Technische Universität München
TUM School of Management

Ethical Perspectives on the Use of Artificial Intelligence in Hiring

Anna Lena Hunkenschroer

Vollständiger Abdruck der von der TUM School of Management der Technischen Universität München zur Erlangung einer Doktorin der Volkswirtschaftslehre (Dr. oec. publ.) genehmigten Dissertation.

Vorsitz:
Prof. Dr. Benedikt Schnurr

Prüfer*innen der Dissertation:
1. Prof. Dr. Christoph Lütge
2. Prof. Dr. Alwine Mohnen

Die Dissertation wurde am 03.06.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Management am 15.09.2022 angenommen.

Abstract

Companies are increasingly using artificial intelligence (AI) technologies to automate their hiring processes and make them more efficient. Because the results of hiring decisions have serious consequences for individuals' lives and careers, application of AI in this area also raise ethical concerns. Opinions about whether such important decisions should be outsourced to AI have led to controversy. However, given the novelty of AI applications in hiring, the topic has not been widely addressed in the academic literature.

This dissertation consists of three self-contained studies with which I[1] contribute to scientific research on ethical considerations related to the use of AI in hiring. Whereas the first study provides an overarching ethical perspective on the topic, the second and third studies each focus on one ethical subtopic—namely, fairness and bias. In the first study, I systematically review the existing literature and outline the ethical opportunities, risks, and ambiguities associated with AI-based hiring. Moreover, I identify research gaps that should be further explored in the future.

The second study examines the question of how companies should design their AI hiring practice so that people perceive them as fair. I find that the positioning of an AI interview within the overall process and people's sensitization to AI's potential to reduce human bias have a significant effect in their fairness perceptions, and thus, overall organizational attractiveness.

In the third study, I examine the actual preferences of individuals, particularly women, for an AI- versus a human-based evaluation procedure, considering expected biases. I find that although individuals generally prefer a human to an AI evaluator, women's belief that AI can reduce bias and the extent to which they have perceived discrimination in the past positively influence their preference for an AI evaluator. In addition, women are more likely to choose AI evaluators when competing against men versus when competing with only women, but only if they believe in AI's potential to reduce biases.

[1] Please note that all studies in my dissertation are based on research that I jointly conducted with my coauthors. Hence, throughout this dissertation, whenever I refer to my research, "I" stands for "my coauthors and me." Of course, all errors are mine.

Zusammenfassung

Unternehmen nutzen zunehmend Technologien der künstlichen Intelligenz (KI), um ihre Einstellungsprozesse zu automatisieren und effizienter zu gestalten. Da die Ergebnisse von Einstellungsentscheidungen schwerwiegende Folgen für das Leben und die Karriere des Einzelnen haben, wirft dieser Anwendungsbereich von KI auch ethische Bedenken auf. So gibt es kontroverse Meinungen darüber, ob so wichtige Entscheidungen an KI ausgelagert werden sollten. Angesichts der Neuartigkeit von KI-Anwendungen in der Personalauswahl wurde das Thema in der akademischen Literatur jedoch noch nicht umfassend behandelt.

Diese Dissertation besteht aus drei in sich abgeschlossenen Studien, mit denen ich einen Beitrag zur wissenschaftlichen Forschung über ethische Überlegungen im Zusammenhang mit dem Einsatz von KI bei der Personalauswahl leiste. Während die erste Studie eine übergreifende ethische Perspektive auf das Thema bietet, konzentrieren sich die zweite und dritte Studie jeweils auf ein ethisches Unterthema, nämlich Fairness und Voreingenommenheit (Bias). In der ersten Studie gebe ich einen systematischen Überblick über die vorhandene Literatur und skizziere die ethischen Chancen, Risiken und Ambiguitäten, die mit KI-basierten Einstellungsverfahren verbunden sind. Darüber hinaus zeige ich Forschungslücken auf, die in Zukunft weiter erforscht werden sollten.

Die zweite Studie geht der Frage nach, wie Unternehmen ihren KI-Einstellungsprozess gestalten sollten, damit er von Menschen als fair empfunden wird. Ich finde heraus, dass die Positionierung eines KI-Interviews innerhalb des Gesamtprozesses und die Sensibilisierung der Menschen für das Potenzial von KI, menschliche Voreingenommenheit zu reduzieren, einen signifikanten Einfluss auf ihre Fairnesswahrnehmung und damit auf die Attraktivität des Unternehmens insgesamt haben.

In der dritten Studie untersuche ich die tatsächlichen Präferenzen von Einzelpersonen, insbesondere von Frauen, für ein KI- versus ein Mensch-basiertes Bewertungsverfahren unter Berücksichtigung von erwarteten Vorurteilen. Ich stelle fest, dass, obwohl Einzelpersonen im Allgemeinen einen menschlichen Bewerter einem KI-Bewerter vorziehen, der Glaube von Frauen daran, dass KI Bias reduzieren kann, und das Ausmaß, in dem Frauen in der Vergangenheit Diskriminierung wahrgenommen haben, ihre Präferenz für einen KI-Bewerter positiv beeinflussen. Außerdem wählen Frauen KI-Evaluatoren eher, wenn sie mit Männern konkurrieren, als wenn sie nur mit Frauen konkurrieren, aber nur, wenn sie an das Potenzial von KI zur Verringerung von Vorurteilen glauben.

Acknowledgements

Throughout my doctoral studies, I have received a great deal of support and assistance for which I am very grateful. First, I would like to thank my supervisor, Christoph Lütge, who successfully guided me throughout all phases of this dissertation. I have greatly appreciated the freedom and trust I was given during my dissertation time at the Chair of Business Ethics as well as the many learning opportunities I have experienced along the way.

I would also like to thank Alwine Mohnen for being my second supervisor and for her constructive suggestions and encouragement, especially during the latest phase of the dissertation project.

Additionally, I thank my coauthor, Christoph Hohenberger, for his openness to my ideas, for his valuable contributions to our joint research project and for the opportunity to collaborate with his artificial intelligence software development start-up.

Thanks also to Philipp Gerbert for being my mentor during my dissertation project. I perceived our exciting conversations, which have shown me new perspectives, as a great privilege.

Moreover, I would like to thank all my colleagues at the Chair of Business Ethics and the Institute for Ethics in Artificial Intelligence for integrating me into the team and enabling me to immerse myself in the scientific environment at the Technical University of Munich (TUM). I have enjoyed working in this diverse and open-minded team, which has provided me with new ways of working and different perspectives, thus extensively broadening my horizons. In particular, I would like to thank Franziska Poszler and Alexander Kriebitz for the insightful discussions and your support during the last two years.

Furthermore, I am forever thankful to all my friends who supported me during my entire PhD journey. Special thanks go to my friends and TUM colleagues Yasmina Trautmann and Isabella Stojkovski for all their help, for sharing the ups and downs, and for making the time unforgettable.

Most importantly, I am deeply thankful to my family and especially to my parents. Thank you for your unconditional love, for encouraging and supporting me in all my decisions, and for shaping me into who I am today.

Lastly, I truly thank Adrian Schulte Steinberg for accompanying me on this journey. Thank you for always having my back, while always encouraging me to take on new challenges. Thank you for always believing in me and for all your patience. I am incredibly grateful to have you by my side.

Table of Contents

1 Introduction

Artificial intelligence (AI) is already widely used and implemented across organizations’ business functions, such as automated manufacturing, the pricing of goods, or the evaluation of candidates in hiring, as well as across almost all industries. According to Kaplan and Haenlein (2019), *AI* can be defined as “a system’s ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” (p. 17). Thus, AI not only includes complex machine learning approaches, such as deep neural networks, but also covers simple algorithms relying on regression analyses as well as other kinds of algorithms, such as natural language processing or voice recognition. Because of the inherent characteristics of AI that distinguish it from other technologies, and because of the influence AI applications may have, the new use cases of AI lead to new ethical challenges and questions (Kriebitz & Lütge, 2020). For example, ethical questions arise about whether AI alone should make decisions that can have a major effect on people’s lives (e.g., promotions or granting credit), even though the reasons for doing so cannot be accurately explained due to the black-box nature of AI (Bloomberg, 2018).

To address these emerging ethical risks, various stakeholder groups from both governmental institutions and the private sector have developed ethical guidelines in recent years (e.g., High-Level Expert Group on Artificial Intelligence, 2019; Microsoft, 2018; University of Montreal, 2018). In these documents, normative principles are developed that should guide the ethical and trustworthy development and use to harness disruptive potentials and to tackle potential cases of misuse of AI technologies. The High-Level Expert Group on AI (2019), for example, has declared four leading principles that should guide trustworthy AI development: respect for human autonomy, prevention of harm, fairness, and explicability. Nevertheless, these ethical guidelines face criticism for being very high-level, general, and even superficial. What is deemed an appropriate action may depend on the domain in which AI is used and may differ across application contexts and business functions, thus revealing the need for domain-specific works in the field of AI ethics (Hagendorff, 2020; Mittelstadt, 2019).

One business function in which companies increasingly make use of AI technology is human resources (HR); more specifically, one process in HR uses AI—namely, hiring. By using the term *AI-enabled hiring*, I refer to any procedure that makes use of AI for the purposes of assisting organizations during the recruitment and selection of job candidates.[2] In this context, AI tools can speed up the process of hiring applicants and make it more efficient. In large companies especially, such as Vodafone, KPMG, BASF, and Unilever, AI tools are already well-established to handle the large numbers of incoming applications (Daugherty & Wilson, 2018; Köchling & Wehner, 2020). Similarly, the supply of AI software vendors, often in the form of software startups that develop AI-based solutions for personality profiling, is increasing. They advertise their products with the claim of not only being more efficient but also being less biased and more objective than human-based recruiting practices are due to the products' reliance on data-driven analyses instead of human intuition (Polli et al., 2019).

However, hiring constitutes one application of AI that is particularly controversial in both public and academic discourse due to its close relations to ethical norms and values as well as the effect it may have on the lives of affected stakeholders—namely, recruiters and applicants (e.g., Giermindl et al., 2021). Criticism has been leveled that important decisions affecting people's future careers are outsourced to AI, which is especially problematic if mistakes are made (Raghavan et al., 2020). One of the best-known real-world examples is the case of Amazon in 2018, in which a tested AI systematically discriminated against women in the hiring

[2] Although the academic literature often distinguishes between recruiting, selecting, and hiring candidates, with recruitment and selection describing two different phases within the hiring process, I use the terms *AI recruiting* and *AI hiring* as synonyms in the context of this dissertation. In doing so, I always refer to the use of AI in the entire hiring process.

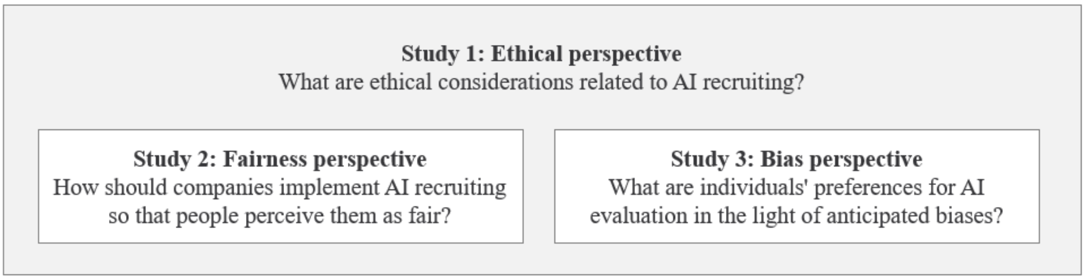
process (Dastin, 2018). Although scientific research on AI recruiting has substantially increased in recent years, the subject is still an emerging topic in academic literature and lags behind the technological developments and current business practices (Köchling & Wehner, 2020). Thus, my dissertation advances and contributes to this research stream by studying the ethical use of AI in recruiting, addressing the call for more domain-specific work in the field of AI ethics.

In this dissertation, I study AI recruiting from three perspectives—namely, an overarching ethical perspective, a fairness perspective, and a bias perspective covering two ethical subtopics. First, I review existing academic literature about the general ethical considerations related to AI recruiting and develop an ethical framework that outlines its ethical benefits, risks, and ambiguities. Moreover, I identify current research gaps and find that due to the phenomenon's novelty, empirical work on the topic of AI recruiting in particular remains insufficient and that findings on people's fairness perceptions of AI recruiting are inconsistent, revealing room for further empirical research in the field to derive implications for practice. In the subsequent two studies of my dissertation, I address these identified gaps and ambiguities. To this end, in the second study, I empirically examine people's fairness perceptions of different designs of AI-powered selection processes and derive insights for companies on ways to implement AI recruiting. Thus, this study provides a fairness perspective on the topic. In the third study, I apply a bias perspective, focusing on another ethical subdimension relevant in the context of AI recruiting. More precisely, I experimentally investigate individuals' revealed preferences about an AI-based evaluation compared to a human-based evaluation in a competitive situation in the light of anticipated biases.

1.1 Motivation and research questions

The purpose of this dissertation is to study AI recruiting from three ethical perspectives. Whereas the first study applies an overarching ethical perspective, the subsequent studies focus on one ethical subdimension and examine AI recruiting from a fairness and a bias perspective. The corresponding guiding questions of this research endeavor are as follow: What are ethical considerations related to AI recruiting? How should businesses implement AI recruiting practices so that applicants perceive them as fair? What are individuals' preferences for AI evaluation in the light of anticipated biases? To answer these questions, this dissertation comprises three self-contained studies. See Figure 1 for an illustration of the scope and guiding research questions of this dissertation.

Figure 1.1: Scope and guiding research questions of this dissertation



In my first study, presented in Chapter 2, I apply an overarching ethical perspective and explore the general ethical considerations related to the use of AI in hiring. Although academic research on AI recruiting has increased substantially in recent years, a comprehensive ethical understanding of recruiting as an expanding application context of AI is still lacking. Although the subject of algorithmic bias in hiring decisions has attracted broad interest among researchers, especially from the legal (Bornstein, 2017; Kim, 2017) and technical perspectives (Chwastek, 2017; Lin et al., 2020; Mujtaba & Mahapatra, 2019), there remain additional ethical concerns related to AI recruiting, such as data privacy, transparency, and accountability, which are worth discussing. To establish a common foundation for future research in the field, I therefore review existing academic literature and map the ethical opportunities, risks, and ambiguities, as well as the proposed ways to mitigate those ethical risks. Further, I identify the gaps in the literature that call for deeper exploration in future research.

In my second study, presented in Chapter 3, I focus on one ethical subdimension related to AI recruiting—namely, perceived fairness. The perceived fairness of recruiting practices by individuals is highly relevant to businesses because it can have meaningful effects on people's attitudes, intentions, and behaviors. For example, it has been shown that perceptions of selection practices directly influence organizational attractiveness and people's intentions to accept job offers (McCarthy et al., 2017). To this end, in my study, I examine the question of how to implement AI recruiting tools so that people perceive them as fair. More specifically, I investigate whether adjusting process design factors—namely, (a) the positioning of an AI interview, (b) applicants' sensitization to AI's potential to reduce human bias, and (c) human oversight of the decision-making process—may help businesses improve people's perceptions of AI interviews as fair as well as overall organizational attractiveness.

In my third study, presented in Chapter 4, I focus on another ethical subdimension related to AI hiring—namely, the existence of bias. It has been shown that the labor market suffers from both systematically biased hiring practices against women in male-dominated occupations (Hoover et al., 2019; Sinclair & Carlsson, 2021) and the self-selection of women into fields where they anticipate less-biased hiring practices (Carlsson & Sinclair, 2018; Pinel & Paulin, 2010), which still lead to large gender imbalances today. New AI-powered hiring practices postulate fairer and less-biased hiring processes. AI tools could thus have the potential to eliminate both actual and anticipated biases against female applicants in hiring. To obtain an initial assessment of whether AI could currently exploit this described potential, my study examines individuals', especially women's, preferences with respect to an AI- versus a human-led evaluation process in a competitive setting in light of anticipated biases. More specifically, I examine whether these preferences are affected by women's beliefs in AI's potential to reduce bias, the gender composition of their competitor pool, and the extent to which they have perceived personal discrimination in the past.

1.2 Data and methodology

To answer my research questions, I conducted three separate studies using different kinds of data and methodologies, each tailored to the specific research question.

In the first study, I conducted a systematic literature review on the ethicality of AI-enabled recruiting and selection practices to provide an overview of existing research on the topic and identify current research gaps. More precisely, I identified 51 distinct articles dealing with the topic, which I synthesized in four stages: First, to show how the ethicality of AI recruiting is assessed in the research, I categorize the identified literature according to the assumed perspectives. Here I differentiate among theoretical, practitioner, legal, technical, and descriptive perspectives. Second, I give an overview of AI applications in recruiting as mentioned in the articles. Third, I map the ethical considerations in the form of ethical opportunities, ethical risks, and ethical ambiguities. Fourth, I outline the mentioned approaches to mitigate ethical risks in practice. Based on this analysis, I identify the shortcomings of current research and outline moral topics and questions that call for a deeper exploration in both theoretical and empirical future research.

In the second study, I used data from an online vignette study to examine the question of how to improve fairness perceptions of AI recruiting. I study whether three process design factors—namely, (a) the positioning of the AI interview throughout the overall selection process, (b) applicants' sensitization to AI's potential to reduce human bias, and (c) human oversight of the AI-based decision-making process—affect applicants' perception of fairness in a hiring process. In the vignette study, I applied a 2 × 2 × 2 between-subjects design, in which I varied the three factors: positioning of the AI interview (initial stage vs. final stage), the sensitization of participants to bias reduction potential (sensitization vs. no sensitization), and human oversight of the AI decision (human oversight vs. no human oversight). This resulted in a total of eight experimental groups. After reading one of the vignettes, which described a company that uses an AI interview in its selection process, the participants responded to items measuring their fairness and organizational attractiveness perceptions. Based on the data from this experiment, I analyzed whether adjusting process design factors has an effect on people's perception of the fairness of AI interviews. For this purpose, I applied a factorial analysis of variance (ANOVA) as well as a mediation analysis using a structural equation modeling (SEM) approach.

In the third study, I used data from an online experiment to study the question of individuals', especially women's, preferences with respect to an AI- versus a human-led evaluation process. More specifically, I examined how the gender composition of the competitor pool and women's perceived personal discrimination affect their preference. In the experiment, participants were assigned to groups of four, and they participated in an incentivized tournament. In the tournament, participants conducted a video interview, and afterwards, they indicated whether they wanted their video to be evaluated by an AI program or by a person. I exogenously varied the gender composition of an individual's competitor pool by randomly assigning participants—based on their gender—to either single- or mixed-gender groups. This economic experimental design with monetary incentives allowed me to study participants' true preferences. As my methodology, I used regression analyses, binomial probability tests (BPTs), and nonparametric tests to analyze the data from the experiment.

1.3 Related literature

Although all studies of this dissertation deal with ethical topics related to AI use in hiring, they relate to and thereby combine different strands of the scientific literature. They are embedded in the existing literature as follows.

The first study deals with the general ethical considerations related to AI use in hiring and can be positioned in the broader discourse on AI ethics. In recent years, various stakeholder groups have contributed to this research stream by releasing several ethics guidelines. These include, for instance, the Montreal Declaration for Responsible AI (University of Montreal, 2018), the Ethics Guidelines for Trustworthy AI of the High-Level Expert Group on AI established by the European Commission (High-Level Expert Group on Artificial Intelligence, 2019), and AI4People's principles for AI ethics (Floridi et al., 2018; see Jobin et al., 2019, for a meta-analysis). Although many of these guidelines offer high-level guidance for AI applications in general, my review article focuses on the ethical use of AI in the hiring context, constituting a domain-specific work in this stream of literature.

Aside from the ethical guidelines, research on AI ethics is multi- and interdisciplinary, combining research from technical, engineering, and social sciences, in which researchers across disciplines often aim to inform on how ethical principles can be put into practice (Tolmeijer et al., 2020). The review article itself organizes extant research on ethical considerations of AI recruiting and provides an overview of five different substreams within this research field: theoretical papers (e.g., Simbeck, 2019; Yarger et al., 2020) assessing AI-powered recruiting practices from an ethics theory perspective; practitioner-oriented articles (e.g., Chamorro-Premuzic et al., 2016; Florentine, 2016) that focus on the implications most relevant for managers and corporations; legal works (e.g., Bornstein, 2017; Kim, 2017) that deal with the legal framework to guide employment decisions under the use of AI; technical papers (Fernández-Martínez & Fernández, 2020; Vasconcelos et al., 2018) that develop technical solutions to implement ethical principles into algorithmic code or design; and descriptive studies (Langer et al., 2018; e.g., M. K. Lee, 2018), which are mainly experimental and assess people's reactions to AI-powered recruiting practices.

My second study ties into this latter substream of empirical research on AI recruiting by studying the question of whether adjusting process design factors may help improve people's perceptions of AI interviews' fairness. The study thereby links the research on AI ethics with the research on applicant reactions to selection procedures, which is largely based on Gilliland's (1993) justice model, and assesses applicants' fairness perceptions in different selection situations. By studying applicant reactions to AI use, I build on the applicant reaction literature dealing with technology-enhanced recruiting practices in a broad sense. This literature stream emerged in the early 2000s and investigates perceptions of technology in personnel selection and job interviews (Bauer et al., 2006; see Blacksmith et al., 2016, for a meta-analysis; Chapman et al., 2003; Wiechmann & Ryan, 2003). Various empirical studies have investigated technology-mediated recruiting procedures, such as telephone and video interviews, by testing technology-related factors' effects on the interviews and on applicant reactions. For example, a couple of studies examined applicants' fairness perceptions of online selection practices (Konradt et al., 2013; Thielsch et al., 2012). Moreover, by investigating ways to improve perceptions of AI interviews by adjusting the process design, the study relates to research on contextual influences on applicant reactions (Hausknecht et al., 2004; Ryan & Ployhart, 2000).

The third study relates to and thereby combines several research streams. First, just as the second study, it contributes to the empirical research on applicant reactions to AI recruiting by examining the question of what individuals (especially women) prefer—an AI-led evaluation process or a human-led evaluation process. In doing so, I extend the research stream on contextual factors and individual differences in applicant reactions (Hiemstra et al., 2019; Langer et al., 2018) by investigating how women's beliefs in AI's potential to reduce bias, competitors' gender composition, and perceived personal discrimination as factors impact applicant preferences. Second, on a more general level, this paper relates to the growing research field investigating humans' behavioral responses to AI-based decision-making as well as their aversion to and trust in algorithms (e.g., Castelo et al., 2019; Dietvorst et al., 2015, 2018; Jauernig et al., 2022). Although prior research on humans' behavioral responses to AI-based decision-making has shown that in many situations, people prefer human decision-makers over those that are AI-based because of algorithmic aversion (Castelo et al., 2019), I examine how contextual and individual factors may affect individuals' reliance on algorithms in the hiring context.

However, this study goes further and, by studying women's preferences for an AI evaluation in a competitive situation, also relates to the extensive literature on experimentally elicited gender differences in competition behavior (e.g., Balafoutas & Sutter, 2012; Berger et al., 2020; Maggian et al., 2020; Niederle et al., 2013). Research has shown that, regardless of their actual performance, women generally tend to shy away from competition against male competitors (e.g., Buser et al., 2014; Niederle et al., 2008; Niederle & Vesterlund, 2008). I relate to this stream of research by exploring women's preferences for an AI evaluation in competition, including the extent to which women's preferences change between competing against women and competing against men.

The third study finally links to the literature on women's self-selection in the job market—that is, their avoidance of applying for male-dominated and competitive high-profile jobs, resulting in gender-related occupational segregation (e.g., Ceci et al., 2009; Parker et al., 2012; Sinclair et al., 2019; Sinclair & Carlsson, 2013). Next to women's competition behavior, research has identified anticipated gender-related biases in hiring as another driver for women's self-selection into certain occupations (Carlsson & Sinclair, 2018; Heilman & Okimoto, 2007; Moss-Racusin et al., 2012). Our study expands this literature stream by examining the potential of AI as a new determinant to address the anticipated bias in human recruitment as well as gender-based self-selection into certain career paths.

1.4 Results and contributions

Although in practice, people often focus on efficiency considerations when it comes to the use of AI in business processes, my dissertation provides ethical perspectives on the topic—more specifically, on the use of AI in hiring. More precisely, I contribute to the scientific literature as follows.

In the first study, I review existing literature on the ethicality of AI-enabled recruiting and selection. I find that ethical considerations related to the use of AI in hiring are diverse and can be clustered into ethical opportunities, risks, and ambiguities. Moreover, I find that the AI recruiting research field suffers from four shortcomings: First, only a few theoretical papers provide the foundation for an ethical discussion on the topic. Second, many papers focus on algorithmic bias, and other ethical concerns, such as accountability and human autonomy in the AI recruiting context, seem neglected. Third, the approaches to mitigate ethical risks related to the use of AI, which can be found in current research, are rather general and lack concrete domain-specific implementation guidelines and therefore tangible impact for the recruiting context. Fourth, I find that empirical research on the topic is remarkably limited.

The article's contributions are threefold. First, I comprehensively organize the extant research on ethical considerations of AI recruiting. Second, I provide researchers and HR professionals with an overview of the ethical considerations in AI recruiting by providing an ethical framework on the ethical opportunities, risks, and ambiguities of AI recruiting. Third, I identify current research gaps and raise moral topics and questions that call for deepening the exploration in both theoretical and empirical future research.

In the second study, I address one of these identified research gaps by empirically examining people's fairness perceptions of different designs of AI-powered selection processes. I find that two process design factors—the positioning in the overall process and applicants' sensitization to AI's potential to reduce bias—are critical to people's fairness perceptions of AI interviews, which in turn affect overall organizational attractiveness. If properly designed, these design factors can help improve applicants' reactions to AI interviews to prevent negative outcomes for organizations that use such interviews. I do not find significant differences in people's fairness perceptions depending on human oversight of the AI decision-making process.

The study's contributions are threefold. First, the study links the research on applicant reactions to selection procedures with research on AI ethics. Thereby, it addresses the call for empirical research on both applicant reactions to AI-based recruiting practices (e.g., Langer et al., 2017) and AI's ethical implementation in a domain-specific context (e.g., Hagendorff, 2020). Second, by identifying ways to improve perceptions of AI interviews, the study advances research on contextual influences on applicant reactions. I extend the current theories of procedural fairness (e.g., Hausknecht et al., 2004; Ryan & Ployhart, 2000) by experimentally demonstrating how the positioning of the AI interview, as well as candidates' sensitization to AI's potential to reduce human bias, can influence people's fairness perception of this tool. Third, the paper has practical implications. It highlights how the process surrounding AI interviews should be designed to lead to better applicant perceptions. This is an important concept for anyone designing AI for or implementing AI in hiring, especially employers whose hiring practices may be subject to public scrutiny (Gelles et al., 2018).

In the third study, I examine individuals', especially women's, preferences for AI evaluations. I find that individuals generally prefer a human evaluator to an AI evaluator in a competitive setting—but only if the human rater is female. Focusing on women, my results indicate a direct effect of women's belief in AI's potential to reduce bias and perceived personal discrimination but no direct effect of competitors' genders on women's preferences for AI evaluation. Moreover, I show that individuals' belief in AI to reduce bias moderates the other two relationships. Women with strong beliefs in AI's potential to reduce bias are more likely to choose AI when competing in a mixed-sex group than when competing in a single-sex group. However, if women believe that AI is more biased than humans, the effect is reversed. Regarding the effect of perceived discrimination, women's belief in AI has a reinforcing effect; for women with strong beliefs about AI, the positive influence of perceived personal discrimination on their preference for AI is even stronger. Overall, the study provides preliminary evidence that the use of AI evaluation can reduce expected biases in the hiring context and thus encourage women to apply for jobs in male-dominated fields.

The study makes several contributions. By identifying new factors driving individuals' preferences related to AI assessment, it contributes to the experimental literature on applicant reactions to AI-enabled hiring practices (Acikgoz et al., 2020; Langer et al., 2018; Newman et al., 2020) as well as to the research field investigating algorithm aversion (Castelo et al., 2019; Dietvorst et al., 2015; Jauernig et al., 2022). However, in contrast to most existing studies, which examine stated preferences about peoples' perceptions of AI hiring tools (Acikgoz et al., 2020; Langer et al., 2018; Langer et al., 2020; Langer, König, Sanchez, & Samadi, 2019), my study is focused on individuals' revealed preferences regarding AI evaluation techniques in an economic experimental design. Moreover, this research contributes to the literature on gender differences in competition, by exploring women's preferences in competition and applying a new paradigm—namely, by employing video interviews as tasks in the tournament. Furthermore, my research expands the literature on gender-related occupational segregation in the labor market by examining AI's potential as a new determinant to address anticipated bias in human recruitment and gender-based selection into certain career paths. In practice, the insights from this study can share with companies information about potential applicants' preferences regarding AI evaluations. The results reveal that AI evaluation tools may especially attract women, encouraging them to apply for jobs in male-dominated fields.

1.5 Dissertation outline and summary

This dissertation's structure reflects the three individual research studies. Following this introductory chapter, the next three chapters cover the three self-contained studies. The final chapter provides a discussion of the three studies and provides a summary of the findings as well as academic and managerial contributions before concluding the dissertation. Table 1.1 provides an overview of the three research studies that constitute this dissertation.

Table 1.1: Overview of the research studies

	Study 1	**Study 2**	**Study 3**
Title	"Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda"	"How to Improve Fairness Perceptions of AI in Hiring: The Crucial Role of Positioning and Sensitization"	"Can AI Close the Gender Gap in the Job Market? Individuals' Preferences for AI Evaluations"
Research question	What are the ethical considerations related to AI recruiting? What are gaps in the extant literature that call for deeper exploration in future research?	To what extent does adjusting process design factors affect people's fairness perception of AI interviews and overall organizational attractiveness?	What are individuals', especially women's, preferences for an AI- versus a human-based evaluation process? How do women's belief in AI's potential to reduce bias, competitors' gender composition, and perceived personal discrimination affect women's preferences?
Data and method	• Systematic literature review	• Online vignette study • ANOVA • Mediation analysis	• Online experimental study • Binomial probability tests • Nonparametric tests • Regression analyses
Related literature	• AI ethics	• AI ethics • Applicant reactions to recruiting practices	• Applicant reactions to recruiting practices • Algorithm aversion • Gender differences in competition • Gender imbalances in the job market
Results	• The use of AI in hiring comes with diverse ethical opportunities, risks, and ambiguities. • The current literature on AI recruiting is rather practitioner oriented and lacks both theoretical and empirical research. • The current research is focused on algorithmic bias but neglects additional ethical topics, such as accountability and human autonomy. • Research on approaches to mitigate ethical risks in AI-based hiring lack domain-specific guidelines.	• People perceive AI interviews as fairer when they are positioned in the initial screening stage of the overall process (vs. the final decision stage) and when they are sensitized to AI's potential to reduce bias (vs. not sensitized to this potential). • People's fairness perception in turn affects overall organizational attractiveness Human oversight in the AI decision-making process does not significantly affect people's fairness perceptions.	• Individuals prefer a human evaluator to an AI evaluation when they have the choice, but only if the human evaluator is female. • The belief in AI's potential to reduce bias and perceived personal discrimination positively affect women's preference for AI evaluation. • Women are more likely to choose an AI evaluation in a mixed-gender group than in a single-gender group—but only if they have a relatively strong belief in AI's potential to reduce bias.

Table 1.1: Overview of the research studies (*continued*)

	Study 1	Study 2	Study 3
Scientific contributions	• Organizes current literature on ethical considerations of AI recruiting and selection • Develops ethical framework on AI recruiting's ethical opportunities, risks, and ambiguities • Identifies current research gaps and proposes an agenda for potential future research	• Links research on applicant reactions to selection procedures with research on AI ethics • Provides causal evidence on process design factors' effects on people's fairness perception of AI interviews • Provides guidance to organizations on how the process surrounding AI interviews should be designed so that people perceive them as fairer	• Provides causal evidence of individual and contextual factors' effects on women's preferences for AI evaluations • Studies subjects' preferences for AI-based evaluations in an economic experiment • Studies women's preferences in competition by applying a new paradigm—namely, video interviews • Studies AI's potential as a new determinant to address anticipated biases in hiring and gender-based self-selection
Coauthors	Christoph Lütge	Christoph Lütge	Christoph Hohenberger
Lead author contributions	• Developing the research question and design was a joint effort by both authors • Conducting the literature search and analysis • Writing the article • Revising the article was a joint effort by both authors	• Developing the research question and design • Deriving the hypotheses was joint effort by both authors • Programming and conducting the experimental vignette study • Conducting the data analyses • Writing the article • Revising the article	• Developing the research question and design—I made the main contribution • Deriving the hypotheses • Programming the experiment was a joint effort by both authors • Conducting the experiment • Conducting the data analyses • Writing the paper
Publication status and outlet	Published, *Journal of Business Ethics,* published online, February 8, 2022	Published, *The AI Ethics Journal*, 2(2)-3 (2021)	Submitted, *Organizational Behavior and Human Decision Processes*

2 Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda

Abstract

Companies increasingly deploy artificial intelligence (AI) technologies in their personnel recruiting and selection process to streamline it, making it faster and more efficient. AI applications can be found in various stages of recruiting, such as writing job ads, screening of applicant resumes and analyzing video interviews via face recognition software. As these new technologies significantly impact people's lives and careers but often trigger ethical concerns, the ethicality of these AI applications needs to be comprehensively understood. However, given the novelty of AI applications in recruiting practice, the subject is still an emerging topic in academic literature. To inform and strengthen the foundation for future research, this paper systematically reviews the extant literature on the ethicality of AI-enabled recruiting to date. We identify 51 articles dealing with the topic, which we synthesize by mapping the ethical opportunities, risks and ambiguities, as well as the proposed ways to mitigate ethical risks in practice. Based on this review, we identify gaps in the extant literature and point out moral questions that call for deeper exploration in future research.

Keywords:	artificial intelligence, algorithmic hiring, employee selection, ethical recruitment, ethics of AI, bias of AI

Authors:	Anna Lena Hunkenschroer, Christoph Lütge

Link to article:	https://doi.org/10.1007/s10551-022-05049-6

3 How to Improve Fairness Perceptions of AI in Hiring: The Crucial Role of Positioning and Sensitization

Abstract

Companies increasingly deploy artificial intelligence (AI) technologies in their personnel recruiting and selection processes to streamline them, thus making them more efficient, consistent, and less human biased. However, prior research found that applicants prefer face-to-face interviews compared with AI interviews, perceiving them as less fair. Additionally, emerging evidence exists that contextual influences, such as the type of task for which AI is used, or applicants' individual differences, may influence applicants' reactions to AI-powered selection. The purpose of our study was to investigate whether adjusting process design factors may help to improve people's fairness perceptions of AI interviews. The results of our 2 x 2 x 2 online study (N = 404) showed that the positioning of the AI interview in the overall selection process, as well as participants' sensitization to its potential to reduce human bias in the selection process have a significant effect on people's perceptions of fairness. Additionally, these two process design factors had an indirect effect on overall organizational attractiveness mediated through applicants' fairness perceptions. The findings may help organizations to optimize their deployment of AI in selection processes to improve people's perceptions of fairness and thus attract top talent.

Keywords: artificial intelligence, algorithmic hiring, employee selection, applicant reactions, fairness perception, trustworthy AI

Authors: Anna Lena Hunkenschroer, Christoph Lütge

Link to article: https://doi.org/10.47289/AIEJ20210716-3

4 Can AI Close the Gender Gap in the Job Market? Individuals' Preferences for AI Evaluations

Abstract

Gender imbalances in the labor market continue to be an economic and social problem that could be reduced by artificial intelligence (AI), which is being promoted as a means for fairer and less biased hiring practices. To examine whether these supposed benefits of AI are perceived as such, we have investigated the preferences of individuals, particularly women, for an AI-based evaluation process in a competitive situation. The results of our experimental study (*N* = 152) show that individuals generally prefer a human evaluator over an AI evaluator—but only if the human evaluator is female. Whereas we demonstrate that women's beliefs in AI to reduce bias and perceived personal discrimination have a positive direct effect, we find no direct effect of the competitors' gender on women's preference for an AI evaluation. However, we find that the belief in AI moderates the other two relationships, which highlights the crucial role of people's general perception of AI tools in realizing AI's full potential and reduce anticipated biases. Our findings provide an initial indication that the use of AI technology in hiring could encourage women to apply for jobs in male-dominated fields and serve as a starting point for future research in this field.

Keywords: artificial intelligence, algorithm aversion, gender, competitive behavior, perceived discrimination

Authors: Anna Lena Hunkenschroer, Christoph Hohenberger

4.1 Introduction

The Global Wage Report (2020) of the International Labour Organization (ILO) reveals that gender imbalances in the labor market in the form of occupational segregation remain a major economic and social concern. For example, they contribute to the gender wage gap and companies risk missing out on the benefits that gender diversity in the workforce can bring in terms of business performance (Sinclair et al., 2019). The lack of women in well-paid, male-dominated fields, such as STEM[3] or in top-level positions in general, cannot be explained by differences in experience, education, or skills (International Labour Organization, 2018). Instead, besides structural overt and covert discrimination (Hoover et al., 2019; Nier & Gaertner, 2012; Sinclair & Carlsson, 2021), an increasingly important determinant arises from women (and men) self-selecting into certain occupations (e.g., Ceci et al., 2009; Parker et al., 2012; Sinclair et al., 2019; Sinclair & Carlsson, 2013). This self-selection is driven by several factors: Aside from internalized gender stereotypes by women that manifest in gender-typical preferences and interests (Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H., 2000; Ceci et al., 2009), the anticipation of a systematically biased hiring process in specific occupations, in which they are discriminated against, is an important factor in this regard (e.g., Carlsson & Sinclair, 2018; Pinel & Paulin, 2010). In addition, research identified another explanation for why women self-select out of applying for male-dominated and competitive high-profile jobs: the behavior of men and women differ in competitive environments. Thus, it was found that, regardless of their actual performance, women generally tend to shy away from competition against male competitors—a finding that can also be applied to the job market (e.g., Buser et al., 2014; Niederle et al., 2008; Niederle & Vesterlund, 2008).

