practitioners should play a much stronger role in the development and assessment of AI/ML-based organizational tools. I-O psychologists obviously do not control the speed of the development and implementation of AI/ML; but we strongly suggest that AI/ML companies will obtain a competitive advantage by involving those I-O psychologists who are trained experts in developing and implementing their employment-based technologies in a practical, legal, valid, and fair manner.

Together, I-O psychologists and data scientists can contribute a diverse and relevant range of knowledge, skills, and abilities to improve the quality and viability of AI/ML talent assessment technologies (Handler & Landers, 2018). For example, data scientists and AI/ML experts possess unique training and expertise in programming, data mining/ wrangling, linear algebra, statistical theory, and advanced analytical modeling, whereas I-O psychologists offer proficiency in personnel selection and recruiting practices, measure development, and psychometrics (i.e., ensuring stronger measurement "signals" in the data for AI/ML algorithms to use), and Equal Employment Opportunity (EEO) law.

Regarding our findings on applicant reactions, organi-

zations should consider how to frame AI/ML-based selection procedures when communicating with applicants and the context surrounding the assessment process. Organizations may also consider educating applicants up front on how and why they use AI/ML. Even if the algorithms are not always as transparent as would be ideal, the process by which the algorithms are applied can be made clear to applicants. This approach could alleviate applicants' interpersonal concerns because the organization is investing time to ensure that applicants feel respected and treated fairly (Bies & Moag, 1986). Our participants in the hypothetical hiring scenario were not told why AI/ML technology was used or how it would make selection decisions. It would therefore be theoretically and practically meaningful to investigate whether organizations can tailor their messaging, and the AI/ML tools themselves, to proactively influence applicant reactions to AI/ML. However, findings from Langer, König, and Fitili (2018) suggest that such transparency poses a double-edged sword, in that it can positively affect organizational attractiveness indirectly, via perceived open treatment, but transparency can also have a direct negative effect on organizational attractiveness. Communication strategies and the assessment context are thus potentially important factors to examine, representing fruitful areas for partnership between researchers from I-O psychology and computer science.

The current state of affairs in the development of AI/ML talent management tools reflects a yawning gap between I-O psychologists—who study the science and practice of personnel selection and have learned from decades of organizational, legal, and ethical lessons in the selection and workplace contexts—and computer scientists and applied statisticians—who implement AI/ML talent assessment technologies. Although data scientists and AI/ML experts are inventing and advancing talent assessment technologies, the practical experience and scientific knowledge base of I-O psychologists can critically aid in developing and implementing psychometrically reliable and valid measures. Having psychometrically solid measures ultimately provides high-quality data (not just "big data"), thereby enabling AI/ML algorithms to yield more substantively interpretable results that are then more defensible at the legal, organizational, individual, and ethical levels. Thus, we argue that by forming cross-functional partnerships with data scientists and AI/ML experts, I-O psychologists can contribute invaluably to today's exponential progress, substantially raising the value of AI/ML technologies and ensuring that organizations implement fair, practical, and valid AI/ML solutions.

Conclusion

More broadly, we encourage researchers and practitioners across all relevant scientific disciplines (e.g.,

computer sciences, applied statistics, human factors, I-O psychology) and organizational roles (e.g., human resources, line and staff managers, employees) to strengthen their collaborative communities, working together to continue investigating the vast, interrelated, and interesting applications and implications of AI/ML in talent assessment (e.g., recruiting, promotion, retention, diversity and inclusion). The most urgent task, we believe, requires sharing AI/ML development, implementation, and evaluation experiences and developing practical community-wide recommendations for researchers to implement reliable, valid, and fair AI/ML-based solutions that conform to ethical, professional, and legal guidelines. This will be no small feat, but meaningful progress can be made. Such progress for AI/ML tools will require teamwork, expertise, and difficult discussions of the sort that cannot be conducted or solved by AI/ML tools themselves—not just yet.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138-52160. doi:10.1109/ACCESS.2018.2870052.

Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage Publications.

Attali, Y. (2007). Construct validity of e-rater in scoring TOEFL essays. Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2007.tb02063.x

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1992. Journal of Applied Psychology, 77, 836-874. https://doi.org/10.1037/0021-9010.77.6.836

Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. Journal of Applied Psychology, 97, 719-738. http://dx.doi.org/10.1037/a0026078

Bâra, A., Şimonca, J., Belciu, A., & Nedelcu, B. (2015). Exploring data in human resources big data. Database Systems Journal, 6, 3-9.

