

# Examining Trust in AI-enabled recruitment

Alexander Fowle

PSYCH/HCI 521, Dr. Stephen Gilbert

Iowa State University

4 December 2023

## Introduction

Using algorithmic decision support systems (ADSSs) in the recruitment process is nothing new – 2023 surveys show that around 35-45% of all companies rely on AI to assist with screening candidate resumes or even facilitate automated job interviews (PEG Staffing, 2023). Utilizing such automation in the recruitment process relieves the burden on hiring managers that traditionally needed to manually screen candidates while theoretically mitigating bias in recruitment (Oberst et al., 2020). The concern with traditional human-conducted recruitment is that humans are fallible and prone to acting on their own biases (be those biases implicit or constructed by a skewed perception of the candidate's qualifications or the requirements of the role); thus, the AI, with the right programming, would be able to make the "correct" or objective assessment of a candidate given accurate criteria for the role the candidate has applied to. On paper, this is a noble goal with sound reasoning. In practice, it's not so simple.

Perhaps the most famous example of ADSSs in recruitment gone wrong is Amazon's dismissal of its own proprietary AI hiring software in 2015, which, after careful human analysis of its decisions, was found to "downgrade" female applicants in favor of male applicants (Lewis, 2018). The problem here is linear and evident in retrospect: first, such algorithms are usually trained on historical hiring data; Amazon has historically favored male applicants for tech-related roles (whether that was the intention or simply a result of the tech industry being dominated by males is unclear); the AI erroneously correlates applicant success with maleness; the AI recommends males for open positions; the males are hired and the AI is padded with even more dirty training data, thus completing a closed-circuit feedback loop. The result is an AI that increasingly prefers men over women and automatically deprioritizes resumes that include such terms as "women's club" – if left unchecked, the AI would have theoretically grown to a point where all female applicants would be immediately dismissed as potential candidates for any open role (Lewis, 2018).

This example is one of many that have resulted in a growing interest in researching public perception of and trust in ADSS use in the recruitment process. Trust research has typically been consigned to the domains of organizational trust and interpersonal (sociological, i.e., "human-to-human") relationships. Trust between humans and AI agents has recently become an important subdomain to examine namely because AI is becoming more and more "humanlike" not only in its interfaces and applications, but also because of its increasingly powerful ability to make decisions either entirely autonomously or to supplement human decision-making. Measuring and prescribing trust in automation is a complicated and multifaceted formula consisting of details such as personality traits of the human agent (including one's propensity to trust another in general), prior experiences with and exposure to AI, the AI's previous validity and reliability (i.e., "reputation"), the organizational, cultural, and environmental context, and the human agent's understanding of the AI's capabilities, among many other things (Lee & See, 2004). Looking specifically at ADSS use in recruitment, trust (or lack thereof) in these systems can affect public perception of an organization (Gonzalez et al., 2019), damage the diversity of the workforce, contribute to socioeconomic disparities, and create rifts within human resources teams via the resulting use or disuse of ADSSs by individual recruiters. The following section elucidates a few core issues of ADSS use in recruitment examined in existing literature: recruiter's trust in and perceptions of ADSSs, applicants' trust in and perceptions of ADSSs, and the downstream effects of automation bias and algorithmic aversion regarding recruitment ADSS use or disuse, respectively.

## Literature Review

Existing literature posits that recruiters tend to trust human expert recommendations of viable job candidates over recommendations provided by an ADSS, though it remains largely unclear what specific factors influence this phenomenon (Lacroux & Martin-Lacroux, 2022; Oberst et al., 2020; Kupfer et al.,

2023). Studies have examined the Big 5 personality traits against recruiters' propensity to trust in and accept recommendations provided by an ADSS, but results are mixed. For example, Lacroux & Martin-Lacroux (2022) found in their original research that conscientiousness is positively associated with one's likelihood of trusting a human expert recommendation over that from an ADSS, though the authors acknowledge that this is in direct odds with previous research, thus suggesting a need for further experimentation examining the correlation between personality traits and propensity to trust in an ADSS for recruitment purposes. Others have attempted to surface relationships between recruiters' sex/gender and their trust in the ADSS recommendations, but found no significant correlation (Oberst et al., 2020). Oberst et al. (2020) state in their original research that "technology acceptance and use seem to be more challenging for women." Discovering then that there is no statistically significant difference between male and female recruiters' trust in ADSS recommendations suggests that AI use and acceptance is distinct from general technology use, which sets a strong precedent for examining the issue from a specific lens of AI literacy and acceptance.