Due to the negative effects of occupational gender segregation, companies are looking for quick and effective means to achieve greater gender balance in the labor market. To this end, organizations have already widely introduced affirmative action policies such as quotas. However, quotas come with possibly serious caveats, among which are economic losses in terms of effort and efficiency, societal losses in the form of their risk of reinforcing existing narratives, as well as image losses because people tend to react negatively to them (e.g., Balafoutas et al., 2016; Crosby et al., 2003; Neschen & Hügelschäfer, 2021).

In our study, we investigate the recent development of artificial intelligence (AI) tools for hiring processes, which postulate fairer and less biased hiring processes (Polli et al., 2019)

[3] A STEM job is any job in the fields of science, technology, engineering, or math.

and may therefore constitute a promising alternative to quotas, mitigating the undesirable but persistent gender imbalances in the labor market. Based on the claim of AI technology providers that their products reduce biases (e.g., Chamorro-Premuzic, 2019; Hmoud & Laszlo, 2019; Polli et al., 2019), the integration of AI into hiring processes could resolve both actual and anticipated biases against female applicants. If the latter indeed believe that AI assessment procedures are less prone to gender bias, they would not need to fear being discriminated against during the selection process. This could diminish the severity of the gender gap in competitive behavior, as well as the self-selection of women away from male-dominated fields, because gender as a factor would lose in salience.

However, public perception of AI's ability to reduce biases is multifaceted. While many practitioners embrace the use of AI in hiring due to efficiency gains (e.g., Chamorro-Premuzic & Akhtar, 2019; Florentine, 2016), researchers also warn of potential algorithmic biases due to poor data sets with which the AI is trained (e.g., Vasconcelos et al., 2018; Williams et al., 2018). This was illustrated by the case of Amazon in 2018, where a tested AI in personnel selection was quickly reinstated after systematically discriminating against women; this incident resulted in public image damage for both Amazon and AI (Dastin, 2018). However, it is precisely this public perception of AI that is crucial. It determines the extent to which AI can counteract the expectation of a systematically biased hiring process by potential female applicants.

In this paper, as an initial assessment of whether AI could currently exploit this described potential, we study individuals', and especially women's, preferences about an AI- versus human-led evaluation process and examine whether these preferences vary by (i) their belief in AI's potential to reduce human bias, (ii) the gender composition of their competitor pool, and (iii) the extent to which they have perceived personal discrimination in the past. We assume that women will have a higher preference for AI-based evaluation when they perceive AI to reduce biases. Moreover, we expect women to have a higher preference for an AI evaluator when competing in a mixed-gender group with men and facing the possibility of being discriminated against based on their gender than when competing in a single-gender group where they do not face gender discrimination. The same tendency is expected if women have already been discriminated against in the past and are therefore more sensitive to potential bias in their environment. In addition, we also examine whether women's belief about AI moderates the latter two effects.

To investigate our research question, we conducted an online experiment with 152 participants, who were assigned to groups of four and participated in an incentivized tournament. In the tournament, participants conducted a video interview, where the person evaluated as the most autonomous within a group won the competition. After recording the interview, participants were asked to indicate their preferred method of evaluation among (i) an AI, (ii) a female, or (iii) a male human recruiter. Participants' payout was determined solely by their performance as evaluated by their method of choice. We exogenously varied the gender composition of an individual's competitor pool by randomly assigning participants to either single- or mixed-gender groups. This design allowed us to examine the causal impact of the gender composition of the competitor pool on participants' preferences for the method of evaluation.

Because we employed an economic experiment with monetary incentives, the choices elicited in our study reflect participants' true preferences. In this way, our results are less likely to be biased, for example by the expression of social desirability than those of a vignette study (Grimm, 2010). Rather, choosing a decision-making entity (AI or human) comes with actual monetary costs and benefits. In accordance with the standards of economic research laboratories, no deception was used on any of the participants, and the entire procedure of the experiment was made transparent to all participants beforehand (e.g., Hertwig & Ortmann, 2001).

Our study relates to and expands the following research fields: First, it contributes to the literature on humans' behavioral responses to machine-based decision-making as well as their aversion against and acceptance of algorithms (e.g., Castelo et al., 2019; Dietvorst et al., 2015; Jauernig et al., 2022). More specifically, we examine individuals' reliance on algorithms in the context of personality assessment practices and study different factors driving their preference for AI-based decision-making.

Second, our work contributes to the experimental literature on applicants' reactions to technology-based recruiting practices (e.g., Acikgoz et al., 2020; Hunkenschroer & Lütge, 2021; Langer et al., 2018; Newman et al., 2020). While being the first study to elicit participants' incentivized revealed preferences for AI-based assessment, we investigate women's belief in AI's potential to reduce bias, competitors' gender composition, and perceived personal discrimination as contextual factors affecting applicant reactions.

Third, this research relates to an extensive literature on experimentally elicited gender differences in competition, especially against men, and potential means to mitigate this gap in competitiveness (e.g., Balafoutas & Sutter, 2012; Berger et al., 2020; Maggian et al., 2020; Niederle et al., 2013). While quotas offer such a means through a priori determining a higher degree of gender-specific competition, we introduce the extent of expected bias in the method of performance evaluation employed (AI versus human) as a potential novel factor to reduce the gender gap in competition. Moreover, we explore competitive behavior in a new paradigm, namely by employing video interviews as task in the competition tournament.

Lastly, we contribute to the literature on gender-related occupational segregation in the labor market, which identified actual and anticipated (gender-)related biases in (human) personnel selection as one driver for women's self-selection into certain occupations (e.g., Carlsson & Sinclair, 2018; Heilman & Okimoto, 2007; Moss-Racusin et al., 2012). Our study expands this literature stream by examining the potential of AI as a new determinant to address the anticipated bias in human recruitment as well as gender-based selection into certain career paths.

4.2 Background and hypotheses

4.2.1 Preferences for AI evaluators versus human evaluators

Prior literature has found a generally high level of algorithmic aversion among individuals (Castelo et al., 2019; Dietvorst et al., 2015, 2018; Gogoll & Uhl, 2018). For example, Dietvorst et al. (2015) found that individuals lose trust in algorithms more easily and strongly than they reasonably should once they have observed an error. Logg et al. (2019) also coined the term algorithm appreciation by showing in several experiments that people may also rely on algorithmic rather than human advice in making decisions. However, these results apply to people with decision-making power and not to people who are directly affected by such algorithmic decisions. This suggests that people might not doubt the quality of algorithmic decisions, but still reject them for themselves (Jauernig et al., 2022). This was also found in a European survey in 2020, where 64% of respondents agreed with the statement: "Algorithms might be objective, but I feel uneasy if computers make decisions about me. I prefer humans make those decisions."[4]

AI faces social resistance, especially in morally sensitive domains where algorithms make decisions that directly affect peoples' lives, such as hiring (Jauernig et al., 2022). Several studies of applicant reactions towards the use of AI in hiring (Acikgoz et al., 2020; e.g., M. K. Lee, 2018; Newman et al., 2020) revealed a low acceptance of AI-led interviews and evaluations compared to face-to-face-interviews and human raters. For example, M. K. Lee (2018) found that participants felt that AI lacked certain human skills required in the hiring context, e.g., AI lacks human intuition, makes judgments based on keywords, and ignores traits that are hard to quantify. Other studies found that while AI evaluations were not perceived as less fair, they were still less preferred by applicants due to less social presence (Langer, König, Sanchez, & Samadi, 2019) or due to less behavioral control and greater privacy concerns (Langer, König, & Papathanasiou, 2019). Hence, we generally expect a lower preference for an AI evaluator compared to a human evaluator, regardless of whether the latter is a female or male evaluator.

Hypothesis 1: More individuals prefer a human evaluator over an AI evaluator than vice versa.

Hypothesis 1a: When choosing an evaluation method, individuals rank AI evaluators lower than female evaluators.

Hypothesis 1b: When choosing an evaluation method, individuals rank AI evaluators lower than male evaluators.

We do not expect significant differences between women and men in this low preference for AI evaluation, as we expect that different potential effects may come into play here: On the one hand, women report higher technology anxiety than men (e.g., Gilbert et al., 2003). On the other hand, men are more likely to rely on impression management[5] in job interviews (Singh et al., 2012), which is generally used less when interviews are evaluated by an AI (Langer et al.,

[4]https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WhatEuropeKnowsAndThinkAboutAlgorithm.pdf (retrieved on February 21, 2022).

[5] Impression management refers to applicants' attempts to influence interviewers' evaluations and decisions through a variety of tactics, which can be honest but also deceptive, to improve their chances of being hired (Roulin and Bourdage 2017). It includes tactics such as concealing one's negative characteristics, exaggerating one's influence on positive results, or lying about past work experiences (Langer et al. 2020).

2020). Consequently, AI evaluations may appear less attractive to men. Additionally, experiments on applicant reactions to video interviews have not identified any differences in perceptions across women and men (e.g., Basch & Melchers, 2019; Brenner et al., 2016; Suen et al., 2019). Thus, we derive the following hypothesis:

> **Hypothesis 1c:** When choosing an evaluation method, both women and men rank an AI evaluator lower than a human evaluator.

4.2.2 Factors impacting women's preference for AI evaluation

Although we expect a generally low preference for AI evaluation among both women and men, we believe that women's preference for AI evaluations may be positively influenced by several factors. First, we expect that women's belief in AI's ability to reduce human bias will influence their preference for AI as an evaluation method. This is consistent with previous research on algorithm aversion. The putative benefit of reducing AI bias becomes particularly relevant in situations where individuals expect to be discriminated against. Therefore, we examine two additional factors that may influence the extent of expected bias in an assessment situation and how these in turn influence women's preference for an AI evaluator. To this end, we examine how the gender composition of competitors and women's perceived past personal discrimination affect their preference for an AI evaluation. In this way, we also link to the literature on gender differences in competition and research on reactions to perceived discrimination.

4.2.2.1 Belief in AI's potential to reduce bias

Prior literature on algorithmic aversion found that perceived algorithm capabilities were an important driver for aversion or appreciation of AI-based decision-making (Jussupow et al., 2020). AI is often perceived as lacking the human capabilities which are necessary for certain tasks. This perception has been shown to drive people's aversion. For example, in the aforementioned study by M. K. Lee (2018), individuals were shown to deny that AI could have human intuition capabilities. Moreover, people perceive AI to lack having a "mind" (i.e., skills such as thinking or communicating with others (Bigman & Gray, 2018)), the ability to take into account qualitative information and contextualization (Newman et al., 2020), and to account for people's unique characteristics (Longoni et al., 2019). However, it was also shown that the aversion effect was reduced when the appearance of the algorithm was adapted to the required capability (Castelo et al., 2019; Longoni et al., 2019).

In assessment situations related to personnel decisions (e.g., hiring, firing, promotions), where a human is the subject of the decision, it has been shown that fairness is an important aspect for individuals (McCarthy et al., 2017; Weaver & Trevino, 2001). Indeed, scholars have long recognized the importance of understanding and improving the perceived fairness of decision-making procedures (Colquitt et al., 2001). They found that perceived fairness influences, for example, organizational attractiveness and people's behavior, such as whether individuals would recommend the company to others (e.g., Bauer et al., 1998). People have been shown to perceive decision procedures as more fair when they are consistent, based on accurate information, and not influenced by personal biases of decision-makers (Brockner, 2006; Leventhal, 1980). In the context of AI hiring, it has already been shown that people perceive an AI interview to be fairer if they have been sensitized to the fact that AI has the potential to reduce human bias (Hunkenschroer & Lütge, 2021). Fair and bias-free selection processes are especially important to women, who have been shown to be often disadvantaged in hiring processes for male-dominated occupations (e.g., Sinclair & Carlsson, 2021). Therefore, they self-select into occupational fields where they expect less biased selection processes (e.g., Carlsson & Sinclair, 2018; Pinel & Paulin, 2010).

Although we acknowledge that fairness, and in particular bias, is not the only important aspect, we derive from previous research findings that perceived AI bias is a crucial factor in the adoption of or aversion against AI-based evaluation methods of women. Thus, we hypothesize that the more a woman believes that AI has the ability to reduce bias, thus enabling a fair decision-making process, the lower is her aversion to AI.

Hypothesis 2: Women's belief in AI's potential to reduce human bias positively affects women's preference for AI evaluation.

4.2.2.2 Gender composition of competitors

Prior literature on competition behavior has shown that the fact of whether women compete in a single-gender versus a mixed-gender environment affect their behavior, preferences, and performance (Gneezy et al., 2003; Inzlicht & Ben-Zeev, 2000; Niederle et al., 2008). For instance, Gneezy et al. (2003) found a significantly larger gender gap in performance on solving mazes in a mixed-gender competitive environment compared to a single-gender competitive environment. This difference is driven largely by women's failure to perform at a high level when competing against men. Niederle et al. (2008) studied women's behavior in an affirmative action tournament, where for every two winners at least one winner had to be a woman. This made the competition more gender-specific, because under this rule a woman already won the competition if she performed better than her female competitors. The authors showed that women behaved differently in this setting compared to a mixed-gender competition and entered the competition more frequently. On the one hand, this may be due to different beliefs on relative performance within versus across gender. On the other hand, women seem to view competition as more intimidating and less enjoyable when competing against groups with men represented compared to groups without men represented. The latter reason is similar to the argument given for the benefits of single-gender education for girls (Harwarth et al., 1997; Solnick, 1995). In our study, we test whether women have different preferences for an AI evaluation when competing against men than when competing against only women.

There is strong evidence that hiring decisions are frequently biased against women regarding jobs that are traditionally or predominantly held by men (e.g., Heilman & Okimoto, 2007). This also speaks to the prototype perspective of perceived discrimination, which suggests that men do not fit the prototype of a typical discrimination victim (Carlsson & Sinclair, 2018; Inman & Baron, 1996; Sinclair & Carlsson, 2021). Rather, people are more inclined to perceive women as victims of gender discrimination than men (Carlsson & Sinclair, 2018)**.** Thus, we believe that an increasing ratio of men among the competitors leads to a higher degree of anticipated discrimination among women in a selection situation. Addressing this concern of discrimination in applicant assessment and hiring, providers of AI-powered evaluation tools advertise their products with the claim that they reduce human bias (Chamorro-Premuzic, 2019; Polli et al., 2019; Polli, 2019). Therefore, women should rank AI evaluations (versus human evaluations) higher in situations where they compete in mixed-gender groups and anticipate being discriminated against versus in a single-gender environment. Thus, we derive following hypothesis:

> **Hypothesis 3:** Women’s preference for AI is higher when competing in a mixed-gender group versus in a single-gender group.

4.2.2.3 Perceived Personal Discrimination

Over the past two decades, the impact of discrimination on individuals has been increasingly explored in social psychology (Sechrist et al., 2004). Social psychologists have focused on determining how targeted individuals perceive (e.g., Inman & Baron, 1996) and react to discrimination (e.g., Major et al., 2002). In this context, empirical evidence suggests that members of chronically stigmatized groups, such as women, are more sensitive to cues of discrimination in their environment (Rodin et al., 1990). Furthermore, within stigmatized groups, there are individual differences in the extent to which individuals expect others to discriminate against them, which is strongly and positively correlated with perceived personal discrimination and perceived group discrimination (Pinel, 1999).

Crocker and Major's (1989) concept of “attributional ambiguity” describes the uncertainty of stigmatized individuals about whether outcomes are indicative of their own deservingness or of social prejudices that others have. Potential discrimination consequently arises from personal interaction with other people. We therefore derive that women who have perceived personal discrimination in the past are more sensitive to potential discrimination in their environment. Thus, they seek to reduce attributional ambiguity by choosing the evaluation by an AI evaluator instead of a human rater.

Hypothesis 4: Perceived personal discrimination positively affects women’s preference for AI evaluation.

4.2.3 Belief in AI as a moderating factor

The larger extent to which women fear being discriminated against in a competitive situation does not necessarily lead to a higher preference for AI evaluation. In a competitive setting, in which women anticipate discrimination, they will only choose the AI evaluation if they truly believe that AI has the potential to reduce bias and discrimination compared to human evaluators. If this is not the case and women believe that humans are the more objective raters, they will be more likely to choose the human raters in order to avoid being discriminated against. Langer, König, Sanchez, and Samadi's (2019) study results demonstrated that applicants perceive AI interviews as more consistent than human interviews, pointing in the direction that AI might also be perceived as more objective and less biased than human evaluators. However, this belief is not self-evident in our current society.

We therefore hypothesize that the belief in AI’s potential to reduce human bias will not only directly affect women’s preference for an AI evaluator, but that it will also act as a moderator in our conceptual model. More specifically, the positive effect of a mixed-gender group versus a single-gender group on women’s preference for AI only applies to women who believe that AI is less biased than humans.

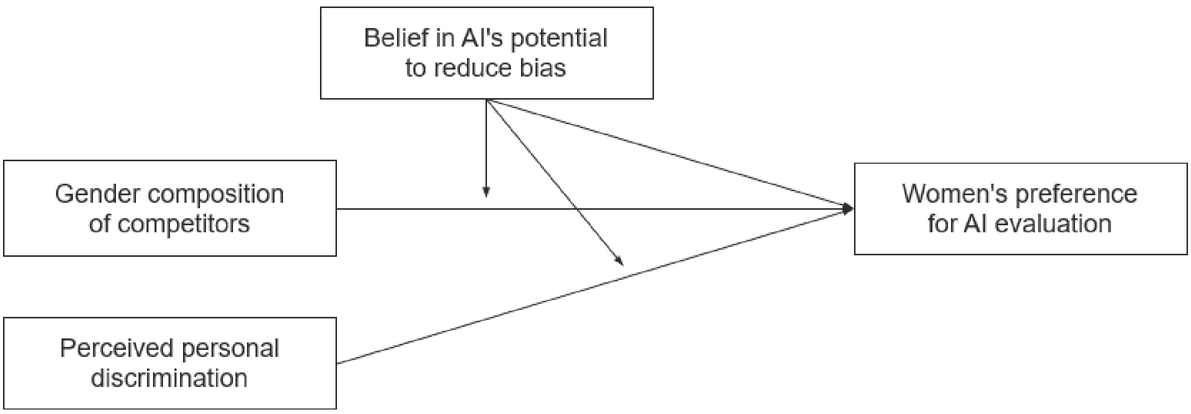
Hypothesis 5: The positive effect of a mixed-gender group versus a single-gender group on women's preference for AI evaluations only applies to women who have a strong belief in AI's potential to reduce bias.

Analogously, we derive that women's belief in AI will moderate the relationship between perceived personal discrimination and women's preference for AI evaluation. More specifically, the positive effect of perceived personal discrimination on women's preference for AI evaluations will be stronger for women with a strong belief in AI's potential to reduce bias.

Hypothesis 6: The positive effect of perceived personal discrimination on women's preference for AI evaluations becomes stronger with an increasing belief in AI's potential to reduce bias.

The conceptual model of our hypotheses is shown in Figure 4.1.

Figure 4.1: The proposed conceptual model



4.3 Method

To test our hypotheses, we conducted a 2 x 2 between-subject online experiment. During the experiment, participants were incentivized to reveal their preference on whether they want to be evaluated by (i) an AI, (ii) a woman, or (iii) a man in a competition. In the experiment, participants competed by solving a one-shot task; namely, recording a video interview of themselves. Afterwards they were evaluated by the evaluator they had chosen.

4.3.1 Sample

We conducted the experiment online with students from the subject pool of experimenTUM, the experimental laboratory at the Technical University of Munich. Prerequisite for participation was that subjects spoke fluent German to ensure that our results were not driven by a lack of linguistic competence. Moreover, participants had to agree to record a video of themselves during the experiment. From December 13–16, 2021, we conducted six online sessions with a total of 175 subjects. Overall, 171 participants finished the experiment. For the evaluation of the tournament and our analysis, we excluded all subjects (13) who did not correctly upload a video recording. Furthermore, we excluded subjects (5) whose video recording could not be evaluated, for example due to no or a too low volume. This left N = 152 participants in the final sample (78 females, 74 males; age: M = 24.1 years, SD = 5.0). 111 Participants (73%) reported having a German nationality. Regarding their highest educational level, 11% had a master's degree, 38% had a bachelor's degree, 1% had completed an apprenticeship, and 50% had a high school diploma. Most participants (90%) were students.

Depending on their gender, the 152 participants were randomly assigned to either single-gender or mixed-gender groups of four[6] (including themselves) and entered a tournament competition, in which only the winner received a bonus of €10.00 on top of a show-up compensation of €4.00. This subgroup design with a group size of four was chosen to better reflect the real recruitment and selection situation, where after an initial CV screening and an optional phone interview a small group of remaining applicants competes for the open position.[7]

4.3.2 Design and procedure

In the tournament competition, participants had to complete a short asynchronous video interview in which they answered the question: "When did you face a difficult challenge and how did you overcome it?" They were informed that the video recording would be evaluated afterwards and that the person who was judged to be the most autonomous person within the group would win the tournament. Autonomous people were characterized by planning and

[6] All mixed-gender groups consisted of two women and two men.

[7] See https://careersidekick.com/interviews-per-job/ (retrieved on February 21, 2022) for a description of a typical selection process and corresponding statistics.

accomplishing tasks independently with little or no guidance, as well as by making decisions on their own. We chose autonomy as the critical attribute because it is often used in job requirement descriptions and because we expected it to be perceived as gender-neutral among subjects based on prior gender stereotype research findings (Spence & Buckner, 2000). Neither gender was therefore a priori disadvantaged in the competition due to the design itself.

After the interview recording, participants were informed about the gender composition of the group they were competing in. Subsequently, they could decide how their own recording should be evaluated for the determination of their payoff. They were asked to rank the following three methods of evaluation according to their preference: i) an AI that evaluates behavioral cues such as facial expression, body language, and voice to derive a behavioral personality profile, ii) a female human recruiter with several years of interview experience, or iii) a male human recruiter with several years of interview experience.[8] Subjects were informed that the evaluation was done by the first ranked entity 90% of the time and based on the second ranked entity 10% of the time. Participants would only win the tournament and receive the payout of €10.00 by coming in at first place based on the evaluation method determined as described. Consequently, a participant's performance was evaluated by all three methods of evaluation, but only the method resulting from a random-draw between first- and second-ranked methods of that participant with a probability distribution of 90%–10% was relevant for determining their payout.[9]

Lastly, participants had to complete a questionnaire in which we asked them about their beliefs, experiences, and demographics.

4.3.3 Performance evaluation

We measured the actual performance of the participants in the tournament competition based on all three mentioned evaluation methods (by an AI software, a female recruiter, and a

[8] We want to note here that there are individuals who do not feel represented by either the female or male gender but consider themselves diverse. However, given our research focus, we wanted to limit our study and focus on preferences for female and male human evaluators as well as on the preferences of the group of female participants.

[9] If the participants in a group chose different preferred evaluation methods, there was a possibility that there were two or even three winners within a group who would receive the winner's payout. An alternative way to deal with this situation was to tell participants that the final evaluation method would be determined by a random drawing from the first ranked evaluation methods of all four group members; however, this option had the drawback of introducing a great deal of uncertainty for participants about which evaluation method would be used in the end, which would have reduced the incentive to make their choice deliberately and thoughtfully.

male recruiter). To this end, a total of four professional recruiters, two male and two female recruiters, took part in the experiment and each participant was individually evaluated in terms of their autonomy by one male and one female recruiter. For the AI rating, we used the AI software of the startup Retorio, a German AI video interview platform provider. Participants' interview performance was evaluated on an absolute scale ranging from 0% (low degree of autonomy) to 100% (high degree of autonomy). Moreover, the recruiters assigned participants a rank in their group of four. All four raters were professional recruiters and interviewers with a business background and professional experience in applicant selection. The average standardized rating across the three evaluation methods per participant was used as an overall score for interviewees' interview performance.

4.3.4 Measures

The additional measures employed applied 5-point Likert scales.

Belief in AI's potential to reduce bias was measured using three self-developed items (Cronbach's α = .75). Participants had to indicate to what degree they agreed with the statements. The following are samples of the items used: "I believe that decisions made by AI are less biased compared to those made by humans," and "I think algorithms can make more objective decisions than humans."

Perceived personal discrimination was measured with four items (Cronbach's α = .90) assessing whether participants had personally experienced discrimination in the past. This measure was adapted from past research (Carvallo & Pelham, 2006; Sechrist et al., 2003). Participants had to indicate to what degree the statements would apply to them. The following are samples of the items used: "I personally have already experienced discrimination," and "I have often been treated unfairly in the past because of my group membership (e.g., religion, gender, etc.)."

4.3.5 Statistical analyses

Data analyses were performed by using RStudio. For analyzing individuals' general preference for the AI evaluation method compared to human evaluation methods, we looked at the entire participant group with male and female subjects ($N = 152$). To test our hypotheses, we conducted binomial probability tests and pairwise Wilcoxon signed-rank tests. In the following analyses, we were specifically interested in women's preference for AI evaluation. Therefore, we focused on the sub-sample of female subjects (N = 78). To test the related hypotheses, we used ordered logistic regression to examine the main effects and interaction effects. As a robustness check, we conducted the same regression analyses using an Ordinary Least Squares (OLS), a binary logistic, and a probit regression model (see Appendix). All analyses showed the same tendencies.

4.4 Results

4.4.1 General preferences for AI evaluation compared to human evaluation

Descriptive results regarding the distribution of participants' preferences for the three different evaluation methods can be inferred from Figure 4.2. Focusing on individuals' first ranked evaluation method, results show that significantly more than half of all participants prefer a human evaluator over the AI evaluator ($p < .001$, *one-sided binomial probability test*[10]), supporting Hypothesis 1. The results of the pairwise Wilcoxon signed-rank tests are shown in Table 4.1. They indicate that the AI evaluator (*Mdn* = 1) is ranked significantly lower than the female recruiter (*Mdn* = 3) with a medium to large effect size[11], $p < .001$, $r = .459$. However, the AI evaluation method is not ranked significantly lower than the male recruiter (*Mdn* = 2), $p = .254$, $r = 0.054$. Thus, Hypothesis 1a is supported, whereas Hypothesis 1b is not supported. It is interesting to note that male evaluators are ranked significantly lower than female evaluators with a medium to large effect size, $p < .001$, $r = .447$.

To test Hypothesis 1c, we conducted the same analyses separately for female and male participants. Results show that among both women ($p < 0.001$, *BPT*) and men ($p < 0.001$, *BPT*), significantly more than half of the participants prefer a human over an AI evaluator. Thus, Hypothesis 1c is supported. Also, the significant differences between the evaluation methods shown by the Wilcoxon tests in the full sample are replicated in both subgroups: Both women and men rank female recruiters significantly higher than AI evaluation and male recruiters; the results indicate medium to large effect sizes.

[10] In the following denoted as *BPT*.

[11] Effect sizes according to Cohen (1992): r = .10 small, r = .30 medium, r = .50 large.

Figure 4.2: Distribution of individuals' preferences for evaluation methods (N = 152)

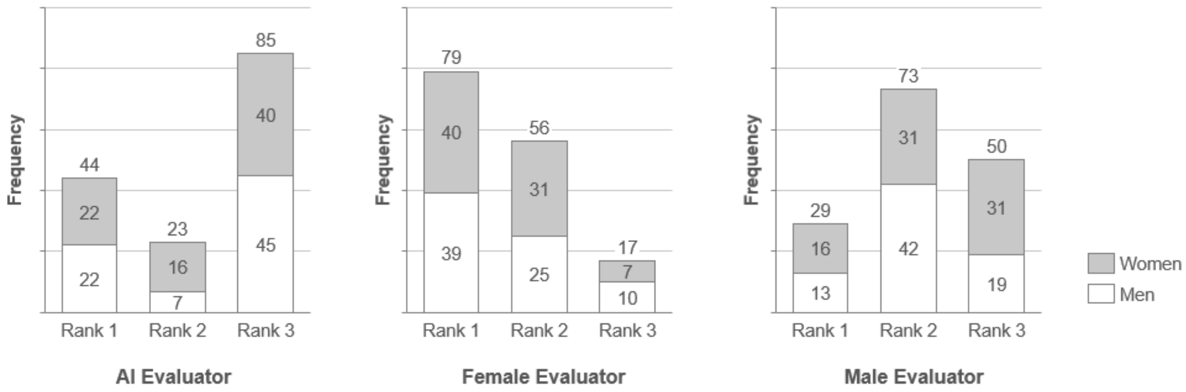


Table 4.1: Results of one-sided *Wilcoxon signed-rank tests*

	Mdn – Mdn	z	p	Pearson's r
Full sample (N=152)				
Pref for AI < Pref for Woman	1–3	-5.66	0.000***	0.459
Pref for AI < Pref for Man	1–2	-0.66	0.254	0.054
Pref for Man < Pref for Woman	2–3	-5.51	0.000***	0.447
Women subgroup (N=78)				
Pref for AI < Pref for Woman	1–3	-4.12	0.000***	0.467
Pref for AI < Pref for Man	1–2	0.09	0.535	0.010
Pref for Man < Pref for Woman	2–3	-4.18	0.000***	0.474
Men subgroup (N=74)				
Pref for AI < Pref for Woman	1–3	-3.87	0.000***	0.450
Pref for AI < Pref for Man	1–2	-1.09	0.138	0.127
Pref for Man < Pref for Woman	2–3	-3.57	0.000***	0.415

Note: Ordinal scale for ranking of evaluation methods was reversed: 1 = Rank 3, 2 = Rank 2; 3 = Rank 1.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

4.4.2 Main effects: Factors impacting women's preference for AI evaluation

Table 4.2 shows the means (M), standard deviations (SD), and correlations among the study variables. To test Hypotheses 2, 3, and 4, we estimated the direct effects of our theoretical model. We performed an ordered logistic regression, in which women's belief in AI, the gender composition of the competitor pool, and perceived personal discrimination were entered as predictor variables, and preference for AI served as the dependent variable. Age and education[12] were included as covariates.[13] The results (see Table 4.3) show that there is a significant positive effect of women's belief in AI's potential to reduce bias on their preference for AI with a small to medium effect size[14] ($p < .01$, $f^2 = 0.095$), supporting Hypothesis 2. Contrary to our expectations, there is no significant impact of the gender composition of the competitor group on women's preference for an AI evaluator ($p = .782$, $f^2 = 0.000$). Thus, Hypothesis 3 is not supported. Moreover, a higher level of perceived personal discrimination leads to a higher ranking of the AI evaluation method with a small to medium effect size ($p < .01$, $f^2 = 0.090$), which supports Hypothesis 4. Overall, the tested explanatory variables collectively explain about $R^2 = 0.19$ of the variance in women's preference for an AI evaluator.

Table 4.2: Means, standard deviations, correlations, and Cronbach's alpha for the study variables

	Variable	M	SD	1.	2.	3.	4.	5.	6.
1.	Age	24.92	6.21	-					
2.	Education	6.15	1.28	.06	-				
3.	Belief in AI	3.32	0.87	.09	-.04	.74			
4.	Gender composition	0.49	0.5	.02	-.06	-.04			
5.	Personal discrimination	3.21	1.07	.00	-.04	-.01	0.13	.88	
6.	Preference for AI	1.77	0.87	-.04	-.06	.27*	-.01	.26*	-

Note: Variable 4 was constructed by dummy coding. Coding of "Gender composition": 1 = Mixed-gender group, 0 = All female group. Ordinal scale for "Preference for AI" was reversed: 1 = Rank 3, 2 = Rank 2, 3 = Rank 1, *N* = 78. Numbers in the diagonal represents Cronbach's alpha of the respective scale.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

[12] In our analyses, we treated "Education" as a metric variable ranging from 1 = "No degree" to 9 = "Doctoral degree." If we treated "Education" as a dummy variable, we yielded the same results (see Appendix).

[13] Results do not change when including participants' nationality as an additional covariate (see Appendix).

[14] Effect sizes according to Cohen (1992): $f^2 = .02$ small, $f^2 = .15$ medium, $f^2 = .35$ large.

Table 4.3: *Ordered logistic regression model* predicting women's preference for AI evaluation—Main effects

	Coef.	SE	Wald Z	p	95% CI Lower B.	95% CI Upper B.	Odds Ratio (OR)	ΔR^2	Cohen's f^2
Age	-0.02	0.04	-0.52	0.601	-0.10	0.05	0.98	0.001	0.002
Education	-0.09	0.18	-0.50	0.618	-0.45	0.27	0.91	0.003	0.003
Belief in AI	0.76	0.29	2.65	0.008**	0.22	1.34	2.13	0.086	0.095
Gender composition	-0.13	0.46	-0.28	0.782	-1.04	0.77	0.88	0.000	0.000
Personal discrimination	0.63	0.24	2.64	0.008**	0.18	1.11	1.87	0.082	0.090
chi^2 (5) = 14.01									
Pseudo R^2 = 0.189									
Pr (> chi^2) = 0.016									
Cohen's f^2 = 0.233									

Note: Coding of "Gender composition": 1 = Mixed-gender group, 0 = All female group. The effect size Cohen's f^2 for an individual variable is calculated as $f^2 = (R^2_{\text{all variables included}} - R^2_{\text{all other variables included}}) / (1-R^2_{\text{all variables included}})$.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

4.4.3 Interaction effects: Belief in AI's potential as moderating factor

To examine whether individuals' belief in the potential of AI to reduce bias moderates the association between gender composition and women's preference for AI evaluation (Hypothesis 5) as well as perceived personal discrimination and women's preference for AI evaluation (Hypothesis 6), we performed two moderation analyses (see Table 4.4 and Table 4.5). As expected, we found a significant moderation effect of the belief in AI on the relationship between a group's gender composition and women's preference for AI ($p < 0.05$, $f^2 = 0.074$), supporting Hypothesis 5. More specifically, results demonstrate that, on the one hand, women with a low level of belief in AI's potential to reduce bias (1SD below the mean), i.e., they believe that humans are less biased than AI, rank AI lower in a mixed-gender group than in a single-gender group, at a 90% significance level (B = -0.47, $p = 0.078$). On the other hand, for women with a high level of belief in AI (1SD above the mean), the effect is reversed. They rank AI higher in a mixed-gender group than in a single-gender group, although not at a significant level (B = 0.36, $p = 0.17$). The moderation effect is illustrated in the left graph of Figure 4.3. The increase in the model fit when adding the moderation to the model indicates a small to medium effect size, resulting in a model fit of $R^2 = 0.25$.

Table 4.4: *Ordered logistic regression model* predicting women's preference for AI evaluation—Interaction effects

	Model 1				Model 2			
	Coef.	SE	ΔR^2	Cohen's f^2	Coef.	SE	ΔR^2	Cohen's f^2
Age	-0.02	0.04	-	-	-0.02	0.04	-	-
Education	-0.07	0.19	-	-	-0.09	0.18	-	-
Belief in AI	0.11	0.41	-	-	-1.19	0.95	-	-
Gender composition	-4.36*	2.07	-	-	-0.27	0.47	-	-
Personal discrimination	0.56*	0.24	-	-	-1.26	0.91	-	-
Gender composition x Belief in AI	1.25*	0.60	0.06	0.074	-	-	-	-
Personal discrimination x Belief in AI	-	-	-	-	0.57*	0.28	0.05	0.071
	chi^2 (6) = 18.73				chi^2 (6) = 18.53			
	Pseudo R^2 = 0.245				Pseudo R^2 = 0.243			
	Pr (> chi^2) = 0.005				Pr (> chi^2) = 0.005			
	Cohen's f^2 = 0.325				Cohen's f^2 = 0.321			

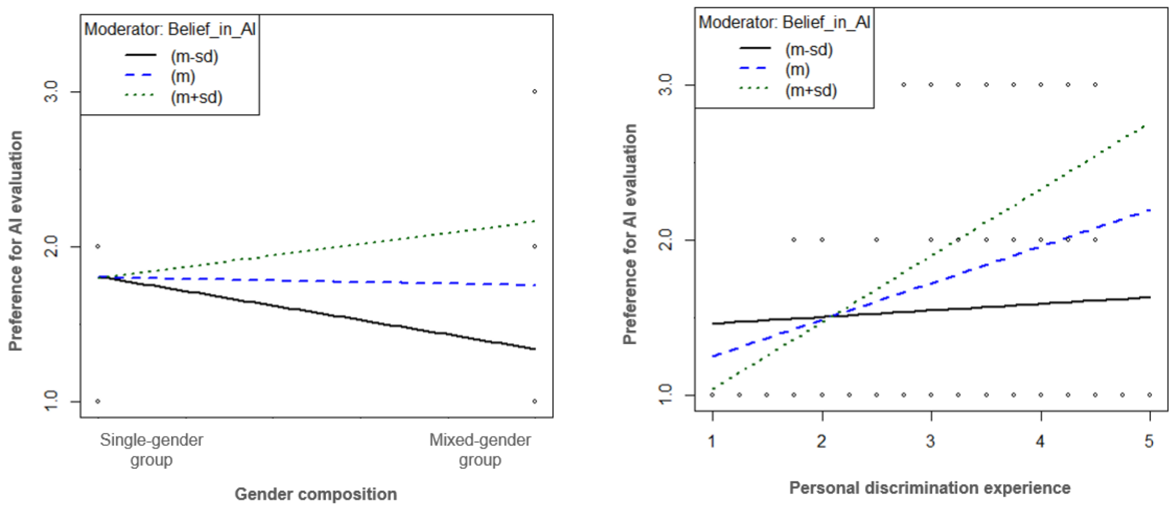
Note: .p < 0.1, *p < 0.05, **p < .01, ***p < .001.