Bauer, T. N., Maertz, Jr., C. P., Dolen, M. R., & Campion, M. A. (1998). Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. Journal of Applied Psychology, 83, 892-903. http://dx.doi.org/10.1037/0021-9010.83.6.892

Bauer, T. N., Truxillo, D., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). Personnel Psychology, 54, 387-419. https://doi.org/10.1111/j.1744-6570.2001.tb00097.x

Bies, R. J., & Moag, J. F. (1986). Interactional justice: Communication criteria of fairness. In R. J. Lewicki, B. H. Sheppard, & M. H. Bazerman (Eds.), Research on negotiations in organizations (Vol. 1, pp. 43-55). Greenwich, CT: JAI/ML Press.

Bologna, M. J. (2019, June 3). 'Hiring Robots' Restrictions Passed

by Illinois Legislature. Bloomberg Law News. Retrieved from https://news.bloomberglaw.com/daily-labor-report/hiring-robots-restrictions-passed-by-illinois-legislature

Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324

Bridgeman, B., Trapani, C., & Attali, Y. (2012). Comparison of human and machine scoring of essays: Differences by gender, ethnicity, and country. Applied Measurement in Education, 25, 27-40. https://doi.org/10.1080/08957347.2012.635502

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), Personnel selection in organizations (pp. 35-70). San Francisco, CA: Jossey-Bass.

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, 958-975. https://doi.org/10.1037/apl0000108

Cao, C. (2019, January 25). George Soros Warns About China's Black Mirror-esque 'Social Credit System' at Davos. Observer. Retrieved from https://observer.com/2019/01/george-soros-china-social-credit-system-worries/

Chambers, E. G., Handfield-Jones, H., & Michaels, E. G. (1998). The war for talent. McKinsey Quarterly, 3, 44-57.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. Journal of Applied Psychology, 86, 386-400. doi:10.1037//0021-9010.86.3.386

Das, P., Pandey, M., & Rautaray, S. S. (2018). A CV Parser Model using Entity Extraction Process and Big Data Tools. I.J. Information Technology and Computer Science, 9, 21-31. doi:10.5815/ijitcs.2018.09.03

Dastin, J. (2018, October 9). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. Journal of Business and Psychology, 29, 1-19. https://doi.org/10.1007/s10869-013-9308-7

Dawson, J. F., & Richter, A. W. (2006). Probing three-way interactions in moderated multiple regression: Development and application of a slope difference test. Journal of Applied Psychology, 91, 917-926. http://dx.doi.org/10.1037/0021-9010.91.4.917

Equal Employment Opportunity Commission (EEOC), Civil Service Commission, Department of Labor & Department of Justice. (1978). Uniform guidelines on employee selection procedures. Federal Register, 43, 38290–39315.

General Data Protection Regulation (GDPR) (2016). Regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46. Official Journal of the European Union, 59, 1–88.

Gilmore, D. C., & Ferris, G. R. (1989). The effects of applicant impression management tactics on interviewer judgments. Journal of Management, 4, 557-564. https://doi.org/10.1177/014920638901500405

Gonzalez, M. F., Capman, J. F., Martin, N. R., Justenhoven, R., &

Preuss, A. (2019, April). Rage Against the Machine: Reactions to Artificial Intelligence in Selection Systems. Poster presented at the 34th annual meeting of the Society for Industrial and Organizational Psychology, National Harbor, MD.

Granville, K. (2018, March 19). Facebook and Cambridge Analytica: What you need to know as fallout widens. The New York Times. Retrieved from https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html

Handler, C., (Producer) & Landers, R. N. (Contributor). (2018, December 17). AI Based Assessments: Science or Science Fiction? [Audio Podcast]. Retrieved from http://rocket-hire.com/podcast/ai-based-assessments-science-or-science-fiction/

Highhouse, S. (1997). Understanding and improving job-finalist choice: The relevance of behavioral decision research. Human Resource Management Review, 7, 449-470. https://doi.org/10.1016/S1053-4822(97)90029-2

Indira, D. N. V. S. L. S., & Kumar, R. K. (2016). Profile screening and recommending using natural language processing (NLP) and leverage Hadoop framework for Bigdata. International Journal of Computer Science and Information Security, 14, 799-811.

Jacobs, R., Murphy, K., & Silva, J. (2013). Unintended consequences of EEO enforcement policies: Being big is worse than being bad. Journal of Business and Psychology, 28, 467-471. https://doi.org/10.1007/s10869-012-9268-3

Johnson, D. G., & Verdicchio, M. (2017). Reframing AI discourse. Minds & Machines, 27, 575-590.

Judge, T. A., Cable, D. M., & Higgins, C. A. (2000). The employment interview: A review of recent research and recommendations for future research. Human Resource Management Review, 10, 383-406. https://doi.org/10.1016/S1053-4822(00)00033-4

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. Journal of Applied Psychology, 98, 1060-1072. doi:10.1047/a0034156

Langer, M., König, C. J., & Fitili, A. (2018). Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection. Computers in Human Behavior, 81, 19-30. https://doi.org/10.1016/j.chb.2017.11.036

Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), Social exchange: Advances in theory and research (pp. 27-55). New York, NY: Plenum.