Looking at recruiters' expertise and experience in their trade, data support the following statements: (a) Recruiters with experience using ADSSs for candidate screening and selection are less likely to trust the output of the ADSS (Lacroux & Martin-Lacroux, 2022); (b) More experienced recruiters tend to rely on and trust inconsistent (i.e., "objectively" incorrect) ADSS recommendations (Lacroux & Martin-Lacroux, 2022); and (c) Recommendations from human experts, based on accrued intuition, are more trustworthy to other recruiters than ADSS recommendations (Oberst et al., 2020). While statements (a) and (b) may seem to contradict one another, evidence suggests that previous exposure to ADSS tools in recruiting negatively influence a recruiter's propensity to trust its recommendations over time. In other words, recruitment experts with extensive experience using only traditional (i.e., non-AI-based) recruitment methods are more likely to blindly accept the ADSS's recommendation at first (see more on automation bias below), but, with time and continued use of the ADSS, report diminished trust in the technology (Lacroux & Martin-Lacroux, 2022). This again warrants further research into the effect of AI literacy and experience on trust in ADSS use for recruitment, positing that recruiters with some understanding of how the AI works will exhibit more skepticism and distrust in its decision-making capabilities. As Oberst et al. (2020) note, most recruiters don't understand the "black box" nature of AI in general and that "there is still little knowledge among [recruitment] professionals about what a hiring algorithm actually does."

Equally worth examining are the perceptions of automated recruitment from those on the other side of the recruitment process: the applicants. Generally, applicants react negatively to AI use in candidate screening processes (Langer et al., 2021; Gonzalez et al., Noble et al., 2021). Existing research tends to examine applicant perceptions with the goal of providing formative advice to organizations that use ADSSs for recruitment; through these arguments, a few trends emerge. First, applicants value some degree of transparency and explainability in screening procedures that leverage AI (Langer et al., 2021; Gonzalez et al., Noble et al., 2021). Applicants with a better understanding of how AI/ML works report fewer negative reactions than those that do not, further supporting an argument that AI literacy is positively related to trust in AI recruitment applications. Interestingly, Langer et al. (2021) found that there is a "sweet spot" for providing procedural information about the AI-enabled screening process to applicants. Providing too much information on the process can make candidates feel overwhelmed and uneasy about the criteria on which they are being judged; providing too little information might decrease perceived trust in the process if an applicant has had negative past experiences with other forms of automation (Noble et al., 2021). Langer et al. (2021) recommend supplementing limited procedural information with process justification, i.e., why the company has opted to use AI in its screening process, thus providing a proxy for trust in the organization overall. It has been found in other research that even applicants that received a favorable outcome from the AI screening process (i.e., those that received job offers) distrusted organizations that embedded AI into their organizational

practices (Gonzalez et al., 2019) and that applicants that have a high degree of AI literacy still prefer judgement and interaction with an expert human recruiter, making the case for an inert sense of algorithmic aversion among applicants regardless of the outcome favorability (Noble et al., 2021).

Examining applicants' perceptions of and trust in AI-enabled candidate screening from the lens of interpersonal justice and fairness theories provides some interesting insights. In an experiment assessing perceived fairness along 9 verticals of organizational justice, Gonzalez et al. (2019) found that perceived fairness was significantly lower than traditional recruitment in all areas of organizational justice but one: consistency. Put in plainer terms, applicants found virtually all aspects of AI-enabled recruiting to be unfair except for the perception that all candidates are treated exactly the same way and judged along the same measures by the AI. Noble et al. (2021) support this finding, arguing that applicants of AI-based recruitment processes do not have any opportunity to reason with human intelligence, appeal decisions, "exercise process control," or make a case for reconsideration should they receive an unfavorable outcome. Candidates don't have a chance to "speak to the operator," so to speak, and this begets diminished trust in the process and, further, the organization.