Table 4.5: *Ordered logistic regression model* predicting women's preference for AI evaluation—Conditional effects of the focal predictor at values of the moderator (Belief in AI)

	Model 1 *Focal predictor: Gender composition*					Model 2 *Focal predictor: Personal discrimination*				
Moderator value	B	SE	p	95% CI Low. B.	95% CI Upp. B.	B	SE	p	95% CI Low. B.	95% CI Upp. B.
Mean - SD	-0.47	0.26	0.078.	-0.991	0.054	0.04	0.11	0.716	-0.184	0.266
Mean	-0.05	0.18	0.774	-0.421	0.315	0.23	0.09	0.008**	0.063	0.407
Mean + SD	0.36	0.26	0.174	-0.163	0.888	0.43	0.12	0.001***	0.185	0.673

Note: .p < 0.1, *p < 0.05, **p < .01, ***p < .001

Figure 4.3: Left: Moderating effect of belief in AI on the interplay between group composition and AI preference; Right: Moderating effect of belief in AI on the interplay between perceived personal discrimination and AI preference



Note: Low level of belief in AI corresponds to one standard deviation below the mean (m-sd); high level of belief in AI corresponds to one standard deviation above the mean (m+sd).

Moreover, results show a significant moderation effect of the belief in AI on the relationship between women's perceived personal discrimination and their preference for AI ($p < 0.05$, $f^2 = 0.071$), supporting Hypothesis 6. More specifically, results demonstrate that the positive effect of personal discrimination on individuals' preference for AI is stronger for women with a strong belief in AI. Thus, the effect is significant for women with high levels (1SD above the mean) of experienced discrimination (B = 0.43, $p < .001$) and for women with medium levels (mean) of experienced discrimination (B = 0.23, $p < .01$). However, it is not significant for those with low levels (1SD below the mean) of perceived personal discrimination (B = 0.04, $p = .71$). The increase in R^2 when adding this moderation to the model indicates a small to medium effect size, resulting in a model fit of $R^2 = 0.24$.

4.4.4 Additional analyses

Analyses on male sub-sample. Given the thematic focus of our study, we were primarily concerned with women's preferences for AI. Still, it is also interesting to see how men's results differ in this regard. Therefore, we also applied the above analyses to the sub-sample of men. Here, we found only one significant result, namely we observed a positive direct effect of men's belief in AI's potential to reduce bias on their preference for AI evaluation ($p < .01$). Apart from that we observed no direct effects of gender composition ($p = .713$) or of perceived personal discrimination ($p = .374$) on the preferences for AI. We note here that overall perceived personal discrimination was significantly lower for men ($M = 2.64$) than for women ($M = 3.21$, $p < .01$, *t-test*). Moreover, no moderation effect was found.

Bias in human evaluation. The idea of our study is based on the underlying assumption that human evaluators are biased and may (unintentionally) disadvantage or discriminate against a certain group of people based on personal characteristics. For the job interview context, this has been shown in several studies and in various contexts. For example, there are studies on gender bias (e.g., Hoover et al., 2019; Isaac et al., 2009; Moss-Racusin et al., 2012), ethnic bias (see Derous & Ryan, 2019 for a review), and bias against overweight applicants (e.g., Pingitore et al., 1994). There are also studies on attractiveness bias (e.g., Sheppard et al., 2011), or bias against applicants with an accent (e.g., Deprez-Sims & Morris, 2010). Thus, we examined whether the general personal characteristics of participants, namely age, education level, gender, and nationality have a significant impact on their performance as evaluated by the human raters and the AI in our experiment. Indeed, results reveal that education level ($\beta = 0.11$, $p < 0.05$), gender ($\beta = 0.28$, $p < 0.05$), and nationality ($\beta = 0.34$, $p < 0.05$) significantly affect individuals' performance as rated by the human evaluators, supporting prior findings. Thereby, individuals with a higher level of education, German participants (versus non-German participants), and women (versus men) are rated significantly better. In contrast, none of these variables significantly affect individuals' performance as rated by the AI software (see Appendix).

Preferences of low- and high-performing candidates. We further investigated whether individuals' self-perceived autonomy and actual performance are related to their preference for a certain evaluation method. On the one hand, we found that individuals who perceive themselves as more autonomous rank the AI evaluator significantly lower than individuals who perceive themselves as less autonomous ($p < .05$). The same applies to individuals who perform better in the experiment. Thus, results show that the actual performance of participants negatively affects their preference for an AI evaluation ($p < .10$). On the other hand, these two variables, self-perceived autonomy ($p < .05$) and actual performance ($p < .05$), positively affect individuals' preference for a male evaluator (see Table 4.6).

Table 4.6: *Ordered logistic regression model* predicting individuals' preference for AI evaluation

	Coef.	SE	Wald Z	p	95% CI Lower B.	95% CI Upper B.	OR
Dependent variable: Preference for AI evaluation							
Age	-0.02	0.03	-0.47	0.63	-0.09	0.05	0.98
Education	-0.08	0.14	-0.60	0.55	-0.36	0.19	0.92
Gender (woman)	0.45	0.34	1.33	0.18	-0.21	1.12	1.57
Avg. performance	-0.50	0.26	-1.91	0.06.	-1.04	0.01	0.60
Self-perceived autonomy	-0.59	0.26	-2.32	0.02*	-1.11	-0.10	0.55
chi^2 (5) = 14.71							
Pseudo R^2 = 0.108							
Pr (> chi^2) = 0.012							
Cohen's f^2 = 0.121							
Dependent variable: Preference for male evaluator							
Age	0.01	0.03	0.47	0.64	-0.05	0.08	1.02
Education	-0.01	0.14	-0.10	0.92	-0.28	0.26	0.99
Gender (woman)	-0.56	0.32	-1.73	0.08.	-1.20	0.07	0.57
Avg. performance	0.61	0.25	2.42	0.02*	0.12	1.12	1.84
Self-perceived autonomy	0.52	0.24	2.15	0.03*	0.05	1.00	1.68
chi^2 (5) = 16.29							
Pseudo R^2 = 0.116							
Pr (> chi^2) = 0.006							
Cohen's f^2 = 0.131							

Note: Coding of Gender (woman): 1 = Woman, 0 = Man. The variable "Average performance" was calculated as the average of the z-standardized absolute scores of the three evaluation methods (female evaluator, male evaluator, AI evaluator) per participant. To collect the variable "Self-perceived autonomy," participants indicated on a 5-Point-Likert scale their self-perceived autonomy from 1 (Not at all autonomous) to 5 (Very autonomous).
$.p < 0.1$, $*p < 0.05$, $**p < .01$, $***p < .001$

4.5 Discussion

4.5.1 Discussion of experimental findings

In line with prior research on algorithm aversion and applicant reactions to AI recruiting, the results of our study confirm that individuals generally prefer a human-based evaluation method over an AI-based method when given the choice. However, we found that this greater preference for human assessment only applies to female evaluators and not to male evaluators.

This suggests that the preference for human evaluation methods is not due to the purely human character of the evaluator, but could be due to a character trait that is primarily attributed to women and not to men. A possible explanation for this could be that women are ascribed a greater empathic capacity than both men and AI.

It is interesting to note that this strong preference for female evaluators is found among both female and male participants. While prior research showed that interviewer-applicant gender similarity has positive effects on applicant reactions, such as perceived organizational attractiveness and intentions to apply (e.g., Harris, 1989; Turban, 1992), the gender similarity seem to have no impact on people's preference for an evaluation method when they have the choice. Other studies (e.g., Saks & McCarthy, 2006; Taylor & Bergmann, 1987) even found that job applicants perceive organizations as less attractive when the interviewer was female compared to when the interviewer was male. This finding also shows that different characteristics of an evaluator are required by applicants when assessing the attractiveness of a company than when selecting one's own evaluator.

More specifically, we examined individual and contextual factors, which affect women's preference for AI assessment. In line with prior research, we found that the more women perceive AI as reducing human bias, the higher is their preference for an AI evaluator. This finding is intuitive, assuming that women are interested in a fair and unbiased evaluation method. Moreover, and contrary to our expectation, we did not find a direct effect of the gender composition of the competitors on women's preference for AI evaluation. We can only observe the tendency for women to prefer an AI assessment in mixed-gender groups versus single-gender groups if they have a strong belief in the potential of AI to reduce bias. However, the effect is reversed if individuals believe that AI is more biased than humans. This finding reveals that the belief that AI generates a less biased evaluation than humans is not yet ingrained in the general female population, but that differing beliefs exist. Thus, depending on this belief, women's preferences for an AI evaluator differ in situations, in which they expect to be discriminated against.

Furthermore, results show that there is a strong positive effect of perceived personal discrimination on women's preference for AI evaluation, which becomes even stronger the more women believe in AI's potential to reduce bias. This result shows that women are more likely to choose an AI evaluator when they are more sensitive to discrimination due to prior experiences. Overall, our results highlight the importance of people's general perception of AI, which not only has a direct effect on women's preference for AI, but can also be considered a prerequisite (or enforcement) for the other effects shown.

Our additional analyses provide further insight into the factors influencing the preference for different assessment methods. Candidates with low self-perceived autonomy and candidates who are performing poorly rate AI higher than candidates with high self-perceived autonomy or strong actual performance. This preference could be explained by the fact that individuals generally have little experience with AI assessment and cannot estimate how they would be evaluated by an AI (only 10% of the participants indicated having done a video interview before). Thus, an AI assessment might be perceived as a lottery, in which individuals with low self-perceived autonomy, knowing that they have little chance of winning in a human process, might give themselves higher odds by choosing an AI evaluator. In contrast, people with high self-perceived autonomy, who are likely to have had good experiences with human evaluations in the past, will not take the risk of using an AI evaluator, which they are unfamiliar with. Another explanation might be rooted in an indirect relationship between our chosen target attribute "autonomy" and people's tendency to use technology. Thus, it was found that autonomous behavior is lower pronounced for people with low levels of self-efficacy (S. Lee & Klein, 2002), which in turn is positively associated with the usage of technology (Joo et al., 2018). In other words, people who are less autonomous and therefore have little belief in their abilities might rather rely on alternative technology-based assessment methods.

Moreover, our results reveal that the preference for a male evaluator increases with raising levels of self-perceived autonomy and improving actual performance. This could be explained by the fact that male evaluators are perceived as the less empathetic and lenient evaluators, compared to women. It is precisely this lenient assessment that top performers seem to reject, preferring instead the opportunity to differentiate themselves from other participants in the context of a supposedly stricter evaluation by a man. This result also ties into the existing literature on biases in subjective performance evaluations (e.g., Bol, 2011; Moers, 2005; Trapp & Trapp, 2019), which found that centrality bias and leniency bias[15] are perceived differently by above-average and below-average performers. For example, Trapp and Trapp (2019) found that a centrality bias has a negative effect on fairness perceptions and motivation for above-

[15] Leniency bias is defined as the tendency to inflate performance ratings; centrality bias leads to compressed ratings, i.e., the variance in the evaluation is lower than the variance in the performance Bol (2011).

average performers, while the effect is positive for below-average performers. Accordingly, in our experiment we could observe a self-selection of the top performers towards a male rater.

4.5.2 Practical implications

The design of selection processes is relevant for companies because of internal benefits such as efficiency gains, but it also influences how applicants perceive the company and their application behavior. Previous studies have shown that women often shy away from competing with men. As a result, they do not even apply for jobs where they expect to compete with men and where they believe they will be discriminated against in the selection process because of their gender. Thus, by designing gender-neutral selection processes that are free of bias—and advertised as such—organizations can influence current gender disparities in STEM industries or in leadership positions.

To counteract the self-selection by women, various organizations have employed quotas or affirmative action policies, which are already widespread. The mechanism of quotas has been shown to increase women's willingness to compete against men by leading to more gender-specific competition, which in turn increase women’s belief about their own abilities or performance (e.g., Balafoutas & Sutter, 2012; Niederle et al., 2013). However, quotas also come with several disadvantages, among which are economic losses in terms of effort and efficiency, as well as societal losses in the form of their risk of reinforcing existing narratives (e.g., Balafoutas et al., 2016; Neschen & Hügelschäfer, 2021).

Our findings show that the use of AI assessment could be an alternative to quotas, encouraging women to apply for jobs in male-dominated fields. More specifically, we showed that women have a higher preference for an AI-based evaluation in situations where they fear being discriminated against than in contexts where they do not have such a fear. This result is promising as it shows the potential of AI to counteract women’s fear of being discriminated against in a competition setting, which constitutes one of the drivers for the self-selection of women in the job market. For companies, this means that the use of AI in hiring, as well as its communication, could have a positive effect on female applicant numbers, as they expect a greater chance of getting the job. This could help companies create greater diversity in their workforce. On a broader level, this could be considered as a way to counteract gender imbalances in the job market.

However, our findings also show that for this to happen, there needs to be a fundamental belief and trust in AI in society. AI has the potential to make certain professions more attractive to women, but this would require improving the image of AI in the public perception. That the positive image of AI is not yet widespread is also shown by the fact that our subjects generally prefer human evaluation methods to AI evaluation. Companies should therefore pay attention to the form and extent to which they integrate AI into the selection process. AI should be used in a trustworthy and ethical way, i.e., its use should be transparent to applicants, and organizations must ensure that it does not lead to hidden bias or discrimination (e.g., Floridi et al., 2018). Only in this way can companies create trust in AI solutions at the societal level, which is necessary for the acceptance of these technological solutions and the development of their full potential.

4.5.3 Limitations and future research

Our study suffers from several shortcomings in our methodological approach, which can be mitigated in future studies as follows. First, we employed an online experiment, in which participants did not see who they were competing against. However, our manipulation in the form of subject's assignment to a single- or mixed-gender group was only communicated in a written form in the experiment instructions. Hence, assessing participants preferences and competition behavior in a lab setting where subjects directly see who and what gender they are competing with, similar to the experiment design of Niederle et al. (2013), might yield more significant results. Consequently, this setting may reveal bigger differences in women's preferences in single- versus mixed-gender groups.

Second, the chosen attribute "autonomy," which participants were rated on, may have moderated the observed effects. Our dataset shows that participating women generally rated women as more autonomous than men ($p < .001$, $N = 78$, one sample t-test[16]). This could mean that in the chosen experimental context, women had high confidence in their own abilities and were therefore less fearful of being disadvantaged or discriminated against in the competition. It would be interesting in future studies to test how the results change when a different attribute is chosen to measure performance in competition, e.g., an attribute which is stereotypically more attributed to men. The fact that women are generally considered to have more autonomy

[16] In In the survey that followed the experiment, participants were asked what they thought in general about autonomy of men and women. Participants answered the question using a Likert scale ranging from [1] "Men are much more autonomous than women" to [5] "Women are much more autonomous than men."

than men could also explain our additional finding that women perform better than men in our experiment according to the human evaluations. At this point, we might even infer a bias against men among the recruiters in our experimental competition setting. However, it should be noted that our analyses are based on a sample of only two raters per subject. Thus, our results are not generalizable.

Third, our results are based on a sample from the subject pool of a technical university. Assuming that the subjects from this sample are tech-savvy individuals who are rather open to the use of algorithms in decision-making, some caution is warranted in generalizing our results. It would be interesting to investigate whether a more technophobic sample would express the same preference toward human- and AI-based evaluation methods.

Overall, in our study we investigated individuals' preference for an assessment method as the main dependent variable. Therefore, our results can only be seen as the first step in drawing conclusions about the impact of the use of AI assessments on individuals' competitive behavior or its potential as a mechanism to counteract self-selection in the job market. While we measure individuals' preferences about AI evaluations, future research could pick up on this point and investigate whether AI evaluations affect individuals', especially women's, competition entry behavior. As such, our results should be seen as a first and promising indication about AI's potential to close the gender gap in the job market, which future research can build on.

4.6 Conclusion

Our study examines individuals‘ preference for an AI evaluation compared to a human evaluation and, more specifically, how (i) women's belief in AI's potential to reduce bias, (ii) the competitors' gender composition, and (iii) personal perceived discrimination affect women's preference for AI evaluation. We found that, in general, individuals prefer a human evaluator over an AI evaluation, but only if the human rater is female. Focusing on women, we found that women are more likely to choose the AI evaluation when they believe in AI's potential to reduce bias as well as when they have perceived discrimination in the past. Against our expectation, we found no direct effect of competitors' genders on women's preferences for AI evaluation. However, we show that individuals' belief in AI to reduce bias moderates the other two relationships. Women with a high belief in AI's potential to reduce bias are more likely to choose AI when competing in a mixed-gender group versus when competing in a single-gender group. This preference is reversed among women who believe that AI discriminates more than humans. In addition, our results reveal that the direct effect of perceived personal discrimination on women's preference for AI assessment further increases the stronger women believe in AI's potential to reduce bias. As such, our experiment highlights the importance of the belief in AI within society, which is a prerequisite for people to choose AI as an evaluation method. Moreover, our results indicate that women are more likely to choose AI evaluations in situations where they expect to be discriminated against, namely when competing in mixed-gender groups or when being sensitized to discrimination due to perceived discrimination in the past. Therefore, our results provide an initial indication that the use of AI could be used in circumventing anticipated discrimination in job selection, encouraging women to apply for jobs in male-dominated fields. We thus also provide a starting point for future research, examining the extent to which the use of AI may affect individuals', especially women's, competitive behavior.

4.7 Appendix

Appendix A: Items of collected variables (original items and English translation)

Original items	English translation
Belief in AI potential	*Belief in AI potential*
KI behandelt unterschiedliche Leute unterschiedlich, was mir Sorgen bereitet. (umgekehrtes Item)	AI treats different people differently, which worries me. (*reverse coded*)
Ich glaube, dass KI weniger diskriminiert als der Mensch.	I believe AI discriminates less than humans do.
Ich bin der Meinung, dass Entscheidungen von KI im Vergleich zu denen von Menschen vorurteilsfrei sind.	I believe that AI decisions are bias-free compared to those of humans.
Perceived personal discrimination	*Perceived personal discrimination*
Vorurteile gegen spezifische Gruppen haben mich schon einmal persönlich betroffen.	Prejudice against specific groups has affected me personally before.
Ich persönlich habe bereits Diskriminierung erfahren.	I personally have already experienced discrimination.
Ich bin aufgrund meiner Gruppenzugehörigkeit (z.B. Religion, Geschlecht, etc.) in der Vergangenheit oft ungerecht behandelt worden.	I have often been treated unfairly in the past because of my group membership (e.g., religion, gender, etc.).
Aufgrund von Diskriminierung wurden mir Möglichkeiten vorenthalten, die anderen zur Verfügung stehen.	Because of discrimination, I have been denied opportunities that are available to others.

Appendix B: Descriptive statistics

Appendix B.1: Descriptive Statistics of the full sample

Because we focus on women's preferences in our study, we conducted most analyses based on the subsample of women. For completeness, we present the descriptive statistics for all study variables of the full sample below, which includes men and women (N = 152).

Table B.1: Means, standard deviations, correlations, and Cronbach's alpha for the study variables of the full sample

	Variable	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1.	Gender (Woman)	0.51	0.5	-													
2.	Age	24.14	5.04	-.16*	-												
3.	Education	6.05	1.22	.09	.24**	-											
4.	German nationality	0.73	0.45	0.6	.08	.01	-										
5.	Belief in AI potential	3.38	0.87	-.07	-.06	-.17*	.17*	0.75									
6.	Gender composition (mixed)	0.5	0.5	-.03	.01	-.03	-.01	0.08	-								
7.	Personal discrimination	2.93	1.19	.24**	.14	.02	-.22**	-.11	.10	0.90							
8.	Pref. for AI evaluator	1.73	0.88	.05	-.07	-.09	-.10	.25**	.04	.14	-						
9.	Pref. for female evaluator	2.41	0.68	.02	.03	.07	.06	-.06	-.06	-.04	-.62**	-					
10.	Pref. for male evaluator	1.86	0.71	-.08	.05	.04	.07	-.26**	-.01	.-15	-.65**	-.20*	-				
11.	Self-perceived autonomy	4.31	0.67	.08	.07	.14	-.10	-.10	.01	.07	-.25**	.07	.24**	-			
12.	AI-rated performance	0.00	1.00	.00	.01	-.08	-.01	-.04	-.15	-.12	-.11	-.01	.14	.15	-		
13.	Human-rated performance	0.00	0.84	.20*	.09	.19*	.19*	.01	-.01	-.08	-.18*	.03	.19*	.26**	.07	-	
14.	Avg. performance	0.00	0.67	.16*	.08	.12	.15	-.01	-.08	-.13	-.20*	.02	.23**	.29**	.55**	.87**	-

Note: Coding of Variable 1: 1 = Woman, 0 = Man; Variable 4 was constructed by dummy coding: 1 = German, 0 = non-German; Coding of Variable 6: 1 = Mixed-gender group, 0 = Single-gender group. Ordinal scales for "Preference for AI evaluator," "Preference for female evaluator," and "Preference for male evaluator" were reversed: 1 = Rank 3, 2 = Rank 2, 3 = Rank 1, Variables 12, 13, and 14 were z-standardized, *N* = 152. Numbers in the diagonal represent Cronbach's alpha of the respective scale.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

Appendix B.2: Balance check

Table B.2 provides an overview of descriptive statistics across the experimental groups.

Table B.2: Means and standard deviations for the study variables across experimental groups

		Condition			
		Female		Male	
		Single-gender group (n = 38)	Mixed-gender group (n = 40)	Single-gender group (n = 36)	Mixed-gender group (n = 38)
	Variable	M (SD)	M (SD)	M (SD)	M (SD)
1.	Age	24.8 (3.68)	25.05 (8.12)	23.25 (3.26)	23.37 (3.25)
2.	Education	6.22 (1.35)	6.08 (1.22)	5.92 (1.23)	5.95 (1.09)
3.	German nationality	0.80 (0.41)	0.71 (0.46)	0.67 (0.48)	0.74 (0.45)
4.	Belief in AI potential	3.35 (0.78)	3.28 (0.96)	3.25 (0.90)	3.61 (0.81)
5.	Personal discrimination	3.07 (1.11)	3.35 (1.02)	2.53 (1.29)	2.75 (1.20)
6.	Pref. for AI evaluator	1.77 (0.86)	1.76 (0.88)	1.61 (0.90)	1.76 (0.91)
7.	Pref. for female evaluator	2.48 (0.55)	2.37 (0.75)	2.42 (0.73)	2.37 (0.71)
8.	Pref. for male evaluator	1.75 (0.81)	1.87 (0.70)	1.97 (0.61)	1.87 (0.70)
9.	Self-perceived autonomy	4.42 (0.64)	4.29 (0.69)	4.17 (0.77)	4.34 (0.58)
10.	AI-rated performance	-0.25 (1.02)	0.27 (0.91)	-0.04 (1.14)	0.04 (0.89)
11.	Human-rated performance	0.08 (0.81)	0.25 (0.90)	-0.10 (0.71)	-0.23 (0.88)
12.	Avg. performance	-0.03 (0.65)	0.25 (0.68)	-0.08 (0.63)	-0.14 (0.68)

Note: Variable 3 was constructed by dummy coding: 1 = German, 0 = non-German. Ordinal scales for ”Preference for AI evaluator,” “Preference for female evaluator,” and “Preference for male evaluator” were reversed: 1 = Rank 3, 2 = Rank 2, 3 = Rank 1. Variables 10, 11, and 12 were z-standardized, *N* = 152.

Appendix C: Robustness checks

Appendix C.1: Stepwise regression analyses—Main effects

To examine the effects of each individual explanatory variable on the dependent variable as well as on the overall model, we performed a stepwise ordinal logistic regression. The change in model fit values between the models indicates the explanatory contribution of each variable. For example, we observe that adding the variable “gender composition” has no effect on R^2. However, adding the variables and “Belief in AI” and “Personal discrimination” increases the model fit value by 0.086 and 0.082 respectively, indicating a small to medium effect size.

Table C.1: Results of *stepwise ordered logistic regression model* predicting women's preference for AI evaluation—Detailed overview of main effects

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Control variables										
Age	-0.01	0.03	-0.02	0.04	-0.01	0.03	-0.00	0.04	-0.02	0.04
Education	-0.07	0.17	-0.09	0.17	-0.08	0.17	-0.07	0.18	-0.09	0.18
Explanatory variables										
Belief in AI			0.66*	0.28					0.76**	0.29
Gender composition					-0.06	0.43			-0.13	0.46
Personal discrimination							0.52*	0.22	0.63**	0.24
Pseudo R^2	0.005		0.091		0.005		0.087		0.189	
Pr (>chi2)	0.857		0.092		0.955		0.104		0.016	
Cohen's f^2	0.053		0.100		0.053		0.095		0.235	

Note: Coding of 'Gender composition': 1 = Mixed-gender group, 0 = All female group.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

Appendix C.2: Treatment of "Education" as dummy variable

Table C.2: Results of *ordered logistic regression model* predicting women's preference for AI evaluation while treating "Education" as dummy variable

	Model 1		Model 2		Model 3		Model 4	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Control variables								
Age	-0.00	0.03	-0.01	0.04	-0.01	0.04	-0.01	0.04
Low education (Dummy)	0.23	0.45	0.14	0.49	0.10	0.50	0.16	0.50
High education (Dummy)	-0.51	0.80	-0.73	-0.73	-0.75	0.89	-0.71	0.87
Explanatory variables								
Belief in AI			0.78**	0.39	0.16	0.41	-1.15	0.95
Gender composition			-0.13	0.47	-4.39*	2.10	-0.27	0.48
Personal discrimination			0.62**	0.24	0.56*	0.56	-1.26	0.91
Interactions								
Gender compos. x Belief in AI					1.26*	0.60		
Personal discrimin. x Belief in AI							0.57*	0.28
Pseudo R^2	0.015		0.199		0.255		0.252	
Pr (>chi2)	0.793		0.021		0.007		0.007	
Cohen's f^2	0.015		0.248		0.342		0.337	

Note: Coding of the dummy variable "Low education": 1 = Lower degree than Bachelor's degree, 0 = Bachelor's degree or higher degree. Coding of the dummy variable "High education": 1 = Master's or doctoral degree; 0 = No Master's or doctoral degree.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

Appendix C.3: Inclusion of “Nationality” as additional control variable

Table C.3: Results of *ordered logistic regression model* predicting women’s preference for AI evaluation while including “Nationality” as additional covariate

	Model 1		Model 2		Model 3		Model 4	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Control variables								
Age	-0.01	0.03	-0.02	0.04	-0.02	0.04	-0.02	0.04
Education	-0.07	0.17	-0.09	0.18	-0.07	0.19	-0.09	0.18
German nationality	-0.63	0.51	-0.49	0.54	-0.42	0.56	-0.36	0.55
Explanatory variables								
Belief in AI			0.77**	0.29	0.14	0.41	-1.12	0.96
Gender composition			-0.15	0.46	-4.27*	2.08	-0.28	0.47
Personal discrimination			0.59**	0.24	0.53*	0.24	-1.23	0.92
Interactions								
Gender compos. x Belief in AI					1.22*	0.60		
Personal discrimin. x Belief in AI							0.57*	0.28
Pseudo R^2	0.027		0.199		0.252		0.248	
Pr (>chi2)	0.602		0.021		0.007		0.008	
Cohen’s f^2	0.028		0.248		0.337		0.330	

Note: “German nationality” was constructed by dummy coding: 1 = German, 0 = non-German.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

Appendix C.4: Alternative regression models

As robustness checks, we conducted the same regression analyses using different regression models. In the following, we present the results of the regression analyses for the main and moderation effects using an OLS regression, a binary logistic regression, and a probit regression model (see Tables C.4, C.5, and C.6).

We find that our results remain constant across the different models used. In all analyses, we find a significant direct effect of women's belief in AI to reduce human bias on their preference for AI as an evaluation method. Similarly, we observe a significant effect of perceived personal discrimination on women's preference for an AI evaluator, although significance is only at a 10% level in the binary logistic and probit regression models. No model shows a direct effect of the gender composition on women's preference for an AI evaluator.

In line with the ordered logistic regression model, all other regression models show significant results for both moderations tested.

Table C.4: Results of *OLS regression model* predicting women's preference for AI evaluation

	Model 1		Model 2		Model 3		Model 4	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Control variables								
Age	-0.01	0.02	-0.01	0.02	-0.01	0.01	-0.01	0.01
Education	-0.04	0.08	-0.03	0.07	-0.02	0.07	-0.03	0.07
Explanatory variables								
Belief in AI			0.27*	0.11	-0.01	0.16	-0.46	0.32
Gender composition			-0.03	0.19	-1.63*	0.74	-0.09	0.18
Personal discrimination			0.22*	0.09	0.20*	0.09	-0.50	0.31
Interactions								
Gender compos. x Belief in AI					0.48*	0.21		
Personal discrimin. x Belief in AI							0.22*	0.09
R^2	0.005		0.150		0.295		0.215	
Adj. R^2	-0.021		0.091		0.138		0.149	
p-value	0.822		0.035		0.010		0.007	

Note: Ordinal scale for dependent variable "Women's preference for AI evaluator" was reversed (1 = Rank 3, 2 = Rank 2, 3 = Rank 1); dependent variable was treated as a metric variable.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

Table C.5: Results of *binary logistic regression model* predicting women's preference for AI evaluation

	Model 1			Model 2			Model 3			Model 4		
	Coef.	SE	OR	Coef.	SE	OR	Coef.	SE	OR	Coef.	SE	OR
Control variables												
Age	-0.02	0.04	0.98	-0.03	0.05	0.97	-0.03	0.05	0.97	-0.04	0.05	0.96
Education	-0.16	0.20	0.85	-0.18	0.22	0.84	-0.20	0.22	0.82	-0.22	0.23	0.80
Explanatory variables												
Belief in AI				0.85*	0.36	2.33	-0.03	0.50	0.97	-2.45*	1.22	0.09
Gender composition				-0.08	0.55	0.92	-7.34*	3.43	0.00	-0.38	0.60	0.69
Personal discrimination				0.51.	0.28	1.67	0.39	0.28	1.47	-2.87*	1.25	0.06
Interactions												
Gender composition x Belief in AI							2.04*	0.93	7.70			
Personal discrimination x Belief in AI										1.01**	0.38	2.74
Pseudo R^2 (Nagelkerke)		0.017			0.185			0.281			0.314	
Pr (>chi2)		0.620			0.057			0.009			0.004	
Cohen's f^2		0.017			0.227			0.391			0.458	

Note: Dependent variable "Women's preference for AI evaluator" was coded as dummy variable: 1 = AI was ranked as 1st preferred evaluation method, 2 = AI was not ranked as 1st preferred evaluation method.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

Table C.6: Results of *binary probit regression model* predicting women's preference for AI evaluation

	Model 1			Model 2			Model 3			Model 4		
	Coef.	SE	OR	Coef.	SE	OR	Coef.	SE	OR	Coef.	SE	OR
Control variables												
Age	-0.01	0.03	0.99	-0.02	0.03	1.03	-0.02	0.03	0.98	-0.02	0.03	0.98
Education	-0.10	0.12	0.90	-0.09	0.13	1.19	-0.11	0.13	0.90	-0.13	0.13	0.87
Explanatory variables												
Belief in AI				0.45*	0.20	2.34	-0.06	0.29	0.95	-1.43*	0.69	0.24
Gender composition				-0.08	0.32	1.73	-4.37*	1.85	0.01	-0.21	0.35	0.81
Personal discrimination				0.27.	0.16	1.79	0.23	0.16	1.26	-1.68*	0.70	0.19
Interactions												
Gender composition x Belief in AI							1.23*	0.51	3.41			
Personal discrimination x Belief in AI										0.59**	0.21	1.81
Pseudo R^2 (Nagelkerke)		0.018			0.172			0.283			0.319	
Pr (>chi2)		0.611			0.078			0.009			0.003	
Cohen's f^2		0.018			0.208			0.395			0.468	

Note: Dependent variable "Women's preference for AI evaluator" was coded as binary variable: 1 = AI was ranked as 1st preferred evaluation method, 2 = AI was not ranked as 1st preferred evaluation method.
.p < 0.1, *p < 0.05, **p < .01, ***p < .001

Appendix D: Additional analyses

Table D.1: Results of *OLS regression model* predicting individuals' performance as rated by the human evaluators and by AI

	Coef.	SE	t	p
Dependent variable: Human-rated performance				
Age	0.00	0.01	0.15	0.88
Education	0.11	0.06	2.05	0.04*
Gender (woman)	0.28	0.13	2.12	0.04*
German nationality	0.34	0.15	2.26	0.03*
$F(4,147) = 4.04$; $R^2 = 0.10$; Adj. $R^2 = 0.07$; p-value = 0.004				
Dependent variable: AI-rated performance				
Age	0.01	0.02	0.32	0.75
Education	-0.07	0.07	-0.98	0.33
Gender (woman)	0.01	0.17	0.07	0.95
German nationality	-0.03	0.19	-0.16	0.87
$F(4,147) = 0.25$; $R^2 = 0.01$; Adj. $R^2 = -0.02$; p-value = 0.91				

Note: The variable "Human-rated performance" was calculated as the average of the z-standardized scores of the two human recruiters (female and male evaluators) per participant. The variable "AI-rated performance" was z-standardized.
$.p < 0.1$, $*p < 0.05$, $**p < .01$, $***p < .001$

Appendix E: Experimental instructions (screenshot and English translation)

Figure E.1: Screenshot of online experiment: Experimental instructions

English translation:

Rules of the competition

You have been assigned to a group of four participants.

All participants need to complete a short asynchronous video interview (i.e., the recording of the video interview and its evaluation are time-delayed), in which they answer a given question and video record the answer. The answer should be 1.5–2 minutes long. Before the recording, all participants have 2 minutes to prepare their answer.

Afterwards, the videos will be evaluated and the person who is evaluated as the most autonomous person within the group of four wins the tournament. Autonomy here includes the ability to plan and accomplish tasks independently, as well as making decisions autonomously.

The winner within your group of four will receive a bonus of 10.00 EUR. The other three participants will not receive a bonus.

Appendix F: Question for video interview (screenshot and English translation)

Figure F.1: Screenshot of online experiment: Question for video interview

Aufgabenstellung:

Wann standen Sie vor einer schwierigen Herausforderung und wie haben Sie diese bewältigt?

Ihre Antwort sollte zwischen **1,5 - 2 Minuten** lang sein.

Sie haben für die **Vorbereitung** Ihrer Antwort jetzt **2 Minuten** Zeit. Wenn die Zeit abgelaufen ist, gelangen Sie automatisch zur Videoaufnahme.

Beachten Sie: Die Person, die als **selbstständigste Person** bewertet wird, gewinnt das Turnier und erhält einen Bonus von **10,00 EUR.** Die Dimension Selbstständigkeit beinhaltet dabei die Fähigkeit, Aufgaben eigenständig zu planen und zu bewältigen, sowie das autonome Treffen von Entscheidungen.

Verbleibende Zeit zur Vorbereitung: 1:56

13% WEITER

English translation:

Task: When did you face a difficult challenge and how did you overcome it?

Your answer should be between 1.5–2 minutes long.

You now have 2 minutes to prepare your answer. When the time is up, you will automatically be taken to the video recording.

Please note: The person who is rated as the most autonomous wins the tournament and receives a bonus of 10.00 EUR. Autonomy here includes the ability to plan and accomplish tasks independently, as well as making decisions autonomously.

Appendix G: Description of evaluation methods (screenshot and English translation)

Figure G.1: Screenshot of online experiment: Description of evaluation methods

Wie sollen die Videos Ihrer Gruppe bewertet werden?

Sie können jetzt entscheiden, wie die Videos Ihrer Gruppe bewertet werden. Es stehen 3 Bewertungsmethoden zur Auswahl:

- **Durch eine künstliche Intelligenz (KI),** die Verhaltensmuster sowie Gesichtsausdruck, Körpersprache und Stimme auswertet, um ein Persönlichkeitsprofil zu erstellen; dafür wurde die KI basierend auf Daten von über >12.000 Personen trainiert und auf etwaige systematische Verzerrungen (z.B. hinsichtlich des ethnischen Hintergrunds oder des Geschlechts einer Person) geprüft und entsprechend bereinigt
- **Durch einen Mann** mit mehrjähriger Erfahrung im Personalbereich, der schon viele Vorstellungsgespräche mit Bewerber/innen geführt hat und auf Basis seiner Ausbildung und Erfahrung darauf trainiert ist, Entscheidungen zur Personalauswahl zu treffen
- **Durch eine Frau** mit mehrjähriger Erfahrung im Personalbereich, die schon viele Vorstellungsgespräche mit Bewerber/innen geführt hat und auf Basis ihrer Ausbildung und Erfahrung darauf trainiert ist, Entscheidungen zur Personalauswahl zu treffen

English translation:

How should your group's videos be evaluated?