Liem, C. C., Langer, M., Demetriou, A., Hiemstra, A. M., Wicaksana, A. S., Born, M. P., & König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In H. J. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlütürk, U. Güçlü, & M. van Gerven, (Eds.) Explainable and interpretable models in computer vision and machine learning (pp. 197-253). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-98131-4

Markoff, J. (2011, February 16). Computer wins on 'Jeopardy!': Trivial, it's not. The New York Times. Retrieved from https://

www.nytimes.com/2011/02/17/science/17jeopardy-wat-son.html

Meos, P., & Roosipuu, K. (2018). Would you trust Artificial Intelligence with your life? Retrieved September 10, 2018, from https://e-estonia.com/trust-artificial-intelligence/

Murphy, K. R. (2008). Explaining the weak relationship between job performance and performance ratings. Industrial and Organizational Psychology: Perspectives on Science and Practice, 1, 148-158. https://doi.org/10.1111/j.1754-9434.2008.00030.x

Oswald, F. L., Ercan, S. McAbee, S. T., Ock, J., & Shaw, A. (2015). Imperfect corrections or correct imperfections?: Psychometric corrections in meta-analysis. Industrial and Organizational Psychology: Perspectives on Science and Practice, 8, 1-4. https://doi.org/10.1017/iop.2015.17

Park, G., Schwarz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. M., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. Journal of Personality and Social Psychology, 108, 934-952. doi: 10.1037/pspp0000020

Putka, D. J., Beatty, A. S., & Reeder, M. C. (2018). Modern prediction methods: New perspectives on a common problem. Organizational Research Methods, 21, 689-732. doi:10.1177/1094428117697041

Rao, A., & Cameron, E. (2018). The future of artificial intelligence depends on trust. Retrieved September 10, 2018, from https://www.strategy-business.com/article/The-Future-of-Artificial-Intelligence-Depends-on-Trust?gko=af118

Redman, T. C. (2018, April 2). If your data is bad, your machine learning tools are useless. Harvard Business Review. Retrieved from https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144. http://dx.doi.org/10.1145/2939672.2939778

Sajjadani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. Journal of Applied Psychology, 104, 1207-1225. http://dx.doi.org/10.1037/apl0000405

Schaeffer, J., & van den Herik, J. (2002). Chips challenging champions: Games, computers and artificial intelligence. Amsterdam, The Netherlands: Elsevier.

Scholz, T. M. (2017). Big data in organizations and the role of human resource management: A complex systems theory-based conceptualization. New York, NY: Peter Lang.

Sessions, V., & Valtorta, M. (2006). The effects of data quality on machine learning algorithms. ICIQ, 6, 485-498.

Society for Industrial and Organizational Psychology (SIOP). (2018). Principles for the validation and use of personnel selection procedures (5th ed.). Cambridge University Press.

Stephan, M., Brown, D., & Erickson, R. (2017, February 28). Talent acquisition: Enter the cognitive recruiter. [Blog post]. Retrieved from https://www2.deloitte.com/insights/us/en/focus/human-capital-trends/2017/predictive-hiring-talent-acquisition.html

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29, 24-54. https://doi.org/10.1177/0261927X09351676

Thibaut, J., & Walker, L. (1975). Procedural justice: A psychological analysis. Hillsdale, NJ: Erlbaum.

Thórisson, K. R., Bieger, J., Thorarensen, T., Sigurðardóttir, J. S., & Steunebrink, B. R. (2016, July). Why artificial intelligence needs a task theory – And what it might look like. In B. Steunebrink, P. Wang, & B. Goertzel (Eds.) Artificial general intelligence. (pp. 118-128). Cham, Switzerland: Springer.

Tonidandel, S., King, E. B., & Cortina, J. M. (2018). Big data methods: Leveraging modern data analytic techniques to build organizational science. Organizational Research Methods, 21, 525-547. https://doi.org/10.1177/1094428116677299

Weintraub, A. (2017, November 14). Consider HIPAA when using A.I. & machine learning. MedStack. Retrieved from https://medstack.co/blog/consider-hipaa-using-machine-learning/

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. Proceedings of the National Academy of Sciences, 112, 1036-1040. https://doi.org/10.1073/pnas.1418680112

Zang, S., & Ye, M. (2015). Human resource management in the era of big data. Journal of Human Resource Management and Sustainability Studies, 3, 41-45. http://dx.doi.org/10.4236/jhrss.2015.31006