One's degree of trust in any automation application results in one of two outcomes: automation use or disuse. Along a similar scale of appropriateness of trust in automation, there are two undesirable effects that one might find themselves adopting: (1) automation bias, which is the blind acceptance of and ADSS's decision even if that decision is faulty, and (2) algorithmic aversion, which is the total disuse of an ADSS despite its potential utility (Lee & See, 2004). Anyone interfacing with any automation should, ideally, find themselves somewhere in the middle: healthily skeptical of its decisions and cautiously confident in its utility. Unfortunately, job applicants' perceptions don't presently make much of a difference: if they distrust the use of an ADSS during the recruitment process, the only real control they may exhibit is to rescind their interest in working for a company that employs ADSS use in recruitment. Thus, the onus falls on organizations to create applications that are trustworthy, asking not "How can we make people trust our systems?" (which puts all the pressure on the applicants to understand and embrace the process) but instead "How can we make trustworthy systems?" I argue that the first step to achieving this goal is to further understand recruiters' trust in and use (or disuse) of ADSS applications in the screening process. Uncovering the reasons why a recruiter chooses to use or disuse—trust or distrust—an ADSS might advise policies and practices that govern the use of ADSSs and contribute to a shared understanding of the technology's limitations and potential utility. The articles referenced in this paper have attempted to make connections between a recruiter's trust in an ADSS and their gender (Oberst et al., 2020), experience in the field (Lacroux & Martin-Lacroux, 2022), and individual personality differences (Lacroux & Martin-Lacroux, 2022), with only little mention of the relationship between a recruiters' AI literacy/prior experiences with ADSSs in recruitment and trust in the technology. In the next section, I propose novel research focused on this relationship.

### **Future research**

To assess how digital literacy might influence trust in ADSS candidate screening systems, I propose a 2 (appropriate ranking vs. inappropriate ranking) x 2 (justification present vs. justification missing) between-subjects experiment with real recruiters ranking resumes against ADSS rank recommendations on an online platform. Participants may be recruited either through an online recruitment platform such as Amazon's Mechanical Turk or via snowballing on social platforms such as LinkedIn and must meet certain criteria such as a minimum amount of experience in recruitment. For each trial, one job description and 5 resumes will be presented to each participant. The job description and resumes, all fictitious but believable, will be the same for each participant. Inspired by Lacroux & Martin-Lacroux's (2022) strategy, I will contract outside help to fine-tune the job description and resumes so that they are believable, but also so that the resumes vary enough on job-relatedness in an effort to create a near-

objective ranking of appropriateness for the position. All resumes will be “scrubbed” of information that might indicate applicant qualities such as age, gender, and race, among other things, to mitigate any potential implicit bias from the participants.

In each of the four cases (which will be randomly assigned to participants that meet the criteria to participate in the study), an “ADSS” ranking of the set of five resumes will be provided, with or without justifications for its decisions. (I have “ADSS” in quotes because, in reality, this will be manipulated by the researchers ahead of time with the help of outside recruitment professionals, which will also serve as the control case.) The job of each participant will be to review the recommendations provided by the ADSS and subsequently rank the resumes on their own based on their perceptions of job-relatedness. Each participant may experience one of four conditions:

1. The ADSS provides an appropriate ranking with justification for its reasoning.
2. The ADSS provides an inappropriate ranking with justification for its reasoning.
3. The ADSS provides an appropriate ranking with no justification for its reasoning.
4. The ADSS provides an inappropriate ranking with no justification for its reasoning.

Before performing the ranking task, participants will be presented with a survey that collects basic demographic information (for normalization and to ensure that personal characteristics such as age and gender will not confound results) and asks each participant to rank their knowledge of, experience with, and comfortability with AI, ML, and ADSSs. Following the ranking task, a post-experiment survey will be administered to measure each participant’s trust in the recommendations provided by the ADSS. This post-task survey will be critical for later data analysis; without it, the experiment only measures whether a recruiter agrees with ADSS recommendations, which does not relate to or measure any degree of trust in the system. To summarize, each trial will follow this design:

1. Demographic and AI literacy survey
2. Trial of one of four conditions
3. Post-task survey on trust in the ADSS recommendations

This experiment differs from other similar experiments not only because it is centered on the effects that AI literacy may have on trust in ADSSs, but also because it focuses on a recruiter’s interaction with an ADSS in isolation. Other similar experiments compare one’s propensity to trust the ADSS recommendation in relation to human expert recommendations; there seems to be enough evidence supporting that recruiters trust human expert recommendations over those from the ADSS. It will be interesting to discover how a recruiter handles ADSS recommendations in lieu of fellow human expertise, which may have practical implications for smaller organizations that do not have large recruitment and/or HR teams.