You can now decide how your group's videos will be evaluated. There are 3 evaluation methods to choose from:

- **By an artificial intelligence (AI)** that evaluates behavioral patterns as well as facial expressions, body language, and voice to create a personality profile; to do this, the AI has been trained based on data from over >12,000 people and checked for any systematic biases (e.g., regarding a person's ethnic background or gender) and adjusted accordingly
- **By a man** with several years of experience in human resources, who has already conducted many interviews with applicants, and is trained to make personnel selection decisions based on his education and experience
- **By a woman** with several years of experience in human resources, who has already conducted many interviews with applicants, and is trained to make personnel selection decisions based on her education and experience

5 Discussion

5.1 Summary of findings

This dissertation's overall objective was to provide ethical perspectives on AI use in hiring. It contributes to scientific literature by examining the questions of what ethical considerations are related to AI hiring, how an AI hiring process can be designed so that people perceive it as fair, and what individuals' actual preferences are for AI-based evaluation—taking into account expected biases. I conducted three studies to respond to these questions. Whereas the first study provided an overarching ethical perspective on the topic, the other two studies each focused on one ethical subtopic—namely, fairness and bias.

In the first study, I examined AI recruiting from an overarching ethical perspective. I performed a systematic literature review and identified 51 articles addressing the topic, which I synthesized. Thereby, I outlined the ethical opportunities, risks, and ambiguities related to AI hiring and provided an overview of the approaches to mitigate these ethical risks in practice. I found that ethical considerations related to AI hiring are diverse. Although most papers have been focused on the ethical risk of algorithmic bias, AI hiring comes with several other ethical components; on the opportunity side, these include increased process consistency and a potential job enhancement for recruiters, and on the risk side, they include the obfuscation of accountability and reduced privacy. Based on the review, I further identified gaps in the extant literature and pointed out moral questions that call for deeper exploration in future research. For example, I identified only limited empirical research on the topic, revealing room for additional empirical research in the field to prove theoretical assumptions and derive implications for practice.

In the second study, I shed light on AI recruiting from a fairness perspective. Addressing the need for more empirical work in the field, I used data from an online vignette experiment (N = 404) to study whether adjusting process design factors may help improve people's fairness perceptions of AI interviews. I found that an AI interview's positioning in the initial assessment stage (versus the final decision stage) and whether participants were sensitized (versus not sensitized) to AI's potential to reduce human bias in the selection process positively influence fairness perceptions. Fairness perceptions, in turn, positively affected overall organizational attractiveness. Moreover, results showed that human oversight (versus no human oversight) over the AI-based decision-making process had no effect on people's fairness perceptions.

In the third study, I focused on expected biases' role in AI hiring. Using data from an online experimental study (N = 152), I investigated individuals', particularly women's, preferences regarding AI evaluations in a competitive setting. The results provide empirical evidence that individuals generally prefer a human evaluator over an AI evaluator when they have the choice—but only if the human rater is female. Focusing on women's preference for AI assessment, I found that the stronger women's belief that AI can reduce human biases and the more discrimination they have experienced in the past, the stronger their preference for an AI evaluation will be. Moreover, women are more likely to choose AI when competing in a mixed-gender group than when competing in a single-gender group—but only if they have a strong belief in AI's ability to reduce biases.

5.2 Overarching contributions

In answering the research questions posed at the outset of this research, each study made its own contribution to research and practice, as summarized in this dissertation's introduction and further outlined in the three preceding chapters. Taken as a whole, this dissertation also makes some overarching contributions to scholars and managers on the meta level.

5.2.1 Academic contributions

A first contribution of this dissertation is expanding the AI ethics literature with domain-specific work focused on the AI application context of hiring. This not only addresses recent calls for domain-specific research in the literature on AI ethics (Hagendorff, 2020; Mittelstadt, 2019) but also offers new avenues for future theoretical and empirical exploration. The first study outlines several ethical concerns specifically related to the hiring context. Future studies can build on this foundation and further explore these specific subtopics in more detail. The second study is focused on the ethical subtopic of fairness, which is especially relevant in the hiring context. By empirically showing how an AI hiring process should be designed so that it is perceived as fair, it also addresses the criticism about general AI ethics guidelines to be too high-level and even superficial (Hagendorff, 2020).

A second overarching contribution can be seen in the literature on algorithm aversion and applicant reactions to technology-based recruiting practices. Whereas the literature on algorithm aversion (e.g., Dietvorst et al., 2015) studies contextual factors affecting people's aversion against or appreciation for AI-based decision-making, the research on applicant reactions (e.g., Langer et al., 2018) investigates contextual and individual factors affecting how people perceive technology-based hiring practices. This dissertation expands both streams by identifying new contextual and individual factors driving individuals' perceptions of and preference for the use of AI in hiring. In the second study, I found that AI interviews' positioning within the overall process as well as people's sensitization for AI's ability to reduce human bias as contextual process design factors significantly influence their fairness perception of AI interviews. The results showed that the process design factor of human oversight has no influence on the fairness perception. In the third study, I identified individuals' beliefs about AI's potential to reduce bias, perceived personal discrimination, and competitors' gender as relevant factors affecting women's preference for an AI evaluation in a competitive setting.

The third contribution is empirical. Whereas in the first review study, I found that especially empirical work on the topic of AI recruiting remains scarce, in the subsequent two studies of my dissertation, I addressed this gap. Thus, I experimentally assessed individuals' fairness perceptions of and preferences for AI-based hiring practices. In addition, the third study takes a unique approach to examine individuals' preferences for AI evaluations. Whereas prior research on applicant reactions has mainly measured stated preferences about an AI-based hiring process (e.g., Acikgoz et al., 2020; Langer et al., 2018; Newman et al., 2020), to our knowledge, this is the first study eliciting participants' preferences for AI-based assessment in a competitive setting.

5.2.2 Managerial contributions

Aside from this dissertation's theoretical influence, it additionally offers implications for practice. First, this dissertation should raise managers' awareness of the ethical concerns related to AI's implementation in hiring. Due to AI's inherent characteristics, such as its use of historic data and its black-box character, numerous ethical issues arise when using AI in hiring. For example, because AI is trained with historical data that may be biased, AI may incorporate bias into its decisions, exacerbating hiring discrimination. Furthermore, the question arises regarding whether it is ethical to outsource important decisions greatly affecting applicants' lives to AI when not even the programmers themselves can provide a qualitative explanation of what factors influence the decision. In addition to algorithmic bias and transparency, there are many other ethical considerations on this topic of which managers and users of AI hiring tools should be aware. Thus, in the first study, I provide an overview of the ethical considerations in AI recruiting, whereas in the second study, I draw attention to the issue of fairness, one of the critical subtopics in this context; the third study highlights the importance of anticipated biases in hiring and their potential effects on individuals' preferences.

Second, this work's results provide managers with advice on how to implement AI-based hiring software in an ethical way. To this end, the first study provides an overview of approaches to mitigate ethical risks in practice. Aside from organizational standards, such as the establishment of an AI ethics board and diverse data scientist teams, these approaches cover technical due diligence (e.g., in the form of auditing methods) and creating awareness for ethical standards among employees. The second study provides managers with guidance on how to specifically design the process around an AI interview in recruiting. Thus, AI interviews should be rather positioned in the initial screening stage and not in the final decision round to be perceived as fair. Furthermore, firms should use explanations that emphasize the advantage of AI interviews regarding their potential to reduce human bias in the process, to prevent applicants' negative reactions.

Third, this dissertation underscores the critical role of public perception of AI in enabling companies to realize its full potential. In particular, people's belief in the potential of AI to reduce human bias is a crucial factor in this regard. Whereas the second study showed that people who are sensitized to AI's potential to reduce human biases perceive AI interviews as fairer than those who are not, the third study revealed that women's belief in AI's potential to reduce bias positively affects their preference for an AI evaluation in a competitive setting. Consequently, companies must help build trust in AI solutions at the societal level, which is necessary for the acceptance and widespread use of these technological solutions. Building this trust requires that companies ensure that AI and the data sets used to train it are free of bias to prevent all kinds of discrimination. Companies should be aware that they have the responsibility of using AI in an ethical and responsible manner because this is the only way to improve the public perception of AI solutions and fully realize their potential.

Fourth, this dissertation appeals to managers to consider the far-reaching effects and consequences of AI use in hiring, which surpass a mere increase in efficiency. For example, the first article highlights the potential implications of AI use on the work of recruiters, for whom AI use may represent an opportunity to reduce their job's repetitive tasks. The second paper shows what impact AI use can have on job applicants' fairness perception. The third paper explores a considerably broad impact of AI use in recruitment—namely, the effect it can have on gender imbalances in the job market, which is based on the fact that AI evaluations could encourage women to apply for jobs in male-dominated fields.

5.3 Avenues for future research

Different avenues for future research originate from this dissertation's studies. Because the initial review study provides an overarching ethical perspective on AI use in hiring, it highlights numerous ethical issues associated with the topic, offering wide-ranging starting points for future research. The study identifies several gaps in extant literature, which could be explored going forward. On the one hand, the study shows that little attention has been given to relevant ethical frameworks in the research on the ethicality of AI-enabled recruiting. Thus, this emerging topic would benefit from being assessed through an established ethical lens such as a utilitarian, deontological, or contract theory perspective. On the other hand, the review calls for additional empirical work. For instance, future research could foster a better understanding of AI recruiting tools' accuracy and validity by investigating the question of whether AI recruiting outperforms traditional selection procedures in terms of validity in any specific situations. Moreover, empirical studies on AI hiring tools' effect on companies' workforce diversity would be beneficial for the ongoing debate. This type of empirical evidence, which should be based on field data, could shed light on whether AI has the potential to overcome biased hiring processes and thus create diverse workforces, as postulated by AI software vendors.

By investigating three process design factors affecting people's fairness perceptions of AI interviews, the second study provides room for future research to better understand additional factors affecting people's perceptions of AI tools. Adverse applicant reactions could have severe impacts for firms because they might lead to negative outcomes, such as public complaints. Thus, further applicant reaction research could offer relevant and practical advice for system designers and recruiters. Building on my findings, future studies could investigate whether the way explanations for AI use are presented plays a role for people's perceptions. For example, companies could show welcome videos to applicants before job interviews. Sensitizing applicants to AI use with a welcome video might even amplify sensitization's beneficial effects compared with written text because it might help ensure that applicants do not overlook the information (Basch & Melchers, 2019). Moreover, the role of the degree of an applicant's interaction with AI might be an interesting topic to examine (M. K. Lee, 2018). Applicants who directly interact with AI (e.g., via a chatbot or a video interview with a virtual AI agent) might perceive the AI-based procedure differently from applicants whose resumes and test results have been analyzed by AI.

The third study is focused on individuals' actual preferences for AI hiring, taking anticipated biases into account. My results provide an initial indication that AI use could be used to circumvent anticipated discrimination in job selection and encourage women to apply for jobs in male-dominated fields. By examining women's preferences for AI evaluations as a first step, my study provides a starting point for future research examining the extent to which the use of AI evaluations in hiring may affect women's competitive behavior. Specifically, a subsequent study could investigate whether women are more likely to enter a competition with men when being evaluated by an AI versus when being evaluated by a human evaluator. In addition, the study's results call for additional research on the relationship between individuals' (perceived) performance and their preference for an AI evaluation. Whereas my results present a first indication that high performers tend to have a relatively low preference for AI evaluations but a relatively high preference for male evaluators, it would be interesting to further explore potential mediators in this relationship.

5.4 Conclusion

Because hiring decisions have far-reaching consequences for individuals, the use of AI in hiring involves many different ethical considerations. This dissertation provides an overview of the ethical issues associated with AI hiring and explores two ethical subtopics in more detail—namely, fairness and bias.

The use of AI in recruiting is associated with ethical opportunities and ethical risks as well as ethical ambiguities that should be considered when implementing AI. Specifically, this work provides insights into how the process surrounding AI interviews should be designed so that applicants perceive these interviews as fair. An AI interview's positioning in the overall recruiting process and applicants' sensitization to AI's potential to reduce human bias are critical factors in this regard.

In general, people's preferences for AI evaluations compared to human evaluations are low. However, women demonstrated a relatively high preference for AI evaluations in competitive situations in which they expected to be discriminated against. Consequently, AI evaluations may encourage women to apply for jobs in male-dominated fields. However, this requires a fundamental belief that AI reduces bias.

Overall, this work highlights the importance of an ethical perspective on the use of AI in hiring. Only when companies and organizations use AI in an ethical manner can people build trust in AI and its capabilities, allowing the public perception of AI to be improved. This improved image of AI in the public perception is crucial for AI to realize its full potential.

6 References

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, *28*(4), 399–416. https://doi.org/10.1111/ijsa.12306

Balafoutas, L., Davis, B. J., & Sutter, M. (2016). Affirmative action or just discrimination? A study on the endogenous emergence of quotas. *Journal of Economic Behavior & Organization*, *127*, 87–98. https://doi.org/10.1016/j.jebo.2016.04.015

Balafoutas, L., & Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science (New York, N.Y.)*, *335*(6068), 579–582. https://doi.org/10.1126/science.1211180

Basch, J., & Melchers, K. (2019). Fair and flexible?! Explanations can improve applicant reactions toward asynchronous video interviews. *Personnel Assessment and Decisions*, *5*(3). https://doi.org/10.25035/pad.2019.03.002

Bauer, T. N., Maertz, C. P., Dolen, M. R., & Campion, M. A [M. A.] (1998). A longitudinal assessment of applicant reactions to an employment test. *Journal of Applied Psychology*, *83*, 892–903.

Bauer, T. N., Truxillo, D. M., Tucker, J. S., Weathers, V., Bertolino, M., Erdogan, B., & Campion, M. A [Michael A.] (2006). Selection in the information age: The impact of privacy concerns and computer experience on applicant reactions. *Journal of Management*, *32*(5), 601–621. https://doi.org/10.1177/0149206306289829

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science*, *11*, 474–480. https://doi.org/10.1111/1467-9280.00291

Berger, J., Osterloh, M., & Rost, K. (2020). Focal random selection closes the gender gap in competitiveness. *Science Advances*, *6*.

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

Blacksmith, N., Willford, J. C., & Behrend, T. S. (2016). Technology in the employment interview: A meta-analysis and future research agenda. *Personnel Assessment and Decisions*, *2*, 12–20.

Bloomberg, J. (2018, September 16). Don't trust artificial intelligence? Time to open the AI black box. *Forbes*. https://www.forbes.com/sites/jasonbloomberg/2018/09/16/

Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, *86*(5), 1549–1575. https://doi.org/10.2308/accr-10099

Bornstein, S. (2017). Reckless Discrimination. *California Law Review*, *105*(4), 1055–1110. https://doi.org/10.15779/Z388P5V86M

Brenner, F. S., Ortner, T. M., & Fay, D. (2016). Asynchronous video interviewing as a new technology in personnel selection: The applicant's point of view. *Frontiers in Psychology*, *7*, Article 863. https://doi.org/10.3389/fpsyg.2016.00863

Brockner, J. (2006). Why it's so hard to be fair. *Harvard Business Review*, *84*, 122–129.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, *129*(3), 1409–1447. https://doi.org/10.1093/qje/qju009

Carlsson, R [Rickard], & Sinclair, S. (2018). Prototypes and same-gender bias in perceptions of hiring discrimination. *The Journal of Social Psychology*, *158*(3), 285–297. https://doi.org/10.1080/00224545.2017.1341374

Carvallo, M., & Pelham, B. W. (2006). When fiends become friends: The need to belong and perceptions of personal and group discrimination. *Journal of Personality and Social Psychology*, *90*(1), 94–108. https://doi.org/10.1037/0022-3514.90.1.94

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: sociocultural and biological considerations. *Psychological Bulletin*, *135*, 218–261. https://doi.org/10.1037/a0014412

Chamorro-Premuzic, T. (2019). Will AI reduce gender bias in hiring? *Harvard Business Review*, *June 10*. https://hbr.org/2019/06/will-ai-reduce-gender-bias-in-hiring

Chamorro-Premuzic, T., & Akhtar, R. (2019). Should companies use AI to assess job candidates? *Harvard Business Review*, *May 17*. https://hbr.org/2019/05/should-companies-use-ai-to-assess-job-candidates

Chamorro-Premuzic, T., Winsborough, D., Sherman, Ryne, A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology*, *9*(3), 621–640.

Chapman, D. S., Uggerslev, K. L., & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews: A field investigation. *The Journal of Applied Psychology*, *88*(5), 944–953. https://doi.org/10.1037/0021-9010.88.5.944

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

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

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, *96*(4), 608–630. https://doi.org/10.1037/0033-295X.96.4.608

Crosby, F. J., Iyer, A., Clayton, S., & Downing, R. A. (2003). Affirmative action. Psychological data and the policy debates. *American Psychologist*, *58*, 93–115. https://doi.org/10.1037/0003-066x.58.2.93

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

Daugherty, P. R., & Wilson, H. J. (2018). *Human + machine: Reimagining work in the age of AI*. Harvard Business Press.

Deprez-Sims, A.-S., & Morris, S. B. (2010). Context Sensitive Links Accents in the workplace: Their effects during a job interview. *International Journal of Psychology*, *45*(6), 417–426. https://doi.org/10.1080/00207594.2010.499950

Derous, E., & Ryan, A. M. (2019). When your resume is (not) turning you down: Modelling ethnic bias in resume screening. *Human Resource Management Journal*, *29*(2), 113–130. https://doi.org/10.1111/1748-8583.12217

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology. General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Fernández-Martínez, C., & Fernández, A. (2020). AI and recruiting software: Ethical and legal implications. *Paladyn: Journal of Behavioral Robotics*, *11*(1), 199–216. https://doi.org/10.1515/pjbr-2020-0030

Florentine, S. (2016). How artificial intelligence can eliminate bias in hiring. *CIO*. https://www.cio.com/article/3152798/how-artificial-intelligence-can-eliminate-bias-in-hiring.html

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Lütge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). Ai4people-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gelles, R., McElfresh, D., & Mittu, A. (2018). *Project report: Perceptions of AI in hiring*. University of Maryland. https://anjali.mittudev.com/content/Fairness_in_AI.pdf

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. https://doi.org/10.1080/0960085X.2021.1927213

Gilbert, D., Lee-Kelley, L., & Barton, M. (2003). Technophobia, gender influences and consumer decision-making for technology-related products. *European Journal of Innovation Management*, *6*(4), 253–263. https://doi.org/10.1108/14601060310500968

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

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quartely Journal of Economics*, *118*(3), 1049–1074.

Gogoll, J., & Uhl, M. (2018). Rage against the machine: Automation in the moral domain. *Journal of Behavioral and Experimental Economics, 74*, 97–103. https://doi.org/10.1016/j.socec.2018.04.003

Grimm, P. (2010). Social desirability bias. *Wiley International Encyclopedia of Marketing*.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Harris, M. (1989). Reconsidering the employment interview: A review of recent literature and suggestions for future research. *Personnel Psychology*, *42*, 691–726.

Harwarth, I., Debra, E., & Maline, M. (1997). *Women's colleges in the united states: History, issues, & challenges*. Diane Publishing.

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

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks? The implied communality deficit. *The Journal of Applied Psychology*, *92*(1), 81–92. https://doi.org/10.1037/0021-9010.92.1.81

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*(3), 383–403.

Hiemstra, A. M. F., Oostrom, J. K., Derous, E., Serlie, A. W., & Born, M. P. (2019). Applicant perceptions of initial job candidate screening with asynchronous job interviews. *Journal of Personnel Psychology*, *18*(3), 138–147. https://doi.org/10.1027/1866-5888/a000230

High-Level Expert Group on Artificial Intelligence. (April 2019). *Ethics guidelines for trustworthy AI*. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Hmoud, B., & Laszlo, V. (2019). Will artificial intelligence take over human resources recruitment and selection? *Network Intelligence Studies*, *7*(13), 21–30.

Hoover, A. E., Hack, T., Garcia, A. L., Goodfriend, W., & Habashi, M. M. (2019). Powerless men and agentic women: Gender bias in hiring decisions. *Sex Roles*, *80*(11-12), 667–680. https://doi.org/10.1007/s11199-018-0964-y

Hunkenschroer, A. L., & Lütge, C. (2021). How to improve fairness perceptions of AI in hiring: The crucial role of positioning and sensitization. *AI Ethics Journal*, *2*(2). https://doi.org/10.47289/AIEJ20210716-3

Inman, M. L., & Baron, R. S. (1996). Influence of prototypes on perceptions of prejudice. *Journal of Personality and Social Psychology*, *70*(4), 727–739. https://doi.org/10.1037/0022-3514.70.4.727

International Labour Organization. (2018). *Global Wage Report 2018/19: What lies behind gender pay gaps*. ILO.

International Labour Organization. (2020). *Global Wage Report 2020–21: Wages and minimum wages in the time of COVID-19*. ILO.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why femalsed are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*(5). https://doi.org/10.5194/gi-2016-11-RC2

Isaac, C., Lee, B., & Carnes, M. (2009). Interventions that affect gender bias in hiring: A systematic review. *Academic Medicine*, *84*(10), 1440–1446. https://doi.org/10.1097/ACM.0b013e3181b6ba00

Jauernig, J., Uhl, M., & Walkowitz, G. (2022). People prefer moral discretion to algorithms: Algorithm aversion beyond intransparency. *Philosophy & Technology*, *35*(1). https://doi.org/10.1007/s13347-021-00495-y

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Joo, Y. J., Park, S., & Lim, E. (2018). Factors influencing preservice teachers' intention to use technology: TPAC, teacher self-efficacy, and technology acceptance model. *Educational Technology & Society*, *21*(3), 48–59.

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *Research Papers*, *168*. https://aisel.aisnet.org/ecis2020_rp/168

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kim, P. T. (2017). Data-driven discrimination at work. *William & Mary Law Review*, *58*(3), 857–937.

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, *13*(3), 795–848. https://doi.org/10.1007/s40685-020-00134-w

Konradt, U., Warszta, T., & Ellwart, T. (2013). Fairness perceptions in web-based selection: Impact on applicants' pursuit intentions, recommendation intentions, and intentions to reapply. *International Journal of Selection and Assessment*, *21*(2), 155–169.

Kriebitz, A., & Lütge, C. (2020). Artificial intelligence and human rights: A business ethical assessment. *Business and Human Rights Journal*, *5*(1), 84–104. https://doi.org/10.1017/bhj.2019.28

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

Langer, M., König, C. J., & Hemsing, V. (2020). Is anybody listening? The impact of automatically evaluated job interviews on impression management and applicant reactions. *Journal of Managerial Psychology*, *35*(4), 271–284. https://doi.org/10.1108/JMP-03-2019-0156

Langer, M., König, C. J., & Krause, K. (2017). Examining digital interviews for personnel selection: Applicant reactions and interviewer ratings. *International Journal of Selection and Assessment*, *25*(4), 371–382. https://doi.org/10.1111/ijsa.12191

Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, *27*(3), 217–234. https://doi.org/10.1111/ijsa.12246

Langer, M., König, C. J., Sanchez, D. R.-P., & Samadi, S. (2019). Highly automated interviews: applicant reactions and the organizational context. *Journal of Managerial Psychology*, *35*(4), 301–314. https://doi.org/10.1108/JMP-09-2018-0402

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718756684

Lee, S., & Klein, H. J. (2002). Relationships between conscientiousness, self-efficacy, self-deception, and learning over time. *Journal of Applied Psychology*, *87*(6), 1175–1182. https://doi.org/10.1037//0021-9010.87.6.1175

Leventhal, G. S. (1980). What should be done with equity theory? New approaches to thestudy of fairness in social relationships. In K. Gergen, M. Greenberg, & R. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27–55). Plenum Press.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

Maggian, V., Montinari, N., & Nicolò, A. (2020). Do quotas help women to climb the career ladder? A laboratory experiment. *European Economic Review*, *123*, 103390. https://doi.org/10.1016/j.euroecorev.2020.103390

Major, B., Quinton, W. J., & McCoy, S. K. (2002). Antecedents and consequences of attributions to discrimination: Theoretical and empirical advances. In *Advances in Experimental Social Psychology*. *Advances in*

Experimental Social Psychology Volume 34 (Vol. 34, pp. 251–330). Elsevier. https://doi.org/10.1016/S0065-2601(02)80007-7

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing “So what?,” “What’s new?,” and “Where to next?”. *Journal of Management*, *43*(6), 1693–1725. https://doi.org/10.1177/0149206316681846

Microsoft. (February 2018). *Microsoft AI Principles*. https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Moers, F. (2005). Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting, Organizations and Society*, *30*(1), 67–80. https://doi.org/10.1016/j.aos.2003.11.001

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109

Neschen, A., & Hügelschäfer, S. (2021). Gender bias in performance evaluations: The impact of gender quotas. *Journal of Economic Psychology*, *85*, 102383. https://doi.org/10.1016/j.joep.2021.102383

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. https://doi.org/10.1016/j.obhdp.2020.03.008

Niederle, M., Segal, C., & Vesterlund, L. (2008). *How costly is diversity? Affirmative action in light of gender differences in competitiveness*. Cambridge, MA. https://doi.org/10.3386/w13923

Niederle, M., Segal, C., & Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, *59*(1), 1–16. https://doi.org/10.1287/mnsc.1120.1602

Niederle, M., & Vesterlund, L. (2008). Gender differences in competition. *Negotiation Journal*, *24*(4), 447–463. https://doi.org/10.1111/j.1571-9979.2008.00197.x

Nier, J. A., & Gaertner, S. L. (2012). The challenge of detecting contemporary forms of discrimination. *Journal of Social Issues*, *68*(2), 207–220. https://doi.org/10.1111/j.1540-4560.2012.01745.x

Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. *Developmental Psychology*, *48*(6), 1629–1642. https://doi.org/10.1037/a0029167

Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology*, *76*(1), 114–128.

Pinel, E. C., & Paulin, N. (2010). Stigma consciousness at work. *Basic and Applied Social Psychology*, *27*(4), 345–352. https://doi.org/10.1207/s15324834basp2704_7

Pingitore, R., Dugoni, B. L., Tindale, R. S., & Spring, B. (1994). Bias against overweight job applicants in a simulated emploment interview. *Journal of Applied Psychology*, *79*(6), 909–917.

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

Polli, F., Dolphin, J., & Kassir, S. (2019). On the basis of brains: How neuroscience and AI advance ethical hiring. *Workforce Solutions Review*, *10*(4), 13–15.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the FAT* '20: Conference on Fairness, Accountability, and Transparency.* Association for Computing Machinery.

Rodin, M. J., Price, J. M., Bryson, J. B., & Sanchez, F. J. (1990). Asymmetry in prejudice attribution. *Journal of Experimental Social Psychology*, *26*(6), 481–504. https://doi.org/10.1016/0022-1031(90)90052-N

Roulin, N., & Bourdage, J. S. (2017). Once an impression manager, always an impression manager? Antecedents of honest and deceptive impression management use and variability across multiple job interviews. *Frontiers in Psychology*, *8*, 29. https://doi.org/10.3389/fpsyg.2017.00029

Ryan, A. M., & Ployhart, R. E. (2000). Applicant perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, *26*, 565–606.

Saks, A. M., & McCarthy, J. M. (2006). Effects of discriminatory interview questions and gender on applicant reactions. *Journal of Business and Psychology*, *21*(2), 175–191. https://doi.org/10.1007/s10869-006-9024-7

Sechrist, G. B., Swim, J. K., & Mark, M. M. (2003). Mood as information in making attributions to discrimination. *Personality & Social Psychology Bulletin*, *29*(4), 524–531. https://doi.org/10.1177/0146167202250922

Sechrist, G. B., Swim, J. K., & Stangor, C. (2004). When do the stigmatized make attributions to discrimination occurring to the self and others? The roles of self-presentation and need for control. *Journal of Personality and Social Psychology*, *87*(1), 111–122. https://doi.org/10.1037/0022-3514.87.1.111

Sheppard, L. D., Goffin, R. D., Lewis, R. J., & Olson, J. (2011). The effect of target attractiveness and rating method on the accuracy of trait ratings. *Journal of Personnel Psychology*, *10*(1), 24–33. https://doi.org/10.1027/1866-5888/a000030

Simbeck, K. (2019). HR analytics and ethics. *IBM Journal of Research and Development, 63*(4/5), 1–12.

Sinclair, S., & Carlsson, R [R.] (2013). Whatwill I be when I grow up? The impact of gender identity threat on adolescents' occupational preferences. *Journal of Adolescence*, *36*, 465–474.

Sinclair, S., & Carlsson, R [Rickard] (2021). Reactions to affirmative action policies in hiring: Effects of framing and beneficiary gender. *Analyses of Social Issues and Public Policy*, *21*(1), 660–678. https://doi.org/10.1111/asap.12236

Sinclair, S., Nilsson, A., & Cederskär, E. (2019). Explaining gender-typed educational choice in adolescence: The role of social identity, self-concept, goals, grades, and interests. *Journal of Vocational Behavior*, *110*, 54–71. https://doi.org/10.1016/j.jvb.2018.11.007

Singh, V., Kumra, S., & Vinnicombe, S. (2012). Gender and impression management: Playing the promotion game. *Journal of Business Ethics*, *37*, 77–89.

Solnick, S. J. (1995). Changes in women's majors from entrance to graduation at women's and coeducational colleges. *Industrial and Labor Relations Review*, *48*(3).

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes: What do they signify? *Psychology of Women Quarterly*, *24*, 44–62.

Suen, H.-Y., Chen, M. Y.-C., & Lu, S.-H. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior*, *98*, 93–101. https://doi.org/10.1016/j.chb.2019.04.012

Taylor, M. S., & Bergmann, T. J. (1987). Organizational recruitment activities and applicants' reactions at different stages of the recruitment process. *Personnel Psychology*, *40*(2), 261–285. https://doi.org/10.1111/j.1744-6570.1987.tb00604.x

Thielsch, M. T., Träumer, L., & Pytlik, L. (2012). E-recruiting and fairness: The applicant's point of view. *Information Technology and Management*, *13*(2), 59–67. https://doi.org/10.1007/s10799-012-0117-x

Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M., & Bernstein, A. (2020). Implementations in machine ethics: A survey. *ACM Computing Surveys*, *53*(6), Article 132. https://doi.org/10.1145/3419633

Trapp, I., & Trapp, R. (2019). The psychological effects of centrality bias: An experimental analysis. *Journal of Business Economics*, *89*(2), 155–189. https://doi.org/10.1007/s11573-018-0908-6

Turban, D. B. (1992). Influences of campus recruiting on applicant attraction to firms. *Academy of Management Journal*, *35*(4), 739–765.

University of Montreal. (December 2018). *Montreal declaration for a responsible development of artificial intelligence*. https://www.montrealdeclaration-responsibleai.com/the-declaration

Vasconcelos, M., Cardonha, C., & Gonçalves, B. (2018). Modeling epistemological principles for bias mitigation in AI systems: An illustration in hiring decisions. In J. Furman, G. Marchant, H. Price, & F. Rossi (Eds.), *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 323–329). ACM. https://doi.org/10.1145/3278721.3278751

Weaver, G. R., & Trevino, L. K. (2001). The role of human resources in ethics/compliance management: A fairness perspective. *Human Resource Management Review*, *11*, 113–134.

Wiechmann, D., & Ryan, A. M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment*, *11*, 215–229.

Williams, B., Brooks, C. F., & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy*, *8*, 78. https://doi.org/10.5325/jinfopoli.8.2018.0078

Yarger, L., Cobb Payton, F., & Neupane, B. (2020). Algorithmic equity in the hiring of underrepresented IT job candidates. *Online Information Review*, *44*(2), 383–395. https://doi.org/10.1108/OIR-10-2018-0334

A Publications

A.1 Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda

Information about the article

Title	"Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda"
Authors	Anna Lena Hunkenschroer, Christoph Lütge
Accepted	January 25, 2022
Journal	*Journal of Business Ethics*
Published online	February 8, 2022
DOI	https://doi.org/10.1007/s10551-022-05049-6

Because we want to attach the article in the form in which it was published, it has no page numbers. The article covers pages 79–109 of this dissertation.

Copyright information by the publisher

Re: Permission letter for article inclusion in my dissertation - Ticket ID [#6553111]

Open Research Support <orsupport@springernature.com>
Do 31.03, 14:26
Hunkenschroer, Anna

Allen antworten

Posteingang

Dear Dr. Hunkenschroer,

Thank you for contacting Springer Nature.

Copyright on any research article in open access (OA) journals published by Springer Nature is retained by the author(s). Authors grant Springer Nature a license to publish the article and identify itself as the original publisher. Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified.

OA articles in Springer Nature journals are predominantly published under Creative Commons Attribution v4.0 International licence (CC BY). For more information please read our journal policy.

If you have any questions, please do not hesitate to contact us quoting your Ticket ID [#6553111].

With kind regards,
--
Joao Cortes
Global Open Research Support Specialist
Author Service

Springer Nature Group
www.springernature.com
--
Visit **Springer Nature Support** for answers to our most frequently asked questions.
If you would like to contact Open Research Support via chat, please visit **BMC Support Portal.**
--
Every day around the globe, our imprints, books, journals, platforms and technology solutions reach millions of people – opening the doors to discovery for our communities by enabling them to access, trust and make sense of the latest research, so that they can improve outcomes, make progress, and benefit the generations that follow.
--
In the Americas: Springer Nature Customer Service Center LLC, 200 Hudson Street, Suite 503, Jersey City, NJ 07311, USA
Registered Agent: Corporation Service Company, 251 Little Falls Drive, Wilmington, DE 19808, USA
State of Incorporation: Delaware, Reg. No. 4538065
Outside the Americas: Springer Nature Customer Service Center GmbH, Tiergartenstraße 15 – 17, 69121 Heidelberg, Germany
Registered Office: Heidelberg | Amtsgericht Mannheim, HRB 336546
Managing Directors: Martin Mos, Dr. Ulrich Vest, Franciscus Vrancken Peeters

Journal of Business Ethics
https://doi.org/10.1007/s10551-022-05049-6

REVIEW PAPER

Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda

Anna Lena Hunkenschroer[1] · Christoph Luetge[1]

Received: 24 September 2020 / Accepted: 25 January 2022
© The Author(s) 2022

Abstract
Companies increasingly deploy artificial intelligence (AI) technologies in their personnel recruiting and selection process to streamline it, making it faster and more efficient. AI applications can be found in various stages of recruiting, such as writing job ads, screening of applicant resumes, and analyzing video interviews via face recognition software. As these new technologies significantly impact people's lives and careers but often trigger ethical concerns, the ethicality of these AI applications needs to be comprehensively understood. However, given the novelty of AI applications in recruiting practice, the subject is still an emerging topic in academic literature. To inform and strengthen the foundation for future research, this paper systematically reviews the extant literature on the ethicality of AI-enabled recruiting to date. We identify 51 articles dealing with the topic, which we synthesize by mapping the ethical opportunities, risks, and ambiguities, as well as the proposed ways to mitigate ethical risks in practice. Based on this review, we identify gaps in the extant literature and point out moral questions that call for deeper exploration in future research.

Keywords Artificial intelligence · Algorithmic hiring · Employee selection · Ethical recruitment · Ethics of AI · Bias of AI

Introduction

Pursuant to advances in technological developments, artificial intelligence (AI) has expanded into various business sectors and workplaces. Along with such fields as the prediction of credit worthiness, criminal justice systems, and pricing of goods, AI-enabled technologies have disrupted companies' personnel recruiting and selection practices, entering the market at exponential rates (Yarger et al., 2020). AI-advanced selection tools are attractive for organizations, due to their higher speed and efficiency gains compared with traditional screening and assessment practices (van Esch & Black, 2019) and are considered a valuable asset in today's "war for talent" (Leicht-Deobald et al., 2019, p. 381). Today's trend toward more remote and home office work further spurs the adoption of alternatives to in-person job interviews to assess candidates remotely (Wiggers, 2020).

We define *AI recruiting* as any procedure that makes use of AI for the purposes of assisting organizations during the recruitment and selection of job candidates, whereas *AI* can be defined as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019, p. 17). We thereby refer to a broad concept of AI that includes complex machine learning approaches such as deep neural networks, but also covers simple algorithms relying on regression analyses as well as other kinds of algorithms, such as natural language processing or voice recognition.

As the outcomes of hiring decisions have serious consequences for individuals, informing where they live and how much they earn (Raghavan et al., 2020), AI recruiting practices are worth considering from an ethical perspective. Vendors of AI recruiting systems advertise that their software makes hiring decisions not only more efficient and accurate, but also fairer and less biased, as they are free of human intuition. However, AI applications in the recruiting context may generate serious conflicts with what society typically considers ethical (Tambe et al., 2019). This was illustrated by the case of Amazon in 2018, when the company abandoned its tested hiring algorithm, which had turned out to be biased and discriminatory against women

✉ Anna Lena Hunkenschroer
Anna.Hunkenschroer@tum.de

Christoph Luetge
Luetge@tum.de

[1] Chair of Business Ethics, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

Published online: 08 February 2022

(e.g., Mujtaba & Mahapatra, 2019). When people are "users" or "consumers" of algorithmic systems (e.g., when following a recommendation of their movie streaming service), issues with the accuracy of recommendations against their taste may be no more than an inconvenience. However, when AI decisions are incorporated into hiring processes, applicants cannot reject them, and inaccuracies become problematic (Lee, 2018).

