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

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.

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).

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

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

"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: *$p < 0.05$, **$p < 0.01$

AI Ethics Journal

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.

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
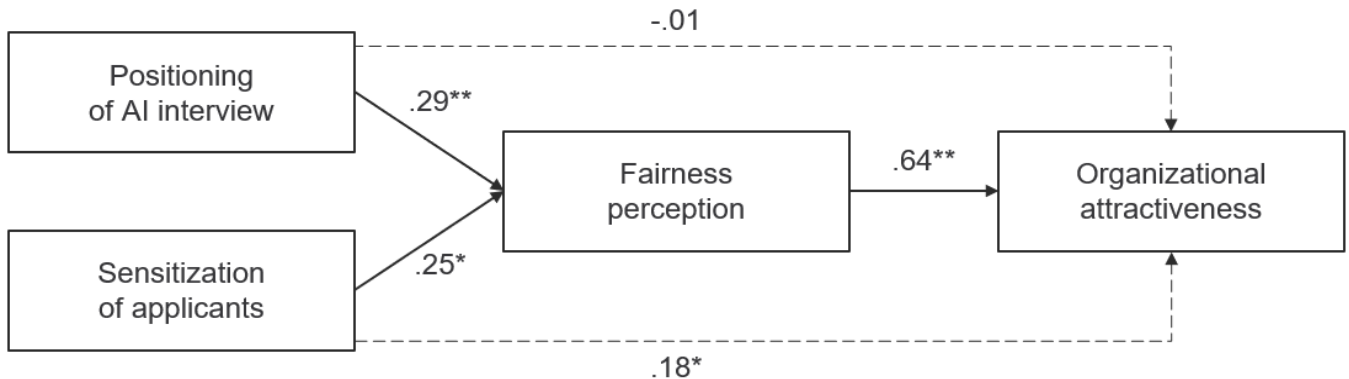
AI Ethics Journal

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).

AI Ethics Journal

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.

AI Ethics Journal

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.