Applicants' Fairness Perceptions of Algorithm-driven Hiring Procedures[1]

Maude Lavanchy[2], IMD (maude.lavanchy@imd.org; +41 216180705)

Patrick Reichert, elea Center for Social Innovation, IMD (patrick.reichert@imd.org)

Jayanth Narayanan, NUS Business School (jayanth.narayanan@nus.edu.sg)

Krishna Savani, The Hong Kong Polytechnic University (krishna.savani@polyu.edu.hk)

Abstract

Despite the rapid adoption of technology in human resource departments, there is little empirical work that examines the potential challenges of algorithmic decision-making in the recruitment process. In this paper, we take the perspective of job applicants and examine how they perceive the use of algorithms in selection and recruitment. Across four studies on Amazon Mechanical Turk, we show that people in the role of a job applicant perceive algorithm-driven recruitment processes as less fair compared to human only or algorithm-assisted human processes. This effect persists regardless of whether the outcome is favorable to the applicant or not. A potential mechanism underlying algorithm resistance is the belief that algorithms will not be able to recognize their uniqueness as a candidate. Although the use of algorithms has several benefits for organizations such as improved efficiency and bias reduction, our results highlight a potential cost of using them to screen potential employees during recruitment.

Keywords: Algorithms, Organizational Justice, Fairness, Applicant Reactions to Selection, Selection, Recruitment

JEL Codes: O15, J20, L20, M12, M51

Running Head: Fairness perceptions of algorithm-driven hiring procedures

---

[1] Conflict of Interest: The authors declare that they have no conflict of interest.
This article does not contain any studies involving animals or humans performed by any of the authors.
No financial support was received to carry out this research.

[2] Corresponding author

Applicants' Fairness Perceptions of Algorithm-driven Hiring Procedures

Abstract

Despite the rapid adoption of technology in human resource departments, there is little empirical work that examines the potential challenges of algorithmic decision-making in the recruitment process. In this paper, we take the perspective of job applicants and examine how they perceive the use of algorithms in selection and recruitment. Across four studies on Amazon Mechanical Turk, we show that people in the role of a job applicant perceive algorithm-driven recruitment processes as less fair compared to human only or algorithm-assisted human processes. This effect persists regardless of whether the outcome is favorable to the applicant or not. A potential mechanism underlying algorithm resistance is the belief that algorithms will not be able to recognize their uniqueness as a candidate. Although the use of algorithms has several benefits for organizations such as improved efficiency and bias reduction, our results highlight a potential cost of using them to screen potential employees during recruitment.

Keywords: Algorithms, Organizational justice, Fairness, Applicant reactions to selection, Selection, Recruitment
JEL Codes: O15, J20, L20, M12, M51

1 INTRODUCTION

From deciding who gets approved for loans to which ads are shown to an internet user, algorithms have become increasingly ubiquitous in contemporary society (O'Neil, 2016). One domain where the use of algorithms and artificial intelligence (AI) is becoming more prevalent, and controversial, is within human resource management (HRM) (Ball, 2010). Because AI technologies can make recruiting and selection processes faster and more efficient, companies have started to integrate data analytics tools and algorithmic decision-making into various steps of the recruitment process, as well as to track employees' satisfaction and determine employees training and compensation (Leicht‐ Deobald et al., 2019; Hunkenschroer and Lütge, 2022; Kelley, 2022).[1]

However, the increased use of metrics and algorithms to optimize HRM practices also carries substantial risks and ethical implications (Giermindl et al., 2021; Islam and Greenwood, 2022). A central challenge is to ensure that AI-enabled recruitment processes are fair and inclusive.[2] Proponents of AI-enabled HRM claim that it can optimize and objectivize the hiring process, thereby alleviating human bias in hiring decisions (Polli, 2019; IBM, 2019). Indeed, extensive research has documented discrimination against women and ethnic minorities in traditional recruitment processes (e.g., Bertrand and Mullainathan, 2004; Feng et al., 2020; Johnson et al., 2016). Nevertheless, scholars have raised legal and ethical concerns with the use of AI-enabled recruitment, including privacy loss, power asymmetry, algorithmic bias, lack of transparency, obfuscation of accountability, and potential loss of human oversight (Morse et al., 2021; Figueroa‐ Armijos et al., 2022; Hunkenschroer and Lütge, 2022). To bridge the

[1] In a 2022 survey, 66% of recruiters use automated processes in the recruitment process, and more than 70% of millennial applicants suspected that companies used AI in the recruitment process (Stefanowicz, 2022).

[2] By "AI-enabled hiring," we refer to systems in which an algorithm processes information provided by candidates, evaluates candidates based on criteria set by the recruiter, and forwards this evaluation to the recruiter. The algorithm may not necessarily collect additional data about the candidate, nor process the data using machine learning. As a result, "Algorithm-driven", "AI-assisted", and "AI-enabled" are used interchangeably in this article.

gap between these two positions, scholars have called for normative assessments of the ethical status of AI-enabled HRM (Kriebitz and Lütge, 2020; Hunkenschroer and Kriebitz, 2022).

In this paper, we build on the normative foundation of organizational justice (Greenberg, 1987; 1990). Considering that job applicants are on the receiving end of salient organizational HRM decisions, it seems reasonable that job candidates would ask if the selection process (i.e. procedural justice) and outcomes (i.e. distributive justice) were fair (Hausknecht et al., 2004). We empirically examine how people perceive the use of algorithms in recruitment using an experimental design. We focus on fairness perceptions of the recruitment process, in particular on AI-assisted resume screening, as an outcome because it is an important area of ethical concerns with important human rights implications (Greenwood, 2013; Kriebitz and Lütge, 2020) and has long been recognized as a core construct for the effective functioning of organizations (e.g., Colquitt et al., 2001; Alder and Gilbert, 2006; Rupp et al., 2017).

Our findings indicate that those being evaluated perceive AI-enabled recruitment processes as less fair compared to human only or AI-assisted human processes (Study 1). This effect persists regardless of whether the outcome is favorable to them or not (Study 2). In Studies 3 and 4, we find that the perception that algorithms are unable to identify a candidate's uniqueness is a potential mechanism to explain these preferences. Together, these findings extend the applicant reaction model to a digital context and suggest that firms should consider not only how algorithms might implicitly induce bias into the recruitment process, but also how they communicate information about the use of algorithms in the hiring process to prospective applicants (Hunkenschroer and Kriebitz, 2022).

2 THEORETICAL BACKGROUND

Organizational justice and human resource management

We take a normative approach to HRM, suggesting that organizations have a moral responsibility towards all their stakeholders (Greenwood, 2002; Demuijnck, 2009). This moral responsibility, referred to as organizational justice, bestows an obligation upon firms to treat stakeholders with respect, equality, and fairness (Greenberg, 1990; Cropanzano et al. 2007; O'Connor and Crowley-Henry, 2019). With regards to HRM, organizations have a responsibility to provide a fair and just process to all job applicants (Arvey and Renz, 1992; Gilliland, 1993). In this context, fairness can be defined as treating everyone equally or equitably based on people's performance or needs (Leventhal, 1980; Alder and Gilbert, 2006).

The organizational justice literature discusses different types of justice, including procedural justice and distributive justice (Greenberg, 1987). Procedural justice refers to the fairness associated with the processes for resource allocation within the workplace. Distributive justice refers to the fairness associated with the actual allocation of resources in the workplace (Greenberg, 1987, 1990).[3] Organizational justice theory has been applied widely across a variety of areas of research in industrial-organizational psychology (e.g., Colquitt et al., 2001, 2013) and has been particularly foundational for much of the research in the field of job applicant reactions (Ployhart et al., 2017). When applied to applicant reactions research, procedural justice refers to the fairness associated with the hiring procedures and distributive justice refers to the fairness associated with the outcome of the hiring decision.

A great deal of academic attention on applicant reactions research is based on Gilliland's (1993) theoretical model of applicant reactions to employment selection systems. Blending the theoretical nuances of the organizational justice literature on applicant selection with the practical context and implications that applicants and recruiters face in the field, the applicant reaction model focuses on the two key aspects of organizational justice (Greenberg, 1987; Gilliland, 1993). Procedural justice within the model refers to fairness perceptions of the way that individuals are treated. In other words, "Am I being treated fairly?" Distributive justice

[3] A third type of justice, interactional justice, refers to the fairness of the interpersonal treatment employees receive from organizational decision makers, such as their supervisors (Kwon et al., 2008). In the recruiting context, it could mean whether candidate believed that were they treated fairly in their interactions with recruiters. We do not focus on this form of justice in this research as our participants did not interact with any recruiter.

refers to fairness perceptions with respect to the outcome of an organizational decision. In other words, “Am I getting the outcome I deserve?” Put simply, while a fair outcome is important to job applicants, as has been consistently observed in the literature, the use of fair hiring procedures also matters to applicants (Ployhart et al. 2017).

Digital recruiting: From periphery to center stage

Although organizations have a responsibility to provide a fair and just process to all job applicants, they routinely undervalue the ethical perspective of HRM. Empirical research indicates that, when screening candidates, hiring managers are often influenced by factors such as age, gender, sexual orientation, race, obesity, facial attractiveness, which trigger unethical treatment of applicants (Bertrand and Mullainathan, 2004; dos Santos et al. 2017). Since algorithms benefit from superior accuracy in predictions and consistency in judgment (Grove et al. 2000), they could provide fairer evaluations and eliminate human biases (Lee, 2018). Additionally, the use of AI-enabled tools in recruiting further enhances the speed and efficiency of the recruiting process (Black and van Esch, 2020).

Algorithms used in resume screening can be grouped into two categories. A first set focuses on the automation of resume screening. Natural Language Processing (NLP) is used to scan resumes to identify semantic matches between the resume and the job description, which is used to quantify the candidate’s fit. A second set of AI recruiting techniques augment the traditional paper resume screening approach by processing additional information, including candidates’ video resumes and recordings of their behavior in cognitive games. Machine learning algorithms built to process text, image, and voice data are used to infer candidates’ skills and to derive a short-list of the most promising candidates. These techniques raise ethical risks, including concerns about validity, autonomy of applicants and recruiters, non-discrimination, privacy, and transparency (see Hunkenschroer and Lütge, 2022, and Hunkenschroer and Kriebitz, 2022, for a detailed review and implications for organizations).

Whether algorithms lead to fairer decisions remains an open question. Nevertheless, there is mounting evidence that people perceive algorithms' recommendations as less discriminatory than human ones (Jago and Laurin, 2021), even if overcoming biases may be insufficient to ensure a process is perceived as fair (Newman et al., 2020; Kim and Routledge, 2022). As a result, research in computer science and AI ethics has been active in trying to embed notions of fairness into the design of algorithms. Scholars have proposed a variety of metrics measuring fairness with reference to groups of candidates (e.g., applicants from different demographic groups should be treated similarly) as well as to individual candidates (e.g., similar individual should be treated similarly) (Haas, 2019; John-Mathews et al., 2022). Yet, these narrow definitions have been criticized for failing to take the societal context into account (John-Mathews et al., 2022; Fazelpour and Lipton, 2020; Selbst et al., 2019; Morse et al., 2021).

Human rights and ethical AI-enabled HRM

When businesses implement new technologies, they need to assume ethical responsibility for their actions (Donaldson and Dunfee, 1995; Martin and Freeman, 2004; Wettstein, 2015; Martin et al., 2019). This notion is reflected in the UN Guiding Principles on Business and Human Rights, which indicates that organizations are "required to comply with all applicable laws and to respect human rights" (UN, 2011; Hunkenschroer and Kriebitz, 2022). Several human rights and freedoms linked to international labor standards, such as human dignity, occupational choice, equality, privacy, education, and favorable conditions of work are codified in the International Bill of Human Rights while others such as collective bargaining, forced labor, child labor, and nondiscrimination have been addressed by the International Labor Organization's Declaration on Fundamental Principles and Rights at Work (Enderle, 2021).

As a result, human rights serve as a boundary that corporate actions must not cross. Acts that violate these principles, such as discrimination and violation of the dignity of employees or job applicants, are deemed morally reprehensible (Hunkenschroer and Kriebitz, 2022). An obligation to safeguard human rights is relevant to the recruiting process given the divergent interests of companies and job applicants. Companies have a legitimate interest to identify and filter out the best job applicants. They also have a right to information by checking whether an applicant fulfills the job qualifications (Hunkenschroer and Kriebitz, 2022). Companies must therefore balance their quest to gain insights into the qualities of a potential employee while also respecting applicants' rights to privacy. In addition to safeguarding applicants' rights in hiring decisions, the human rights perspective suggests that organizations also have a moral duty in how they treat applicants during the selection process (Alder and Gilbert, 2006; Demuijnck, 2009).

Invasive assessment techniques, such as personality tests and drug testing, particularly when administered without lack of informed consent or under coercion, encroach upon applicants' rights to dignity. For example, the US Employee Polygraph Protection Act forbids private employers from using most lie detector tests, which are considered disrespectful and demeaning (Hunkenschroer and Kriebitz, 2022). Thus, HR managers have a duty to preserve the privacy rights of applicants by protecting their personal information and exercising discretion when conducting background checks (Adler and Gilbert, 2006). By extension, the right to privacy suggests that applicants may choose to withhold information on topics as marriage, pregnancy, or religious affiliation, which could potentially be used for purposes of discrimination (Hunkenschroer and Kriebitz, 2022).