Although research on AI recruiting has increased substantially in recent years, a comprehensive ethical understanding of recruiting as an expanding application context of AI is still lacking. While the subject of algorithmic bias in hiring decisions has attracted broad interest among researchers, especially from a legal and technical perspective, there are more ethical concerns related to AI recruiting, such as data privacy, transparency, and accountability, which are worth discussing. To establish a common foundation for future research in the field, it is crucial to synthesize extant theoretical and empirical approaches to assess the ethicality of AI-powered recruiting. We address this need in our paper.

Through a systematic review of extant literature, we take a first step to offer an overview of the various ethical considerations in AI-enabled recruiting and selection. The inherent multidisciplinary nature of AI recruiting has led to a broad view of the phenomenon. Thus, we categorize extant research that considers the ethicality of AI recruiting from theoretical, practitioner, legal, technical, and descriptive perspectives. Furthermore, we provide an overview of the different AI applications along the recruiting stages, show where major ethical opportunities and risks arise and outline the proposed ways of mitigating such risks in practice. Due to the huge impact of recruiting decisions on people's lives, it is crucial that companies understand both the opportunities and the potential risks that AI recruiting technologies may create and how algorithmic decisions may be in complete disagreement with what they want to achieve.

We observe that the field of AI recruiting suffers from four shortcomings: First, there are only a few papers that provide a theoretical foundation for the ethical discussion, leaving many arguments unfounded. Additional theoretical, normative work in this area could prove beneficial to managers and organizations by providing guidance beyond mere casuistry in determining the right course of action. Second, most papers focus on the challenge of algorithmic bias, neglecting other ethical concerns, such as accountability and human autonomy in the AI recruiting context. Third, the established approaches to mitigate ethical risks are rather general and lack concrete domain-specific implementation guidelines for the recruiting context. However, a domain-specific focus is desirable, as general normative guidelines often lack tangible impact due to their superficiality (Hagendorff, 2020). Finally, we identified only limited empirical research on the topic, as well as inconsistent findings on people's fairness perceptions of AI recruiting, revealing room for further empirical research in the field to prove theoretical assumptions and derive implications for practice.

Our paper contributes to the literature in three ways. First, we comprehensively organize the extant research on ethical considerations of AI recruiting by identifying and summarizing the different perspectives taken. Second, we make accessible to researchers and human resources (HR) professionals an overview of the ethical considerations in AI recruiting by synthesizing extant research. We thereby ethically evaluate these considerations and classify them into ethical opportunities, risks, and ambiguities, developing an ethical framework of AI recruiting. Third, we identify current research gaps and propose moral topics and questions that call for a deeper exploration in both theoretical and empirical future research.

The remainder of our paper is organized as follows. First, we specify our research method and selection criteria. Afterwards, we systematically review the literature on the ethicality of AI recruiting, organizing the identified literature according to underlying perspectives. We then outline the different AI applications in the recruiting process and map ethical considerations in the form of ethical opportunities, risks, and ambiguities. In the "Discussion" Section, we summarize and discuss our review findings, by outlining considerations for future theoretical and empirical research and highlighting implications for practice. Finally, we provide concluding thoughts.

Research Method

In our review, we employ an interpretative approach according to Noblit and Hare (1988), like other literature reviews with aims similar to ours (e.g., Seele et al., 2019; Suddaby et al., 2017). In contrast to integrative reviews, which are most appropriate to summarize quantitative studies with data and construct similarity, interpretative reviews are thematic and applicable to a diverse body of literature, constituting of qualitative, quantitative, and conceptual works (Noblit & Hare, 1988). We systematically review the literature on the ethicality of AI-enabled recruiting and selection practices in four stages: First, to show how the ethicality of AI recruiting is assessed in extant research, we categorize the identified literature according to the perspectives assumed. Second, to afford a profound understanding of the underlying research topic, we give an overview of AI applications in recruiting. Third, we map the ethical considerations found in extant literature in the form of ethical opportunities, ethical risks, and ethical ambiguities. Fourth, we outline the mentioned approaches to mitigate ethical risks in practice. Figure 1 outlines the research design of this review paper.

Criteria for Selection, Inclusion, and Exclusion

On January 4, 2021, we performed a structured keyword-based literature search in the major online databases: Business

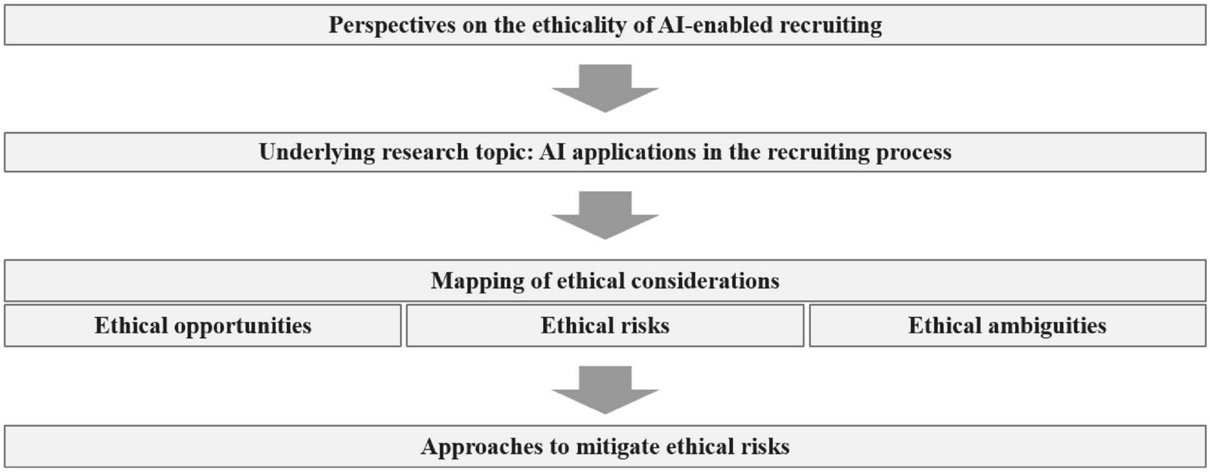


Fig. 1 Research design

Source Complete, Web of Science, and Scopus. Due to the novelty and interdisciplinarity of research on AI recruiting, we adopted a broad literature search strategy. We therefore decided in favor of openness of the sample and against an inclusion criterion such as publication in a top-tier journal of a specific field. Instead, we included all articles from academic peer-reviewed journals, conference proceedings, and practitioner-oriented articles (e.g., magazine articles) that study the ethicality of AI-powered practices in a company's recruiting and selection context. We ran the search by combining keywords from three topics: recruiting, ethics, and AI. Furthermore, we searched for articles in English without limiting the timespan. This initial search resulted in 756 hits after removing duplicates. The titles, abstracts, and full texts of all these articles were reviewed to determine the articles' relevance to our research scope, leading to 33 relevant articles. We then applied a backward search (by reviewing the references of the articles yielded from the keyword search) as well as a forward search (by reviewing additional sources that have cited the articles yielded from the keyword search), which resulted in a total of 51 distinct articles in scope for our review.

Our review excludes literature with a sole focus on a technical assessment of *algorithmic fairness*. Recently, a new body of literature emerged across such disciplines as law, policy, and computer science on fairness and bias in machine learning (ML) models, as well as their societal consequences (Barocas & Selbst, 2016; Lepri et al., 2018). Many works within this stream have proposed different definitions of fairness and non-discrimination (e.g., Dwork et al., 2012; Hardt et al., 2016) and focus on technical options to identify, measure, and mitigate discrimination in ML models (e.g., Corbett-Davies et al., 2017; Zafar et al., 2017). Only if an article explicitly discussed the application field of recruiting, as well as ethical implications did we include it in our review. Furthermore, we set a narrow scope of AI-enabled recruiting and excluded all literature dealing with *technology-enhanced recruiting* practices in a broader sense. This literature stream had already emerged in the early 2000s and investigates perceptions of technology in personnel selection and job interviews (Wiechmann & Ryan, 2003; Bauer et al., 2006; Chapman et al., 2003; see Blacksmith et al., 2016 for a meta-analysis). In various empirical studies, technology-mediated recruiting procedures, such as telephone and video interviews were investigated by testing the effects of technology-related factors on the interviews and the applicant reactions. For example, a couple of studies examined the fairness perceptions of applicants in online selection practices (Konradt et al., 2013; Thielsch et al., 2012). Nevertheless, we only included articles on recruiting practices that make use of AI techniques. Table 1 provides an overview of the data collection and selection criteria.

Table 1 Criteria for literature search and selection

Search terms	"Artificial Intelligence"; "AI"; "Algorithm*"; "Machine Learning"; "Robot*" AND "Ethic*"; "Moral*"; "Responsib*"; "CSR"; "Philosoph*"; "Fair*"; "Bias*"; AND "Recruit*"; "Hiring"; "Talent"; "Employee screening"; "Employee selection"; "Job interview*"; "Applicant screening"; "Employment interview*"
Search procedure	Initial keyword search; backward search; forward search
Language	English
Time frame	No limitation
Databases	Business Source Complete, Web of Science, Scopus
Inclusion	Articles from journals or databases, primarily related to ethicality of AI-enhanced recruiting, accessible in full text
Exclusion	Letters to the editor, commentaries, interviews, reviews, conference abstracts, and articles and studies without direct relation to the ethicality of AI-enabled recruiting

Structural Analysis of the Literature and Categorization

The chronological development of the identified literature underscores the novelty and increasing importance of our research topic in the last couple of years, with the first articles published in 2016. The 51 papers were published variously across 40 journals from different research fields, such as law, management, organizational psychology, robotics, and computer science. Table 2 shows the distribution of articles per journal and per year.

Based on our literature analysis, we identified five perspectives for categorizing the articles found: theoretical, practitioner, legal, technical, and descriptive. Two researchers independently performed the grouping according to these categories.[1] In case of disagreement, agreement was reached through discussion. A comprehensive overview of the collected literature and its categorization is provided in Table 3.

Table 2 Distribution of articles per journal and year

Journal Title	2016	2017	2018	2019	2020	Total
Academic journals						26
ABA Journal of Labor & Employment Law		1				1
Annual Review of Org. Psychology & Org. Behavior					1	1
Big-Data and Society			1			1
Business Horizons				1		1
Business & Information Systems Engineering					1	1
California Law Review		1				1
California Management Review				1		1
Computers in Human Behavior			1	1		2
Current Opinion in Behavioral Sciences		1				1
IBM Journal of Research and Development				1		1
IEEE Access					1	1
Industrial and Organizational Psychology	1					1
International Journal of Selection and Assessment				1	1	2
Journal of Information Policy			1			1
Journal of Managerial Psychology				1	1	2
Journal of Work and Organizational Psychology				1		1
Management Systems in Production Engineering				1		1
Online Information Review				1		1
Organizational Behavior and Human Decision Processes					1	1
Paladyn: Journal of Behavioral Robotics					1	1
Philosophy & Technology					1	1
Saint Louis University Law Journal			1			1
William & Mary Law Review		1				1
Practitioner journals						15
CIO Magazine	1					1
Fast Company			1			1
Harvard Business Review	1			7		8
Recruiter					1	1
SHRM			1			1
Training Journal				1		1
Workforce Solutions Review				1		1
Conference Proceedings						11
AAMAS (Int. Conf. on Autonomous Agents and Multiagent Systems)					1	1

Table 2 (continued)

Journal Title	2016	2017	2018	2019	2020	Total
Academy of Management Proceedings				1		1
AIES (AAAI/ACM Conf. on AI, Ethics, and Society)			1			1
FAT (Conf. on Fairness, Accountability, and Transparency)*					2	2
BESC (Int. Conf. on Behavioral, Economic Advance in Behavioral, Economic, and Sociocultural Computing)		1				1
CVPRW (Conf. on Computer Vision and Pattern Recognition Workshops)					1	1
ICIS (Int. Conf. on Information Systems)				1		1
IFIP Advances in Information & Communication Technology	1					1
IMC (Int. Mgmt. Conf. on Mgmt. Strategies for High Performance)				1		1
International Symposium on Technology and Society				1		1
Total	4	5	7	22	13	51

Table 3 Summary of articles on ethical considerations of AI-enabled recruiting

Authors	Type	Topic	Perspective	Theory/framework included	Ethical opportunities					Ethical risks					Ethical ambiguities			
					Reduction of human bias	Process consistency	Timely feedback for applicants	Efficiency gains for organizations	Job enhancement for recruiters	Introduction of algorithmic bias	Privacy loss & power asymmetry	Lack of transparency & explainability	Obfuscation of accountability	Potential loss of human oversight	Effect on workforce diversity	Informed consent & use of personal data	Impact on assessment validity & accuracy	Perceived fairness
Acikgoz et al. (2020)	Empirical: Experiment	Candidates' reactions to AI recruiting	Descriptive	Procedural justice	X	X		X		X		X						O
Bîgu and Cernea (2019)	Conceptual	Algorithmic bias in hiring	Practitioner, Technical*	None	X			X		O		X		X	X			
Bogen (2019)	Conceptual	Risk of bias in hiring algorithms	Practitioner	None	X	X		X		O								
Bornstein (2017)	Conceptual	Algorithmic bias and antidiscrimination law	Legal	None	O					X			X		X			
Cappelli (2019)	Conceptual	Challenges of AI recruiting	Practitioner	None						X	X					X	O	
Chamorro-Premuzic (2019)	Conceptual	AI recruiting to overcome bias	Practitioner	None	O	X											X	
Chamorro-Premuzic and Akhtar (2019)	Conceptual	AI recruiting to overcome bias	Practitioner	None	O	X		X		X	X				X	X	X	
Chamorro-Premuzic et al. (2017)	Conceptual	Technological applications in recruiting	Practitioner	None	X	X		X		X	X	X				X	O	

Table 3 (continued)

Authors	Type	Topic	Perspective	Theory/framework included	Ethical opportunities					Ethical risks					Ethical ambiguities			
					Reduction of human bias	Process consistency	Timely feedback for applicants	Efficiency gains for organizations	Job enhancement for recruiters	Introduction of algorithmic bias	Privacy loss & power asymmetry	Lack of transparency & explainability	Obfuscation of accountability	Potential loss of human oversight	Effect on workforce diversity	Informed consent & use of personal data	Impact on assessment validity & accuracy	Perceived fairness
Chamorro-Premuzic et al. (2016)	Conceptual	Assessment of technology-based hiring methods and their risks	Practitioner	None		X		X			X					X	O	
Chamorro-Premuzic et al., (2019)	Conceptual	Ethical implementation of AI recruiting	Practitioner	None	O	X		X		X		X			X		X	
Chwastek (2017)	Conceptual	AI applications in recruiting and ethical aspects	Technical	None				X		O	X					X		
Dattner et al. (2019)	Conceptual	Candidate privacy	Practitioner	None			X	X			O					X	X	
Fernández-Martínez and Fernández (2020)	Conceptual	Auditing of AI recruiting	Technical, Legal*	None				X		O	X	X		X		X	X	
Florentine (2016)	Conceptual	AI recruiting to overcome bias	Practitioner	None	O			X		X					X			
Giang (2018)	Conceptual	Ethical implementation of AI recruiting	Practitioner	None				X		O								
Hipps (2019)	Conceptual	AI recruiting to push diversity	Practitioner	None	X			X		X					O		X	

Table 3 (continued)

Authors	Type	Topic	Perspective	Theory/framework included	Ethical opportunities					Ethical risks					Ethical ambiguities			
					Reduction of human bias	Process consistency	Timely feedback for applicants	Efficiency gains for organizations	Job enhancement for recruiters	Introduction of algorithmic bias	Privacy loss & power asymmetry	Lack of transparency & explainability	Obfuscation of accountability	Potential loss of human oversight	Effect on workforce diversity	Informed consent & use of personal data	Impact on assessment validity & accuracy	Perceived fairness
Jayaratne and Jayatilleke (2020)	Empirical: Quant. analysis	AI's ability to measure applicants' personality	Technical	None	O												X	
Kaibel et al. (2019)	Empirical: Experiment	Candidates' reactions to AI recruiting	Descriptive	Procedural justice	X			X										O
Kim (2017)	Conceptual	Algorithmic bias and antidiscrimination law	Legal	None	X					O								
Kim and Scott (2018)	Conceptual	Online targeted advertising and antidiscrimination law	Legal	None				X		O								
Köchling et al. (2020)	Empirical: Quant. analysis	Bias in algorithmic video analysis	Technical	None	X	X		X		O		X	X	X	X		X	
Langer et al. (2020)	Empirical: Experiment	Candidates' reactions to AI recruiting	Descriptive	Procedural justice				X			X					X	X	O
Langer et al. (2019a)	Empirical: Experiment	Candidates' reactions to AI recruiting	Descriptive	Procedural justice	X	X		X			X					X		O

Table 3 (continued)

Authors	Type	Topic	Perspective	Theory/framework included	Ethical opportunities					Ethical risks					Ethical ambiguities			
					Reduction of human bias	Process consistency	Timely feedback for applicants	Efficiency gains for organizations	Job enhancement for recruiters	Introduction of algorithmic bias	Privacy loss & power asymmetry	Lack of transparency & explainability	Obfuscation of accountability	Potential loss of human oversight	Effect on workforce diversity	Informed consent & use of personal data	Impact on assessment validity & accuracy	Perceived fairness
Langer et al. (2019b)	Empirical: Experiment	Candidates' reactions to AI recruiting	Descriptive	Procedural justice		X		X			X					X		O
Langer et al. (2018)	Empirical: Experiment	Candidates' reactions to AI recruiting	Descriptive	Procedural justice	X	X		X			X	X				X	X	O
Lee (2018)	Empirical: Experiment	Perceived fairness of algorithmic decisions	Descriptive	Procedural justice	X	X		X		O				X				O
Lewis (2018)	Conceptual	AI recruiting to push diversity	Practitioner	None	X					X					O			
Lin et al. (2020)	Conceptual	Assessment of AI approaches to reduce implicit bias	Technical	None	O	X				X	X	X	X	X		X		
Mann and O'Neil (2016)	Conceptual	Ethical implementation of AI recruiting	Practitioner	None	X			X		O								
Mujitaba and Mahapatra (2019)	Conceptual, Mathematical	Algorithmic fairness and bias mitigation in hiring	Technical	None		X		X		O		X						
Newman et al. (2020)	Empirical: Experiment	Candidates' reactions to AI recruiting	Descriptive	Procedural justice	X	X		X				X					X	O

Table 3 (continued)

Authors	Type	Topic	Perspective	Theory/framework included	Ethical opportunities					Ethical risks					Ethical ambiguities			
					Reduction of human bias	Process consistency	Timely feedback for applicants	Efficiency gains for organizations	Job enhancement for recruiters	Introduction of algorithmic bias	Privacy loss & power asymmetry	Lack of transparency & explainability	Obfuscation of accountability	Potential loss of human oversight	Effect on workforce diversity	Informed consent & use of personal data	Impact on assessment validity & accuracy	Perceived fairness
Oswald et al. (2020)	Conceptual	Use of big-data in organizations and ethical implications	Practitioner, Technical, Legal*	None				X			O					X	X	
Pena et al. (2020)	Empirical: Experiment	Biases in multimodal machine learning and prevention	Technical	None						O	X	X				X		
Persson (2016)	Empirical: Qual. explorative analysis	Challenges of data mining and profiling in recruiting	Technical, Practitioner*	None	X			X		O	X				X	X		
Polli (2019)	Conceptual	AI recruiting to overcome bias	Practitioner	None	O										X			
Polli et al. (2019)	Conceptual	Video games to overcome bias in hiring	Practitioner	None	O			X							X		X	
Rąb-Kettler and Lehnervp (2019)	Conceptual	Humanistic recruiting	Theoretical	Humanistic management	X				O									

Table 3 (continued)

Authors	Type	Topic	Perspective	Theory/framework included	Ethical opportunities					Ethical risks					Ethical ambiguities			
					Reduction of human bias	Process consistency	Timely feedback for applicants	Efficiency gains for organizations	Job enhancement for recruiters	Introduction of algorithmic bias	Privacy loss & power asymmetry	Lack of transparency & explainability	Obfuscation of accountability	Potential loss of human oversight	Effect on workforce diversity	Informed consent & use of personal data	Impact on assessment validity & accuracy	Perceived fairness
Raghavan (2020)	Empirical: Qual. analysis	Bias mitigation in practice	Legal, Technical*	None						O		X		X			X	
Recruitment and Employment Confederation (2020)	Conceptual	AI recruiting to push diversity	Practitioner	None	X					X					O			
Ryan and Derous (2019)	Conceptual	Pitfalls of technology in recruiting	Technical	None		X		X	X	X							O	
Sánchez-Monedero et al. (2020)	Empirical: Qual. analysis	Bias mitigation in practice	Legal, Technical*	None	X			X		O		X	X	X	X			
Savage and Bales (2017)	Conceptual	Video games to overcome bias in hiring	Practitioner	None	O			X		X								
Schumann et al. (2020)	Conceptual	(Technical) research challenges in algorithmic hiring	Technical	None						O		X		X				

Table 3 (continued)

Authors	Type	Topic	Perspective	Theory/framework included	Ethical opportunities					Ethical risks					Ethical ambiguities			
					Reduction of human bias	Process consistency	Timely feedback for applicants	Efficiency gains for organizations	Job enhancement for recruiters	Introduction of algorithmic bias	Privacy loss & power asymmetry	Lack of transparency & explainability	Obfuscation of accountability	Potential loss of human oversight	Effect on workforce diversity	Informed consent & use of personal data	Impact on assessment validity & accuracy	Perceived fairness
Simbeck (2019)	Conceptual	Ethical implications of HR analytics	Theoretical	Ethics of AI, Medical ethics, Ethics in learning analytics, Ethics in coaching						X	X	X				X	X	
Suen et al. (2019)	Empirical: Experiment	Candidates' reactions to AI recruiting	Descriptive	Procedural justice	X			X				X						O
Tambe et al. (2019)	Empirical: Qual. analysis	Challenges of AI recruiting and practical responses	Practitioner	None	X			X		O		X	X		X		X	X
van den Broek (2019)	Empirical: Case study	Notions of fairness in AI recruiting	Descriptive	Organizational justice	X	X						X		X	X		X	O
van Esch and Black (2019)	Empirical: Experiment	Candidates' reactions to AI recruiting & practical implications	Descriptive	None	X		X	X	X	X								O

Table 3 (continued)

Authors	Type	Topic	Perspective	Theory/framework included	Ethical opportunities					Ethical risks					Ethical ambiguities			
					Reduction of human bias	Process consistency	Timely feedback for applicants	Efficiency gains for organizations	Job enhancement for recruiters	Introduction of algorithmic bias	Privacy loss & power asymmetry	Lack of transparency & explainability	Obfuscation of accountability	Potential loss of human oversight	Effect on workforce diversity	Informed consent & use of personal data	Impact on assessment validity & accuracy	Perceived fairness
Vasconcelos et al. (2018)	Conceptual, Mathematical	Bias mitigation in AI recruiting	Technical, Theoretical*	Popper's epistemological principles/Problem of induction	X			X		O	X	X		X	X	X		
Williams et al. (2018)	Conceptual	Algorithmic discrimination	Technical	None						O	X					X		
Yarger et al. (2020)	Conceptual	Equity in algorithmic hiring	Theoretical	Feminist design thinking	X	X		X		O	X	X		X	X	X		

Explanatory legend: O – Focus topic; X – Topic mentioned.

*Several perspectives are applied, of which the first-mentioned is the predominant one; for all statistics we only count the first-mentioned perspective to avoid double counting.

Overall, the articles falling into the practitioner perspective clearly constitute the majority of extant research on the ethicality of AI recruiting (39%), followed by the technical, descriptive, and legal perspectives (24%, 22%, and 10%), and lastly, the theoretical perspective (6%).

Findings

Perspectives on the Ethicality of AI-Enabled Recruiting and Selection

We start by reviewing the different perspectives from which AI-enabled recruiting and selection practices are investigated and ethical considerations are articulated.

Theoretical Perspective

The first group of papers assessed AI-powered recruiting practices from an ethics theory perspective. We identified three articles that applied a theoretical framework to AI recruiting and thereby provide a theoretical foundation for discussion: First, Simbeck (2019) referred to ethical frameworks from other disciplines, such as medicine, robotics, and AI, and applied them to the HR context. She proposed the transfer of key ethical concepts from the other fields that should be implemented when applying new AI technologies in HR analytics. She identified five key ethical principles: privacy and confidentiality, opportunity to opt out, institutional review, transparency, and respect for the dynamic nature of personal development.

Second, Yarger et al. (2020) referred to feminist thinking and methods, arguing that these should guide the design of AI hiring systems. Feminist approaches shed light on the extent to which algorithms may perpetuate disadvantage for historically marginalized groups when equity is not considered in their design. The authors presented a feminist design justice framework, which includes prompts that commit the architects of AI systems to engage with the design process in ways that support an ethic of equity.

Third, Rąb-Kettler and Lehnervp (2019) assessed AI recruiting from a humanistic perspective, in which people were placed at the center. The authors presented humanistic recruiting as an answer to the current technological developments. They argued that technology and automation can be implemented in a way that improves the experience for both the recruiters and candidates in the process. They concluded that both humanistic insight and sophisticated technology are important to adjust to today's dynamic reality. Reviewing these three theoretical papers reveals that a detailed assessment of AI recruiting from the standpoint of one of the traditional ethics theories, such as utilitarianism or deontology, and a discussion of potential implications for the hiring practice has not been done yet.

Practitioner Perspective

The second and largest category of papers assumed a practice-oriented perspective and focused on implications that are most relevant for managers and corporations. Most of the identified papers fall into this group, the common aim of which was to raise practitioners' awareness of the strengths and limitations of AI technologies implemented in the recruiting process. From an experience-based perspective, some papers (Florentine, 2016; Polli et al., 2019) underlined the problematic nature of traditional candidate assessment methods and presented the use of AI as a promising alternative; others (Bogen, 2019; Dattner et al., 2019) rather warned of AI-powered hiring practices by raising many yet-unanswered questions about their accuracy, as well as the ethical, legal and privacy implications that they introduce. Furthermore, some papers (Bîgu & Cernea, 2019; Chamorro-Premuzic et al., 2019; Giang, 2018; Mann & O'Neil, 2016) provided practical recommendations for managers on how to ethically implement AI for recruiting, aiming to guide organizations to take the right steps and make the right investments.

Legal Perspective

The third group of papers looked at AI recruiting from a legal viewpoint. The importance of employment decisions to individuals, as well as to broader society, has led to the design of an extensive legal framework to guide these decisions. For example, in the US, Title VII of the Civil Rights Act protects people from discrimination in any employment decision that would result in disparate treatment[2] or disparate impact[3]. It also assigns liability and legal responsibility to employers to ensure that the tools used do not create such results. However, the identified literature (Bornstein, 2017; Kim, 2017; Kim & Scott, 2018) has claimed that, so far, the law of Title VII lags behind current scientific knowledge and modern business practices: Kim and Scott (2018) discussed that targeted advertising may result in unfair exclusions that are not covered by current law, Bornstein (2017) argued that current regulation does not go far enough and argued for liability when an employer acts with reckless disregard for the consequences of implicit bias in employment decisions, and Kim (2017) claimed that Title VII should be broadened, requiring employers to prove that the data created by their algorithms are accurate and do not discriminate, instead of requiring victims of discrimination to prove its occurrence. We further identified two qualitative analyses that embraced both a legal and a technical perspective, while investigating how bias mitigation methods are used in practice. While Raghavan et al. (2020) evaluated the efforts of AI software vendors to mitigate bias, focusing on the employment laws in the US, Sánchez-Monedero et al. (2020) analyzed three recruiting software vendors from the perspective of UK law, addressing concerns over both discrimination and data protection.

Technical Perspective

Moreover, we identified a group of articles that established ethical considerations on AI recruiting, while taking a technical perspective. Some papers (Chwastek, 2017; Köchling et al., 2020; Lin et al., 2020; Mujtaba & Mahapatra, 2019; Persson, 2016; Williams et al., 2018) explained emerging ethical problems by looking at the mechanisms of algorithms used. Others (Fernández-Martínez & Fernández, 2020; Pena et al., 2020; Vasconcelos et al., 2018) presented technical solutions to implement ethical principles into algorithmic code or design. For instance, Fernández-Martínez and Fernández (2020) found that there is a lack of regulation and a need for external and neutral auditing of the used AI technologies, and consequently, they presented a multi-agent software architecture to support auditing the recruiting processes. Furthermore, Vasconcelos et al. (2018) proposed a computational framework to mitigate discrimination and unfairness caused by bias in AI systems, inspired by epistemological principles. Lastly, one paper (Schumann et al., 2020) outlined several technical challenges for future research in algorithmic hiring that must be overcome to make it fairer and more intelligible.

Descriptive Perspective

Covering the field of descriptive ethics, the last category comprises several experimental studies (e.g., Langer et al., 2018; Lee, 2018; van Esch & Black, 2019), as well as a case study (van den Broek et al., 2019) that assessed people's reactions to AI-powered recruiting practices. A couple of studies compared applicants' fairness perceptions of AI-enabled interviews vs. traditional interviews with a human recruiter, revealing contrasting findings. Whereas a group of papers (Acikgoz et al., 2020; Lee, 2018; Newman et al., 2020) found that people perceived algorithm-driven decisions as less fair than human-made decisions, another group of papers (Langer et al., 2019a, 2019b, 2020; Suen et al., 2019) found no difference in fairness perception between decisions made by an AI or a human. Other studies (Gelles et al., 2018; Kaibel et al., 2019; Langer et al., 2018; van Esch & Black, 2019) examined different contextual and procedural factors, such as the level of information given to applicants regarding the used AI or the level of computer experience of applicants, and how they affect applicant reactions to the use of AI in hiring.

In summary, this overview attests to the overall heterogeneous perspectives applied to ethical considerations of AI-based recruiting and selection. It also reveals that only a few theoretical articles exist, and that extant literature is rather practitioner oriented.

Underlying Research Topic: AI Applications in the Recruiting and Selection Process

In the following, we provide an overview of AI applications used in the recruiting and selection process and addressed in the identified literature. An understanding of where AI-powered tools and practices are applied can assist in understanding where ethical opportunities and risks may arise. Our review shows that AI-enabled practices are relevant in each stage of the recruiting process and can include different types of AI and algorithms. Table 4 gives an overview of the different AI applications across the recruiting and selection stages: outreach, screening, assessment, and facilitation, which we further expand on below.

Table 4 AI applications per recruiting stage

Stage	Outreach	Screening	Assessment	Facilitation
Objective	Identify possible candidates & persuade them to apply	Derive shortlist of most promising candidates	Identify which candidate is most appropriate for the job	Coordinate with applicants throughout the process
AI applications	Formulation of job ads (e.g., gender-neutral wording) Targeted advertisement of open positions (e.g., via social media) Notification of job seekers Identification of active or passive candidates (e.g., via LinkedIn or ATS[a])	Scanning of resumes (beyond keywords) to score or rank candidates Matching of candidates & job openings to identify best fit	Analysis of video interviews with AI technology (voice/face recognition) Simulation/games/tests to assess certain skills, capabilities and traits Scraping & analytics of social media postings for psychological profiles Linguistic analysis of writing samples & web activity	Use of NLP[b] to parse CVs & extract relevant information to fill-in application forms automatically Transparency on where applicants stand in the process & elucidation of next steps Scheduling of interviews & sending of job offers Communication with applicants & answering of questions by chatbot

[a]Automated tracking system

[b]Natural language processing

Outreach

Several articles deal with AI technologies applied in the outreach stage, in which businesses try to detect talent and attract applicants. By leveraging algorithms for targeted communication across online platforms and social media or for the automated notification of job seekers, companies can expand their outreach to potential candidates (Bogen, 2019). Furthermore, AI bots are used to identify the pool of active and passive candidates (e.g., via LinkedIn) or to (re-)discover top talents in the pool of former candidates via their internal automated tracking system (ATS) (van Esch & Black, 2019). Sometimes, the challenge is not just finding the right candidates but persuading them to apply via appealing job descriptions. AI software vendors, such as Textio, use AI in the form of text-mining techniques to predict the attractiveness of a job listing based on the hiring outcomes of several millions of job posts. The software thereby scans the job ad for key phrases that will statistically impact its performance. Additionally, a tone meter can determine whether the overall tone of the writing is likely to attract more men or more women and make suggestions on how to improve the inclusiveness of the language used (Lewis, 2018; Yarger et al., 2020). This is how AI can help businesses de-bias the wording of job ads, making them gender neutral to attract a diverse pool of applicants, or customize them for a specific target group (Rąb-Kettler & Lehnervp, 2019).

Screening

Notably, most articles that deal with the ethicality of AI recruiting focus on the application of AI technology in an initial resume screening. AI systems are used to filter applicants to derive a shortlist and a ranking of the most promising candidates (Bornstein, 2017; Fernández-Martínez & Fernández, 2020; Vasconcelos et al., 2018). For many years, companies have used traditional algorithms to scan resumes for preselected key words or phrases; however, today's AI technology goes beyond that. Now, chatbots and resume-parsing tools look for semantic matches and related terms determining a candidate's qualification. Other tools go even further and use ML to make predictions about a candidate's future job performance based on signals related to tenure or productivity, or the absence of signals related to tardiness or disciplinary action (Bogen, 2019). Based on the initial screening, algorithms can also suggest the best matching job opening for a given candidate (Rąb-Kettler & Lehnervp, 2019). These screening tools are considered highly efficient to streamline the process, especially for top employers who receive huge numbers of applications for each open position; however, concerns have been raised that highly qualified applicants may be overlooked (Persson, 2016).

Assessment

Although screening algorithms are not new in practice, there has been a recent trend toward video-interview analysis in recruiting. In such structured video interviews, AI technology replaces a human interviewer and asks the candidate a short set of predetermined questions (Chamorro-Premuzic et al., 2016; Fernández-Martínez & Fernández, 2020). Moreover, the AI technology can not only evaluate the actual responses, but also make use of audio and facial recognition software to analyze additional factors such as the tone of voice, microfacial movements, and emotions to provide insights on certain personality traits and competencies

(Köchling et al., 2020; Tambe et al., 2019; van Esch & Black, 2019).

Besides interviews, AI-powered skill tests, simulations, and neuroscience video games are used to assess further qualities, for example, applicants' risk attitude, planning abilities, persistence or motivation. Thereby, target variables need not be predefined by the company (Giang, 2018; Polli et al., 2019; Raghavan et al., 2020), but ML algorithms can analyze the data of a company's current top performers and derive which applicant characteristics and skills have been associated with better job performance (Tambe et al., 2019). In this way, data-driven assessment tools have changed talent signals and the criteria by which candidates are evaluated (Chamorro-Premuzic et al., 2016). For example, the software vendor Pymetrics uses ML and psychometric training data based on current top performers to predict an applicant's fit for a specific role. To this end, first, the top-performing incumbent employees in that role play a series of online games, which are gamified assessments that measure numerous cognitive and social traits. The data collected from these games are then used to establish a "success profile" for the job at hand. Second, the candidates applying to the job play the same games, and the ML model predicts their likelihood of success in the role (Polli et al., 2019).

Other software vendors offer AI technologies that analyze a person's digital records such as social media posts to construct a psychological profile of a candidate. Based on linguistic analyses of candidates' Web activities, new technologies infer talent, personality, and other important individual differences and compare them against the culture of the hiring company (e.g., Chamorro-Premuzic et al., 2016, 2017; Vasconcelos et al., 2018).

Facilitation

Finally, AI is used to facilitate the recruiting process, taking over administrative tasks. For instance, AI tools address the problem of long online questionnaires for applicants via natural language processing (NLP) techniques. These are used to parse unstructured documents, such as candidates' CVs, and extract relevant information to automatically complete a company's application form (Chwastek, 2017). Furthermore, AI-powered assistants can be used to interact and communicate with candidates: They can guide candidates through the different steps of the recruitment process, from answering company and process-related questions to scheduling interviews (Rąb-Kettler & Lehnervp, 2019; van Esch & Black, 2019). Today, many companies also use programs to create offers automatically and have them signed electronically (Sánchez-Monedero et al., 2020).

Mapping of Ethical Considerations

This rise of new AI recruiting practices comes with new ethical quandaries for organizations and society. In what follows, we examine extant research literature and map the ethical considerations established. This mapping of ethical considerations can be understood as a summary of areas in which society may have ethical concerns about the use of AI, which is derived from extant literature. In mapping the ethical considerations, we distinguish between aspects that are, on the one hand, clearly characterized as morally good and thus as ethical opportunities, and on the other hand, aspects that are clearly characterized as morally bad and thus ethical risks. In addition, we outline issues that are controversially discussed in the literature and thus reflect ethical ambiguities that require deeper exploration. Table 5 provides a structured overview of this ethical evaluation.

Human and Algorithmic Bias

The most-discussed topic in extant literature on AI-enabled recruiting is the occurrence of bias. Although there is broad agreement that the practices currently in place are far from effective and unbiased (e.g., Chamorro-Premuzic & Akhtar, 2019; Persson, 2016; Polli, 2019), there are two differing ways, in which AI-powered tools may effect the scope of bias.

