Ensembling vs. Delegating: Different Types of AI-Involved Decision-Making and Their Effects on Procedural Fairness Perceptions

Research Paper

Christopher Diebel[1], Akylzhan Kassymova[2], Mari-Klara Stein[3], Martin Adam[3], and Alexander Benlian[1]

[1] Technical University of Darmstadt, Information Systems and Electronic Services, Darmstadt, Germany
{diebel, benlian}@ise.tu-darmstadt.de
[2] Technical University of Darmstadt, Darmstadt, Germany
aklyzhan.kassymova@stud.tu-darmstadt.de
[3] Tallinn University of Technology, Department of Business Administration, Tallin, Estonia
mari-klara.stein@taltech.ee
[4] Georg-August-University Göttingen, Smart Services, Göttingen, Germany
martin.adam@uni-goettingen.de

Abstract. Various approaches exist for incorporating artificial intelligence (AI) systems into decision-making processes to improve the accuracy and efficiency of decisions in organizations. However, involving AI systems in decision-making can provoke adverse reactions from decision recipients. Therefore, it is essential to deepen our understanding of how different types of AI-involved decision-making processes affect decision recipients' perceptions. To do so, we drew on fairness heuristic theory and conducted a vignette-based online experiment with 79 participants. Our results show that in a performance evaluation context, receiving a decision from a fully delegated AI decision-making process is perceived as less fair than one made solely by a manager. Moreover, the reduced perception of fairness undermines the perceived trustworthiness of the manager. Importantly, these negative effects do not occur when an ensemble decision-making process is used, in which both an AI system and a manager are equally involved.

Keywords: *Decision-Making, AI Systems, Procedural Fairness, Ensemble, Delegation*

1 Introduction

AI systems have emerged as increasingly relevant for decision-making within organizations across several industries. They can enhance decision-making in hiring processes by evaluating job candidates (Choudhary et al., 2023), provide credit scores in financial services (Strich et al., 2021), and assist human resource managers in evaluating employees by generating performance reviews (Macorva, 2025). This development aligns with rapid advancements in AI systems' decision-making capabilities. AI systems have improved significantly during the last few years in accuracy and efficiency (Anthropic, 2024), enabling them even to outperform humans in certain tasks such as image classification (Fuegener et al., 2022). Furthermore, AI systems have become more autonomous and are increasingly capable of interacting in natural and human-like ways (Baird & Maruping, 2021; Dwivedi et al., 2023). This has led to the emergence of many different types of decision-making processes (i.e., the process of selecting actions from alternatives to solve problems (Lunenburg, 2010)) involving AI systems. For instance, managers can delegate the task of deciding to an AI system, which is referred to as *delegated AI decision-making* (DAIDM) process (Baird & Maruping, 2021). In contrast, in a *human-AI ensemble decision-making* (HAEDM) process, a manager and an AI system make the same decision independently. After that, the decisions of the manager and the AI system are aggregated to form a final decision (Choudhary et al., 2023). Both types of decision-making processes have the potential to increase the accuracy of prediction by mitigating the effect of human biases or utilizing higher accuracy of AI systems (Choudhary et al., 2023; Fuegener et al., 2022).

Given the increasing utilization of different types of AI-involved decision-making processes in organizations, it is critical to understand how such processes affect decision recipients. In an employee-manager relationship, recent studies have shown that AI-involved decision-making processes are often perceived as less fair compared to decision-making processes with only a human manager involved (Gonzalez et al., 2022; Köchling et al., 2024). Such perceptions of unfairness can lead to distrust in the manager responsible (Sholihin & Pike, 2009). However, trust is essential, as employees who distrust their manager tend to underperform at work and are more likely to leave the organization (Brower et al., 2009). Thereby, trust is primarily driven by employees' perception of managers' trustworthiness (i.e., perception of managers' ability, benevolence, and integrity) (Colquitt et al., 2007). Existing studies have largely focused on a limited set of AI-involved decision-making processes and their effects on the perception of fairness and its consequences (e.g., Gonzalez et al., 2022; Köchling et al., 2024). However, emerging types of decision-making processes, such as HAEDM and DAIDM, have hardly been given any attention in this relationship so far despite their potential to increase the accuracy and efficiency of decision-making (Choudhary et al., 2023; Fuegener et al., 2022). Due to the potential advantages of these emerging decision-making processes, the critical importance of trust in an employee-manager relationship, and the significant gap in the literature, this study aims to answer the following research questions: (1) How do HAEDM and DAIDM affect decision recipients' perceived fairness of the decision processes? (2) How do these perceptions influence their perception of the managers' trustworthiness?