Properties of AI-enabled HRM, such as automated decision-making, use of historic data, and access to private data represent potential challenges with the human rights obligations of firms (Mittelstadt et al., 2016; Pasquale, 2015). Although scholars have recently argued that AI-enabled HRM does not inherently conflict with human rights relevant to the recruitment context (e.g., validity, autonomy, non-discrimination, privacy and transparency), organizations still need to responsibly use algorithms in the hiring process (Hunkenschroer and Kriebitz, 2022). Indeed, scholars have argued that organizations even have a moral duty inform applicants that their application materials will be processed using an algorithm (Adler and Gilbert, 2006).

The intersection of organizational justice, human rights and digital recruiting

Contrary to the common belief that technological choices are free of moral compunctions, algorithmic decisions have moral consequences and are value-laden (Johnson, 2015; Martin, 2019). Getting a job has significant – and potentially long-lasting – consequences for peoples' lives, and unfair processes can directly harm those unfairly treated, for instance by reducing their ability to participate in the economy and society (Taggar and Kuron, 2016).

Our faculty of mind perception, which can be decomposed into our capacity to intend and act (agency) and our capacity to sense and feel (experience), is central to our judgments about morality (Gray et al. 2012). Although algorithms may be seen to have some level of agency, they are perceived to be devoid of experience. As jobseekers are considered to have both high agency and experience, this distinction may lead people to perceive algorithms as lacking a "human mind" (Gray et al. 2007; Bigman and Gray, 2018).

Given that individuals engage in motivated reasoning when formulating justice perceptions, replacing a human-generated process by an automated one could represent a violation of an individual's moral mandate (i.e. a strong moral conviction about how things should be). In response to this violation, individuals tend to look for flaws in the procedure to justify being upset (Skitka 2002; Rupp et al. 2017). Hence, when resume screening is undertaken by an algorithm, those being evaluated are less likely to perceive the process as fair, compared to when resume screening is undertaken by a human:

> **Hypothesis 1:** Perception of procedural fairness is lower when an algorithm oversees the resume screening process compared to when a human oversees the process.

This hypothesis is consistent with prior research showing that humans are perceived as more procedurally fair than automated decision agents (Dineen et al. 2004; Newman et al. 2020) as well as research on the prevalence of algorithm aversion – human distrust of algorithmic output (e.g., Dietvorst et al. 2014; Yeomans et al., 2019).

Firms have well understood concerns over algorithms and regularly put forward the importance of "human involvement" in their marketing campaigns and communications. A common refrain, embodied by a quote from the Managing Director for Randstad Singapore, suggests that "by applying digital innovation, Randstad empowers its consultants to focus on what matters most: bringing in their personal touch and delivering a better human experience" (Randstad, 2019). Prior research on algorithm aversion has also found that people are more likely to accept algorithmic errors when they can exert some control over it, for instance by being able to modify the algorithm (Dietvorst et al. 2016). Human involvement has also been shown to improve people's opinions of algorithm authenticity (Jago, 2019). Nevertheless, research in the context of complex automated decision-making, such as self-driving cars, has also highlighted that humans are often poorly equipped to identify problems and take corrective actions (Elish, 2019). As a result, we expect human involvement to improve perceptions of fairness, but not fully overcome the initial aversion described in hypothesis 1:

> **Hypothesis 2:** Perception of procedural fairness is higher when a human is involved in the resume screening process compared to a fully algorithmic process.

Fairness perceptions are also likely to differ between algorithmic and human resume screening depending on the recruitment outcome. According to the social monitoring system (SMS) model (Pickett and Gardner, 2005), people have a drive to belong with others and exclusion induces them to pay more attention to social information that could imply rejection. Thus, after receiving a positive outcome, job applicants may not pay attention to the fact that the recruitment process was undertaken by a human or an algorithm, as this positive experience of acceptance is enough to satisfy people's need to belong with others (Lucas et al. 2010). As a result, job applicants react in the same manner to both an algorithm and human recruiter (Ho et al. 2018; Reeves and Nass 1996). However, in the case of a negative outcome, the SMS model predicts that people pay more attention to social information related to rejection, making the identity of the recruiter more salient to the candidate.

Therefore, the identity of the recruiter should more strongly affect people's perception of fairness in the case of a negative outcome. Prior studies have also shown that negative outcomes are more likely to turn opinions against algorithms (Bigman and Gray, 2018). Moreover, given that people tend to see positive outcomes as fair and unfavorable outcomes as unfair (Diekmann et al. 1997; Messick and Sentis, 1979), our third hypothesis suggests that:

Hypothesis 3a: Following a positive outcome, people do not perceive fairness of the recruitment process differently between the human and algorithmic recruiter.

Hypothesis 3b: Following a negative outcome, the identity of the recruiter is more salient and affects more negatively the perception of fairness for the algorithmic recruiter.

Finally, we investigate potential mechanisms that could explain the resistance toward the use of algorithms in recruitment. A potential mechanism underlying this aversion relates not only to the impersonal nature of the process, but also to the fact that algorithm-driven processes could possibly screen out good candidates and overlook intangible qualities.

Although new technologies provide sophisticated resume-filtering (see Cowgill (2021) for a description of recent resume-screening algorithms), job applicants may still have the traditional and widely used applicant-tracking systems (ATS) in mind when thinking about resume-screening algorithms. ATS are automated systems programmed to scan for keywords (e.g., number of years of experience, former employers, degree, etc.) to identify candidates that match the job description. Thus, having a non-traditional profile or the 'wrong' keywords could significantly affect the likelihood of selection (e.g., someone applying for a statistician job could be rejected for using the term "numeric modeler" rather than statistician, see Cowgill (2021) and Weber (2012) for additional examples).

Thus, an algorithm's perceived inability to detect less tangible or unique characteristics might be a reason behind the aversion hypothesized in *H1* and *H2*. Given that being valued for uniqueness is a fundamental human need (Brewer, 1991), the perceived inability to detect unique characteristics might also affect people's sense of individuality (Fromkin and Snyder, 1980) and thus their perception of fairness within the recruitment process.

The importance of a process' ability to identify unique characteristics has already been shown to drive resistance to algorithms in the medical context (Longoni et al., 2019). We believe this mechanism could be equally important in the recruiting context since candidates applying to a job need to differentiate themselves from the pool of applicants by showcasing some unique characteristics. The prospect of being selected by an algorithm is more likely to evoke fairness concerns because applicants might believe that their unique characteristics/circumstances will be overlooked because algorithms are trained to recognize patterns in data and often fail to capture exceptions (Hannen, 2020). Our fourth hypothesis therefore suggests that:

> **Hypothesis 4:** The belief that algorithms are unable to identify a candidate's unique characteristics mediates their lower perceptions of fairness.

3 OVERVIEW OF STUDIES

We conducted four studies using scenario experiments to examine how people perceive the use of computer algorithms in a recruitment context. In Study 1, we examine if the use of algorithms affects job applicants' fairness perceptions of the selection process (*H1* and *H2*). Study 2 tests whether these perceptions are affected by the outcome of the recruitment process (i.e. whether or not a job applicant is shortlisted for a position) (*H3a* and *H3b*). Finally, Study 3 investigates the mechanism underlying the aversion to the use of algorithms found in Study

1 and Study 2 (*H4*) while Study 4 explores boundary conditions of our mediating mechanism to better understand the robustness of our results.

All four studies measure participants' perceptions of whether the recruiting procedure is fair and focus on the first stage of recruitment: the resume-screening process. Thus, we do not examine mathematical formulations of fairness (e.g., algorithmic fairness), which is typically narrower and has been criticized for failing to capture the full meaning of fairness in the eyes of lay people, thereby being susceptible to the formalism trap (Selbst et al., 2019). As a result, throughout the manuscript, we rely on the two definitions of fairness: procedural fairness, which is defined as the perceived fairness associated with the hiring procedures, and distributive fairness, which is defined as the perceived fairness associated with the outcome of the hiring decision. Our studies' sample size is in line with other experimental studies (e.g., Giroux et al. 2022), and is sufficient to ensure the detection of a medium effect size (80% power). Additional details on the power analysis are presented in the Web Appendix.

4 STUDY 1: ALGORITHMS AND FAIRNESS PERCEPTIONS

Sample and Procedure

A sample of 409 participants was recruited from Amazon's Mechanical Turk (59.4% female, M_{age} = 35.2). 160 participants failed to pass the set of attention check questions. Our final sample consisted of 249 participants who completed the survey (58.2% female, M_{age} = 35.5). Participants failing the attention check questions were not completely excluded from the analysis and are investigated as a robustness check.

We conducted a between-subjects scenario study where participants are presented with a job offer in a hypothetical company. They were asked to imagine that their profile matched the requirements of the job and that they were interested in applying for this job. We randomly manipulated the information given to participants about the recruitment process. Specifically, participants were either told that the resume screening process will be undertaken by: (i) the hiring manager (*human condition*); (ii) a computer algorithm, that will scan all resumes submitted and automatically select the most relevant applicants to be interviewed, without any intervention from the hiring manager (*AI condition*); (iii) the hiring manager, assisted by a computer algorithm (*AI-assisted human condition*). We then asked participants to rate the fairness of the recruitment process. We did not provide participants with details about how the recruitment algorithm works because companies offering pre-employment assessment systems rarely provide job candidates with details about how their product works (Raghavan et al., 2020). The full scenario texts are presented in the Web Appendix.

Measures

Perceived fairness. Similar to operationalizations of fairness used in Lee (2018), Conlon et al. (2004) and Dineen et al. (2004), we measured perceptions of fairness by asking 'How fair is this recruitment process?' (from 0 = "extremely unfair" to 10 = "extremely fair").

Control variables. We measured the following demographic variables: participant age, gender, race, English proficiency, political orientation, country of residence, citizenship, religiosity, education and social class. We ran the analyses both with and without these controls; the sign and significance level of the coefficients are unchanged and the magnitudes are similar (see columns (1) and (2) in Table 2 for estimation details). In addition to reading and attention check questions regarding the job offer, we further asked participants to indicate how attractive they found the job ad and how likely they were to apply for the job. Unless indicated, the response scale for all measures – except the demographic variables – was an 11-point Likert scale (from 0 = "strongly disagree" to 10 = "strongly agree"). No significant differences were observed between conditions. Table 1 reports details about the measures and control variables.

INSERT TABLE 1 ABOUT HERE

Results

We first conducted a between-subjects ANOVA using perceived fairness as the dependent variable. Results indicated large and significant differences between our three conditions. Participants in the human condition reported higher perceptions of fairness (M = 7.95, SD = 1.85) compared to the AI condition (M = 3.79, SD = 2.77, F[1, 246] = 112.06, p = 0.000) and the AI-assisted human condition (M = 4.97, SD = 3.01, F[1, 246] = 52.69, p = 0.000). Figure 1 provides an illustration of these differences. Similar results were found when those who failed attention check questions were kept in the sample of analysis.

INSERT FIGURE 1 ABOUT HERE

The results support Hypothesis 1 in that participants perceived the use of algorithms in hiring decisions as significantly less fair than the more traditional human process. The aversion is particularly large in magnitude (average perceived fairness decreases from 7.95 to 3.79). Having a human take the final decision, based on a list of applicants determined by an algorithm, significantly increases fairness perceptions (mean difference = 1.185, F[1, 246] = 8.66, p = 0.004), in support of Hypothesis 2. Nevertheless, an AI-assisted recruitment process (AI-assisted human condition) only partially mitigates the aversion to the use of algorithms.

We further examined the relationship between perceived fairness and the use of AI within the recruitment process by conducting regression analysis with perception of fairness as the dependent variable, the identity of the recruiter as the independent variable and the set of demographics as control variables. Adding covariates to the regression function typically improves precision without jeopardizing consistency (Imbens and Wooldridge, 2009). Table 2 reports the regression results (with bootstrapped standard errors). Controlling for demographics marginally decreases the magnitude of the treatment effects. All coefficients of interest remain significant at the 1% level. We also found no statistical differences between our three conditions across variables measuring the attractiveness of the job offer, likelihood to apply for the job and the chances to be shortlisted.

INSERT TABLE 2 ABOUT HERE

Discussion

Study 1 provides preliminary evidence that the use of algorithms in hiring negatively affects candidates' perceptions of fairness on the recruitment process and offers support for our first and second hypotheses (*H1* and *H2*). These results are consistent with previous research suggesting that humans are perceived as more procedurally fair than automated decision agents (Dineen et al. 2004; Newman et al. 2020) and the general preference for a human expert over an algorithm (Promberger and Baron, 2006). The results are also consistent with Dietvorst et al. (2016) since human control over the algorithm significantly improves its acceptance. These results indicate that the intervention of a human in the final decision increases perceptions of fairness (*H2*). Nevertheless, having a human being involved in the resume-screening process did not fully overcome the participants' aversion to algorithms.

5 STUDY 2: ALGORITHMS, FAIRNESS AND OUTCOME FAVORABILITY

Study 1 provided support for our first two hypotheses. In Study 2, we explored whether the aversion to algorithm-driven recruitment processes found in Study 1 was invariant to the outcome of a recruitment process (i.e. whether a job applicant is shortlisted for a position).

Sample and Procedure

A sample of 451 participants were recruited from Amazon's Mechanical Turk (62.7% female; M_{age} = 35.8). 179 participants failed the set of attention checks. Our final sample consisted of

272 participants. As in Study 1, participants who failed the attention checks were used as a robustness check. The final sample was 65.8% female with a mean age of 36.4 years.