Table 5 Overview of ethical evaluation: Ethical opportunities, risks and ambiguities

Ethical opportunities	Ethical risks	Ethical ambiguities
Reduction of human bias	Introduction of algorithmic bias	Effect on workforce diversity
Process consistency	Privacy loss & power asymmetry	Informed consent & use of personal data
Timely feedback for applicants	Lack of transparency & explainability	Impact on assessment validity & accuracy
Efficiency gains for organizations	Obfuscation of accountability	Perceived fairness
Job enhancement for recruiters	Potential loss of human oversight	

On the one hand, the use of AI may reduce human bias in different stages of the recruiting process and should therefore be considered a huge ethical opportunity (e.g., Chamorro-Premuzic & Akhtar, 2019; Savage & Bales, 2017). In the outreach stage, AI can address bias in the form of gendered language in job descriptions that dissuades certain candidates from applying for a role by creating inclusive job descriptions (Mann & O'Neil, 2016; Recruitment & Employment Confederation, 2020). In the screening procedure, subjectivity can be reduced by using algorithms that screen all applicants against the same criteria. AI is thereby able to assess the entire pipeline of candidates rather than forcing time-constrained humans to shrink the pool from the start, based on a biased process. Instead, AI can shrink the initial pipeline so a recruiter with a constrained capacity can manually handle it (Polli, 2019). Especially in the assessment stage, the use of AI technology can remove human bias from the process – or at least reduce it substantially. Human intuition can be very good and accurate, but it is nevertheless based on subjective value assessment (Persson, 2016). In contrast, via a digital interview or a video game assessment, AI automatically captures many data points of the applicants' behavior, such as what they say, their language use or their body language, for an objective, data-driven assessment of personality (Jayaratne & Jayatilleke, 2020). Moreover, human bias (e.g., related to applicants' physical appearance or other attributes) can be reduced, as AI can be taught to ignore people's personal attributes and focus only on specified skills and behaviors (e.g., Bîgu & Cernea, 2019; Chamorro-Premuzic & Akhtar, 2019; Fernández-Martínez & Fernández, 2020). Lastly, human bias can be removed from the process, as the required skills and qualities for successful candidates are not determined by bias-prone intuitions from recruiters, but based on analyzing the characteristics of the company's top performers (Lin et al., 2020).

On the other hand, AI-enabled recruiting also bears the risk of introducing different types of algorithmic bias (e.g., Bogen, 2019; Yarger et al., 2020). Yarger et al. (2020) cited three factors that may lead to biased decisions: bias in the model design principles, bias in the feature selection, and bias in the training data. A biased design, for example, may be manifested in online job platforms that make superficial predictions, not focusing on who will be successful in the role, but on who is most likely to click on the job ad. This can lead to a reinforcement of gender and racial stereotypes. A study found that targeted ads on Facebook for supermarket cashier positions were shown to an audience of 85% women, indicating that adverse impact can also occur in sourcing algorithms (Bogen, 2019). Moreover, critics are concerned that algorithms derived from information about current employees will unintentionally discriminate against underrepresented groups if existing employees are not proportionately representative of the broader application pool; this would constitute a case of biased training data (Kim, 2017). A known example from practice is the Amazon case, in which a hiring algorithm (in test mode) discriminated against women, assigning lower scores to resumes of women when ranking candidates. The algorithm was trained on data of current top performers, of which the majority were male. Thus, the algorithm penalized female attributes (e.g., Mujtaba & Mahapatra, 2019). In all these cases, algorithms can introduce bias and even magnify discrimination, affecting entire classes of individuals (Bogen, 2019; Tambe et al., 2019). The occurring discrimination may thereby be direct or indirect via proxy attributes. In the latter case, a protected group (e.g., a specific race) is discriminated against but based on legitimate grounds (e.g., a zip code) (Bîgu & Cernea, 2019; Fernández-Martínez & Fernández, 2020).

Proponents of AI recruiting tools admit that adverse impact can occur; however, they state that, compared with human biases, algorithmic biases are much easier to detect and remove (Florentine, 2016; Polli, 2019). Often, the fear of biased AI ignores the fact that the original source of algorithmic bias is the human behavior it is simulating (e.g., the biased data set used to train the algorithm). Thus, if people criticize what the AI is doing, they should criticize human behavior even more because AI is purely learning from humans (Polli, 2019).

Although there is an ongoing debate on the potential occurrence of algorithmic bias in AI recruiting, there is no ambiguity on the topic itself but general agreement that all kinds of bias and discrimination should be prevented. Therefore, AI recruiting can be classified as ethically preferable, as long as it seeks to reduce interpersonal bias in the process. However, current research suggests that the usage of AI can reduce bias but is never completely free of bias and carries the risk of algorithmic discrimination, even without bad intentions on the part of the programmers, which should be morally denounced. Thus, technical due diligence regarding algorithmic design and implementation is crucial to keep this risk low (see Sect. 3.4).

Effect on Workforce Diversity

A topic closely related to the occurrence of bias in the selection process is its impact on diversity: On the one hand, a reduction in human bias could lead to diversification of a company's workforce (Chamorro-Premuzic & Akhtar, 2019; Recruitment & Employment Confederation, 2020). For example, the use of bias-neutral job posts created through AI may result in a more diverse pool of applicants (Lewis, 2018). Furthermore, the data-driven assessment leads to hiring of "nontraditional" candidates who might typically not make it through a hiring process (e.g., from a non-elite college, but with other strong skills). In this way, AI-enhanced recruiting tools can provide people from a wider range of socioeconomic backgrounds access to better jobs, expanding diversity, and socioeconomic inclusion (e.g., Florentine, 2016; Hipps, 2019). Moreover, case studies have shown that, for example, the aforementioned AI-powered video games by Pymetrics have a clear positive impact on companies' gender diversity (Polli et al., 2019).

On the other hand, a systematic bias through AI could result in more homogeneity in organizations (Chamorro-Premuzic et al., 2019; Vasconcelos et al., 2018; Yarger et al., 2020). As a single decision-making algorithm, which selects candidates based on certain profiles and traits, replaces several human decision makers with potentially differing views, this may also imply a loss in diversity (Vasconcelos et al., 2018; van den Broek et al., 2019; Bîgu & Cernea, 2019). Further, Fernández-Martínez and Fernández (2020) warned that the use of AI leads to increased racial bias: Given that emotional recognition software may not consider different intonations in different languages or that emotions are differently expressed in different cultures, it may systematically disadvantage specific races or ethnic groups, which could lead to a decrease in workforce diversity.

This research question about the influence of AI on diversity has also been discussed in general diversity scholarship. Ozkazanc-Pan (2019) outlined how advanced technological shifts impact diversity scholarship, underlining the importance of bias, ethical considerations and digital inequalities in this context. She also thereby referred to the recruiting context and, for example, pointed out how the creation of employee profiles that are based on behavioral preferences, when not implemented carefully, can lead to HR managers hiring the same groups over and over again, which can hinder a company's diversity efforts.

Overall, there is no clear understanding of what impact the use of AI has on the diversity of corporate workforces, but the topic is controversially discussed in extant literature. Therefore, relevant empirical studies would be desirable in future. It must be noted that diversity is related to, but different from, non-discrimination, and more textured efforts are needed to explore the balance between diversity and non-discrimination (Schumann et al., 2020). An interesting question in this context may be whether it is ethical to promote diversity even if it discriminates against historically advantaged groups.

Privacy and Informed Consent

Another ethical consideration raised is the concept of privacy and informed consent. In this context, businesses must account for government regulations, which differ across countries. The European General Data Protection Regulation (GDPR), which came into effect in May 2018, is one of the strictest. It aims to protect EU citizens' rights by regulating how to collect, store, and process personal data and requires informed consent for any personal data processing (i.e., applicants must have the opportunity to agree or not agree to the use of their data). However, the informed consent requirement is not yet well implemented in the big-data and AI-regulation context, rendering the protection of personal privacy an ethical challenge (Oswald et al., 2020). An ethical dilemma emerges at this point as applicants in the job market generally hold less power than employers. Even if applicants are informed enough to consent to the process, they may not be able to opt out without being disadvantaged in the process. It is therefore difficult to give explicit consent in the context of hiring anyhow (Sánchez-Monedero et al., 2020).

Moreover, there is active debate about the extent to which it is ethically appropriate to use social media information for personnel selection purposes (Chamorro-Premuzic et al., 2016; Oswald et al., 2020). Legally, social media content is public data, but it is questionable whether it is ethical to mine social media data for hiring purposes when users generally use those platforms for other purposes and may not have provided their consent for data analysis (Dattner et al., 2019; Tambe et al., 2019). Also, the extent to which social media posts are a valid and reliable indicator of personality or job performance is doubtful (Vasconcelos et al., 2018; Yarger et al., 2020). Chamorro-Premuzic et al. (2016) argued that it is naive to expect online profiles to be more authentic than resumes, but they can offer a wider set of behavioral samples. Prior empirical findings on the validity of social media data have been mixed (Ryan & Derous, 2019). Whereas some studies found connections to job performance (e.g., Kluemper et al., 2012), van Iddekinge et al. (2016) showed that recruiter ratings of applicants' Facebook information were unrelated to their subsequent job performance and lead to subgroup differences, by favoring female and Caucasian applicants. This discussion on the use of social media information in the hiring context is not new and only connected to the use of AI. A study in Sweden showed that at least half of the interviewed recruiters had scanned applicant social media profiles themselves at some point before hiring (Persson, 2016). However, the new AI techniques make the analysis of social media profiles easier and even more tempting.

There are further AI-enabled ways to discern applicants' private information indirectly. For example, image and voice recognition techniques can predict applicants' sexual orientation, race, and age, as well as their physical attractiveness (Chamorro-Premuzic et al., 2016; Dattner et al., 2019). Other prediction algorithms may forecast who is more likely to become pregnant (Oswald et al., 2020; Simbeck, 2019). This greater access to candidates' personal attributes can not only increase the risk of misuse and intentional discrimination (Fernández-Martínez & Fernández, 2020), but also might further an information and power asymmetry between candidates and potential employers, leaving applicants with less room to negotiate (Sánchez-Monedero et al., 2020).

Overall, extant research has agreed that AI recruiting practices constitute a potential privacy loss for applicants attended by a greater power imbalance between applicants and employers; this poses an ethical risk. In addition, the use of more personal data, which may lead to more accurate predictions, is controversial (see also the next section). Thus, it is currently an unresolved normative question the extent to which a company may legally and ethically collect, store, and use personal data from applicants, such as the information available on social media platforms (Lin et al., 2020).

Consistency, Accuracy and Validity

There is broad agreement in extant literature that AI enables companies to make decisions more consistently across candidates and time (van den Broek et al., 2019). Whereas traditional assessment techniques such as analogue interviews are difficult to standardize, AI-based practices allow firms to put all applicants through exactly the same experience, resulting in an increase in the consistency of candidate assessment (Chamorro-Premuzic, 2019).

However, the accuracy and validity of the new AI assessment methods are controversially discussed. Today, employers do not necessarily know exactly which characteristics make an applicant a good fit for a given role. Studies have shown a very small correlation between a person's academic grades and their professional performance; still, many companies make above average grades a requirement for application. In contrast, some articles (e.g., Chamorro-Premuzic et al., 2019; Polli et al., 2019) argued that the new AI technologies have the potential to make the selection process more accurate as hiring algorithms predict a candidate's work-related behavior and performance potential based on the data of current top performers. AI may thereby outperform human inferences of personality in accuracy because it can process a much larger range of behavioral signals (Chamorro-Premuzic & Akhtar, 2019; Chamorro-Premuzic et al., 2016; Polli et al., 2019). In this way, the use of AI improves both the possibilities of "what" and "how" skills and abilities are measured (Ryan & Derous, 2019).

One article pointed to the accuracy–fairness trade-off in recruiting decisions and stated that AI technologies constitute the opportunity to overcome it (Chamorro-Premuzic et al., 2019). Historically, research has shown that traditional cognitive ability tests have led to discrimination of underrepresented groups, such as candidates with a lower socioeconomic status. Thus, to increase diversity and create an inclusive culture, companies have often de-emphasized cognitive tests in hiring (Chamorro-Premuzic et al., 2019). However, AI may overcome this fairness–accuracy trade-off by deploying more dynamic and personalized scoring algorithms that can optimize for both (Chamorro-Premuzic et al., 2019; Raghavan et al., 2020).

Nevertheless, critics have raised concerns about the technical robustness and validity of AI-powered assessment methods. First, many of the newly offered AI tools have emerged as technological innovations, rather than from scientifically derived methods or research programs. Although there has been broad psychological research on the validity of traditional methods for candidate assessment, such as job interviews, assessment centers, or cognitive ability tests, the newly emerging AI tools have not been sufficiently scientifically validated, with regard to the underlying criteria for the prediction of job performance (Chamorro-Premuzic et al., 2016; Dattner et al., 2019; Raghavan et al., 2020). This means that firms may reject candidates based on unexplained correlations and make decisions based on factors with no clear causal connection to job performance (Cappelli, 2019; Kim, 2017). When AI links the tone of voice to differences in job performance, it raises the additional ethical question of whether it is appropriate to screen out people based on physically determined and rather unchangeable attributes (Dattner et al., 2019). Moreover, the indirect measurement of personality itself is still an open and discussed topic (De Cuyper et al., 2017).

Second, technical implementation bears some risks. For example, Tambe et al. (2019) argued that good employees are hard to measure, as it is difficult to disentangle individual from group performance. Further, introducing technological context, such as video games or avatar interviewers, to the recruiting process may add noisy variance to applicants' performance and, thus, measurement error (Ryan & Derous, 2019). Therefore, a constant re-validation and control of the algorithmic tools is crucial. However, AI software vendors often do not publicly communicate whether or how they conduct validation studies on their models (Raghavan et al., 2020).

Third, Fernández-Martínez and Fernández (2020) brought up the risk that AI might not work equally for many people, undermining its accuracy. Along these lines, several studies (Buolamwini & Gebru, 2018; Raji & Buolamwini, 2019; Rhue, 2018) have shown that facial recognition software performs rather poorly, suffering from disparities in error rates across gender and race. Finally, Tambe et al. (2019) reported that AI recruiting faces the challenge of making trade-off decisions between accuracy and other ethical principles. For example, the authors stated that more "complicated" algorithms are more accurate, but they are also harder to explain, resulting in a trade-off between accuracy and explainability (we discuss the latter in the next paragraph).

These concerns about potential lack of validity and accuracy result in the question of whether it is ethical to use these new AI tools compared with more longstanding psychometric assessments that have been scientifically derived and validated. Which features are predictive, which are not, and which are protected? In particular, the selection of the features that define a good candidate is an ethically laden decision, about which current literature is ambivalent and further scientific validation is necessary (Schumann et al., 2020).

Transparency and Explainability

Another ethical opportunity mentioned in extant literature is the ability to establish transparency by providing applicants with updates and feedback throughout the process and in a timely fashion (e.g., via chatbots and AI technology), which can be considered one element of fair treatment (van Esch & Black, 2019). Often, firms fall short of providing relevant information in a timely manner, or they provide no information other than confirmation that a candidate's application has been received. This can be very frustrating for candidates. However, next to progress updates, AI further enables firms to generate detailed feedback and give millions of job applicants data-driven insights on their strengths and development needs (Dattner et al., 2019).

However, the use of AI can also lead to a lack of transparency toward applicants, when the use of AI and automated systems is not proactively communicated to candidates (Sánchez-Monedero et al., 2020). Moreover, the predictive and decision-making processes of algorithms are often opaque, even for the programmers themselves. When algorithms take millions of data points for the assessment of a candidate, it becomes difficult to provide a qualitative explanation of which attributes are driving the decisions (Raghavan et al., 2020; Simbeck, 2019). This is ethically critical in the personnel selection context, due to its high relevance for people's lives, and because this kind of black-box system may remain unchallenged, thereby obscuring discrimination (e.g., Tambe et al., 2019; Vasconcelos et al., 2018). Therefore, the GDPR also warrants a "right to explanation," by which people can ask for explanations about (algorithmic) decisions made about them (Pena et al., 2020).

Overall, the ethicality of AI recruiting depends highly on the mode in which it is implemented and used. On the one hand, it offers a huge ethical opportunity in the form of timely feedback for applicants; on the other hand, it bears the ethical risk of omitting transparency and explainability. Extant literature agrees that companies and recruiters should not rely on information produced by a black-box algorithm they do not fully understand. This is an open technical challenge to solve: building algorithms and AI applications that lead to explainable results (Schumann et al., 2020).

Accountability

Closely related to the issue of explainability is the topic of accountability in the hiring decision-making context. When automated AI technologies are used for decision-making, a question arises of whose job it is to adhere to ethical norms and labor laws, and who can be held responsible and accountable for the decisions made: the data scientists, the hiring managers or the company as a whole? This question becomes even more difficult when firms are not developing the AI themselves, but instead buying the technology from third-party vendors who want to protect their intellectual property and may not be willing to grant full transparency into the algorithms used (Sánchez-Monedero et al., 2020; Tambe et al., 2019). Lin et al. (2020) outlined that in the recruiting process, agents with different roles in the collective decision-making process can have a collective responsibility (i.e., each agent fulfills his or her role and shares a collective responsibility). Thus, when a recruiter makes a morally wrong decision based on a problematic recommendation by an AI, which in turn results from the negligence of a software engineer, both the recruiter and the engineer are collectively responsible and accountable for the wrong decision. Building on this discussion, Bornstein (2017) and Kim (2017) claimed that current regulation should be broadened, making companies that apply AI recruiting practices fully liable for any occurrence of discrimination or implicit bias in employment decisions.

It is clear that the AI itself cannot be held accountable; it should be a human agent who is ultimately responsible for the decision made when selecting an employee (Lin et al., 2020). However, the use of AI results in an obfuscation of responsibilities and accountabilities, which represents an ethical risk and must be clarified.

Human Oversight and Autonomy

The extent to which AI is integrated into the decision-making process varies across businesses. Some papers (Fernández-Martínez & Fernández, 2020; Yarger et al., 2020) have reported that increasingly more tasks are taken over by algorithms, though firms still rely on human recruiters to make the final decision. However, other papers (e.g., Lee, 2018; Vasconcelos et al., 2018) have stated that AI has already taken over the automated decision-making process, forwarding or rejecting candidates. This raises the question of whether it is ethical to base hiring decisions solely on algorithms and without human intervention. Sánchez-Monedero et al. (2020) even raised the point of whether, due to the new GDPR regulation, it is in fact illegal to use a solely automated hiring system in the EU, because the GDPR grants people the right to a "human in the loop." Overall, extant literature agrees that the loss of human oversight should be avoided, but human involvement in the training, validation, and deployment process should be maintained.

Additionally, Lin et al. (2020) raised the question of whether the usage of AI effects human autonomy. When AI applications and analyses shape human decisions by interfering with deliberation processes, the violation of human autonomy can become a serious ethical concern. The authors called this "AI paternalism" (p. 16). However, this topic is not further discussed in the identified literature. Thus, questions regarding how AI impacts the autonomy and dignity of candidates remain open.

Efficiency Gains and Effects on Internal Organization

In the first place, AI-advanced selection tools are attractive for organizations, as they make hiring more cost- and time-efficient (e.g., Lee, 2018; van Esch & Black, 2019). With the help of AI, employers have a greater ability to quickly shortlist candidates with high potential and streamline the selection process (Hipps, 2019; Persson, 2016; Savage & Bales, 2017). For example, AI technology provides firms with the ability to initially screen and process hundreds of applications in a short time frame (Persson, 2016). Moreover, AI-powered video interviews increase efficiency by reducing selection process time as well as candidate time and travel distances (Fernández-Martínez & Fernández, 2020).

However, the use of AI has further effects on the internal organization. The enhancement of recruiters' jobs is thereby considered an ethical opportunity of AI-enabled recruiting practices (Rąb-Kettler & Lehnervp, 2019; van Esch & Black, 2019). Daily, recruiters are confronted with numerous repetitive tasks, such as screening resumes, scheduling interviews and conducting similar conversations. When these tasks are taken over by AI, it results in a more meaningful job, as recruiters can undertake activities of higher value for the company. For instance, they can adapt better engagement techniques to ensure that a leading candidate accepts a job offer (Hipps, 2019; van Esch & Black, 2019) and can better focus on the individual candidates, stepping from a pure head hunter role into a career guide role (Rąb-Kettler & Lehnervp, 2019). Although the identified articles evaluated the effects of AI recruiting on the internal organizational members very positively, they must be studied in greater detail. For example, it needs to be tested whether a greater volume of candidates may prevent any gains in work time for recruiters (Ryan & Derous, 2019). Further, potential job losses of recruiters are not yet part of the discussion.

Perceived Fairness

Although the research on applicant reactions to technology-powered recruiting processes has increased in recent years (see Woods et al., 2020 for a review on applicant reactions to digital selection procedures), there is limited understanding of how people perceive AI recruiting and contrasting findings exist. Several studies of applicant reactions to AI interviews provide some cause for concern as they reveal that applicants perceived AI interviews as less fair and less favorable than face-to-face interviews with humans (Acikgoz et al., 2020; Lee, 2018; Newman et al., 2020). For example, Lee (2018) found that participants believe that AI lacks certain human skills that are required in the recruiting context: It lacks human intuition, makes judgments based on keywords, ignores qualities that are hard to quantify and is not able to make exceptions. Furthermore, some participants felt that using algorithms and machines to assess humans is demeaning and dehumanizing (Lee, 2018). In contrast to those findings, another group of papers (Langer et al., 2019a, 2019b, 2020; Suen et al., 2019) found no differences in perceived fairness between interviews with an AI and interviews with a human among job applicants, although most of them exhibited lower favorability to AI interviews.

Other studies (Gelles et al., 2018; Kaibel et al., 2019; Langer et al., 2018; van Esch & Black, 2019) examined the effect of different contextual factors on applicant reactions to the use of AI in hiring. For instance, Langer et al. (2018) found that applicants with a computer science background did not perceive AI recruiting differently from non-computer science applicants. Another study by Kaibel et al. (2019) examined the moderating effect of applicants' discrimination experience and uniqueness. They found that applicants who have experienced discrimination before perceive selection processes as fairer when an algorithm instead of a human makes the decision, whereas the negative effect of AI-based selection decisions on organizational attractiveness was stronger for individuals with a high sense of personal uniqueness. Underlining the relevance of perceived fairness, a study (van Esch & Black, 2019) found that the more job candidates perceive the AI-enabled recruiting system as providing fair treatment, the likelier they are to engage in and complete the recruiting process.

In a case study, van den Broek et al. (2019) found that different stakeholder groups may hold different and clashing notions of fairness, which may even be reconsidered during the implementation of AI recruiting in practice. For example, although AI tools are introduced to make the process fairer and decisions consistent across the company, it was observed that some recruiters did not use the algorithmic results consistently, but made exceptions, which they perceived as fairer.

Overall, there is no clear answer to the question of how AI recruiting is perceived. What is perceived as fair in one context may be judged differently in another. Although we found several studies examining the fairness perceptions of applicants, the perspective of current employees and HR managers on AI recruiting tends to be neglected. This leaves open the question of the extent to which HR managers trust and accept AI recruiting.

Approaches to Mitigate Ethical Risks

As shown in the previous section, the new AI technologies pose new challenges to regulation and governments, especially as they are being applied in recruiting. Some approaches to mitigating the emerging ethical risks in the AI recruiting context are discussed in extant literature.

Governmental Regulation

In the identified literature, it has been broadly claimed that more governmental regulation is needed to respond to the new developments in hiring: Whereas Kim (2017) argued for a legal response to what she called classification bias, Fernández-Martínez and Fernández (2020) called for governments to track selection processes and check for any infringement of fundamental employment laws or human rights. In their recent analysis, Raghavan et al. (2020) found that currently, vendors' practices in bias mitigation are heterogeneous. This suggests that evolving industry norms are sensitive to bias concerns but lack clear guidance on how to respond. However, as current regulation leaves room for unethical behavior of firms, today, employers need to think beyond governmental law when developing and using predictive hiring tools (Bogen, 2019).

Organizational Standards

Extant literature refers to various organizational standards that firms may and should implement to ensure ethical use of AI in recruiting. First, it is suggested that companies applying AI tools in the personnel selection process *comply with privacy laws* just as they would in traditional hiring. On the one hand, this means that organizations should fully protect and keep safe all sensitive data. On the other hand, recruiters should not use or predict any private or sensitive candidate information in the recruiting process. In addition, firms should proactively and fully brief candidates that their data will be analyzed by AI systems and obtain their consent (e.g., Chamorro-Premuzic & Akhtar, 2019; Simbeck, 2019). Second, firms should proactively and explicitly provide meaningful information on the hiring decision-making process, including information about the algorithmic techniques and data sets used, to ensure *transparency* and craft effective policy (Köchling et al., 2020; Raghavan et al., 2020; Sánchez-Monedero et al., 2020). Additionally, it should be always transparent to applicants whether they are communicating with another human or with AI (Simbeck, 2019). Third, several papers (e.g., Chamorro-Premuzic & Akhtar, 2019; Köchling et al., 2020) also suggested *human oversight on AI* as a standard for organizations. The authors encouraged a human review, in which experienced recruiters oversee the selection and evaluation made by AI. They argued that decisions should be made by an algorithm-informed human, rather than by an algorithm alone. Fourth, to further ensure and audit the implementation of these ethical standards, various authors have referred to compliance instruments companies should establish, such as an *AI ethics board* with an oversight function, consisting of representatives of relevant stakeholders who debate the data and ethical dimensions of AI algorithms and agree on boundaries for AI technology in the company (Simbeck, 2019; Tambe et al., 2019). In addition, Tambe et al. (2019) recommended specifying a *code of ethics* for AI-related initiatives within the company. Lastly, authors have encouraged *diverse data scientist teams* in organizations to foster inclusion and equity in AI (Giang, 2018; Yarger et al., 2020). In particular, in the ML algorithm development process, diverse voices across gender and race must be present to raise questions and check implicit assumptions.

Technical Due Diligence

Next to approaches on the governmental and organizational level, the identified literature also discusses technical methods to ensure ethical application of AI tools in recruiting. First, authors mentioned the *data literacy* of programmers, as well as the knowledge of hiring managers on how to use the AI solutions as a first prerequisite. Given that any data concerns can have a life-changing impact on applicants, companies need to have adequate levels of data and statistical skills to assure the accuracy and validity of the developed algorithms (Fernández-Martínez & Fernández, 2020; Lewis, 2018; Simbeck, 2019). Second, if companies do not develop the algorithms in-house, but buy more innovative skill tests or games from external vendors, practitioners are strongly encouraged to refer to *professional test standards* and obtain critical information about the tools: for example, evidence that informs psychometric reliability, criterion-related validity and bias implications (Oswald et al., 2020).

Third, the *ethicality of the AI tool design,* which should include bias mitigation techniques, plays a crucial role. For instance, some AI software vendors remove any wording or phrases that can unconsciously predict the gender of a candidate from CVs to circumvent unconscious bias and improve equity (e.g., Lin et al., 2020; Yarger et al., 2020). A different approach suggested by Williams et al. (2018) is to proactively gather and use social category data to illuminate and combat discriminatory practices. The authors argued that only when data are labeled with social categories can data scientists detect, understand, or remediate patterns of discrimination. Furthermore, open-source tools and technical frameworks for data scientists (e.g., IBM's "AI Fairness

360") can facilitate systematic bias checks and assist developers in embedding fairness in their algorithms (see Mujtaba & Mahapatra, 2019 for an overview of open-source toolkits). However, Sánchez-Monedero et al. (2020) pointed to the computational limitations of bias mitigation techniques and further argued that most bias mitigation systems aim at meeting the constraints of US law, which makes them not directly applicable in EU markets. In the context of ethical AI, Polli (2019) further referred to the movement among AI practitioners to develop a set of design principles for making AI ethical and fair (i.e., beneficial to everyone). She thereby emphasized the key principle according to which AI should be designed so that it can be easily audited. Rather than just assuming that algorithms yield accurate results, employers must regularly check the technology used for discrimination, as well as data errors and biases (e.g., Fernández-Martínez & Fernández, 2020; Hipps, 2019; Polli, 2019). Efforts must be made to constantly improve the robustness of any AI tool and, thus, *proactive auditing methods* should be implemented (Köchling et al., 2020). For example, outside professionals can be hired to build an internal auditing team to look at the AI decisions and audit key algorithms (Giang, 2018; Mann & O'Neil, 2016). They can carry out random spot checks on algorithmic recommendations, investigating in detail which candidates the algorithm has been selecting and why. To this end, Fernández-Martínez and Fernández (2020) developed an automated multi-agent software architecture to support auditing the recruiting process.

Lastly, companies need to be able to explain why a candidate has been selected and the causality regarding which specific attributes can be associated with their success in a role (Chamorro-Premuzic et al., 2019; Lewis, 2018). Thus, employers should not rely on black-box models, but develop AI applications that are interpretable (Lin et al., 2020). Transparency on algorithmic assumptions and models (e.g., in the form of *explainability reports*) is key in the mitigation of bias and when addressing trade-off decisions data scientists have to make (e.g., Mujtaba & Mahapatra, 2019; Tambe et al., 2019).

Awareness Among Employees

AI plays a critical role in technology to attack the diversity problem. It is therefore crucial that companies invest not only in AI technology, but also in people who are aware of both the opportunities and the risks that attend AI-powered recruiting practices (Chamorro-Premuzic et al., 2019). The awareness and sensibility of recruiters and data scientists about the potential bias and shortcomings of their algorithms is key to address the accompanying ethical challenges (Simbeck, 2019). When regulation is not enough to guide human behavior, ethical thinking and awareness of conscious use of predictive AI tools must be further promoted beyond regulation (Persson, 2016).

Discussion

Overall, we make four observations from structuring and synthesizing the current literature. First, this review indicates that there are various streams addressing ethical considerations of AI-based recruiting, but that insufficient attention has been given to ethical theories in this context. As various extant articles have a practitioner, legal, technical or descriptive focus, they tend to mention ethical considerations, but avoid normatively assessing them from a theoretical perspective. We identified only three theoretical articles, underlining the lack of a theoretical foundation within this field of research to date. However, by exploring the ethicality of AI recruiting, additional work based on ethical theory could prove beneficial to managers and organizations. Our review has shown that ethicality underlies the law and regulations in this area, but goes beyond them as well. Thus, more theoretical and normative papers are needed to provide organizations with a set of perspectives and suggested actions that may be taken to enhance morality in hiring (Alder & Gilbert, 2006).

Second, we found that some ethical concerns are prevalent in extant research, whereas others have not been sufficiently discussed. Most articles focus on human and algorithmic bias, whereas, for example, critical thoughts about accountability for AI-based recruiting practices, which were only mentioned in five of the 51 papers reviewed, are underrepresented. Thus, the field lacks an explicit discussion of the accountability of organizations for AI applications in recruiting, although this is a fundamental concern of today's research on ethics of AI, which treats the responsibility gap that may arise when an AI technology makes decisions independently and without direct human control (Johnson, 2015; Martin, 2018; Mittelstadt et al., 2016). Similarly, AI's impact on human autonomy and dignity, which is often considered an important principle for ethical AI (e.g., Floridi et al., 2018; University of Montreal, 2018), has only been briefly mentioned by one article and has not been assessed in detail.

Third, the identified solution approaches to mitigating ethical risks of AI applications are rather general and not specifically tailored to the recruiting context. They resemble the recommendations given in extant AI ethics guidelines. For instance, some of the mentioned solution approaches can be similarly found in the methods proposed by the High-Level Expert Group (High-Level Expert Group on Artificial Intelligence, 2019) to help implement trustworthy AI. However, a domain-specific focus would be desirable, because general normative guidelines do not have a tangible impact in many cases, precisely because of their generality and superficiality (Hagendorff, 2020). Instead, concrete implementation guidelines should be sensitive to contextual details and speak to the domain-specific regulation (High-Level Expert Group on Artificial Intelligence, 2019). In addition, ethical guidelines would benefit from being supplemented with detailed technical explanations. This would bridge the gap between abstract ethical principles and concrete technological implementations, for example by defining what it really means to implement privacy or transparency in AI systems in a given context (Hagendorff, 2020).

Finally, our systematic review reveals a predominance of non-empirical work. Only 18 articles (35% of papers reviewed) are empirical. Most examine perceived fairness and provide contrasting findings. Therefore, the ways in which people react to decisions made by AI in the recruiting context are still not well understood. Additional empirical research in this area is desirable, because respective findings may guide organizations on how to best use AI in selection to attract and retain top talent. Our mapping of ethical considerations reveals that there are other research topics, besides perceived fairness, that remain ambiguous. Topics such as the use of personal data or AI's impact on workforce diversity and assessment validity could also benefit from empirical evaluation.

In line with the findings of Chamorro-Premuzic et al. (2016), we observed that academic research struggles to keep pace with the rapidly evolving technology, allowing firms and vendors of recruiting technology to push the boundaries of justifiable selection practices. Addressing the identified research gaps, the following sections provide a more detailed roadmap for theoretical and empirical directions to advance research as well as a discussion of the practical implications of our findings.

Implications for Future Theoretical Research

Our review shows that insufficient attention has been given to relevant ethical frameworks in the research on the ethicality of AI-enabled recruiting. In fact, we did not find any article that provided a normative ethical analysis of AI recruiting by linking it to an established ethical lens such as a utilitarian, deontological or contract theory perspective (see Table 3). Thus, the emerging topic would benefit from being assessed through the lens of ethics theory, showing how these ethical schools would characterize morally relevant aspects of AI recruiting.

For example, one topic mentioned in our review that may be discussed controversially across the traditional schools is the ambit of privacy. As scraping social media platforms or face recognition techniques can gather and assess highly personal information on applicants' personality, health status, or sexual orientation, AI recruiting tools can be considered quite invasive technologies. On the one hand, contractarians, who pride themselves on their defense of private freedom from outward intrusion should find this problematic. Deontologists, who find morality in adherence to universal obligations, might agree with them in this assessment, especially when the respective information is gathered via an untransparent process. On the other hand, utilitarians, who see the greatest good or happiness for the greatest number of people as the most important value, are skeptical toward a strict differentiation between "the private" and "the public" (Seele et al., 2019). Thus, they might prefer to base their assessment on the practical consequences of employing applicant profiling based on private data. If this practice leads to the outcome that each position is filled by the best candidate, then the greatest good for the greatest number will usually be accomplished (Alder & Gilbert, 2006): The company will benefit from high productivity and enhanced competitiveness, customers and society will benefit from better products and services, shareholders will benefit from increasing profits, and also the employees will benefit from a higher job satisfaction.

A similar argumentation may apply to the topic of transparency and explainability. The implementation of transparent processes and explainable decision-making may not be important for utilitarians as long as the best candidates are hired. However, the deontological view may argue that the greatest good for the greatest number does not justify violating individuals' rights. Whereas, at this point, we have exemplified possible lines of argumentation of the classical ethical schools, future research should comprehensively delve deeper into each of the identified topics in our review and draw on major streams of ethical thinking, to mark and classify instances where AI recruiting would be approved or rejected by that ethical school.

Furthermore, AI recruiting should reconnect to the applied ethics fields of business ethics and AI ethics, where stronger theoretical contributions may be generated. Within business ethics research, the social contracts theory (Donaldson & Dunfee, 1994, 1999, 2000) perspective might be worth employing. Martin (2016) applied this approach to technology and online privacy, recognizing that people develop micro-social contracts with each provider, technological artifact, and circumstances as they navigate the increasingly interconnected world. By building on and extending Nissenbaum's framework of contextual integrity (2004, 2009), Martin (2016) further argued that stakeholder complaints about privacy violations are often due to changes in social contracts without consultation and approval. Future research could draw from this work to advance the field of AI recruiting using this micro-social contract narrative as a theoretical construct. This requires a detailed examination of expectations about ethical standards in the recruiting process, such as privacy and transparency, from an applicant perspective, as well as an analysis of whether, in this context, new technologies come into conflict with underlying norms.

Moreover, the field of AI recruiting and selection ethics can be positioned in the broader ongoing discourse on AI ethics. Current advances in the development and application of AI have, in recent years, been accompanied by the release of several ethics guidelines by various stakeholder groups. These include, for instance, the Montreal Declaration for Responsible AI (University of Montreal, 2018), the Ethics Guidelines for Trustworthy AI of the High-Level Expert Group on AI set up by the European Commission (High-Level Expert Group on Artificial Intelligence, 2019), and the AI4People's principles for AI ethics (Floridi et al., 2018; see Fjeld et al., 2020; Hagendorff, 2020; or Jobin et al., 2019 for a meta-analysis). In these documents, normative principles are developed to harness the disruptive potential and to tackle potential cases of misuse of AI technologies. Although these guidelines offer high-level guidance for AI applications in general, they need to be further tailored to the domain-specific use cases of AI, such as the recruiting context (High-Level Expert Group on Artificial Intelligence, 2019). What is deemed an appropriate action may depend on the domain in which AI is used and may differ between recruiting and other domains. To this end, there is a strong need for domain-specific works (Tolmeijer et al., 2020).

Thus, future research could build upon the general AI ethics frameworks and derive detailed guidelines for their operationalization in the recruiting context. For instance, concrete guidelines could be built on the AI4People's work (Floridi et al., 2018), which first proposed five principles to guide AI ethics: beneficence, non-maleficence, autonomy, justice, and explicability. Within each principle, concrete guidelines and ethical questions related to AI recruiting could be outlined. Thereby, the input from domain experts would be as important as the input from AI developers, implying the need for close collaboration between disciplines. Computer scientists and philosophers as well as domain experts and social science experts would have to work together to ensure the desired effects of ethical AI (Tolmeijer et al., 2020). Drawing from AI ethics frameworks could inform a more holistic view of ethical considerations in AI-enabled recruiting practices. Moreover, detractors' critiques that AI ethics initiatives provide few practical recommendations because they are vague and high level (Hagendorff, 2020; Mittelstadt, 2019) would thereby be rebutted.