A major limitation of this experiment—and one that is cited in virtually all articles referenced in the present paper—is that this experiment is lab-controlled and may have limited implications for real-world scenarios. While I agree that situated, real-world experimentation is eventually necessary, I do not believe there is currently enough evidence of factors that inspire trust in recruiters for any situated experimentation to have meaningful results. Recruitment is a complex process with a lot of potential for confounding variables. Without stronger, more evidence-based hypotheses, it would be difficult to make any correlations in a real-world setting.

I hypothesize that recruiters with higher self-proclaimed AI literacy will have higher trust in ADSS systems that provide justification information and lower trust in ADSS systems that do not provide justification information. I believe recruiters that understand how AI works will understand its limitations and treat its recommendations with a degree of skepticism, and that justification information will increase trust in the system by making its decision criteria both explainable and transparent relative to obfuscating its justification. While this is not my immediate goal with the experiment, I believe an

interesting effect that may be observed is how AI literacy might influence automation bias and algorithmic aversion. I hypothesize that recruiters that self-report low AI literacy and high trust in the ADSS will be more influenced by inappropriate recommendations (automation bias) and, conversely, that recruiters that self-report low AI literacy and low trust in the ADSS will reject appropriate ADSS recommendations (algorithmic aversion).

## References

- Lee, J. D., See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80. doi:10.1518/hfes.46.1.50\_30392
- Lacroux, A., Martin-Lacroux, C. (2020). Should I trust the artificial intelligence to recruit? Recruiters' perceptions and behavior when faced with algorithm-based recommendation systems during resume screening. *Frontiers in Psychology*, 13:895997. doi:10.3389/fpsyg.2022.895997
- Oberst, U., De Quintana, M., Del Cerro, S., Chamarro, A. (2020). Recruiters prefer expert recommendations over digital hiring algorithm: A choice-based conjoint study in pre-employment screening scenario. *Management Research Review*, 44(4), 625-641. <https://doi.org/10.1108/MRR-06-2020-0356>
- Kupfer, C., Prassl, R., Fleiss, J., Malin, C., Thalmann, S., Kubicek, B. (2023). Check the box! How to deal with automation bias in AI-based personnel selection. *Frontiers in Psychology*, 14:1118723. doi:10.3389/fpsyg.2023.1118723
- Gonzalez, M., Capman, J., Oswald, F., Theys, E., Tomczak, D. (2019). "Where's the I-O?" Artificial intelligence and machine learning in talent management systems. *Personnel Assessment and Decisions*, 5(3), 33-44. doi:10.25035/pad.2019.03.005
- Noble, S. M., Foster, L. L., Craig, S. B. (2021). The procedural and interpersonal justice of automated application and resume screening. *International Journal of Selection and Assessment*, 29(2), 139-153. doi:10.1111/ijisa.12320
- Langer, M., Baum, K., Koenig, C. J., Haehne, V., Oster, D., Speith, T. (2021). Spare me the details: How the type of information about automated interviews influences applicant reactions. *International Journal of Selection and Assessment*, 29(2), 154-169. doi:10.1111/ijisa.12325
- Lewis, N. (2018, November 12). *Will AI Remove Hiring Bias?* SHRM. <https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/will-ai-remove-hiring-bias-hr-technology.aspx>
- PEG Staffing. (2023, August 4). *How often is AI used in recruiting?* PEG Staffing & Recruiting. <https://www.pegstaff.com/how-often-is-artificial-intelligence-used-in-hiring/#:~:text=A%20research%20study%20done%20by,AI%20for%20the%20hiring%20process>