We conducted a similar between-subjects scenario study as in Study 1 but extended it by manipulating both the outcome of the recruitment process (shortlisted vs. rejected for an interview) and the information about who took the decision (human agent vs. computer algorithm,) yielding a 2 x 2 factorial design. All participants were shown the same job offer as in Study 1 and were told they had applied and were waiting for an answer from the company. The story surrounding this fictional job application was written in a narrative way to make the participant feel like a candidate receiving the result of an important job application. Participants were then randomly assigned to one of the four conditions: (i) *positive outcome & human condition*, (ii) *negative outcome & human condition,* (iii) *positive outcome & AI condition*, or (iv) *negative outcome & AI condition*. We then asked participants to rate their perception of fairness of the recruitment process. The outcome result was presented as an email from the company's HR. The information about the company's recruitment process (i.e. whether the resume screening process is undertaken by the hiring manager or a computer algorithm) was disclosed in a second step. Full scenario texts are included in the Web Appendix.

Measures

Unless indicated, the response scale for all measures – except the demographic variables – was an 11–point Likert scale (from 0 = "strongly disagree" to 10 = "strongly agree").

Perceived fairness. As in Study 1, we measured perceptions of fairness by asking "How fair is this recruitment process?" (from 0 = "extremely unfair" to 10 = "extremely fair").

Locus of control. The measure of control beliefs was measured using Rotter's (1966) locus of control 13-item scale. For each of the 13 questions, participants had to choose between an item expressing an internal control belief (e.g. "People's misfortunes result from the mistakes they make.") or external control belief (e.g. "Many of the unhappy things in people's lives are partly due to bad luck”). Scores were added to determine each participant’s locus of control. A high score indicates an external locus of control while a low score indicates an internal locus of control. These questions were asked before condition randomization to avoid any potential post-treatment bias (Montgomery et al. 2018).

Exposure to artificial intelligence. We measured familiarity with artificial intelligence by asking ‘What is your level of exposure to artificial intelligence?’ (from 0 = “not at all exposed” to 10 = “extremely exposed”).

Knowledge of artificial intelligence. Knowledge of AI was measured using five multiple choice questions (with four answers for each question, see the Web Appendix for more details). The score of each question was added to determine a final score. A high score indicates a good knowledge of AI.

Control variables. We measured the same set of demographic variables as in Study 1 and further added a check to inquire about the plausibility of the scenario (i.e. “In your opinion, how plausible is this scenario?” (from 0 = “not at all plausible” to 10 = “extremely plausible”)).

Results

Using perception of fairness as the dependent variable, we conducted a 2 x 2 design between-subjects ANOVA where outcomes are either positive or negative and made by one of two types of decision-making processes (an algorithm versus a human). Consistent with Study 1, and irrespective of whether the outcome was positive or negative, participants who were told that a computer algorithm screened resumes rated the recruitment process as significantly less fair compared to human-screened processes, ($F[1, 268] = 24.93$, $p = 0.000$). We also observed a main effect of the recruitment outcome ($F[1, 268] = 10.77$, $p = 0.001$), consistent with the organizational justice literature on outcome favorability (Diekmann et al. 1997; Messick and Sentis 1979) and our third hypothesis (*H3a* and *H3b*). Finally, the interaction between our two manipulations (outcome and recruiter) is significant at the 10% level ($F[1, 286] = 2.80$, p = 0.095). Figure 2 illustrates this interaction.

INSERT FIGURE 2 ABOUT HERE

We further examine the mean levels and statistical differences for the four conditions (see Table 3). Although positive outcomes by a computer algorithm significantly increased perceived fairness, it remains significantly lower than the positive outcome–human decision condition ($M = 7.30$ vs. $M = 5.11$, $p = 0.000$). The only non-significant condition comparison occurs between the negative outcome–human decision and the positive outcome–AI decision. Results are robust to the inclusion/exclusion of participants who failed attention check questions, demographics, locus of control, the level of exposure and knowledge of AI (see Table 2 for estimation details).

INSERT TABLE 3 ABOUT HERE

Discussion

Study 2 replicated the algorithm aversion found in Study 1 and further demonstrated that algorithm aversion persists regardless of whether the outcome is favorable to applicants. We also found a weak interaction effect between outcome and recruiter identity. The presence of an interaction effect between the outcome of the recruitment process and the identity of recruiter in Study 2 provide partial support for our third hypothesis (*H3a* and *H3b*). While a positive outcome alleviates the adverse effect of algorithmic resume screening on fairness perception, it was not enough to completely resolve fairness perceptions of algorithmic selection processes.

Although we attempted to control for participants' level of AI knowledge, our measure might have been too broad; even participants scoring high on the AI knowledge scale might not have a good idea about concerns related to the fairness of AI-enabled recruiting systems. Nevertheless, our main findings remain consistent irrespective of whether we control this variable in the analyses. The results from this estimation are available in the Web Appendix.

6 STUDY 3: ALGORITHMS, FAIRNESS AND UNIQUENESS NEGLECT

Across our first two studies, we found evidence that people have an aversion to the use of algorithms in hiring, regardless of the outcome of the recruitment process. Study 1 also revealed that the intervention of a human in the hiring process was not enough to overcome the negative perception of computer algorithms. In this study, we explored uniqueness neglect as a potential mechanism to explain the aversion to algorithms in the recruitment process.

Sample and Procedure

A sample of 421 participants were recruited from Amazon's Mechanical Turk (56.3% female; M_{age} = 36.9). Of these, 139 participants failed to pass the set of attention checks. Thus, our final sample consisted of 282 participants. As in Study 1 and Study 2, participants failing the attention checks are only investigated as a robustness check. The final sample was 56.0% female with a mean age of 37.6 years. This study used the same scenario as in Study 1, except that we did not include the AI-assisted human *condition*.

Measures

Unless indicated, the response scale for all measures – except the demographic variables – was an 11–point Likert scale (from 0 = "strongly disagree" to 10 = "strongly agree").

Perceived fairness. As in Study 1 and 2, we measured perceptions of fairness by asking "How fair is this recruitment process?" (from 0 = "extremely unfair" to 10 = "extremely fair").

Ability of recruitment process to identify unique characteristics. We measured the ability of the recruitment process to identify unique characteristics of job applicants by asking, "In your opinion, the hiring manager/computer algorithm is able to identify your unique characteristics."

Control variables. We measured the same set of demographic variables as in Study 1 and 2.

Results

Using perception of fairness as the dependent variable, we examined the link between the identity of the decision-maker (i.e. human vs. AI) and the ability to identify candidates' unique characteristics by conducting a between-subjects ANOVA. Consistent with results of Study 1 and Study 2, participants who were told that an algorithm screened the resumes rated the recruitment process as significantly less fair (M = 3.48, SD = 2.53) compared to the human condition (M = 7.24, SD = 1.86, F[1, 281] = 203.14, p = 0.000).

As a first visual inspection, Figure 3 depicts the average perceived fairness for distinct levels of the perceived ability of the recruitment process to identify unique characteristics (values below 4 categorized as coded as "no", values between 4 and 6 coded as "indifference" and values above 6 are coded as "yes"). We observe a positive relationship between the perceived ability to identify unique characteristics and perceived fairness for both the human and AI conditions. To further investigate, we conducted a between-subjects ANOVA using perceived fairness as the dependent variable and by adding an interaction term between the decision-maker identity and the ability of the recruitment process to identify unique characteristics. We found a main effect for the decision maker identity (F[1, 278] = 27.54, p = 0.000), a main effect for the ability to identify unique characteristics (F[1, 278] = 37.26, p = 0.020) and a significant positive interaction (F[1, 278] = 5.51, p = 0.020) (see Figure 4). Results are robust to the inclusion of participants who failed attention check questions. However, when including both those who failed reading checks and attention checks about the job ad, the interaction term loses some statistical significance (F[1, 417] = 2.06, p = 0.152).

INSERT FIGURE 3 AND 4 ABOUT HERE

To determine whether the ability of the human/algorithm hiring manager to identify applicant uniqueness mediates the aversion to algorithms in hiring, we conducted a mediation analysis. We computed a 5,000-iteration bootstrapped mediation model using decision-maker identity (0 = human, 1 = algorithm) as the independent variable, perceived fairness as the dependent variable, ability to identify uniqueness as the mediator and the set of demographic variables as controls (Hayes, 2013). Results indicated that an increase in the perceived ability of the recruitment process to identify uniqueness partially mediated participants' perception of fairness (indirect effect = -1.77, 95% bias-corrected *CI* [-2.42, -1.20], representing 48% of the total treatment effect). Partial mediation is consistent with recent work suggesting algorithm aversion is multi-dimensional (Binns et al. 2018; Grgić-Hlača et al. 2018).

Nevertheless, one should exercise caution when applying causal arguments. Although we controlled for demographic variables, treatment assignment and participants' assessments of the ability of the recruitment process to identify unique characteristics, our measure of interest might be correlated with unobserved variables. As a robustness to this traditional mediation approach, we also computed the treatment effect decomposition proposed by Heckman and Pinto (2015), and applied in Heckman, Pinto and Savelyev (2013). Results indicate a more conservative estimate (indirect effect = -1.18, 90% bias-corrected *CI* [-2.25, -0.16]), accounting for 31% of the total treatment effect (more details available in the Web Appendix). Results are robust to the inclusion of participants who failed attention checks.

Discussion

Study 3 provides support for our fourth hypothesis (*H4*): that the belief that algorithms will not be able to see how unique they are as a candidate significantly contributes to their lower fairness perception. These results are consistent with comments provided by participants in Study 1 and

2 (both studies included an open-ended question asking participants to describe their feelings about the recruitment process, see Table W2 in the Web Appendix for a summary). Two representative comments include the following:

'I feel that a computer program could not choose an employment candidate as well as a human could. I feel that the emotions I put into applying for the position were wasted since computers can not relate to emotions in the same way humans can.'

A participant in the positive outcome & AI condition (Study 2)

'I feel like the recruitment process is unfair. It really takes luck to get your resume to a real person. A computer program might not pick up on something on my resume that would help me land the job.'

A participant in the negative outcome & AI condition (Study 2)

Most of the objections to the use of hiring algorithms espoused by participants relate to the impersonal nature of the process, but also to the fact that AI-enabled processes could possibly screen out good candidates and overlook important qualities, thus neglect their uniqueness.

Overall, these results suggest that organizations adopting computer algorithms to facilitate hiring decisions might benefit by highlighting the capacity of their technology to capture the uniqueness of potential candidates, even though this may not entirely eliminate people's aversion to algorithms, compared to humans.

7 STUDY 4: ALGORITHMS, FAIRNESS AND UNIQUENESS NEGLECT

Following the results of Study 3 and its limitations, this study employs an alternative method to identify mediation. Using the same scenario as Study 1 and Study 3, we employed a 2 x 2 factorial design by manipulating both the information about who took the decision (human agent vs. computer algorithm) and the framing of the recruiter (being able to identify unique characteristics vs. no information).

Sample and Procedure

A sample of 421 participants were recruited from Amazon's Mechanical Turk (59.6% female; M_{age} = 36.6). 151 participants failed to pass the set of attention checks. Our final sample consisted of 270 participants. As in the other studies, participants failing the attention checks were investigated as a robustness check. The final sample was 64.4 % female with a mean age of 38.1 years.

Participants started by answering four questions designed to measure their self-attributed need for uniqueness, based on Lynn and Harris' (1997) SANU four-item scale. All participants were then shown the same job offer as in the previous studies. Participants were then randomly assigned to one of four conditions: (i) *AI condition*, (ii) *AI with detailed description condition*, (iii) *human condition*, or (iv) *human with detailed description condition*. In condition (ii) and (iv), the following text was added after the information about who would undertake the resume screening process:

> *'MangoPick Inc.<u>'s hiring manager has been trained in understanding</u> (uses a technology that is able to understand) the meaning and nuances of a resume, extract relevant information and evaluate the match of a candidate with the job description. <u>The hiring manager takes into account</u> (The program incorporates) information from existing employee's past performance, skills, tenure and experience and applies this knowledge to improve the screening process.'*

where the underlined text is replaced by the text in parentheses for condition (ii). The included information reflects the current capabilities of technologies deployed in human resources (e.g. Mya Systems, BeWorkHappy, Ideal, etc.). The remainder of the study is then like Study 3, where participants are asked to rate the fairness of the recruitment process.

Measures

Unless indicated, the response scale for all measures – except the demographic variables – was an 11–point Likert scale (from 0 = "strongly disagree" to 10 = "strongly agree").

Perceived fairness. As in the previous studies, we measured perceptions of fairness by asking "How fair is this recruitment process?" (from 0 = "extremely unfair" to 10 = "extremely fair").

Ability of recruitment process to identify unique characteristics. As in Study 3, we measured the ability of the recruitment process to identify unique characteristics by asking, "In your opinion, the hiring manager/computer algorithm would be able to identify your unique characteristics" (from 0 = "strongly disagree" to 10 = "strongly agree").

Self-attributed need for uniqueness (SANU). SANU is based on the four-item scale from Lynn and Harris (1997) (see Table 1 for more details about each item). The score of each question (5-point Likert scale) was added to determine the participant's cumulative score. A high score indicates a higher need for uniqueness.

Unique characteristics hurt/help in front of a human/computer algorithm interviewer. We further included two questions on whether participants think their unique characteristics (if any) would hurt or help them in front of a human interviewer and a computer algorithm interviewer (both questions were asked across conditions; from 0 = "hurt you a lot" to 10 = "help you a lot").

Control variables. We measured the same set of demographic variables as in the previous studies (refer to Table 1).