Implications for Future Empirical Research

In summary, our review revealed several ethically ambivalent topics related to AI recruiting, which should be addressed by future empirical research. First, future research needs to better understand the accuracy and validity of AI recruiting tools (Woods et al., 2020). In this context, relevant questions are, for example: What are the criterion validities of different forms of AI in recruiting? Does AI recruiting outperform traditional selection procedures in terms of validity in any specific situations? To answer these questions, it may not be enough to establish measurement equivalence with traditional methods, which has been undertaken in the past, for example, when evaluating web-based assessment tools (e.g., Ployhart et al., 2003). Instead, research needs to approach the validation of AI assessment tools in their own right, rather than benchmarking it against traditional formats (Woods et al., 2020). To this end, quantitative studies that examine the validity of AI tool predictions, for example based on some measures of job performance, should be conducted. This has also been a common research design in the field of industrial, work and organizational (IWO) psychology for the examination of traditional recruiting methods (see, for example, Aguado et al., 2019).

Second, in our review, we touched on critical concerns related to informed consent and the use of personal data, which could be explored empirically in greater detail. In line with North-Samardzic (2019), we propose that future research could build on the findings of Hoofnagle et al. (2010) and Park (2013) by examining whether candidates are sufficiently informed about how AI is used during their application process and whether they understand the implications of AI technologies to be able to consent properly. Research is needed to clarify fundamental questions about the factors that determine applicants' privacy concerns. There may be differences between countries and cultures, attributable to differences in cultural and contextual factors, as well as privacy and data protection laws. A quantitative research design, e.g., in the form of online surveys, may be a suitable research design in this context (see, for example, Jeske & Shultz, 2019). The implications of research on these issues would help hiring managers shape their recruiting process and improve related privacy policies, ensuring an effective recruiting procedure (Woods et al., 2020).

Third, empirical studies on AI's effect on workforce diversity would be highly beneficial for the ongoing debate. This kind of empirical evidence would finally determine whether algorithms have the potential to overcome bias in hiring to establish diverse workforces. To this end, experimental research designs that examine the differences in decisions made by recruiters compared to AI decisions may be applicable. In addition, field data will be needed to increase external validity and make a final judgment on whether AI-based recruiting represents this huge ethical opportunity of more diverse workforces.

Finally, future research should further investigate and better understand the perceived fairness of AI recruiting. Adverse applicant reactions could have severe impacts for firms, as they might lead to negative outcomes, such as public complaints. Thus, applicant reaction research can offer relevant and practical advice for system designers and recruiters (Yarger et al., 2020). People's attitudes toward technologies have changed throughout history, and the same is expected to happen with the perception of AI applications, including AI-based recruiting tools (Lee, 2018). Thus, it would be interesting to study people's attitudes and perceptions of fairness over time, while increasingly more companies deploy AI tools in their recruiting processes and develop a current and up-to-date view on applicant reactions. Future research should thereby further shed light on the contextual and interactional factors that influence people's perception of AI-based recruiting decisions, because ethical concerns are often related to context (North-Samardzic, 2019; van den Broek et al., 2019). For instance, the role of the degree of an applicant's interaction with the AI could be examined (Lee, 2018). Applicants who directly interact with AI (e.g., via a chatbot or a video interview with a virtual AI agent) might perceive the AI-based procedure differently from applicants who do not interact with the AI, but whose CVs and test results have been analyzed by AI. Furthermore, the design features of gamified AI assessments (e.g., ease-of-use, mobile hosting or the nature of games themselves) and the positioning of AI tools in different stages of the hiring process could similarly affect reactions (Woods et al., 2020). Moreover, the type of job, the industry context, the cultural background, and other individual or demographic differences might affect an applicant's perception and are worth studying in greater detail.

While studying applicant reactions, it seems appropriate to primarily use a survey experiment methodology based on hypothetical situations or, alternatively, a lab design, in this early stage of research in this area (Woods et al., 2020). Nevertheless, these must be complemented with field studies involving people's actual experiences in high-stake selection situations to increase the external validity and generalizability of the findings (Acikgoz et al., 2020; Lee, 2018). Future research could thereby benefit from building upon Gilliland's (1993) theories of organizational justice, which explain factors that affect the perceived fairness of a selection system. However, it needs to go beyond that to re-define the changing nature of procedural justice in the context of AI recruiting, as well as the associated impacting factors and outcomes (Woods et al., 2020). Alternative models from the field of technology acceptance, such as Davis's (1989) technology acceptance model (TAM) or Venkatesh et al.'s (2003) unified theory of acceptance and use of technology (UTAUT), may also contribute to a better understanding of reactions to new technology in selection (Brenner et al., 2016). These models identify the core determinants of people's acceptance of new technologies, which may also be good predictors of applicants' reactions to the use of AI in recruiting.

Implications for Practice

By synthesizing and evaluating the ethical considerations around AI recruiting in the extant literature, our review provides implications for practice. We identified the core opportunities and risks of AI-enabled recruiting and selection practices, as well as a set of practical approaches to mitigate the latter. On the one hand, our review shows the ethical opportunities AI offers, such as the reduction of human bias in hiring or the ability to give timely and detailed feedback to applicants, which could help managers attain greater legitimacy within their organizations, as well as society, for their recruitment practices. On the other hand, our work stresses the importance of companies being aware of ethical risks that accompany the implementation of AI in recruiting. Even if AI software vendors advertise the avoidance of human bias, algorithms may be biased due to technical shortcomings, such as biased training sets or algorithmic design. Problems become even more complex when algorithms are based on ML and develop individually, so that developers are no longer able to explain how the AI has come to its decisions. Moreover, companies should be aware that the validity of the decisions made is not only determined by the AI itself, but also the underlying criteria used to predict job performance, which may not be scientifically validated (Chamorro-Premuzic et al., 2016; Dattner et al., 2019; Raghavan et al., 2020).

Overall, we observed contrasting views in the identified literature on the ethicality of AI recruiting. Even if we cannot offer a conclusive evaluation of whether the ethical opportunities outweigh the risks, managers need to understand the ethical concerns AI technologies might create and that algorithmic decisions might contradict what they aim to do with their workforce (Hickok, 2020). Thus, they must consider approaches to address those ethical concerns. In our review, we provide an overview of such practical approaches mentioned in the identified literature, although this list does not claim to be exhaustive.

As governmental regulation currently leaves room for unethical behavior of companies, firms should think and act beyond regulation and establish organizational standards to ensure the ethical use of AI recruiting tools. These might include compliance with privacy laws, transparency on AI usage, and human oversight on the AI in place. In addition, organizational compliance mechanisms, such as AI ethics boards or a code of ethics, could help to ensure ethical use of AI within firms. Indeed, in his study, Somers (2001) found that the presence of a corporate code of ethics is associated with less perceived wrongdoing in organizations. However, the author also pointed out that formal ethics codes should be considered as "one component of a milieu that encourages and supports high standards of ethical behavior" (p.

194) and that such codes need to be reinforced by supportive measures and values. Thus, to mitigate the ethical risks of AI applications in practice, a multi-tier approach is needed that includes all kinds of measures mentioned in our review, covering organizational standards, as well as technical due diligence and awareness among employees. It is crucial to anchor ethics competencies at the team and individual levels within organizations, e.g., via the implementation of diverse data scientist teams. Given that manifold ethical questions may arise in the development of algorithms, diverse voices and people who are aware of the potential shortcomings of recruiting algorithms are needed to check implicit assumptions and foster inclusion and equity.

Only by proactively tackling the ethical concerns, both in implementation and in external communication, can practitioners create new forms of AI recruiting practices that are both efficient and effective, and which also have the potential to manifest a competitive advantage and financial payoff (Bartneck et al., 2021).

Conclusion

AI tools have already become part of today's recruiting and selection practices. Our review of the literature on ethical consideration of AI-enabled recruiting organizes the extant research, which is still in an emerging stage. The topic is addressed from theoretical, practitioner, legal, technical and descriptive perspectives. By synthesizing the identified articles and ethically evaluating the considerations made, we provide researchers with guidance on the current state of the literature and establish a common basis for future research in the field. Furthermore, we identify gaps in extant research and reveal future research opportunities. A need exists for theoretical and empirical research bridging the gap between business ethics and AI recruiting applications in practice. Because the development and deployment of AI recruiting practices are increasing and come with a variety of ethical risks and ambiguities, we hope that our review will stimulate research to address the many remaining unstudied areas of AI-enabled recruiting and selection.

Notes

1. Grouping the 51 articles into the five categories yielded an average agreement of 86%. Cohen's kappa (κ) was 0.815 between the two raters, indicating "almost perfect" agreement (Landis & Koch, 1977, p. 165).
2. *Disparate treatment* discrimination refers to intentional discrimination based on protected attributes, where the employer intentionally treats people of a class protected under Title VII less variably than others.
3. *Disparate impact* or *adverse impact* discrimination refers to employment practices that appear neutral but have a discriminatory effect on a class protected under Title VII. The rule of thumb is the four-fifths rule: The selection rate for a protected group should not be less than four-fifths of the group with the highest selection rate.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare they have no conflict of interest.

Ethical Approval This article does not contain studies with human participants performed by the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment, 28*, 399–416. https://doi.org/10.1111/ijsa.12306

Aguado, D., Andrés, J. C., García-Izquierdo, A. L., & Rodríguez, J. (2019). LinkedIn "big four": Job performance validation in the ICT Sector. *Revista De Psicología Del Trabajo y De Las Organizaciones, 35*, 53–64. https://doi.org/10.5093/jwop2019a7

Alder, G. S., & Gilbert, J. (2006). Achieving ethics and fairness in hiring: Going beyond the law. *Journal of Business Ethics, 68*, 449–464. https://doi.org/10.1007/s10551-006-9039-z

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*, 671–732. https://doi.org/10.15779/Z38BG31

Bartneck, C., Luetge, C., Wagner, A., & Welsh, S. (2021). *An introduction to ethics in robotics and AI*. Berlin: Springer.

Bauer, T. N., Truxillo, D. M., Tucker, J. S., Weathers, V., Bertolino, M., Erdogan, B., et al. (2006). Selection in the information age: The impact of privacy concerns and computer experience on applicant reactions. *Journal of Management, 32*, 601–621. https://doi.org/10.1177/0149206306289829

Bîgu, D., & Cernea, M.-V. (2019). Algorithmic bias in current hiring practices: An ethical examination. In *13th international*

management conference (IMC) on management strategies for hiph Performance, Bucharest, Romania, October 31-November 1.
Blacksmith, N., Willford, J. C., & Behrend, T. S. (2016). Technology in the employment interview: A meta-analysis and future research agenda. *Personnel Assessment and Decisions, 2*, 12–20.
Bogen, M. (2019). All the ways hiring algorithms can introduce bias. *Harvard Business Review, May 6*.
Bornstein, S. (2017). Reckless discrimination. *California Law Review, 105*, 1055–1110. https://doi.org/10.15779/Z388P5V86M
Brenner, F. S., Ortner, T. M., & Fay, D. (2016). Asynchronous video interviewing as a new technology in personnel selection: The applicant's point of view. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.00863
Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. Friedler & C. Wilson (Eds.), *Conference on fairness, accountability, and transparency* (Vol. 81, pp. 1–15). ACM.
Cappelli, P. (2019). Data science can't fix hiring (yet). *Harvard Business Review, May-June*.
Chamorro-Premuzic, T. (2019). Will AI reduce gender bias in hiring? *Harvard Business Review, June 10*.
Chamorro-Premuzic, T., & Akhtar, R. (2019). Should companies use AI to assess job candidates? *Harvard Business Review, May 17*.
Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & Sherman, R. A. (2017). The datafication of talent: How technology is advancing the science of human potential at work. *Current Opinion in Behavioral Sciences, 18*, 13–16. https://doi.org/10.1016/j.cobeha.2017.04.007
Chamorro-Premuzic, T., Polli, F., & Dattner, B. (2019). Building ethical AI for talent management. *Harvard Business Review, November 21*.
Chamorro-Premuzic, T., Winsborough, D., Sherman, R., & A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology, 9*(3), 621–640.
Chapman, D. S., Uggerslev, K. L., & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews: A field investigation. *The Journal of Applied Psychology, 88*, 944–953. https://doi.org/10.1037/0021-9010.88.5.944
Chwastek, R. (2017). Cognitive systems in human resources. In Y. Demazeau, J. Gao, G. Xu, J. Kozlak, K. Muller, I. Razzak, et al. (Eds.), *4th international conference on behavioral, economic advance in behavioral, economic, sociocultural computing, Krakow, Poland, October 16–18, 2017*. New York: IEEE.
Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. In S. Matwin, S. Yu, & F. Farooq (Eds.), *International conference on knowledge discovery and data mining, Halifax, NS, Canada, August 13–17* (pp. 797–806). New York: ACM. https://doi.org/10.1145/3097983.3098095.
Dattner, B., Chamorro-Premuzic, T., Buchband, R., & Schettler, L. (2019). The legal and ethical implications of using AI in hiring. *Harvard Business Review, April 25*.
Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*, 319–340. https://doi.org/10.2307/249008
De Cuyper, J., De Houwer, J., Vansteelandt, K., Perugini, M., Pieters, G., Claes, L., et al. (2017). Using indirect measurement tasks to assess the self–concept of personality: A systematic review and meta–analyses. *European Journal of Personality, 31*, 8–41. https://doi.org/10.1002/per.2092
Donaldson, T., & Dunfee, T. W. (1994). Toward a unified conception of business ethics: Integrative social contracts theory. *Academy of Management Review, 19*(2), 252–284.
Donaldson, T., & Dunfee, T. W. (1999). *Ties that bind: A social contracts approach to business ethics*. Harvard Business Press.
Donaldson, T., & Dunfee, T. W. (2000). Precis for ties that bind. *Business and Society Review, 105*, 436–443. https://doi.org/10.1111/0045-3609.00092
Dwork, C., Hardt, M., Pitassi, T., & Reingold, O. (2012). Fairness through awareness. In *3rd conference on innovations in theoretical computer science, Cambridge, MA, USA, January 8–10* (pp. 214–226).
Fernández-Martínez, C., & Fernández, A. (2020). AI and recruiting software: Ethical and legal implications. *Paladyn: Journal of Behavioral Robotics, 11*, 199–216. https://doi.org/10.1515/pjbr-2020-0030
Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. C., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center for Internet & Society*.
Florentine, S. (2016, December 22). How artificial intelligence can eliminate bias in hiring. *CIO*. . Retrieved Jul 20, 2020 from https://www.cio.com/article/3152798/how-artificial-intelligence-can-eliminate-bias-in-hiring.html.
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, 689–707. https://doi.org/10.1007/s11023-018-9482-5
Gelles, R., McElfresh, D., & Mittu, A. (2018). Project report: Perceptions of AI in hiring. University of Maryland. Retrieved Dec 18, 2020 from https://anjali.mittudev.com/content/Fairness_in_AI.pdf.
Giang, V. (2018, May 8). The potential hidden bias in automated hiring systems. *Fast Company*. . Retrieved Aug 13, 2020 from https://www.fastcompany.com/40566971/the-potential-hidden-bias-in-automated-hiring-systems.
Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*, 99–120. https://doi.org/10.1007/s11023-020-09517-8
Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. In *30th conference on neural information processing systems (NIPS), Barcelona, Spain, Dec 4–9* .
Hickok, M. (2020). Why was your job application rejected: Bias in recruitment algorithms? (Part 1). Retrieved Aug 13, 2020 from https://montrealethics.ai/why-was-your-job-application-rejected-bias-in-recruitment-algorithms-part-1/.
High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. Retrieved Aug 13, 2020 from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
Hipps, C. (2019). Can AI reduce bias in talent selection? *Training Journal, December*, 22–24.
Hoofnagle, C. J., King, J., Li, S., & Turow, J. (2010). How different are young adults from older adults when it comes to information privacy attitudes & policies? Retrieved Aug 13, 2020 from http://repository.upenn.edu/ascpapers/399.
Jayaratne, M., & Jayatilleke, B. (2020). Predicting personality using answers to open-ended interview questions. *IEEE Access, 8*, 115345–115355. https://doi.org/10.1109/ACCESS.2020.3004002
Jeske, D., & Shultz, K. S. (2019). Social media screening and content effects: Implications for job applicant reactions. *International Journal of Manpower, 40*, 73–86. https://doi.org/10.1108/IJM-06-2017-0138
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2
Johnson, D. G. (2015). Technology with no human responsibility? *Journal of Business Ethics, 127*, 707–715. https://doi.org/10.1007/s10551-014-2180-1

Kaibel, C., Koch-Bayram, I., Biemann, T., & Mühlenbock, M. (2019). Applicant perceptions of hiring algorithms - Uniqueness and discrimination experiences as moderators. *Academy of Management Proceedings*. https://doi.org/10.5465/AMBPP.2019.210

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*, 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kim, P. T. (2017). Data-driven discrimination at work. *William & Mary Law Review, 58*(3), 857–937.

Kim, P. T., & Scott, S. (2018). Discrimination in online employment recruiting. *Saint Louis University Law Journal, 63*, 93.

Kluemper, D. H., Rosen, P. A., & Mossholder, K. W. (2012). Social networking websites, personality ratings, and the organizational context: More than meets the eye? *Journal of Applied Social Psychology, 42*, 1143–1172. https://doi.org/10.1111/j.1559-1816.2011.00881.x

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2020). Highly accurate, but still discriminatory: A fairness evaluation of algorithmic video analysis in the recruitment context. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-020-00673-w

Konradt, U., Warszta, T., & Ellwart, T. (2013). Fairness perceptions in web-based selection: Impact on applicants' pursuit intentions, recommendation intentions, and intentions to reapply. *International Journal of Selection and Assessment, 21*(2), 155–169.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

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

Langer, M., König, C. J., & Hemsing, V. (2020). Is anybody listening? The impact of automatically evaluated job interviews on impression management and applicant reactions. *Journal of Managerial Psychology, 35*, 271–284. https://doi.org/10.1108/JMP-03-2019-0156

Langer, M., König, C. J., & Papathanasiou, M. (2019a). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment, 27*, 217–234. https://doi.org/10.1111/ijsa.12246

Langer, M., König, C. J., Sanchez, D.R.-P., & Samadi, S. (2019b). Highly automated interviews: Applicant reactions and the organizational context. *Journal of Managerial Psychology, 35*, 301–314. https://doi.org/10.1108/JMP-09-2018-0402

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*. https://doi.org/10.1177/2053951718756684

Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., et al. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics, 160*, 377–392. https://doi.org/10.1007/s10551-019-04204-w

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology, 31*(4), 611–627.

Lewis, N. (2018, November 12). Will AI remove hiring bias? Strategic HR Review. Retrieved July 20, 2020 from https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/will-ai-remove-hiring-bias-hr-technology.aspx.

Lin, Y.-T., Hung, T.-W., & Huang, L.T.-L. (2020). Engineering equity: How AI can help reduce the harm of implicit bias. *Philosophy & Technology*. https://doi.org/10.1007/s13347-020-00406-7

Mann, G., & O'Neil, C. (2016). Hiring algorithms are not neutral. *Harvard Business Review, December 9*.

Martin, K. (2016). Understanding privacy online: Development of a social contract approach to privacy. *Journal of Business Ethics, 137*, 551–569. https://doi.org/10.1007/s10551-015-2565-9

Martin, K. (2018). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*, 835–850. https://doi.org/10.1007/s10551-018-3921-3

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*, 501–507. https://doi.org/10.1038/s42256-019-0114-4

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*, 205395171667967. https://doi.org/10.1177/2053951716679679

Mujtaba, D. F., & Mahapatra, N. R. (2019). Ethical considerations in AI-based recruitment. In M. Cunningham & P. Cunningham (Eds.), *IEEE International Symposium on Technology in Society, Medford, MA, USA, November 15–16* (pp. 1–7). IEEE. https://doi.org/10.1109/ISTAS48451.2019.8937920.

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. https://doi.org/10.1016/j.obhdp.2020.03.008

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*(1), 119–158.

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University.

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Sage.

North-Samardzic, A. (2019). Biometric technology and ethics: Beyond security applications. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-019-04143-6

Oswald, F. L., Behrend, T. S., Putka, D. J., & Sinar, E. (2020). Big data in industrial-organizational psychology and human resource management: Forward progress for organizational research and practice. *Annual Review of Organizational Psychology and Organizational Behavior, 7*, 505–533. https://doi.org/10.1146/annurev-orgpsych-032117-104553

Ozkazanc-Pan, B. (2019). Diversity and future of work: Inequality abound or opportunities for all? *Management Decision*. https://doi.org/10.1108/MD-02-2019-0244

Park, Y. J. (2013). Digital literacy and privacy behavior online. *Communication Research, 40*, 215–236. https://doi.org/10.1177/0093650211418338

Pena, A., Serna, I., Morales, A., & Fierrez, J. (2020). Bias in multimodal AI: Testbed for fair automatic recruitment. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, June 14–19* (pp. 129–137). IEEE. https://doi.org/10.1109/CVPRW50498.2020.00022.

Persson, A. (2016). Implicit bias in predictive data profiling within recruitments. In A. Lehmann, D. Whitehouse, S. Fischer-Hübner, L. Fritsch, & C. Raab (Eds.), *Privacy and identity management. Facing up to next steps* (Vol. 498, pp. 212–230). Cham: Springer.

Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgement tests comparable? *Personnel Psychology, 56*, 733–752.

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

Polli, F., Dolphin, J., & Kassir, S. (2019). On the basis of brains: How neuroscience and AI advance ethical hiring. *Workforce Solutions Review, 10*(4), 13–15.

Rąb-Kettler, K., & Lehnervp, B. (2019). Recruitment in the times of machine learning. *Management Systems in*

Production Engineering, 27, 105–109. https://doi.org/10.1515/mspe-2019-0018

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Conference on fairness, accountability, and transparency, Barcelona, Spain, January 27–30*. New York: Association for Computing Machinery.

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing. In V. Conitzer, G. Hadfield, & S. Vallor (Eds.), *AAAI/ACM conference on AI, ethics, and society, Honolulu, HI, USA, January 27–28* (pp. 429–435). New York: ACM. https://doi.org/10.1145/3306618.3314244.

Recruitment & Employment Confederation. (2020). Indiscriminate robots. *Recruiter - News & Business Intelligence for Recruiters, 17*(04), 2020.

Rhue, L. (2018). Racial influence on automated perceptions of emotions. *Social Science Research Network. SSRN 3281765*.

Ryan, A. M., & Derous, E. (2019). The unrealized potential of technology in selection assessment. *Journal of Work and Organizational Psychology, 35*(2), 85–92.

Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In *Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27–30* (pp. 458–468). New York: Association for Computing Machinery. https://doi.org/10.1145/3351095.3372849.

Savage, D. D., & Bales, R. (2017). Video games in job interviews: Using algorithms to minimize discrimination and unconscious bias. *ABA Journal of Labor & Employment Law, 32*, 34.

Schumann, C., Foster, J. S., Mattei, N., & Dickerson, J. P. (2020). We need fairness and explainability in algorithmic hiring. In B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, & G. Sukthankar (Eds.), *19th international conference on autonomous agents and multiagent systems (AAMAS 2020), Auckland, New Zealand, May 9–13* .

Seele, P., Dierksmeier, C., Hofstetter, R., & Schultz, M. D. (2019). Mapping the ethicality of algorithmic pricing: A review of dynamic and personalized pricing. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-019-04371-w

Simbeck, K. (2019). HR analytics and ethics. *IBM Journal of Research and Development, 63*(4/5), 1–12.

Somers, M. J. (2001). Ethical codes of conduct and organizational context: A study of the relationship between codes of conduct, employee behavior and organizational values. *Journal of Business Ethics, 30*, 185–195.

Suddaby, R., Bitektine, A., & Haack, P. (2017). Legitimacy. *Academy of Management Annals, 11*, 451–478. https://doi.org/10.5465/annals.2015.0101

Suen, H.-Y., Chen, M.Y.-C., & Lu, S.-H. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior, 98*, 93–101. https://doi.org/10.1016/j.chb.2019.04.012

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*, 15–42. https://doi.org/10.1177/0008125619867910

Thielsch, M. T., Träumer, L., & Pytlik, L. (2012). E-recruiting and fairness: The applicant's point of view. *Information Technology and Management, 13*, 59–67. https://doi.org/10.1007/s10799-012-0117-x

Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M., & Bernstein, A. (2020). Implementations in machine ethics: A survey. *ACM Computing Surveys, 53*, Article 132. https://doi.org/10.1145/3419633.

University of Montreal. (2018). Montreal declaration for a responsible development of artificial intelligence. https://www.montrealdeclaration-responsibleai.com/the-declaration.

van den Broek, E., Sergeeva, A., & Huysman, M. (2019). Hiring algorithms: An ethnography of fairness in practice. In *40th international conference on information systems, Munich, Germany, December 15–18* (Vol. 6).

van Esch, P., & Black, J. S. (2019). Factors that influence new generation candidates to engage with and complete digital, AI-enabled recruiting. *Business Horizons, 62*, 729–739. https://doi.org/10.1016/j.bushor.2019.07.004

van Iddekinge, C. H., Lanivich, S. E., Roth, P. L., & Junco, E. (2016). Social media for selection? Validity and adverse impact potential of a Facebook-based assessment. *Journal of Management, 42*, 1811–1835. https://doi.org/10.1177/0149206313515524

Vasconcelos, M., Cardonha, C., & Gonçalves, B. (2018). Modeling epistemological principles for bias mitigation in AI systems: An illustration in hiring decisions. In J. Furman, G. Marchant, H. Price, & F. Rossi (Eds.), *AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA, February 2–3* (pp. 323–329). New York: ACM. https://doi.org/10.1145/3278721.3278751.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478.

Wiechmann, D., & Ryan, A. M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment, 11*, 215–229.

Wiggers, K. (2020, March 13). AI weekly: Coronavirus spurs adoption of AI-powered candidate recruitment and screening tools. *VentureBeat.* Retrieved July 20, 2020 from https://venturebeat.com/2020/03/13/ai-weekly-coronavirus-spurs-adoption-of-ai-powered-candidate-recruitment-and-screening-tools/.

Williams, B., Brooks, C. F., & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy, 8*, 78. https://doi.org/10.5325/jinfopoli.8.2018.0078

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology, 29*, 64–77. https://doi.org/10.1080/1359432X.2019.1681401

Yarger, L., Cobb Payton, F., & Neupane, B. (2020). Algorithmic equity in the hiring of underrepresented IT job candidates. *Online Information Review, 44*, 383–395. https://doi.org/10.1108/OIR-10-2018-0334

Zafar, M. B., Valera, I., Gomez Rodriguez, M., & Gummadi, K. P. (2017). Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In R. Barrett, R. Cummings, E. Agichtein, & E. Gabrilovich (Eds.), *26th International World Wide Web Conference, Perth, Australia, April 3-7* (pp. 1171–1180). Geneva, Switzerland: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3038912.3052660.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A.2 How to Improve Fairness Perceptions of AI in Hiring: The Crucial Role of Positioning and Sensitization

Information about the article

Title	"How to Improve Fairness Perceptions of AI in Hiring: The Crucial Role of Positioning and Sensitization"
Authors	Anna Lena Hunkenschroer, Christoph Lütge
Accepted	July 12, 2021
Journal	*AI Ethics Journal*
Volume	2
Number	2
Year	2021
DOI	https://doi.org/10.47289/AIEJ20210716-3

Because we want to attach the article in the form in which it was published, it has no page numbers. The article covers pages 112–129 of this dissertation.

Copyright information by the publisher

Re: Permission letter

The AI Ethics Journal <editors@aiej.org>

Heute, 19:04

Hunkenschroer, Anna

Allen antworten |

Posteingang

Hi Anna,

Thank you for your email. As the author of the article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Please let us know if you have any further questions.

Best wishes,
The *AIEJ* Editorial Board

The AI Ethics Journal

AI Ethics Journal

Volume 2, Issue 2*
Spring 2021
*Special Conference Edition

AIRES 2021 Research Conference: *Data and Privacy Applications*

How to Improve Fairness Perceptions of AI in Hiring: The Crucial Role of Positioning and Sensitization

Anna Lena Hunkenschroer[1], Christoph Lütge[1]
[1]Department of Ethics, Technical University of Munich, Germany

DOI: https://doi.org/10.47289/AIEJ20210716-3

Abstract

Companies increasingly deploy artificial intelligence (AI) technologies in their personnel recruiting and selection processes to streamline them, thus making them more efficient, consistent, and less human biased. However, prior research found that applicants prefer face-to-face interviews compared with AI interviews, perceiving them as less fair. Additionally, emerging evidence exists that contextual influences, such as the type of task for which AI is used, or applicants' individual differences, may influence applicants' reactions to AI-powered selection. The purpose of our study was to investigate whether adjusting process design factors may help to improve people's fairness perceptions of AI interviews. The results of our 2 x 2 x 2 online study (N = 404) showed that the positioning of the AI interview in the overall selection process, as well as participants' sensitization to its potential to reduce human bias in the selection process have a significant effect on people's perceptions of fairness. Additionally, these two process design factors had an indirect effect on overall organizational attractiveness mediated through applicants' fairness perceptions. The findings may help organizations to optimize their deployment of AI in selection processes to improve people's perceptions of fairness and thus attract top talent.

History

Received **1 April 2021**

Accepted **12 July 2021**

Published **16 July 2021**

Keywords

Artificial Intelligence, Algorithmic Hiring, Employee Selection, Applicant Reactions, Fairness Perception, Trustworthy AI

Contact

Anna Lena Hunkenschroer
PhD Candidate,
Technical University of Munich
Munich, Germany
(anna.hunkenschroer@tum.de)

Professor Dr. Christoph Lütge
Chair Holder and Full Professor,
Chair of Business Ethics,
Technical University of Munich,
Munich, Germany
(luetge@tum.de)

Disclosure of Funding

The cost of data collection for the study was covered by the Technical University of Munich; the sponsor had no further involvement in the study.

AI Ethics Journal

1. Introduction

Organizations are increasingly utilizing artificial intelligence[1] (AI) in the recruiting and selection processes. By screening applicant resumes via text mining and analyzing video interviews via face recognition software, AI techniques have the potential to streamline these processes. AI thereby allows companies to process large numbers of applications and to make the candidate selection process faster, more efficient, and ideally, less prone to human bias (Acikgoz, Davison, Compagnone, & Laske, 2020).

However, research has fallen behind the rapid shift in the organizational usage of new selection processes, as well as applicants' perceptions of these processes (Woods, Ahmed, Nikolaou, Costa, & Anderson, 2020). Former research implies that novel technologies can detrimentally affect applicants' reactions to selection procedures (e.g., Blacksmith, Willford, & Behrend, 2016). Applicants' perceptions of recruiting processes are important, as they have meaningful effects on people's attitudes, intentions, and behaviors. For example, it has been shown that perceptions of selection practices directly influence organizational attractiveness and people's intentions to accept job offers (McCarthy et al., 2017).

The increasing incorporation of AI in the hiring process raises new questions about how applicants' perceptions are shaped in this AI-enabled process. A particular question involves the perception of fairness (Acikgoz et al., 2020): what does "fair" mean in this new context, and how are fairness perceptions of AI shaped? Although the amount of research on applicant reactions to technology-powered recruiting processes has increased in recent years (see Woods et al., 2020 for a review), there is still a limited understanding of whether people view recruiting decisions that AI makes as fair. Several studies on applicant reactions to AI recruiting practices provide some cause for concern, as they revealed that applicants perceived AI interviews as less fair and less favorable compared with face-to-face (FTF) interviews with humans (Acikgoz et al., 2020; Lee, 2018; Newman, Fast, & Harmon, 2020). In contrast, another group of papers (Langer, König, & Hemsing, 2020; Langer, König, & Papathanasiou, 2019; Langer, König, Sanchez, & Samadi, 2019; Suen, Chen, & Lu, 2019) found no differences in the perceived fairness between AI interviews and FTF interviews among job applicants, although most of them exhibited less favorability to AI interviews.

Unlike previous work that compared AI-based recruiting procedures with traditional ones, our study focused on ways in which to improve people's fairness perceptions of AI used in hiring. In the study, we explored how participants perceived the different process designs of AI recruiting procedures, rather than contrasting AI with humans. In line with the study by Gelles, McElfresh, and Mittu (2018), we focused on teasing out participants' feelings about how AI decisions are made, rather than focusing on their opinions about whether they should be made at all.

We thereby narrowed our focus to only one application of AI decision-making, which we considered to be particularly important, as it is increasingly used in practice: AI interviews (Fernández-Martínez & Fernández, 2020). AI interviews are structured video interviews where AI technology replaces a human interviewer and asks the candidate a short set of predetermined questions (Chamorro-Premuzic, Winsborough, Sherman, Ryne, A., & Hogan, 2016; Fernández-Martínez & Fernández, 2020). Then, the AI technology evaluates the actual responses and also makes use of audio and facial recognition software to analyze additional factors, such as the tone of voice, micro-facial movements, and emotions, to provide insights into certain applicant personality traits and competencies (Tambe, Cappelli, & Yakubovich, 2019).

AI Ethics Journal

Building on Gilliland's (1993) justice model, which assumes that the formal design factors of the selection process are crucial for applicants' fairness perceptions, we derived our research question regarding whether adjusting process design factors may help to improve people's fairness perceptions of AI interviews. We therefore selected and investigated three process design factors of AI interviews that previous evidence suggests may have the greatest influence on applicants' perceptions of fairness: (a) the positioning of the AI interview throughout the overall selection process; (b) applicants' sensitization to AI's potential to reduce human bias; and (c) human oversight of the AI-based decision-making process. We then proceeded to study the extent to which these factors affected participants' fairness perceptions. In addition, we examined the mechanism through which these process factors may affect overall organizational attractiveness.

We made three key contributions to the literature. First, our study linked the research on applicant reactions to selection procedures with research on AI ethics. Whereas research on applicant reactions is largely based on Gilliland's (1993) justice model assessing applicants' fairness perceptions in different selection processes, the discourse on AI ethics addresses how to implement fair and ethical AI. It is based on several ethics guidelines that provide very general normative principles to ensure the ethical implementation of AI technologies (e.g., High-Level Expert Group on Artificial Intelligence, 2019). Our study can be positioned at the intersection of these two streams: on the one hand, it addresses calls for empirical research on applicant reactions to new recruiting practices that involve the use of AI (Blacksmith et al., 2016; Langer, König, & Krause, 2017). On the other hand, we address the call for research on the ethical and fair implementation of AI in a domain-specific context (Hagendorff, 2020; High-Level Expert Group on Artificial Intelligence, 2019; Tolmeijer, Kneer, Sarasua, Christen, & Bernstein, 2020).

Second, our study was aimed at identifying ways in which to improve perceptions of AI interviews by adjusting the process design, thereby advancing research on contextual influences on applicant reactions. We extended the current theories of procedural fairness (e.g., Hausknecht, Day, & Thomas, 2004; Ryan & Ployhart, 2000) by experimentally demonstrating how the positioning of the AI interview, as well as candidates' sensitization to AI's potential to reduce human bias, can influence people's fairness perception of this tool.

Third, our work has practical implications, as it highlights how the process around AI interviews should be designed to lead to better applicant perceptions. This is an important question for anyone designing and implementing AI in hiring, especially employers whose hiring practices may be subject to public scrutiny (Gelles et al., 2018).

2. Background and Hypotheses

2.1 Applicant Reactions Towards the Use of AI in Recruiting

Although our work is the first to empirically examine how the process design factors of AI interviews may impact applicants' perceptions, it is not the first to examine people's reactions to AI recruiting in general. Building on research on applicant reactions to technology-based recruiting processes, several studies have investigated the use of AI tools for recruiting and selection. A couple of studies compared applicants' perceptions of fairness for AI-enabled interviews with traditional interviews with a human recruiter and found contrasting findings.

One group of papers (Acikgoz et al., 2020; Lee, 2018; Newman et al., 2020) provided some cause for concern, as they revealed that applicants perceived AI interviews as less fair and less favorable compared with FTF interviews with humans. For example, Lee (2018) found that

AI Ethics Journal

participants believed that AI lacks certain human skills that are required in the recruiting context: it lacks human intuition, makes judgments based on keywords, ignores qualities that are hard to quantify, and is not able to make exceptions. Furthermore, some participants felt that using algorithms and machines to assess humans is demeaning and dehumanizing (Lee, 2018). Similarly, Acikgoz et al. (2020) found that AI interviews are viewed as less procedurally and interactionally just, especially due to the fact that they offer fewer opportunities to perform.

In contrast to those findings, another group of papers (Langer et al., 2020; Langer, König, & Papathanasiou, 2019; Langer, König, Sanchez, & Samadi, 2019; Suen et al., 2019) found no differences in the perceived fairness between interviews with AI and interviews with a human, although most of them exhibited lower favorability to AI interviews. For instance, Langer, König, Sanchez, and Samadi (2019) found that participants thought that the organization using the highly automated interviews was less attractive because they perceived less social presence: however, they found that people perceive machines to be more consistent than humans are.