Results

Table 2 columns (5) and (6) report the regression estimations (with bootstrapped standard errors). Consistent with the findings of our previous studies, participants who were told that an algorithm would screen resumes perceived the recruitment process as significantly less fair. Manipulation of the recruiter's framing only slightly changed the results (the coefficient for the interaction between AI decision and information manipulation is β=1.114, p = 0.063). To further illustrate these results, Figure 5 depicts fairness perceptions across the four conditions, distinguishing between respondents who believed the recruitment process was (1) able, (2) indifferent or (3) unable to identify unique characteristics. While fairness perceptions qualitatively increase for participants in the "no detailed description condition" when they believe the recruitment process was able to identify unique characteristics, the increase was significantly larger in the "detailed description condition" for both the human and AI condition. Interestingly, the additional information had a negative effect on participants' perception of the human recruiter, indicating that, in order to identify their unique characteristics, participants expected something different than what was described in this extended paragraph.

INSERT FIGURE 5 ABOUT HERE

Discussion

Although the condition in which AI is framed as being able to identify unique characteristics only partially improve participants' perceptions, the combined results of Study 3 and 4 support the hypothesis that the belief that algorithms will not be able to see how unique they are as a candidate is a significant contributor to algorithm aversion (*H4*). Another interesting result of Study 4 relates to information disclosure. People might have different knowledge and understanding of what a computer algorithm can do, and thus have different reference points when judging it. As a result, disclosing details about a computer algorithm might reduce the discrepancy between people's knowledge. Prior research has shown that negative applicant reactions can be mitigated by providing information (McCarthy et al. 2017b). In this study, even though similar information was shown for both the human and AI conditions, participants reacted positively in the AI scenarios and negatively in the human scenarios.

8 GENERAL DISCUSSION

Across four studies, we found strong evidence that people find algorithms as being a less fair process for making selection decisions during hiring. The belief that algorithms will not be able to identify the unique characteristics of an individual contributes to this aversion.

8.1 Theoretical contributions

Our findings offer a number of theoretical contributions. First, our work extends procedural justice theory by examining new and emerging decision-making tools that are becoming the *de facto* method to screen applicant resumes in organizations. Specifically, we investigated job applicants' reactions to algorithms used to make decisions in the recruitment process. In doing so, we add to existing theory that emphasizes the importance of human involvement, which is perceived as improving the use of accurate information (Newman et al. 2020).

Additionally, our work complements normative theories about whether AI-enabled HRM inherently conflicts with human rights (Hunkenschroer and Kriebitz, 2022). Although it remains unclear whether job applicants feel that AI-enabled HRM violates their human rights related to personal dignity, privacy, and discrimination, our findings suggest that if firms wish to be perceived as fair, they need to increase the transparency of AI-enabled HRM procedures. Our studies indicate that job seekers likely question the validity of AI within the HRM process over distributive and procedural justice concerns. Thus, firms need to allay candidates' concerns about distributive and procedural justice if they are using algorithm-enabled tools in the hiring process, for instance by explaining the ability of the hiring algorithm to assess qualitative factors such as charisma, human values or soft factors like team fit, motivation or truthfulness in application documents (Giermindl et al., 2021; Hunkenschroer and Kriebitz, 2022).

We also found that participants viewed the hiring process as more fair when it involved human recruiters, potentially because human recruiters are better equipped to evaluate qualitative information concerning a candidate's uniqueness. This finding suggests that firms can prioritize involving human recruiters in the hiring process whenever possible. This finding also complements research in consumer behavior on uniqueness neglect, i.e. the concern that algorithms are less able than humans to account for unique characteristics and circumstances

(Longoni et al., 2019). To this end, our work identifies a theoretically and managerially relevant moderator (applicants' concerns about perceived uniqueness) that has been underexplored in the existing literature on algorithmic decision making in human resource management.

8.2 Managerial implications

Our findings have several implications for human resource divisions' choice of using algorithms as a part of the selection and recruitment process. On the one hand, algorithms can make organizations agile by responding to candidates quickly and possibly removing bias from the selection process (Kleinberg et al., 2018; Cowgill, 2021; for a review, see Miller, 2018). However, they also introduce ethical, legal and privacy implications, that human resource managers must consider. As technology advances, algorithms may increasingly be able to discern private attributes (such as family status, political orientation, or whether a candidate is mentally ill) indirectly and without proper consent. As a result, if applicants perceive algorithm-driven recruitment as unfair, or infringing upon their right to privacy, they might be less likely to apply to the firm (Uggerslev et al. 2012). Additionally, research has shown that applicants likely perceive unfair hiring processes as a symptom of a dysfunctional company. As a result, applicants who feel unfairly treated are likely to form negative attitudes toward the company, which can reduce the acceptance of job offers and/or lower commitment to the organization and job satisfaction, discourage others from applying or purchasing the firm's goods or services, or increase the likelihood of a lawsuit (Alder and Gilbert, 2006).

Taking applicant perceptions to AI-enabled recruiting systems into account can help generate a better understanding of "how algorithm-based HR decision-making can be both efficient, and at the same time ethically sound" (Leicht-Deobald et al. 2019, p. 388). Since procedural justice has been shown to be relevant for predicting general job attitudes (e.g. trust, commitment), perceptions of a fair process could mitigate the tendency to react negatively to unfavorable outcomes (Thibaut and Walker, 1975, 1978; Folger and Konovsky, 1989).

Researchers across academic disciplines have suggested alternative solutions to improve algorithm acceptance, such as anthropomorphizing algorithms (Waytz et al. 2014), increasing their seeming authenticity (Jago, 2019), explaining how they work (Yeomans et al., 2019) or providing other types of information (Bigman and Gray, 2018). Although some of these solutions are promising, these findings also highlight the lack of a clear-cut recipe to completely eliminate algorithm aversion, especially in certain contexts such as HRM.

Our findings also introduce several practical contributions. First, they indicate that job applicants are likely to push back on hiring algorithms as unfair. Yet at the same time, there are a number of reasons why companies may wish to use AI-enabled decision procedures. It may be possible for managers to roll out algorithms in a manner that allows for positive responses from employees. Viewed from a Bayesian perspective, data subjects (such as job applicants) should be entitled to both an ex-ante explanation about how the company is going to use their information as well as an ex-post explanation that meaningfully communicates the key factors in the decision process (Kim and Routledge, 2022). For instance, framing techniques could highlight how relying on analytics is objectively more, rather than less, fair than relying on human decision makers. Organizations could also pair algorithms with a human decision maker who assesses qualitative factors.

8.3 Limitations and future directions

We recognize that our studies have certain limitations that offer avenues for future research. First, this study was administered through online panels providing experimental scenarios portraying potential situations. Although the use of hypothetical examples using online panels such as Amazon Mechanical Turk and Prolific have been shown to be an effective method (e.g., Crump et al., 2013; Giroux et al., 2022), future studies should replicate the effects with participants in a more controlled setting via real interactions with AI-enabled HRM technologies. An additional caveat of our empirical analysis is the representativeness of our findings. Given that our sample is relatively more female, white, and educated than the US population (see Table W4 in the Web Appendix for more details), future research could also investigate whether our results hold in population with different demographic as well as for people from specific industries and experience level (e.g., entry-level vs. senior staff).

Second, we acknowledge that the choice to rely on a single-item fairness measure may limit the generalizability of our results. Yet, a recent review of fairness perceptions of algorithmic decision-making indicate that our single-item operationalization of fairness is consistent with past studies (Starke et al. 2021). Scarpello and Campbell (1983) found that a global measure of job satisfaction provides a more accurate assessment than the sum of multiple facet scales. Jordan and Turner (2008) used a similar single-item fairness measure to test the predictive validity of single-item with multiple-item measures and found that the single-item measure was a reliable and valid measures of the construct. Single-item measures also allow respondents to determine which facets are the most important to them, whereas multi-items measures tend to attribute equal weight to each scale items when aggregating values to obtain an overall measure. Although we believe that the use of a single-item measure is appropriate to our research design and for the measurement of perceived fairness, the field would benefit from the development of a validated multi-dimensional scale measuring fairness perceptions of algorithmic decision-making.

Third, we have focused on uniqueness neglect as the mechanism by which algorithms affect the perceived fairness of hiring procedures. However, distrust of algorithms may arise for many reasons. Recent work studying justice perceptions of algorithms has found that there is 'no best approach' in explaining algorithmic decisions (Binns et al., 2018). This could be attributable to the multi-dimensionality of unfairness concerns (Grgić-Hlača et al., 2018). Future research might therefore explore perceptions of algorithmic fairness by drawing upon theoretical frameworks that extend Gilliland's (1993) justice approach, for instance by manipulating how algorithms are explained to potential job candidates. Theoretical perspectives such as expectations theory (Sanchez et al. 2000) and fairness heuristic theory (Lind, 2001) have been suggested as avenues to direct academic attention because applicant reactions may be directly influenced by their beliefs and circumstances, as well as situational factors such as procedural and organizational characteristics (McCarthy et al., 2017a). Future research might also investigate empirically the influence of different information disclosure about fairness, such as whether the algorithm was audited (Wilson et al., 2021) or the algorithmic fairness metric applied (Morse et al., 2021). We believe that these theoretical perspectives can enable the advancement of past work by helping to explain applicant perceptions and reactions during the recruitment process.

Another open question delves deeper into people's tendency to seek algorithmic evaluation when they fear discrimination. An interesting avenue for future research could be the study of the consequences of algorithm's aversion to commitment to the job and job-related outcomes. The issue of discrimination against minorities is also crucial in the recruiting context (Hunkenschroer and Lütge, 2022; Kim and Routledge, 2022). Even though we did not find any effect of age or gender in our studies, if a demographic category ends up being (un-) favored by a specific algorithm, this might also affect how they are generally perceived.

9 CONCLUSION

This research found evidence for the proposition that people have an aversion to the use of algorithms in hiring and selection. We showed that this aversion persists regardless of whether the outcome is favorable to applicants or not. Although several reasons may underlie the aversion to algorithms, we demonstrated that the belief that algorithms will not be able to identify unique characteristics of an individual is a potential mechanism behind this aversion.

Given the proliferation of algorithms in many aspects of our life and in the workplace, this paper offers a word of caution to the blind adoption of these practices by organizations.

REFERENCES

Alder, G. S., & Gilbert, J. (2006). Achieving ethics and fairness in hiring: Going beyond the law. *Journal of Business Ethics*, *68*(4), 449–464.

Arvey, R. D., & Renz, G. L. (1992). Fairness in the selection of employees. *Journal of Business Ethics, 11*(5), 331–340.

Ball, K. (2010). Workplace surveillance: An overview. *Labor History, 51*(1), 87–106.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*(4), 991–1013.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21–34.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. In *Proceedings of the 2018 Chi Conference on human factors in computing systems* (pp. 1–14).

Black, J. S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it? *Business Horizons, 63*(2), 215–226.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin, 17*(5), 475–782.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., & Porter, C. O. L. H. (2001). Justice at the millennium : A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology, 86*(3), 425–445.

Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Conlon, D. E., & Wesson, M. J. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology, 98*(2), 199–236.

Conlon, D. E., Porter, C. O., & Parks, J. M. (2004). The fairness of decision rules. *Journal of Management*, 30(3), 329–349.

Cowgill, B. (2021). Bias and Productivity in Humans and Algorithms: Theory and Evidence from Resume Screening. Working Paper, Columbia Business School.

Cropanzano, R., Bowen, D. E., & Gilliland, S. W. (2007). The management of organizational justice. *Academy of Management Perspectives, 21*(4), 34–48.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one, 8*(3), e57410.

Demuijnck, G. (2009). Non-discrimination in human resources management as a moral obligation. *Journal of Business Ethics, 88*(1), 83-101.

Diekmann, K. A., Samuels, S. M., Ross, L., & Bazerman, M. H. (1997). Self-interest and fairness in problems of resource allocation: Allocators versus recipients. *Journal of Personality and Social Psychology, 72*(5), 1061–1074.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 143*(6), 1–13.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170.

Dineen, B. R., Noe, R. A., & Wang, C. (2004). Perceived fairness of web-based applicant screening procedures: Weighing the rules of justice and the role of individual differences. *Human Resource Management, 43*(2–3), 127–145.

Donaldson, T., & Dunfee, T. W. (1995). Integrative social contracts theory: A communitarian conception of economic ethics. *Economics & Philosophy, 11*(1), 85–112.

dos Santos, N. R., Pais, L., Leitão, C. C., & Passmore, J. (2017). Ethics in recruitment and selection. In H. Goldstein, E. Pulakos, J. Passmore, & C. Semedo (Eds.). *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention* (pp. 91–112). West Sussex: John Wiley & Sons.

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society*, *5*, 40–60.

Enderle, G. (2021). *Corporate Responsibility for Wealth Creation and Human Rights*. Cambridge: Cambridge University Press.

Fazelpour, S., & Lipton, Z. C. (2020, February). Algorithmic fairness from a non-ideal perspective. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 57-63).

Feng, Z., Liu, Y., Wang, Z, & Savani, K. (2020). Let's choose one of each: Using the partition dependence bias to increase diversity in hiring decisions. *Organizational Behavior and Human Decision Processes, 158*, 11–26.

Figueroa-Armijos, M., Clark, B. B., & da Motta Veiga, S. P. (2022). Ethical perceptions of AI in hiring and organizational trust: The role of performance expectancy and social influence. *Journal of Business Ethics*, 1-19.

Folger, R., & Konovsky, M. A. (1989). Effects of procedural and distributive justice on reactions to pay raise decisions. *Academy of Management Journal, 32*(1), 115–130.

Fromkin, H. L., & Snyder, C. R. (1980). The search for uniqueness and valuation of scarcity. In K. Gergen, M. Greenberg, & R. Willis (Eds.) *Social Exchange* (pp. 57–75). Boston: Springer.

Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2021). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, 1–26.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*(4), 694–734.

Giroux, M., Kim, J., Lee, J. C., & Park, J. (2022). Artificial intelligence and declined guilt: Retailing morality comparison between human and AI. *Journal of Business Ethics*, *178*, 1027–1041.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619–619.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*(2), 101–124.

Greenberg, J. (1987). A taxonomy of organizational justice theories. *Academy of Management Review, 12*(1), 9–22.

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management, 16*(2), 399–432.

Greenwood, M. (2002). Ethics and HRM: A review and conceptual analysis. *Journal of Business Ethics, 36*(3), 261–278.

Greenwood, M. (2013). Ethical analyses of HRM: A review and research agenda. *Journal of Business Ethics, 114*(2), 355–366.

Grgic-Hlaca, N., Redmiles, E. M., Gummadi, K. P., & Weller, A. (2018, April). Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In *Proceedings of the 2018 World Wide Web Conference* (pp. 903-912).

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19–30.

Haas, C. (2019). The price of fairness-A framework to explore trade-offs in algorithmic fairness. In *40th International Conference on Information Systems*, ICIS 2019.

Hannen, T. (2020). What went wrong with the A-level algorithm? Financial Times. Retrieved February 8, 2022 from: https://www.ft.com/video/282ecd1f-8402-4bf4-8ee7-3d179ce5fcc2.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*(3), 639–683.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach*. New York: The Guilford Press.

Heckman, J. J., & Pinto, R. (2015). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Reviews, 34*(1-2), 6–31.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review, 103*(6), 2052-86.

Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication, 68*(4), 712–733

Hunkenschroer, A. L., & Lütge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics*, 178, 977–1007.

Hunkenschroer, A. L., & Kriebitz, A. (2022). Is AI recruiting (un) ethical? A human rights perspective on the use of AI for hiring. *AI and Ethics*, 1-15.

IBM. (2019). The role of AI in mitigating bias to enhance diversity and inclusion. Retrieved April 8, 2022 from: https://www.ibm.com/downloads/cas/2DZELQ4O.

Imbens, G. W., & Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5–86.

Islam, G., & Greenwood, M. (2022). The metrics of ethics and the ethics of metrics. *Journal of Business Ethics, 175*(1), 1–5.

Jago, A. S. (2019). Algorithms and authenticity. *Academy of Management Discoveries, 5*(1), 38–56.

Jago, A. S., & Laurin, K. (2021). Assumptions about algorithms' capacity for discrimination. *Personality and Social Psychology Bulletin*, 1–14.

John-Mathews, J. M., Cardon, D., & Balagué, C. (2022). From reality to world. A critical perspective on AI fairness. *Journal of Business Ethics*, *178*, 945–959.

Johnson, D. G. (2015). Technology with No Human Responsibility? *Journal of Business Ethics, 127*(4), 707–715.

Johnson, S. K., Hekman, D. R., & Chan, E. T. (2016). If there's only one woman in your candidate pool, there's statistically no chance she'll be hired. *Harvard Business Review, 26*(4), 1-7.

Jordan, J. S., & Turner, B. A. (2008). The feasibility of single-item measures for organizational justice. *Measurement in Physical Education and Exercise Science*, 12(4), 237-257.

Kelley, S. (2022). Employee perceptions of the effective adoption of AI principles. *Journal of Business Ethics*, 1–23.

Kim, T. W., & Routledge, B. R. (2022). Why a right to an explanation of algorithmic decision-making should exist: A trust-based approach. *Business Ethics Quarterly, 32*(1), 75-102.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *Quarterly Journal of Economics, 133*(1), 237-293.

Kriebitz, A., & Lütge, C. (2020). Artificial intelligence and human rights: A business ethical assessment. *Business and Human Rights Journal, 5*(1), 84-104.

Kwon, S., Kim, M. S., Kang, S. C., & Kim, M. U. (2008). Employee reactions to gainsharing under seniority pay systems: The mediating effect of distributive, procedural, and interactional justice. *Human Resource Management, 47*(4), 757-775.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 1–16.

Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics, 160*(2), 377–392.

Leventhal, G. S. (1980). What should be done with equity theory?. In K. Gergen, M. Greenberg, & R. Willis (Eds.) *Social Exchange* (pp. 27–55). Boston: Springer.

Lind, E. A. (2001). Fairness heuristic theory: Justice judgments as pivotal cognitions in organizational relations. *Advances in Organizational Justice, 56*(8), 88–96.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629–650.

Lucas, G. M., Knowles, M. L., Gardner, W. L., Molden, D. C., & Jefferis, V. E. (2010). Increasing social engagement among lonely individuals: The role of acceptance cues and promotion motivations. *Personality and Social Psychology Bulletin, 36*(10), 1346–1359.

Lynn, M., & Harris, J. (1997). Individual differences in the pursuit of self-uniqueness through consumption. *Journal of Applied Social Psychology, 27*(21), 1861–1883.

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*(4), 835–850.

Martin, K., & Freeman, R. E. (2004). The separation of technology and ethics in business ethics. *Journal of Business Ethics, 53*(4), 353–364.

Martin, K., Shilton, K., & Smith, J. (2019). Business and the ethical implications of technology: Introduction to the symposium. *Journal of Business Ethics, 160*(2), 307–317.

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S.M. (2017a). Applicant perspectives during selection: A review addressing "So what?," "What's new?," and "Where to next?". *Journal of Management, 43*(6), 1693–1725.

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Campion, M. C., Van Iddekinge, C. H., & Campion, M. A. (2017b). Using pre-test explanations to improve test-taker reactions: Testing a set of "wise" interventions. *Organizational Behavior and Human Decision Processes, 141*, 43–56.

Messick, D. M., & Sentis, K. P. (1979). Fairness and preference. *Journal of Experimental Social Psychology, 15*(4), 418–434.

Miller, A. P. (2018, July 26). Want Less-Biased Decisions? Use Algorithms. *Harvard Business Review*. Retrieved December 12, 2018 from: https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 2053951716679679.

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on post-treatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3), 760–775.

Morse, L., Teodorescu, M. H. M., Awwad, Y., & Kane, G. C. (2021). Do the ends justify the means? Variation in the distributive and procedural fairness of machine learning algorithms. *Journal of Business Ethics*, 1-13.

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149–167.

O'Connor, E. P., & Crowley-Henry, M. (2019). Exploring the relationship between exclusive talent management, perceived organizational justice and employee engagement: Bridging the literature. *Journal of Business Ethics, 156*(4), 903-917.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.

Pickett, C. L., & Gardner, W. L. (2005). The social monitoring system: Enhanced sensitivity to social cues as an adaptive response to social exclusion. In K. D. Williams, J. P. Forgas, & W. von Hippel (Eds.), *The Social Outcast: Ostracism, Social Exclusion, Rejection, and Bullying* (pp. 213–226). East Sussex: Psychology Press.

Polli, F. (2019). Using AI to eliminate bias from hiring. *Harvard Business Review*, 29.

Ployhart, R. E., Schmitt, N., & Tippins, N. T. (2017). Solving the Supreme Problem: 100 years of selection and recruitment at the Journal of Applied Psychology. *Journal of Applied Psychology, 102*(3), 291–304.

Promberger, M., & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making, 19*(5), 455–468.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020, January). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 469-481).

Randstad. (2019). Randstad aims to deliver a more human experience with new human forward. Retrieved October 2, 2019 from: https://www.randstad.com.sg/about-us/news/randstad-aims-to-deliver-a-more-human-experience-with-new-human-forward-approach/.

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. Cambridge, UK, 10, 236605.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1–28.

Rupp, D. E., Folger, R., & Skarlicki, D. P. (2017). A critical analysis of the conceptualization and measurement of organizational justice: Is it time for reassessment? *Academy of Management Annals, 11*(2), 919–959.

Sanchez, R. J., Truxillo, D. M., & Bauer, T. N. (2000). Development and examination of an expectancy-based measure of test-taking motivation. *Journal of Applied Psychology, 85*(5), 739–750.

Scarpello, V., & Campbell, J. P. (1983). Job satisfaction: Are all the parts there? *Personnel Psychology*, *36*, 577–600.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 59-68).

Skitka, L. J. (2002). Do the means always justify the ends, or do the ends sometimes justify the means? A value protection model of justice reasoning. *Personality and Social Psychology Bulletin, 28*(5), 588–597.

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2021). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. arXiv preprint arXiv:2103.12016.

Stefanowicz, B. (2022, March 23). *AI Recruitment: The Future of Hiring or HR's Nightmare?* Tidio. https://www.tidio.com/blog/ai-recruitment/

Taggar, S., & Kuron, L. K. J. (2016). The toll of perceived injustice on job search self-efficacy and behavior. *Career Development International, 21*(3), 279–298.

Thibaut, J. W., & Walker, L. (1975). *Procedural Justice: A Psychological Analysis*. New Jersey: Erlbaum.

Thibaut, J. W., & Walker, L. (1978). A theory of procedure. *California Law Review, 66*, 541–566.

Uggerslev, K. L., Fassina, N. E., & Kraichy, D. (2012). Recruiting through the stages: A meta-analytic test of predictors of applicant attraction at different stages of the recruiting process. *Personnel Psychology, 65*(3), 597–660.

United Nations: Guiding principles on business and human rights. https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf (2011). Accessed 17 Aug 2022

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113–117.

Weber, L. (2012, January 24). Your Résumé vs. Oblivion. The Wall Street Journal. Retrieved April 8, 2022 from: https://www.wsj.com/articles/SB10001424052970204624204577178941034941330.

Wettstein, F. (2015). Normativity, ethics, and the UN guiding principles on business and human rights: A critical assessment. *Journal of Human Rights, 14*(2), 162-182.

Wilson, C., Ghosh, A., Jiang, S., Mislove, A., Baker, L., Szary, J., ... & Polli, F. (2021, March). Building and auditing fair algorithms: A case study in candidate screening. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 666-677).

Yeomans, M., Shah, A. K., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making, 32*(4), 403–414.

Table 1: Measured variables across all studies

	Study 1	Study 2	Study 3	Study 4
Outcome:				
- Please describe your feeling about this recruitment process in 2-3 complete sentences. (open-ended)	✓	✓	✓	✓
- How fair is this recruitment process? (0-10)	✓	✓	✓	✓
- PANAS – Negative Affect score (1-5)		✓		
- Please explain why you think you have (not) been shortlisted in 2-3 complete sentences. (open-ended)		✓		
Controls:				
- Demographics (age, gender, race, English proficiency, political orientation, country of residence, citizenship, religiosity, education and social class)	✓	✓	✓	✓
- Attention check questions about the job offer (1-4)	✓	✓*	✓*	✓*
- Reading attention check questions (open-ended)	✓	✓	✓	✓
- How attractive is the job offer? (0-10)	✓			
- How likely are you to apply for this job? (0-10)	✓			
- Given your profile, what are your chances to be short-listed and get an interview? (0-10)	✓			
- Locus of control 13-item (0-1)		✓*		
- What is your level of exposure to Artificial Intelligence? (0-10)		✓		
- 5 multiple choice questions about artificial intelligence knowledge (1-4)		✓		
- In your opinion, how plausible is this scenario? (0-10)		✓		
- In your opinion, the hiring manager/computer would be able to identify your unique characteristics. (0-10)			✓	✓
- In your opinion, the hiring manager/computer would be unable to detect things that make you special. (0-10)			✓	
- In your opinion, your unique characteristics (if any) would hurt/help you in front of a human interviewer. (0-10)				✓
- In your opinion, your unique characteristics (if any) would hurt/help you in front of a computer algorithm interviewer. (0-10)				✓
- SANU 4-item scale (1-5): I prefer being __ different from other people; Being distinctive is __ important to me; I __ intentionally do things to make myself different from those around me; I have a __ need for uniqueness.				✓*

Notes. ✓*= the measure was collected before the random assignment of condition. All other measures were collected after the assignment of participant to the condition. For Study 3b, the question "*In your opinion, the hiring manager/computer would be able to identify your unique characteristics*" was slightly rephrased and read "*In your opinion, the recruitment process would be able to identify your unique characteristics.*"

Table 2. Study 1, 2 and 4 regression tables

	Study 1		Study 2		Study 4	
DV:	Fairness	Fairness	Fairness	Fairness	Fairness	Fairness
	(1)	(2)	(3)	(4)	(5)	(6)
AI decision	-4.163***	-4.098***	-2.186***	-1.955***	-3.928***	-2.811***
	(0.357)	(0.384)	(0.444)	(0.459)	(0.340)	(0.482)
AI-assisted human decision	-2.978***	-3.008***				
	(0.396)	(0.396)				
Negative outcome			-1.470***	-1.268***		
			(0.462)	(0.449)		
AI x Negative			-1.044*	-1.170*		
			(0.637)	(0.633)		
Detailed description					-0.672**	-2.122**
					(0.320)	(0.875)
AI x detailed desc.					1.136**	1.114*
					(0.539)	(0.594)
Locus of control				-0.087		
				(0.061)		
AI exposure				0.116		
				(0.076)		
AI knowledge				-0.591***		
				(0.190)		
Identify uniqueness						0.295***
						(0.098)
Uniqueness x Info						0.293**
						(0.118)
SANU						0.002
						(0.043)
Constant	7.952***	11.06***	7.297***	5.723**	8.262***	5.665***
	(0.203)	(2.187)	(0.286)	(2.060)	(0.169)	(1.719)
Demographic controls		✓		✓		
Observations	249	246	272	246	270	266
Adj. R-squared	0.317	0.333	0.325	0.380	0.361	0.55
Chi-squared statistic	154.67	198.54	147.19	261.15	184.34	390.62
# bootstrap iterations	5000	4993	5000	5000	1000	995

Notes. Bootstrapped standard errors shown in parentheses. Stars indicating: * p<0.1, ** p<0.05 and *** p<0.01. Human decision is the reference dummy variable. Control variables included: gender dummy, non-white dummy, class dummies, age, age squared, education dummy variables, political orientations dummy variables (left/right; democrat/republican, liberal/conservative), and degree of religiosity. Participants failing attention check questions have been excluded from the analysis. In Study 2, attention check questions were asked before condition randomization to avoid any potential post-treatment bias related to the exclusion of participants who failed them. Participants failing the attention check might not be a random subset of the population and conditioning on a post-treatment measure can imbalance the sample with respect to observed or unobserved confounders (Montgomery et al. 2018).

Table 3. Study 2: Results of post-hoc subgroup analysis

DV: Perceived fairness of recruitment process	*Mean*	*SD*	*Post-hoc comparisons*
1. Positive outcome and human decision (N=74)	7.30	2.46	1>2***, 1>3***, 1>4***
2. Negative outcome and human decision (N=58)	5.83	2.73	2<1***, 2>4***
3. Positive outcome and AI decision (N=63)	5.11	2.68	3<1***, 3>4***
4. Negative outcome and AI decision (N=77)	2.60	2.40	4<1***, 4<2***, 4<3***

Notes. Participants failing attention check questions have been excluded, but similar results are found when keeping them. Post-hoc comparisons were performed using Tukey's HSD tests.

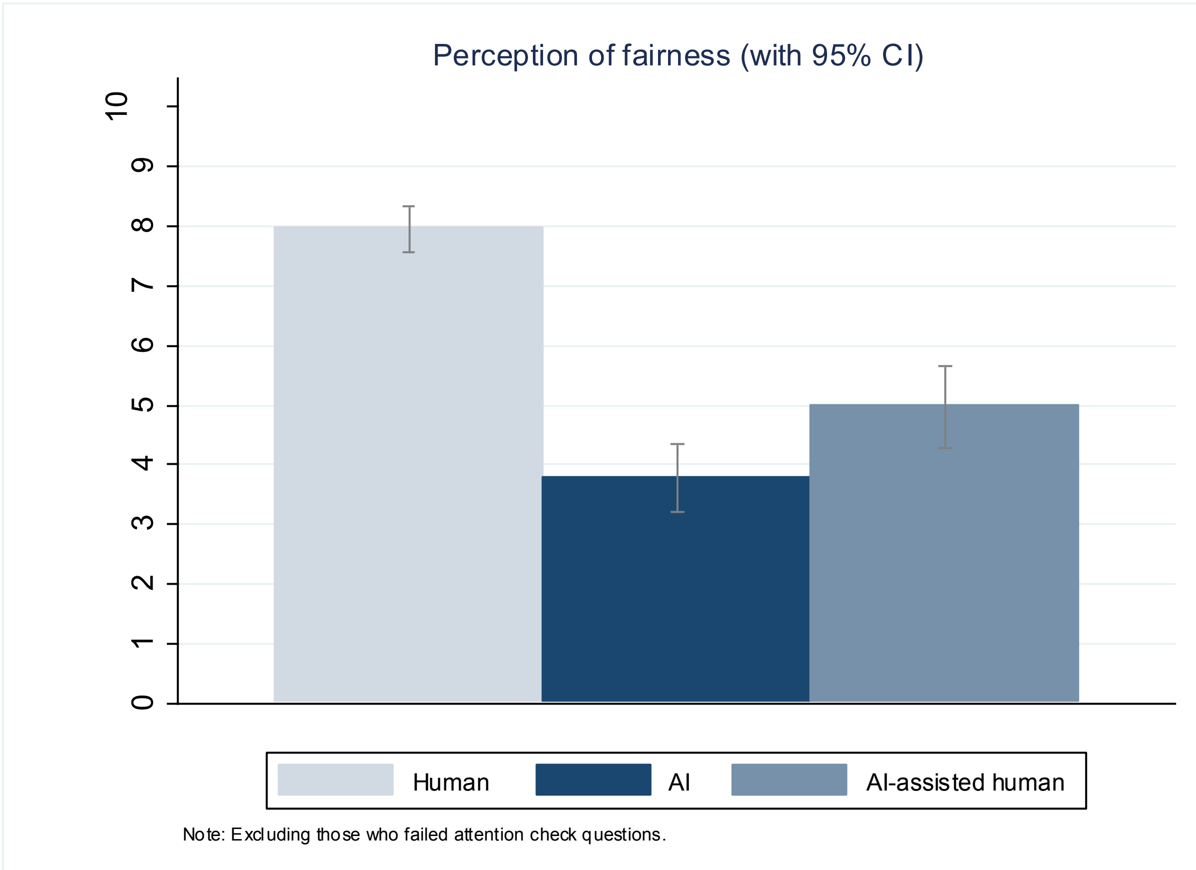

Figure 1. Study 1: Perception of fairness mean score

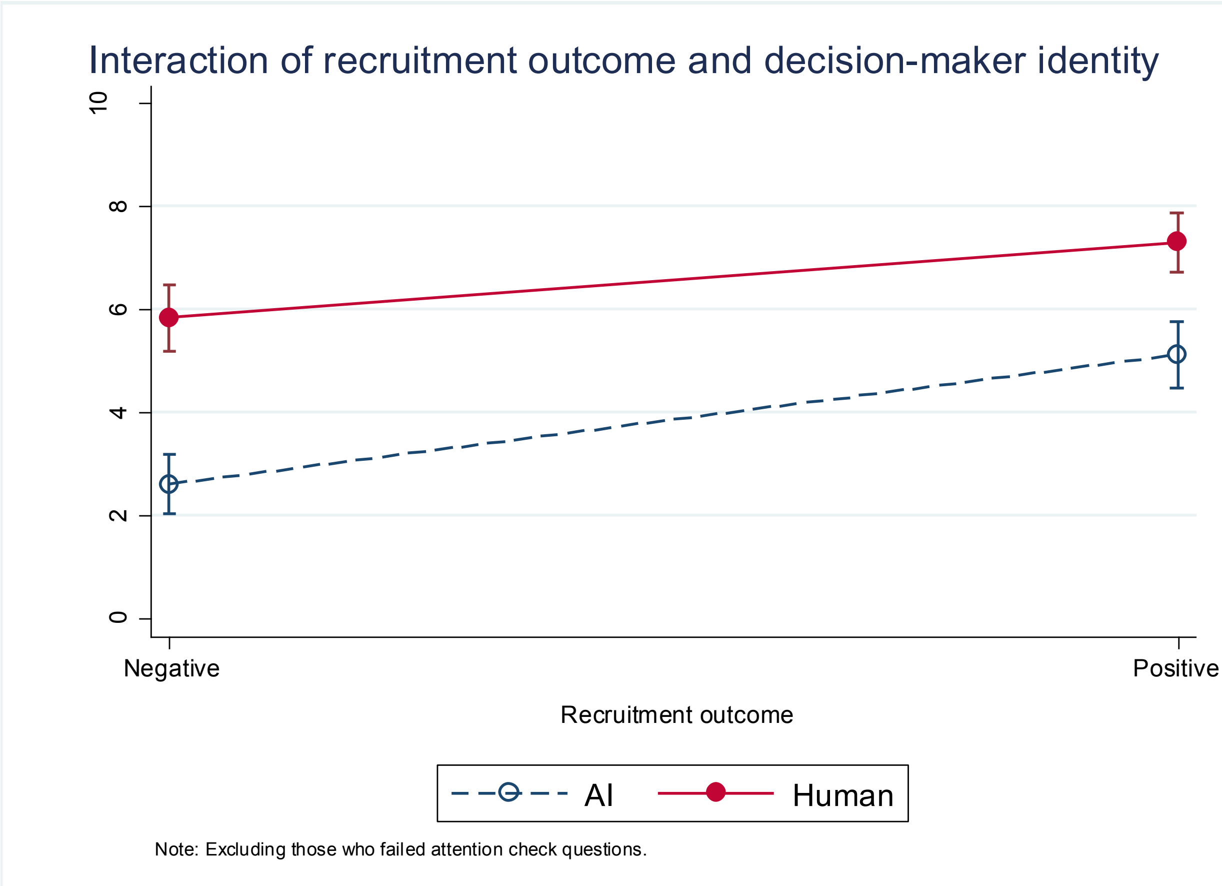

Figure 2. Study 2: Effects of outcome and recruiter identity on perceived fairness

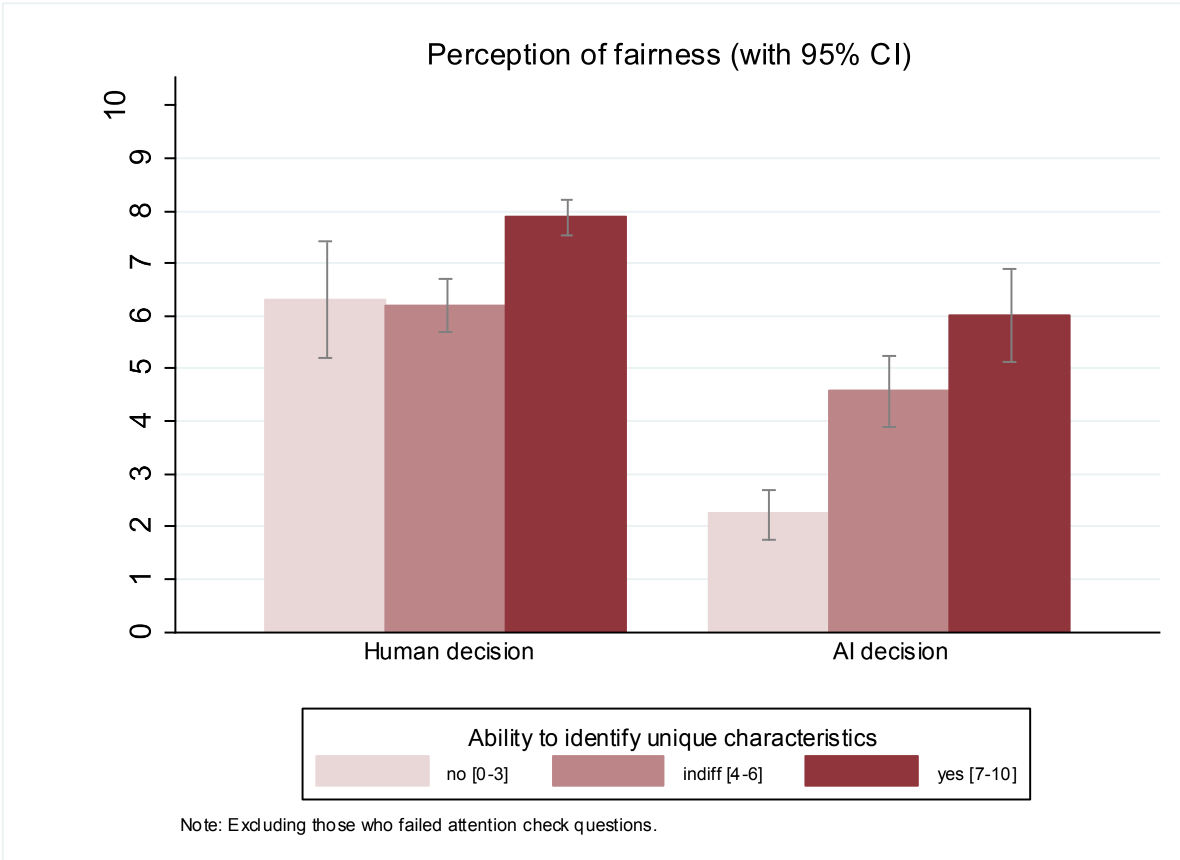

Figure 3. Study 3: Perception of fairness mean score, split in three categories depending on respondents' answer to the question "In your opinion, the hiring manager/computer algorithm would be able to identify your unique characteristics."