2.2 The Influence of Process Design Factors on Fairness Perceptions

In searching for conceptual reasons for differences in fairness perceptions, prior research referred to Gilliland's (1993) theoretical justice model, the most influential model to describe perceptions of the selection process (Basch & Melchers, 2019). It explains factors that affect the perceived fairness of a selection system, such as formal aspects of the selection process, candidates' opportunities to perform, or interpersonal treatment.

Growing evidence exists that contextual factors also play a role in applicant reactions to AI interviews (Langer, König, Sanchez, & Samadi, 2019). For example, it has been shown that the types of tasks for which AI is used (Lee, 2018), as well as an applicant's age (Langer, König, Sanchez, & Samadi, 2019) have significant impacts on the applicants' perceptions. Gelles et al. (2018) examined different designs of AI-enabled recruiting processes: specifically, they investigated whether the transparency or the complexity of algorithms as decision-makers impacted people's fairness perception or trust; they found no significant results. However, we aim to advance this stream on contextual influences in the form of process design factors by applying the underlying theory to the new AI recruiting context. Building on Gilliland's (1993) assumption that the formal characteristics of the selection process play an important role in perceptions of fairness, we selected three process design factors that previous evidence suggests may have the greatest influence on applicants' perceptions of fairness. Thus, we considered three process factors, namely: (a) the AI interview's positioning in the overall recruiting process; (b) people's sensitization to AI's potential to reduce human bias; and (c) human oversight in the AI decision-making process.

2.2.1 The Effect of the Positioning of AI in the Selection Process on Applicant Reactions

Gilliland's (1993) justice model identifies applicants' opportunity to perform, or show their job skills, as a crucial factor in their perceptions of procedural fairness. This implies that applicants view a selection process as fairer if they are better able to demonstrate their skills. This, in turn, means that if AI interviews could be positioned in the overall selection process in a way that gives applicants better opportunities to show their skills, they may increase people's perceptions of fairness.

Traditionally, an applicant submits a written cover letter and resume during the initial stage of the selection process, the screening stage. However, compared with these written application documents, an AI interview gives candidates the opportunity to demonstrate aspects of themselves as well as a variety of additional skills that

AI Ethics Journal

cannot be automatically derived from a resume, such as their personalities and their verbal communication skills (Chamorro-Premuzic et al., 2016). Therefore, when AI interviews are used as additional screening tools and not as decision tools later in the process, they could lead to increased chances for applicants to perform. In the context of situational judgement tests (SJTs), which are also used for assessing applicants, previous research found similar results (Lievens, Corte, & Westerveld, 2015; Patterson et al., 2012; Woods et al., 2020). Lievens et al. (2015) compared two response formats for a SJT: a video-based response that an applicant records, and a written response that a candidate provides; they found that applicants favored the digitally enhanced assessment for communicating their replies over the written response mode.

Prior studies on applicants' reactions to AI interviews (Langer, König, Sanchez, & Samadi, 2019; Lee, 2018) found that AI interviews were perceived as less fair than FTF interviews due to a lack of personal interaction. However, when the AI interview is used as additional tool in the initial screening stage rather than as a final decision-making tool substituting FTF interviews, this justification is no longer valid. Giving applicants FTF interviews later in the process should further reduce the negative impact resulting from the lack of a personal touch.

A qualitative study by Guchait, Ruetzler, Taylor, and Toldi (2014) was aimed at highlighting the appropriate uses of asynchronous video interviews, and found that applicants perceived this interview form to be ideal for screening large groups of applicants. However, they found video interviews to be less accepted among candidates for making final job offers. Because video interviews resemble AI interviews in that they lack interpersonal interaction with applicants, this finding could also be applicable to AI interviews. Considering this line of argumentation, we provide the following:

Hypothesis 1: Applicants perceive AI interviews to be fairer when used in the initial screening stage than when used in the final decision stage of the selection process.

2.2.2 The Influence of Explanations and Sensitization on Applicant Perceptions

According to previous research, applicant reactions can be positively affected by providing information and explanations on the selection procedure, which is also a central point of the selection justice model by Gilliland (1993). The information provided could thereby include diverse topics and may reduce uncertainty, increase transparency, or pronounce the job validity of the selection process, thus improving people's fairness perceptions. This has been shown for several selection procedures (Basch & Melchers, 2019; McCarthy et al., 2017; Truxillo, Bodner, Bertolino, Bauer, & Yonce, 2009).

However, in the context of AI recruiting, the effect of information seems to be complicated and may not always lead to better acceptance. We are aware of two studies (Gelles et al., 2018; Langer, König, & Fitili, 2018) that examined the effects of providing additional information about an AI-enabled interview on applicant reactions. Both did not find purely positive influences of the information given. Langer et al.'s (2018) investigation of the level of information revealed ambiguous findings: they showed that more detailed information positively impacts the perception of overall organizational attractiveness via higher transparency and open treatment, but also a direct negative effect on the overall organizational attractiveness. These two opposing effects indicate that applicants are, on the one hand, thankful that they are being treated honestly, but on the other hand, they might be somehow intimidated by the technological aspects of the selection procedure, or they may question it (Langer et al., 2018). Similarly, Gelles et al. (2018) studied the effect of transparency based on a higher level of the information

AI Ethics Journal

provided on applicant reactions, but they did not find a significant impact on applicants' fairness perceptions.

In the two studies, the information provided did not emphasize any specific advantages of the AI interview, but rather explained its specific features. In contrast, explanations that sensitize applicants to the opportunities of such interviews, such as a high degree of consistency and the reduction of human bias in the selection process, should evoke more positive reactions. According to Gilliland's model (1993), fairness perceptions relate to aspects of standardization, such as the independence of biases, or the same opportunity for all applicants to show their qualifications.

In the context of asynchronous video interviews, Basch and Melchers (2019) showed that explanations emphasizing the advantages of the standardization of these interviews can have positive effects on fairness perceptions. From this, we can infer the following for the context of AI interviews, where AI makes the recruiting decisions: an explanation that sensitizes people to AI's potential to reduce human bias in the process should improve how fairly they are perceived compared with an explanation that refers to the efficiency gains that AI has achieved. We suggest the following hypothesis:

> **Hypothesis 2:** Applicants perceive AI interviews to be fairer if they are sensitized to the potential of reducing human bias in the selection process compared with not being sensitized to this advantage of AI interviews.

2.2.3 Perceptions of AI Decision Agents and Human Oversight

Today, the extent to which AI is integrated into the recruiting decision-making process varies across businesses. In some organizations, AI is increasingly over more tasks, thus providing recruiters with additional information and analyses about applicants: however, they still rely on human recruiters to make the final decisions (Fernández-Martínez & Fernández, 2020; Yarger, Cobb Payton, & Neupane, 2020). In other firms, AI has already taken over the automated decision-making process, including forwarding or rejecting candidates (Vasconcelos, Cardonha, & Gonçalves, 2018). This variation in design across organizations raises the question of whether who the ultimate decision-maker in the selection process is might also have an impact on people's fairness perceptions.

Some empirical evidence exists that decision-makers prefer to rely on AI if they have the opportunity to adjust the AI's decision (Dietvorst, Simmons, & Massey, 2018; van den Broek, Sergeeva, & Huysman, 2019). In their case study, van den Broek et al. (2019) identified human resource (HR) managers' preferences to be able to make exceptions and to adjust AI-made decisions depending on the context. For them, the ability to differentiate between situated contexts and temporary changes in supply and demand is important to their perceptions of a fair selection process. Whereas these findings apply to people vested with decision-making authority, prior research on the question of how people who are affected by such decisions and who lack opportunities for control revealed ambiguous findings. Although Lee (2018) qualitatively found that most participants did not trust AI due to its inability to accommodate exceptions, Newman et al. (2020) found no significant impact of the human oversight of AI-made decisions on applicants' perceptions. They stated that only when a human, rather than an algorithm, is the default decision-maker will the decision be perceived as fair as one that is made purely by a human.

Moreover, a potential risk concerning HR managers' oversight and their option of human intervention in AI-made decisions is that this may raise concerns about the consistency of the process. The central advantage of AI

AI Ethics Journal

interviews related to the reduction of favoritism and human biases may thereby be undermined.

However, applicants assume that an organization considers them to be potential future employees (Chapman, Uggerslev, Carroll, Piasentin, & Jones, 2005), and applicants might thus expect organizations to invest the time and efforts of employees into the selection process. Thus, if applicants perceive that an organization does not invest time to hire personnel and instead relies on the automatic assessment and selection of applicants, this might violate applicants' justice expectations, therefore leading to a decrease in fairness perceptions (Langer et al., 2020). Therefore, we suggest the following hypothesis:

> **Hypothesis 3:** Applicants perceive AI interviews to be fairer when the hiring decision is AI-made with human oversight than without human oversight.

2.3 Impact of Fairness Perceptions on Organizational Attractiveness

Overall organizational attractiveness is an important outcome of applicant reactions to a selection method (Gilliland, 1993; Highhouse, Lievens, & Sinar, 2003). Evidence exists that whenever applicants take part in a selection process, they form perceptions about the organizations through their perceptions of the selection procedure (Rynes, Bretz, & Gerhart, 1991). Thus, when candidates perceive the selection procedure to be fairer, this could evoke better evaluations of the organizations' overall attractiveness (Bauer et al., 2006; Hausknecht et al., 2004). Accordingly, we argue that the three process design factors, namely positioning in the screening stage, sensitization to the potential to reduce human bias, and human oversight in the decision-making process of AI interviews might indirectly affect applicant reactions via their effect on fairness perceptions. Thus, we submit the following hypothesis:

> **Hypothesis 4:** Fairness perceptions will mediate the relationship among the three factors: (a) positioning in screening phase, (b) applicants' sensitization and (c) human oversight, and organizational attractiveness.

3. Method

To test our hypotheses, we conducted an online vignette study with an experimental 2 x 2 x 2 between-subject design in March 2021. The three factors were the positioning of the AI interview (initial stage vs. final stage), the sensitization of participants to bias reduction potential (sensitization vs. no sensitization), and human oversight of the AI decision (human oversight vs. no human oversight). This scenario-based method is commonly used in social psychology and ethics research to study the perceptions of decisions, particularly in the recruiting and selection contexts (see, for example, Acikgoz et al., 2020; Gelles et al., 2018; Lee, 2018).

3.1 Participants

As hiring is a process that affects most people at some point in their lives, we were not targeting a specific audience for our study but were rather interested in reaching a large population (Gelles et al., 2018). Therefore, participants (N = 450) were recruited on the platform of CloudResearch (powered by MTurk), similarly to the studies of, for example, Langer, König, Sanchez, and Samadi (2019) and Lee (2018). We exclusively recruited United States residents over the age of 18 as participants.

We collected answers from 450 participants who had passed an initial attention check and completed the survey. For the data analysis, we excluded participants who did not pass the second attention check (N = 14) or filled out the survey in less than 120 seconds (N = 14). Additionally, we excluded participants who did not appear to have taken the experiment seriously (N = 18) (e.g., due to answering

AI Ethics Journal

"strongly agree" to all items, including the reverse-coded items). This procedure left 404 participants in the final sample (62% female). The sample was 84% Caucasian, 4% Hispanic or Latino, 5% Black or African American, 4% Asian American, and 3% other.

3.2 Design and Procedure

Participants were randomly assigned to one of eight experimental groups. After reading one of the vignettes, which described a company that uses an AI interview in its selection process, the participants responded to items measuring their fairness and organizational attractiveness perceptions. The descriptions were equal in length and type of information, except for the three experimental manipulations. As mentioned above, the scenarios differed in three conditions: (a) whether AI was positioned in the initial screening or in the final decision stage, (b) whether people were sensitized to the bias reduction potential of AI, and (c) whether the decision was made with human oversight.

In creating the scenarios, we used a projective, general viewpoint rather than one that put the reader directly into the scenario, as we aimed to capture people's general perceptions of fairness rather than their personal preferences for particular procedures, which may vary. The scenario description can be found in the Appendix.

3.3 Measures

The participants responded to the items on a scale from 1 (Strongly disagree) to 5 (Strongly agree), which were presented in random order.

Perceived fairness was measured with three items from Warszta (2012). The three statements were: "I believe that such an interview is a fair procedure to select people," "I think that this interview itself is fair," and "Overall, the selection procedure used is fair."

Organizational attractiveness was measured using five items from Highhouse et al. (2003). Sample items were "For me, this company would be a good place to work," "This company is attractive to me as a place for employment," and "I am interested in learning more about this company."

4. Results

Table 1 and Table 2 provide an overview of descriptive statistics and correlations. To test our hypotheses, we conducted a factorial analysis of variance (ANOVA) for a simultaneous evaluation of main and potential interaction effects. As we were not expecting any significant interaction effects, and in line with the argumentation of Langsrud (2003), we performed the ANOVA based on Type II sums of squares. We included all three independent variables stated in Hypotheses 1–3, the three two-way interactions, and the three-way interaction between the three factors. Table 3 on the following page shows the results of the ANOVA.

AI Ethics Journal

Table 1
Correlations and Cronbach's Alpha for the Study Variables

	Scale	M	SD	1.	2.	3.	4.	5.
1.	Perceived fairness	3.23	1.15	0.95				
2.	Organizational attractiveness	3.04	1.14	0.66**	0.95			
3.	Positioning	0.49	0.5	0.13**	0.09	-		
4.	Sensitization	0.49	0.5	0.11*	0.15*	0.03	-	
5.	Oversight	0.52	0.2	0.02	-0.02	-0.04	0.05	-

Note: Variables 3 to 5 were constructed by dummy coding. Coding of Positioning: 1 = Positioning in initial screening stage, 0 = Positioning in final decision stage. Coding of Sensitization: 1 = Sensitization, 0 = No Sensitization. Coding of Oversight: 1 = Human oversight, 0 = No human oversight. N = 404. Numbers in the diagonal represents Cronbach's alpha of the scales. *p < 0.05, **p < 0.01

Table 2
Means and Standard Deviations for the Dependent Variables Across Experimental Groups

	Condition							
	Positioning in screening stage				Positioning in final decision stage			
	Sensitization		No sensitization		Sensitization		No sensitization	
	Oversight (n=53)	No Ov. (n=48)	Oversight (n=45)	No Ov. (n=52)	Oversight (n=55)	No Ov. (n=43)	Oversight (n=56)	No Ov. (n=52)
Variable	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Perceived fairness	3.45 (1.01)	3.61 (0.98)	3.38 (1.11)	3.1 (1.26)	3.1 (1.15)	3.29 (1.16)	3.09 (1.22)	2.86 (1.18)
Organizational attractiveness	3.16 (1.09)	3.4 (0.97)	2.99 (1.15)	3.02 (1.16)	3.3 (1.13)	3.32 (1.24)	2.91 (1.2)	2.6 (1.02)

Table 3
Results for the Factorial ANOVA (Type II Test) Including Effect Sizes (Partial η^2 and Cohen's f)

Response: Perceived Fairness	Df	Sum Sq	F-value	Pr (>F)	ηp^2	Cohen's f
Positioning	1	8.83	6.81	0.009**	0.02	0.13
Sensitization	1	6.29	4.85	0.028*	0.01	0.11
Oversight	1	0.18	0.14	0.711	$3.55e^{-4}$	0.02
Positioning:Sensitization	1	0.11	0.09	0.769	$3.84e^{-4}$	0.02
Positioning:Oversight	1	0.04	0.03	0.862	$2.51e^{-5}$	$5.01e^{-3}$
Positioning:Oversight	1	4.72	3.64	0.057	$9.10e^{-3}$	0.10
Positioning:Sensitization:Oversight	1	0.00	0.00	0.957	$7.22e^{-6}$	$2.69e^{-3}$
Residuals	396	513.59				

Note: *p < 0.05, **p < 0.01

As expected, we could not identify any significant interaction effect between the independent variables. Therefore, we focused on the analysis of the main effects of the three examined factors on perceived fairness.

Hypothesis 1 stated that participants would evaluate the AI interview as fairer when positioned in the initial screening stage than when positioned in the final decision stage. The results of the ANOVA indicated that overall, a significant difference existed between AI interviews in the screening versus the final decision stage, F (7, 396) = 6.81, $p < 0.01$, supporting Hypothesis 1. The observed effect size ($\eta_p{}^2 = 0.02$) indicated a small effect size.

Hypothesis 2 proposed that participants would perceive selection procedures as fairer when they received additional information on AI's potential to reduce human bias. The results of the ANOVA indicated that overall, a significant difference was found between the groups who were sensitized and not sensitized, F (7, 396) = 4.85, $p < 0.05$, supporting Hypothesis 2. The observed effect size ($\eta_p{}^2 = 0.01$) indicated a small effect size.

Hypothesis 3 posited that participants would perceive selection procedures as fairer when AI made the selection decision under the supervision of a human who would be able to adjust the AI's decision. The results of the ANOVA indicated that overall, no significant difference was found between AI decision-making with oversight and without oversight. Therefore, Hypothesis 3 was not supported.

Although we did not identify any significant interaction effects as already mentioned, we could observe a light effect, F (7, 396) = 3.64, $p < 0.1$, of the interaction between sensitization and human oversight on fairness perception. This means that sensitization has a stronger effect on fairness perceptions when no human oversight is involved in the process. When human oversight is involved, sensitization had only a small effect. This finding is intuitive because any option of human interference in the decision-making process may undermine the potential of AI to reduce human bias in the process.

Hypothesis 4 suggested that fairness perceptions would mediate the positive relation between the three process factors and overall organizational attractiveness. Therefore, a mediation analysis was conducted using a structural equation modeling (SEM) approach (Breitsohl, 2019). We tested the path from the three factors to organizational attractiveness through perceived fairness using a path analysis while also allowing a direct effect between the factors and organizational attractiveness. Mediation results are shown in Table 4.

The results indicated that the positive indirect effects of the positioning in the screening stage and people's sensitization through perceived fairness on organizational attractiveness were significant. This means that participants perceived organizational attractiveness resulting from these two factors to be higher because it conveyed higher perceived fairness. No significant indirect effect of human oversight on organizational attractiveness through perceived fairness was found. Hence, Hypothesis 4 was partially supported. The resulting model is presented in Figure 1 on the following page.

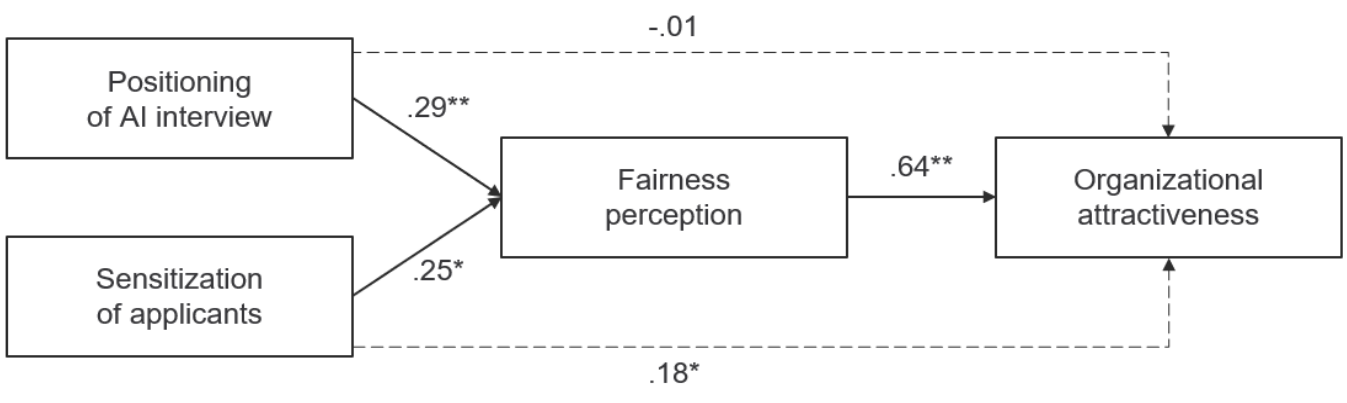
AI Ethics Journal

Table 4
SEM Results for the Mediation Analysis (Direct and Indirect Effects)

	Coefficient	SE	z-value	P(>|z|)
Direct effects				
Positioning → Org. Attractiveness	-0.01	0.09	-0.09	0.93
Sensitization → Org. Attractiveness	0.18	0.09	2.05	0.04*
Oversight → Org. Attractiveness	-0.08	0.09	-0.96	0.34
Positioning → Perceived fairness	0.29	0.11	2.61	0.01**
Sensitization → Perceived fairness	0.25	0.11	2.22	0.03*
Oversight → Perceived fairness	0.04	0.11	0.38	0.71
Perceived fairness → Org. Attractiveness	0.64	0.04	17.17	0.00**
Indirect effects				
Positioning → Perceived fairness → Org. Attractiveness	0.19	0.07	2.58	0.01**
Sensitization → Perceived fairness → Org. Attractiveness	0.16	0.07	2.20	0.03*
Oversight → Perceived fairness → Org. Attractiveness	0.03	0.07	0.38	0.59

Note: *p < 0.05, **p < 0.01

Figure 1
The Proposed Conceptual Model



AI Ethics Journal

5. Discussion

The aim of the current study was to identify ways in which to improve the fairness perceptions of AI interviews by examining the influences of three process design factors, namely (a) their positioning in the overall process, (b) applicants' sensitization to their potential to reduce human bias, and (c) human oversight of the AI decision-making process. The study thereby responded to the call for research on novel technologies for personnel selection (e.g., Blacksmith et al., 2016), as well as to the call for domain-specific work on the implementation of fair AI (e.g., Tolmeijer et al., 2020).

The results of our study showed that the positioning of AI in the initial screening stage as well as people's sensitization to the bias reduction potential of AI can have a positive effect on perceived fairness and thereby also indirectly on applicant reactions. We could not find significant differences in people's fairness perceptions depending on human oversight of the AI decision-making process.

Our results confirmed the qualitative findings of prior research (Guchait et al., 2014), validating the hypothesis that applicants perceive AI interviews as appropriate for screening large groups of applicants, but they are less accepted for making final job offers. Giving applicants the perspective of having an FTF interview may also reduce negative perceptions that may be driven by the lack of personal interaction in the selection process.

Furthermore, our results are more encouraging than the findings by Langer et al. (2018) who found that providing more information on the technological aspects of AI interviews may lead to both positive and negative effects. In line with previous evidence on the beneficial effects of explanations concerning other selection procedures (e.g., Basch & Melchers, 2019), we found that an explanation stressing AI's potential to reduce human bias can help to mitigate applicants' skeptical views of AI interviews. It should be noted that this effect can be stronger when AI is the sole decision-maker and human recruiters can make no exceptions.

Finally, this study investigated whether human oversight in the process affects fairness perceptions. Although prior research has shown that FTF interviews with a human recruiter led to an overall higher perceived fairness (e.g., Acikgoz et al., 2020), it appears that the mere opportunity for human agents to adjust the AI decision was inadequate for improving perceptions of fairness. Therefore, the assumption that human oversight of the AI decision-making process has a positive impact on fairness perceptions has to be currently dismissed, as fairness seems to require a high level of human discretion (Newman et al., 2020).

5.1 Limitations

A limitation of our study is that we used CloudResearch to recruit participants, which allowed us broad recruitment. However, this recruitment panel has known data biases (Difallah, Filatova, & Ipeirotis, 2018; Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010). Thus, our participant pool is, for example, more Caucasian and more female than the general US population is. Therefore, our sample may be considered to be a convenience sample, which may limit the external validity of our results.

Moreover, our use of an experimental design, which allows for greater internal validity but may lack the fidelity of an actual job application situation, is another limitation. Given the early stage of research in the area of AI recruiting, it seems appropriate for us to use this type of survey experiment methodology. Nevertheless, this form of studies must be complemented with field studies involving people's actual experiences in high-stake selection situations to increase the external validity and generalizability of the findings (Acikgoz et al., 2020; Lee, 2018).

5.2 Practical Implications

This study has important practical implications. Even if the implementation of AI in hiring enhances efficiency, organizations should pay attention to the possible detrimental effects on their applicant pools. This is especially true in times of a tense labor market where every applicant is a potential market advantage because applicants might withdraw their applications if they perceive the selection procedure to be unfair (Langer, König, Sanchez, & Samadi, 2019). Therefore, companies should think about ways in which to improve applicants' fairness perceptions. The current study fills an important gap in the literature and provides empirical evidence addressing the question of how to improve people's fairness perceptions of AI interviews.

First, our paper might provide guidance to firms on how to position an AI interview in the overall selection process. Our findings suggest that firms should use AI interviews as additional screening tool in an early stage of the recruiting process rather than as final decision-making tool. Organizations should consider to complement AI interviews with FTF interviews in a later stage of the selection process to ensure a certain level of human interaction, as well as to ensure that applicants feel that they are valued as individuals rather than as data points only (Acikgoz et al., 2020).

Moreover*,* to prevent negative reactions by applicants, organizations should use explanations that emphasize the advantages of AI interviews regarding their potential to reduce human bias in the process. Underlining this potential is a cost-effective way to give applicants an understanding of the reasons for the usage of these interviews and to make their advantages more salient to applicants. When doing so, organizations should apply AI interviews consistently and prevent exceptions that human recruiters make so as not to undermine this potential of AI again. As an increasing number of companies adopt AI interviews, industrial educators or universities may also consider educating future applicants about this new form of interview, including its advantages and risks (Suen et al., 2019).

5.3 Future Research

Regarding the role of explanations and applicants' sensitization, it would be interesting to examine how sensitization to a topic might occur and how explanations are presented. For example, companies could show welcome videos before the actual applicant interviews. Sensitizing applicants with a welcome video might even amplify its beneficial effects compared with written text, as this might help to ensure that applicants do not overlook it (Basch & Melchers, 2019).

Additionally, future research might investigate other contextual influences on reactions to AI tools in the selection process (Langer, König, Sanchez, & Samadi, 2019). For instance, the role of the degree of an applicant's interaction with AI might be an interesting topic (Lee, 2018). Applicants who directly interact with AI (e.g., via a chatbot or a video interview with a virtual AI agent) might perceive the AI-based procedure differently from applicants who do not interact with AI but whose resumes and test results have been analyzed by AI. Furthermore, the design features of gamified AI assessments (e.g., ease of use, mobile hosting, or the nature of the games themselves) could similarly affect reactions (Woods et al., 2020). Moreover, the type of job, the industry context, the cultural background, and other individual or demographic differences might affect an applicant's perception and thus are worth studying in greater detail.

Finally, additional research that goes beyond applicant reactions is necessary (Basch & Melchers, 2019). For instance, further research needs to foster a better understanding of the accuracy and validity of AI recruiting tools (Woods et al., 2020). In this context, relevant

AI Ethics Journal

questions are, for example: What are the criterion validities of different forms of AI in recruiting? Does AI recruiting outperform traditional selection procedures in terms of validity in any specific situations? For answering these questions, it may not be enough to establish measurement equivalence with traditional methods, which has been undertaken in the past, for example, when evaluating web-based assessment tools (e.g., Ployhart, Weekley, Holtz, & Kemp, 2003). Instead, research needs to approach the validation of AI assessment tools in their own right, rather than benchmarking them against traditional formats (Woods et al., 2020).

6. Conclusion

In our study, we aimed to find ways in which to improve people's fairness perceptions of AI interviews. To this end, we examined three process design factors, namely the positioning of the AI interview throughout the selection process; the sensitization of participants to the potential of AI to reduce human bias; and human oversight of the AI decision-making process, as well as their influence on people's perception of fairness. We found that two of these factors –positioning and sensitization– are critical to people's perception of AI interviews. If properly designed, they can help to improve applicants' reactions to AI interviews to prevent negative effects on organizations that use such interviews. We believe this work could be valuable for organizations that implement AI in their hiring processes to make better decisions about how to use AI interviews that people will find trustworthy and fair.

Notes

We refer to a broad concept of AI, which can be defined as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019, p. 17). AI thereby includes complex machine learning approaches such as deep neural networks, but also covers simple algorithms relying on regression analyses as well as other kinds of algorithms, such as natural language processing or voice recognition.

AI Ethics Journal

References

1. Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, *28*(4), 399–416. https://doi.org/10.1111/ijsa.12306
2. Basch, J., & Melchers, K. (2019). Fair and flexible?! Explanations can improve applicant reactions toward asynchronous video interviews. *Personnel Assessment and Decisions*, *5*(3). https://doi.org/10.25035/pad.2019.03.002
3. Bauer, T. N., Truxillo, D. M., Tucker, J. S., Weathers, V., Bertolino, M., Erdogan, B., & Campion, M. A. (2006). Selection in the information age: The impact of privacy concerns and computer experience on applicant reactions. *Journal of Management*, *32*(5), 601–621. https://doi.org/10.1177/0149206306289829
4. Blacksmith, N., Willford, J. C., & Behrend, T. S. (2016). Technology in the employment interview: A meta-analysis and future research agenda. *Personnel Assessment and Decisions*, *2*, 12–20.
5. Breitsohl, H. (2019). Beyond ANOVA: An introduction to structural equation models for experimental designs. *Organizational Research Methods*, *22*(3), 649–677. https://doi.org/10.1177/1094428118754988
6. Chamorro-Premuzic, T., Winsborough, D., Sherman, Ryne, A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology*, *9*(3), 621–640.
7. Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology*, *90*(5), 928–944. https://doi.org/10.1037/0021-9010.90.5.928
8. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643
9. Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and dynamics of Mechanical Turk workers. In Y. Chang, C. Zhai, Y. Liu, & Y. Maarek (Eds.), *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 135–143). ACM. https://doi.org/10.1145/3159652.3159661
10. Fernández-Martínez, C., & Fernández, A. (2020). AI and recruiting software: Ethical and legal implications. *Paladyn: Journal of Behavioral Robotics*, *11*(1), 199–216. https://doi.org/10.1515/pjbr-2020-0030
11. Gelles, R., McElfresh, D., & Mittu, A. (2018). Project report: Perceptions of AI in hiring. Retrieved from https://anjali.mittudev.com/content/Fairness_in_AI.pdf
12. Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, *18*(4), 694–734.
13. Guchait, P., Ruetzler, T., Taylor, J., & Toldi, N. (2014). Video interviewing: A potential selection tool for hospitality managers – A study to understand applicant perspective. *International Journal of Hospitality Management*, *36*, 90–100. https://doi.org/10.1016/j.ijhm.2013.08.004
14. Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8
15. Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*, 639–683.

16. Highhouse, S., Lievens, F., & Sinar, E. F. (2003). Measuring attraction to organizations. *Educational and Psychological Measurement*, *63*(6), 986–1001. https://doi.org/10.1177/0013164403258403
17. High-Level Expert Group on Artificial Intelligence (April 2019). Ethics guidelines for trustworthy AI. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
18. Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004
19. Langer, M., König, C. J., & Fitili, A. (2018). Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection. *Computers in Human Behavior*, *81*, 19–30. https://doi.org/10.1016/j.chb.2017.11.036
20. Langer, M., König, C. J., & Hemsing, V. (2020). Is anybody listening? The impact of automatically evaluated job interviews on impression management and applicant reactions. *Journal of Managerial Psychology*, *35*(4), 271–284. https://doi.org/10.1108/JMP-03-2019-0156
21. Langer, M., König, C. J., & Krause, K. (2017). Examining digital interviews for personnel selection: Applicant reactions and interviewer ratings. *International Journal of Selection and Assessment*, *25*(4), 371–382. https://doi.org/10.1111/ijsa.12191
22. Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, *27*(3), 217–234. https://doi.org/10.1111/ijsa.12246
23. Langer, M., König, C. J., Sanchez, D. R.-P., & Samadi, S. (2019). Highly automated interviews: applicant reactions and the organizational context. *Journal of Managerial Psychology*, *35*(4), 301–314. https://doi.org/10.1108/JMP-09-2018-0402
24. Langsrud, O. (2003). ANOVA for unbalanced data: Use type II instead of type III sums of squares. *Statistics and Computing*, *13*, 163–167.
25. Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718756684
26. Lievens, F., Corte, W. de, & Westerveld, L. (2015). Understanding the building blocks of selection procedures. *Journal of Management*, *41*(6), 1604–1627. https://doi.org/10.1177/0149206312463941
27. McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing "So what?," "What's new?," and "Where to next?". *Journal of Management*, *43*(6), 1693–1725. https://doi.org/10.1177/0149206316681846
28. 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. https://doi.org/10.1016/j.obhdp.2020.03.008
29. Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O'Neill, P. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education*, *46*(9), 850–868. https://doi.org/10.1111/j.1365-2923.2012.04336.x
30. Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgement tests comparable? *Personnel Psychology*, *56*, 733–752.

AI Ethics Journal

31. Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? Shifting demographics in Mechanical Turk. In *28th Annual CHI Conference on Human Factors in Computing Systems,* Atlanta, GA, USA.
32. Ryan, A. M., & Ployhart, R. E. (2000). Applicant perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, *26*, 565–606.
33. Rynes, S. L., Bretz, R. D., & Gerhart, B. (1991). The importance of recruitment in job choice: A different way of looking. *Personnel Psychology*, *44*, 487–521.
34. Suen, H.-Y., Chen, M. Y.-C., & Lu, S.-H. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior*, *98*, 93–101. https://doi.org/10.1016/j.chb.2019.04.012
35. Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, *61*(4), 15–42. https://doi.org/10.1177/0008125619867910
36. Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M., & Bernstein, A. (2020). Implementations in machine ethics: A survey. *ACM Computing Surveys*, *53*(6), Article 132. https://doi.org/10.1145/3419633
37. Truxillo, D. M., Bodner, T. E., Bertolino, M., Bauer, T. N., & Yonce, C. A. (2009). Effects of explanations on applicant reactions: A meta-analytic review. *International Journal of Selection and Assessment*, *17*, 346–361.
38. Van den Broek, E., Sergeeva, A., & Huysman, M. (2019). Hiring algorithms: An ethnography of fairness in practice. In *40th International Conference on Information Systems,* Munich, Germany.
39. Vasconcelos, M., Cardonha, C., & Gonçalves, B. (2018). Modeling epistemological principles for bias mitigation in AI systems: An illustration in hiring decisions. In J. Furman, G. Marchant, H. Price, & F. Rossi (Eds.), *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 323–329). New York, NY, USA: ACM. https://doi.org/10.1145/3278721.3278751
40. Warszta, T. (2012). *Application of Gilliland's model of applicants' reactions to the field of web-based selection* (unpublished dissertation). Christian-Albrechts Universität Kiel, Kiel.
41. Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, *29*(1), 64–77. https://doi.org/10.1080/1359432X.2019.1681401
42. Yarger, L., Cobb Payton, F., & Neupane, B. (2020). Algorithmic equity in the hiring of underrepresented IT job candidates. *Online Information Review*, *44*(2), 383–395. https://doi.org/10.1108/OIR-10-2018-0334

AI Ethics Journal

Appendix

Scenario Description

Imagine the recruiting and selection process of a company looking for talented employees.

In the [initial screening / final decision] round of the selection process, the company uses a video interview format, which is enabled by Artificial Intelligence (AI). One of the reasons the company has adopted this AI-powered solution is to make the interview process more [consistent across applicants and reduce human bias in / efficient and reduce the time and cost of] the selection process. Therefore, applicants are sent a link via email to start the interview process using a webcam on their computer. Throughout the interview, the AI software asks the applicants structured interview questions, such as "tell me about a time where you had to improve a process and how that has helped you in your career".

The responses are recorded on the computer and then rated by the AI software based on the content, as well as the applicants' vocal tone and non-verbal behavior. The AI software, [under / without] the supervision by an HR manager who is able to adjust the AI's decision, then decides whether an applicant will be [invited to the next round of FTF interviews / offered the position] or not.

The next day, applicants are informed about the company's decision.

B Contributions to Publications

B.1 Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda

Published in the Journal of Business Ethics *(2022), https://doi.org/10.1007/s10551-022-05049-6 with Christoph Lütge*

- Developing the research question and design was a joint effort both authors.
- I was responsible for conducting the literature search and analysis.
- I was responsible for writing the article.
- As corresponding author, I was responsible for the coordination of the submission and revision process.
- Revising the article was a joint effort by both authors.

Anna Lena Hunkenschroer Lead author	Christoph Lütge Coauthor

B.2 How to Improve Fairness Perceptions of AI in Hiring: The Crucial Role of Positioning and Sensitization

Published in the AI Ethics Journal, 2(2)-3 (2022), https://doi.org/10.47289/AIEJ20210716-3 with Christoph Lütge

- I was responsible for developing the research question and design.
- Deriving the hypotheses was a joint effort by both authors.
- I was responsible for programming and conducting the experimental vignette study.
- I was responsible for conducting the data analyses.
- I was responsible for writing the article.
- As corresponding author, I was responsible for the coordination of the submission and revision process.
- I was responsible for revising the article.

______________________________	______________________________
Anna Lena Hunkenschroer	Christoph Lütge
Lead author	Coauthor

B.3 Can AI Close the Gender Gap in the Job Market? Individuals' Preferences for AI Evaluations

Submitted to Organizational Behavior and Human Decision Processes *with Christoph Hohenberger*

- I made the main contribution to developing the research question and design.
- I was responsible for deriving the hypotheses.
- Programming the experiment was a joint effort by both authors.
- I was responsible for conducting the experiment.
- I was responsible for conducting the data analyses.
- I was responsible for writing the article.
- As corresponding author, I was responsible for the coordination of the submission process.

Anna Lena Hunkenschroer
Lead author

Christoph Hohenberger
Coauthor