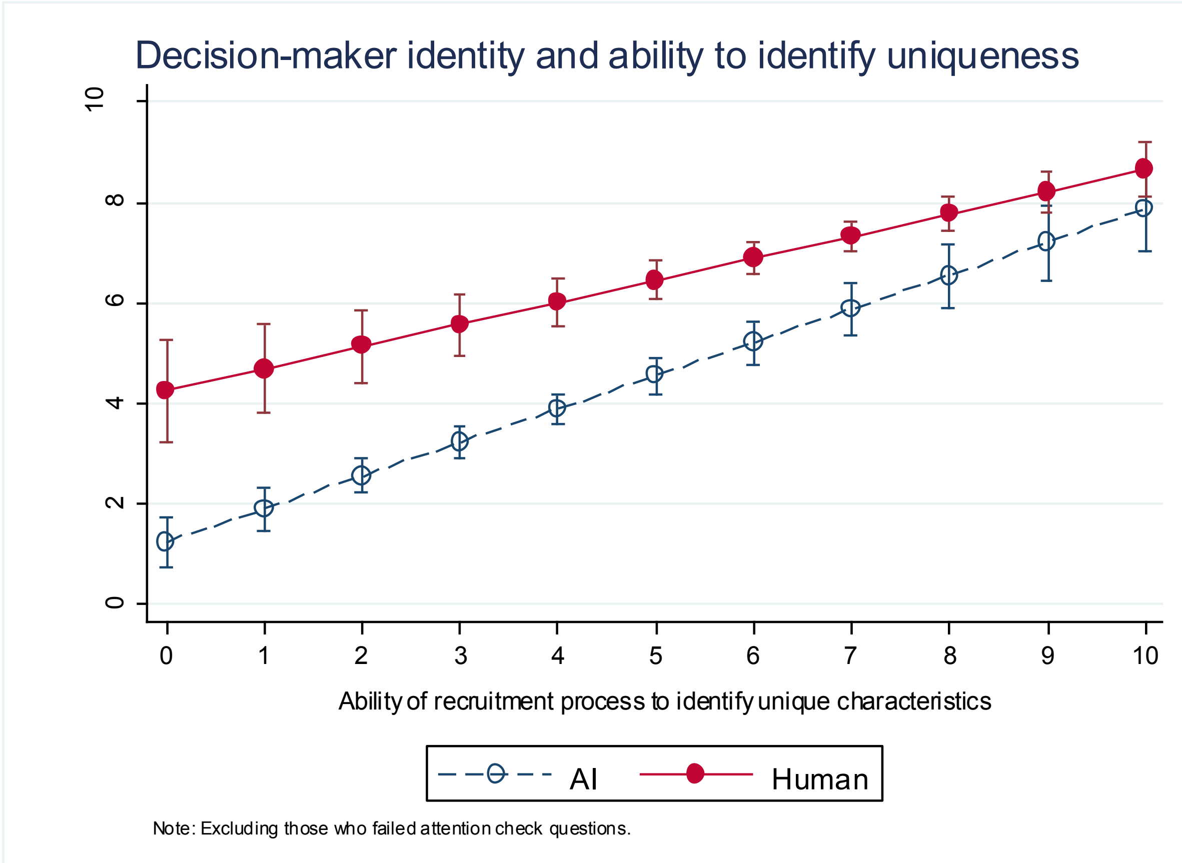

Figure 4. Study 3: Effects of recruiter identity and ability to identify uniqueness on perceived fairness

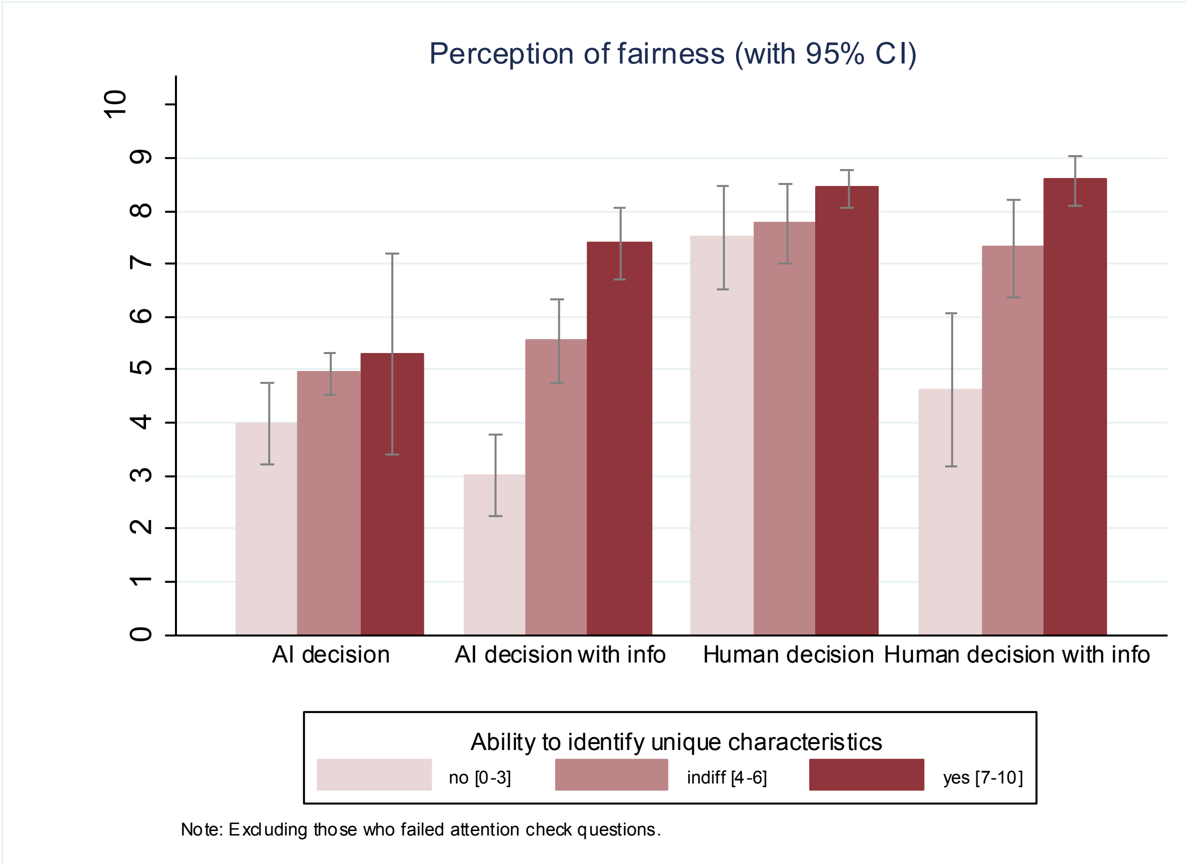

Figure 5. Study 4: Perception of fairness, split in three categories depending on respondents' answer to the question "In your opinion, the recruitment process would be able to identify your unique characteristics." The bar subtitles ending with "with info" refers to the detailed description conditions.

The dataset and replication code of the four studies used to support the findings in this manuscript can be made available on GitHub.

Applicant Fairness Perceptions of AI-enabled Hiring Procedures:

Web appendix

TABLE OF CONTENTS

WEB APPENDIX A: SAMPLE SIZE DETERMINATION

To determine the sample size for our experiments, we used power analysis (the power of a statistical test must be sufficient to detect a statistically significant "true" difference between groups). We based our power analysis on an ANOVA (to provide a more conservative sample size than a t-test) with 2-4 groups depending on the study design and calculated the sample size necessary to detect a medium effect size (f=.25; Cohen, 1988) at 80% power (the customary level use in experimental studies) and a 5% significance level. These calculations indicate a minimum sample size of 159 (Study 1), 180 (Study 2 and 4) and 128 (Study 3). The power analysis reveals the sample sizes of our four studies are sufficient to detect a significant (medium) effect. These sample sizes are also in line with other experimental studies published in the *Journal of Business Ethics* (e.g., Giroux et al., 2022; Li, Jain and Tzini, 2021) as well as other fields such as Human Resources (e.g., Ciancetta and Roch, 2021).

Web Appendix B: Studies scenario

Study 1

Imagine you have five years of sales experience in the food industry. You exhibit great communication and interpersonal skills at your current company. However, you wish to have a change in your working environment. You are now looking for new opportunities in the sales sector, ideally at an environmentally friendly organization.

You discovered a job offer from MangoPick Inc. This company manufactures non-alcoholic beverages and has one of the fastest growing brands in the industry. Please read the job ad on the next screen carefully:

Sales Manager

MangoPick Inc. is currently looking to hire a dynamic leader for the role of Sales Manager. As a Sales Manager, you will be responsible for sales development and brand growth. The specific responsibilities of the Sales Manager include:

- Look for new ways to develop products in the market and formulate action plans to achieve overall goal of the company
- Establish and maintain relationships with retailers
- Communicate our brand message to all current and future accounts
- Actively participate in promos and events

MangoPick Inc. envisions a world where healthy, organic and environmentally friendly beverages will be the standard rather than the exception. If you are ready to work hard, have fun and sell sustainable products, we look forward to receiving your application.

<u>Human decision manipulation:</u>

Before applying for this job, you decide to do some research on MangoPick Inc. After some enquiry about its recruitment process, you learn that the hiring manager would review all resumes submitted and select the most relevant applicants to be interviewed.

<u>AI decision manipulation:</u>

Before applying for this job, you decide to do some research on MangoPick Inc. After some enquiry about its recruitment process, you learn that a computer program would scan all resumes submitted and automatically select the most relevant applicants to be interviewed. There will be no intervention from the hiring manager at this stage.

Human decision assisted by AI manipulation:

Before applying for this job, you decide to do some research on MangoPick Inc. After some enquiry about its recruitment process, you learn that the hiring manager will use a computer program to scan all resumes submitted and automatically provide a list of the most relevant applicants to be interviewed. The hiring manager will then review this subset and call the most relevant applicants in his view.

Study 2

Imagine you are just back from a scuba diving trip in Australia to explore the Great Barrier Reef. However, you have been shocked by the severe bleaching of the coral reefs. Seeing these corals dying has deeply affected you. Even though you like your current job, this trip made you realize how important it was for you to work for a company that shares the same vision of the world as yours. Hence, you decide to search for a new job in an environmentally friendly organization.

After weeks of searching, you discovered the following job offer from MangoPick Inc. This company manufactures non-alcoholic beverages and has one of the fastest growing brands in the industry. MangoPick Inc. is also well-known for investing a significant share of its profits to projects contributing to the preservation of our planet, including a project to save the coral reefs in Australia. In the next screen, you will see the advertisement posted by MangoPick Inc. in a magazine, which you read regularly. Please carefully read the job advertisement.

Sales Manager

MangoPick Inc. is currently looking to hire a dynamic leader for the role of Sales Manager. As a Sales Manager, you will be responsible for sales development and brand growth. The specific responsibilities of the Sales Manager include:

- Look for new ways to develop products in the market and formulate action plans to achieve overall goal of the company
- Establish and maintain relationships with retailers
- Communicate our brand message to all current and future accounts
- Actively participate in promos and events

MangoPick Inc. envisions a world where healthy, organic and environmentally friendly beverages will be the standard rather than the exception. If you are ready to work hard, have fun and sell sustainable products, we look forward to receiving your application.

After reading the advertisement, you immediately submit an application for the job. It's been a week since you submitted your application, and you are eagerly awaiting a response from them. At that instant, you notice a new email in your inbox. This email is from MangoPick's human resources. As you have been looking forward to hearing from them and are excited about it, you open the email immediately.

Positive outcome manipulation:

Below is the email from MangoPick Inc.:

"We have received your application and thank you for your interest in our company. We have rigorously screened your application with great attention. We are delighted to announce that you have been selected to proceed to the next stage of the recruitment process.

We would like to invite you for an interview. Could you send us your availability in the coming weeks so that we can set a date and time convenient for all of us?"

You are extremely excited to be shortlisted to the job! However, you only went through the first stage of the recruitment process. To maximize your chance to get the job, you need to prepare for the interview.

Negative outcome manipulation:

Below is the email from MangoPick Inc.:

"We have received your application and thank you for your interest in our company. We have rigorously screened your application with great attention. Despite your interesting background, we regret to inform you that the skills and qualifications of other candidates correspond more closely to the requirements of this position.

Nonetheless, we hope that you will apply again to future MangoPick positions as they arise. Thank you for your time and consideration."

This news devastates you. You can't believe you are not shortlisted given all your qualifications. To understand what went wrong with your application you decide to do more research on MangoPick's recruitment process.

Human decision manipulation:

After some enquiry about MangoPick's recruitment process, you learn that the hiring manager reviewed all resumes submitted and selected himself the most relevant applicants to be interviewed.

AI decision manipulation:

After some enquiry about MangoPick's recruitment process, you learn that a computer program scanned all resumes submitted and automatically selected the most relevant applicants to be interviewed. There was no intervention from the hiring manager at this stage.

WEB APPENDIX C: ADDITIONAL DETAILS FOR STUDY 2

Removing AI knowledge variable

The regression results of Study 2 without the AI knowledge variable are displayed in the table below. Removing this variable increases the magnitude of the AI decision coefficient; however, the outcome variable and interaction remain significant and with similar signs and magnitudes. In short, the inclusion of the control does not contradict our main findings.

Table W1. Robustness check - Study 2

	Study 2		
DV:	Fairness *as Table 2 col. (3)*	Fairness *as Table 2 col. (4)*	Fairness *w/o AI knowledge*
AI decision	-2.186***	-1.955***	-3.132***
	(0.444)	(0.459)	(0.452)
Negative outcome	-1.470***	-1.268***	-1.500***
	(0.462)	(0.449)	(0.437)
AI x Negative	-1.044*	-1.170*	-1.034*
	(0.637)	(0.633)	(0.615)
Locus of control		-0.087	-0.143**
		(0.061)	(0.058)
AI exposure		0.116	0.097
		(0.076)	(0.076)
AI knowledge		-0.591***	
		(0.190)	
Constant	7.297***	5.723**	3.774*
	(0.286)	(2.060)	(1.984)
Demographic controls		✓	✓
Observations	272	246	269
Adj. R-squared	0.325	0.380	0.372
Chi-squared statistic	147.19	261.15	276.56
# bootstrap iterations	5000	5000	9999

General AI knowledge questions (* indicates correct response):

1. Google achieved what with its deep learning neural networks?
   (a) Code-breaking (b) Weather prediction (c*) Encryption (d) Trend prediction
2. Which one of the below is not a machine learning technique?
   (a) Bayesian (b) Deep learning (c*) Habituation (d) Reinforcement
3. IBM’s Watson AI in best known for what?
   (a) Driverless technology (b*) Cognitive computing (c) IoT network controlling (d) Predictive maintenance
4. What form of processing is ideal for deep learning?
   (a*) Parallel processing (b) Serial processing (c) Sequential processing (d) Data processing
5. Which of the following is a program that allows the computer to simulate conversation with a human being? "Eliza" and "Parry" are early examples of programs that can at least temporarily fool a real human being into thinking they are talking to another person.
   (a) Speech Application Program Interface (b*) Chatterbot (c) Speech recognition (d) Amiga

Summary of Responses to Open-ended Question

Table W2: *Please describe your feeling about this recruitment process*?

	Positive outcome	**Negative outcome**
Human decision	- Normal, fair, happy to be short-listed and that the selection was made by HR/a person, direct involvement from HR, reliable, - Multiple stage process can be long (can be a waste of time) but it also shows that there is competition, rigorous process, nervous (about the interview), too much focus on one person's opinion and risk of bias	- Odd process, doubt, self-esteem, sad, disappointed, unfair because not given a chance, should have more than one person / more checks, potential bias of hiring manager, only hire internal people, biased by preferences, potential conflict of interest. - Normal, good that human hand selects, appreciate that they take the time to read, fair, good that personal touch (make them feel better), much better than when algorithm is used.
AI decision	- Shocked/unsettled, unfair, still need to prove myself, cold and unfeeling, should use human knowledge and emotion (not machinery), irresponsible process, impersonal, selected because use the right keywords (luck), computer makes mistakes (do not trust them), disappointed that a computer chose instead of human, concern about next steps, nervous about interview and the fact that it was not a human who selected, lower the initial excitement of being selected, arbitrary - Computer as good as human, fair, smart and efficient, standard nowadays for large companies, good way to narrow down number of application, doesn't matter who pick as along as picked, good as no visual discrimination/ no bias, neutral or don't care,	- Computers are not reliable, "not rigorous", unfair because eliminate human judgment, waste of time to apply if computer review, can screen out good workers, importance of human touch, upset, disappointed - Common, no reason to be stressed over

WEB APPENDIX D: MEDIATION AND TREATMENT EFFECT DECOMPOSITION

The objective of mediation analysis is to disentangle the average treatment effect on outcome variables that operates through two channels: (i) indirect effect arising from the effect of the treatment on mediating variables; (ii) direct effect that operates through other channels than changes in the measured input. Randomized controlled trials allow for identification of the causal effect of treatment on measured inputs and outputs, but additional assumptions are needed to identify the causal effect of a mediator on outcome variables.

The standard literature on mediation analysis dealt with the problem of confounding effects by invoking different assumptions. Baron and Kenny's (1986) traditional approach assumes that both treatment and mediator variable are exogenous. Imai et al. (2010) and Imai et al. (2011) consider an alternative non-parametric approach, and invoke a *Sequential Ignorability Assumption*, which assumes that all confounding variables are observed, and that there is no unobserved mediator. However, these assumptions are rarely satisfied in practice (Heckman and Pinto 2015; Shaver 2005). It is indeed often not possible to collect data and measure all factors that could have an influence on the equation of interest without error. And our study is no exception. When these underlying assumptions are violated, statistical estimates have undesirable properties and could lead to incorrect conclusions.

Several approaches have been suggested in the literature to limit the adverse effects and provide more meaningful estimates. A common statistical technique developed to deal with this issue is to explicitly model the interdependence between the mediator, the outcome and other variables, using Two-Stage Least Squares (2SLS) or Structural Equation Modeling (SEM), which often require the use of an instrumental variable (IV) (Antonakis et al. 2010, 2014; Shaver, 2005). The downside of this method is that, very often, the "perfect" instrument is not available, and using a weak instrument leads to a similar bias as that of OLS (Bound et al. 1995). As a robustness to the traditional mediation analysis, we follow the methodology proposed by Heckman and Pinto (2015), and applied in Heckman et al. (2013), to decompose treatment effects into direct and indirect effects (i.e. channeled through a mediator). Even though some econometric exogeneity and linearity assumptions are still necessary, the decomposition strategy they propose is to minimize the problems of endogeneity plaguing the mediation methods mentioned above.

Heckman and Pinto (2015) decompose a linear model into measured and unmeasured components as follows:

$$Y_d = \kappa_d + \sum_{j \in \mathcal{J}} \alpha_d^j \theta_d^j + \beta_d X + \tilde{\epsilon}_d$$

$$= \kappa_d + \underbrace{\alpha_d^p \theta_d^p}_{\text{proxied input}} + \underbrace{\sum_{j \in \mathcal{J} \setminus \mathcal{J}^p} \alpha_d^j \theta_d^j}_{\text{unmeasured inputs}} + \beta_d X + \tilde{\epsilon}_d$$

$$= \left(\kappa_d - \sum_{j \in \mathcal{J} \setminus \mathcal{J}^p} \alpha_d^j E\left(\theta_d^j\right) \right) + \alpha_d^p \theta_d^p + \beta_d X$$

$$+ \left[\tilde{\epsilon}_d + \sum_{j \in \mathcal{J} \setminus \mathcal{J}^p} \alpha_d^j \left(\theta_d^j - E\left(\theta_d^j\right) \right) \right]$$

$$= \tau_d + \alpha_d^p \theta_d^p + \beta_d X + \epsilon_d \quad (1)$$

where Y_d is the outcome variable, $\tau_d = \kappa_d + \sum_{j \in \mathcal{J} \setminus \mathcal{J}^p} \alpha_d^j E\left(\theta_d^j\right)$ is a constant, α_d is a $|\mathcal{J}|$-dimensional vector of inputs, θ_d^p is our proxied input – our mediator (i.e. whether the recruitment process is able to identify unique characteristics), β_d is a $|X|$-dimensional vector of pre-treatment variables, X are pre-treatment control variable, $\tilde{\epsilon}_d$ is a zero-mean error term assumed to be independent of regressors θ_d and X, $d \in \{0,1\}$ is the treatment indicator, $\tau_d = \kappa_d + \sum_{j \in \mathcal{J} \setminus \mathcal{J}^p} \alpha_d^j E\left(\theta_d^j\right)$ and $\epsilon_d = \tilde{\epsilon}_d + \sum_{j \in \mathcal{J} \setminus \mathcal{J}^p} \alpha_d^j \left(\theta_d^j - E\left(\theta_d^j\right) \right)$, which is a zero-mean error term. Hence, the error term ϵ_d will be correlated with the proxied/measured input if these measured inputs are correlated with unmeasured inputs.

Then, to decompose the treatment effects into components attributable to change in our proxied inputs ($\Delta\theta = \theta_1 - \theta_0$) and change in parameters ($\Delta\alpha = \alpha_1 - \alpha_0$), it is necessary to assume that changed in unmeasured inputs attributable to the experiment are independent of X:

$$\begin{aligned} E(Y_1 - Y_0|X) &= (\tau_1 - \tau_0) + E\left(\alpha_1\theta_1^p - \alpha_0\theta_0^p\right) + (\beta_1 - \beta_0)X \\ &= \underbrace{(\tau_1 - \tau_0)}_{\text{direct effect}} + \underbrace{(\Delta\alpha + \alpha_0)E(\Delta\theta^p) + (\Delta\alpha)E\left(\theta_0^p\right)}_{\text{indirect effect}} + \underbrace{(\beta_1 - \beta_0)X}_{\text{other}} \end{aligned} \tag{2}$$

where Y_1 and Y_0 represent the outcome variable under the treatment and control conditions, $(\tau_1 - \tau_0)$ is the average difference between the treatment and control groups that is not attributable to measured inputs, θ_d describes our mediator (i.e. whether the recruitment process is able to identify unique characteristics) and X are pre-treatment control variable. This equation can be simplified if the structural invariance or autonomy assumptions is satisfied, that is if $\beta_1 = \beta_0$ and $\alpha_1 = \alpha_0$. As explained in Heckman and Pinto (2015), if measured and unmeasured inputs are independent, these parameters can be consistently estimated by OLS and tested. Wald tests revealed that only the model coefficients associated to pre-treatment variables X are the same for both the treatment and control groups (i.e. $\beta_1 = \beta_0$, $\chi^2(16)=14.98$, $p=0.526$ but $\alpha_1 \neq \alpha_0$, $\chi^2(1)=3.79$, $p=0.052$). Equation (2) then rewrites:

$$E(Y_1 - Y_0) = (\tau_1 - \tau_0) + (\Delta\alpha + \alpha_0)E\left(\theta_1^p - \theta_0^p\right) + (\Delta\alpha)E(\theta_0^p) \tag{3}$$

And the outcome equation to be estimated using standard linear regression, comprising both treatment groups becomes:

$$Y = \tau_0 + \phi D + \alpha\theta^p + \omega\theta^p \cdot D + \beta X + \eta \tag{4}$$

Hence, the treatment effects channeled through the mediator originates from (i) the impact of the mediator on the outcome; (ii) the enhancement of the mediator by the intervention. Given the assumptions that θ^p is measured without error and is independent of the error term ϵ, Heckman and Pinto (2015) showed that least squares estimators of the parameters of equation (4) are unbiased. We estimate all parameters in these decompositions through a series of regression steps:

1. We regress our mediator on the treatment indicator and the vector of pre-treatment variables

$$\theta_i^p = \delta_0 + \delta_1 D_i + \delta_2 X_i + \upsilon_i \qquad i = 1 \dots N \tag{5}$$

   This estimation step yields the mediator mean $E(\theta_0^p) = \hat{\delta}_0 + \hat{\delta}_2 E(X)$ and the expected change in our mediator from the treatment conditional on X: $E(\theta_1^p - \theta_0^p) = \hat{\delta}_1$.

2. We regress equation (4) to obtain:
   a. The direct effect: $(\tau_1 - \tau_0) = \hat{\phi}$
   b. Parameters $\Delta\alpha = \widehat{\omega}$ and $\alpha_0 = \hat{\alpha}$

3. Regression results from the two previous steps are combined to calculate the decomposition of treatment effect:
   a. The direct effect: $(\tau_1 - \tau_0) = \hat{\phi}$
   b. The indirect effect: $(\Delta\alpha + \alpha_0)E(\theta_1^p - \theta_0^p) + \Delta\alpha E(\theta_0^p) = (\widehat{\omega} + \hat{\alpha})\hat{\delta}_1 + \widehat{\omega}[\hat{\delta}_0 + \hat{\delta}_2 E(X)]$

Estimation results are shown in Table W3. As in Heckman et al. (2013), coefficients whose one-sided bootstrap p-values below 0.1 are considered to decompose the treatment effect. Unbiasedness of this decomposition relies on the key assumption that measured and unmeasured input are independent. However, note that, while Heckman and Pinto (2015) used factor analysis to aggregate measures and account for measurement error, we do not have enough indicators of our concept to do so. Hence, our estimates may suffer from attenuation bias induced by uncorrected measurement error.

Table W3. Study 3: Regression table based on Heckman and Pinto's (2015) methodology

DV:	Identify uniqueness	Fairness	
	(1)	(2)	(3)
AI decision	-3.341***	-3.718***	-2.612***
	(0.278)	(0.274)	(0.619)
Identify uniqueness			0.490***
			(0.077)
Identify uniqueness x AI			0.154
			(0.102)
Constant	6.697***	8.214***	5.008***
	(1.598)	(1.643)	(1.380)
Demographic controls	✓	✓	✓
Observations	278	278	278
Adj. R-squared	0.372	0.429	0.631
Chi-squared statistics	213.77	257.28	616.68
# bootstrap iterations	5000	5000	5000

Notes: Bootstrapped standard errors, p-values in parentheses. Stars indicating: * p<0.1 **, p<0.05 and *** p<0.01. Control variables included: gender dummy, non-white dummy, class dummies, age, age squared, education dummy variables, political orientations dummy variables (left/right; democrat/republican, liberal/conservative) and degree of religiosity. Participants failing attention check questions have been excluded from the analysis.

WEB APPENDIX E: SAMPLE SIZE REPRESENTATIVENESS

Table W4. Study samples key demographic summary statistics (average) vs. U.S. population

	Study 1	**Study 2**	**Study 3**	**Study 4**	**US Population (2020)**
Gender (% female)	0.585	0.658	0.564	0.648	0.505
Non-white	0.265	0.203	0.262	0.253	0.384
Age	35.494	36.41	37.56	38.05	38.2
Education:					
• High school	0.092	0.066	0.082	0.09	0.267
• Incomplete college	0.201	0.25	0.228	0.221	0.203
• Associate degree	0.149	0.110	0.146	0.142	0.086
• Bachelor degree	0.406	0.404	0.406	0.401	0.202
• Master or Doctoral degree	0.153	0.169	0.139	0.146	0.127

Notes: The data source for the U.S. population are the U.S. Census Bureau (race, age and education) and the World Bank (gender).

References

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *Leadership Quarterly*, 21(6), 1086–1120.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and endogeneity: Problems and solutions. *Oxford Handbook of Leadership and Organizations, 1*, 93–117.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Pesonality and Social Psychology*, 51(6), 1173–1182.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430), 443–450.

Ciancetta, L. M., & Roch, S. G. (2021). Backlash in performance feedback: deepening the understanding of the role of gender in performance appraisal. Human Resource Management, 60(4), 641-657

Cohen. J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Giroux, M., Kim, J., Lee, J. C., & Park, J. (2022). Artificial intelligence and declined guilt: Retailing morality comparison between human and AI. Journal of Business Ethics, 1-15.

Heckman, J. J., & Pinto, R. (2015). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Reviews*, 34(1-2), 6–31.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6), 2052-86.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1), 51–71.

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(4), 765–789.

Li, S., Jain, K., & Tzini, K. (2021). When Supervisor Support Backfires: The Link Between Perceived Supervisor Support and Unethical Pro-supervisor Behavior. Journal of Business Ethics, 1-19.

Shaver, J. M. (2005). Testing for mediating variables in management research: Concerns, implications, and alternative strategies. *Journal of Management*, 31(3), 330–353.