We conducted a vignette-based online experiment using a mixed within- and between-subjects design, with 79 participants and 237 responses to answer our research questions. Our results show that employees who receive a performance evaluation (PE) that is made automatically by an AI system, following the managers' delegation of the PE task, perceive the decision-making process as less fair compared to when the PE is made exclusively by the manager. Additionally, this lower perceived fairness of the decision-making process (i.e., procedural fairness) leads to a lower employees' perceived trustworthiness of the manager. Notably, our findings show that these adverse effects do not emerge when the decision is made in a manager-AI ensemble, in which both the AI system and the manager are equally involved. Furthermore, we find that the PE outcome (higher vs. lower) does not moderate the negative effect of perceived procedural fairness on the decision recipients' perceived trustworthiness of the manager. Our findings contribute to a better understanding of HAEDM and DAIDM and their effects on decision recipients, offering important implications for managers who seek to integrate AI systems into decision-making processes.

2 Theoretical Background

2.1 Types of AI-Involved Decision-Making Processes

Unlike traditional rule-based decision support systems, AI systems are characterized by increasing autonomy, learning capabilities, and opacity, which makes their decisions less transparent (Berente et al., 2021). However, those characteristics enable AI systems to be applied to increasingly complex decision-making problems (Berente et al., 2021). In practice, AI systems take on many different roles in decision-making processes, with varying degrees of responsibility and human involvement (Baird & Maruping, 2021). For instance, banks are using AI systems to automatically grant loans to customers (Strich et al., 2021), and AI-enhanced recruiting software provides human recruiters with candidate ratings that they can consider when making their own evaluations (Workable, 2024). These examples refer to automated AI decision-making or AI-assisted decision-making. Automated AI decision-making refers to a decision-making process in which an AI system operates without human involvement, acting autonomously, and is responsible for carrying out the act of making the decision (Strich et al., 2021). In contrast, in AI-assisted decision-making, an AI system solves a specific decision task and assists a human decision-maker with its results, who is still responsible for making the decision. While most of the existing literature has focused primarily on these types of AI-involved decision-making (e.g., Köchling et al., 2024; Strich et al., 2021), AI systems can take on other roles in decision-making processes that should be distinguished. In this study, we focus on two of these types of AI-involved decision-making, i.e., – DAIDM and HAEDM. DAIDM refers to decision-making in which a human delegates a decision-making task to an AI system. In doing so, the human transfers the responsibility for making decisions to an AI system, which acts autonomously after the task has been delegated (Baird & Maruping, 2021). DAIDM can be particularly efficient when delegating tasks in which humans have a high degree of uncertainty (Fuegener et al., 2022) or when the AI system has an overall higher performance than the humans involved. In contrast, HAEDM refers to a decision-making process in which a human and an AI system solve the same decision-making task independently. Their individual outcomes are then aggregated (either averaged or with different weights) to form a final decision (Choudhary et al., 2023). As a result, both the human and AI system are responsible for making the decision. This approach has the advantage of counteracting individual biases by combining human and AI system judgments, thereby reducing their respective limitations. HAEDM can be particularly effective in tasks such as project evaluations, in which outcomes are uncertain and not fully predictable (Choudhary et al., 2023). However, while AI-involved decision-making can be more effective than human decision-making in certain tasks (Fuegener et al., 2022), involving AI systems in decision-making can also introduce important challenges. Involving AI systems in decision-making processes can threaten the role identity of the humans involved (Strich et al., 2021) or lead to an overreliance on AI-generated outcomes (Keding & Meissner, 2021). Furthermore, involving AI systems in decision-making processes can lead to negative reactions from humans affected by these decisions, as AI-involved decision-making processes are often perceived as less fair than processes in which decisions are made solely by humans (HUDM) (Gonzalez et al., 2022; Köchling et al., 2024).

2.2 Fairness and AI Systems in Decision-Making

According to the fairness heuristic theory, individuals use fairness judgments as a heuristic that guides their perceptions and behavior (Lind, 2001). For instance, when individuals perceive that they have been fairly treated by an authority, they tend to respond positively toward the authority (Lind, 2001). In contrast, and applied to the context of decision-making, when an individual perceives a decision-making process or outcome made by an authority as unfair, they tend to reject the position of the authority (Lind et al., 1993), decreasing individual trust in the authority (Lind, 2001; Sholihin & Pike, 2009) and increasing their turnover intention (Köchling et al., 2024). Thereby, individuals' fairness judgment can be divided into three dimensions, namely distributive fairness, procedural fairness, and interactional fairness judgments (Lind, 2001). Distributive fairness refers to the individuals' perception of how fair a decision outcome is (Adams, 1965). Procedural fairness refers to individuals' perception of how fair the procedures of the decision-making process were (Lind & Tyler, 1988; Van Den Bos et al., 1996), and interactional fairness refers to individuals' perceptions of how fairly they were treated during the enactment of the decision-making process, including aspects such as respect and honesty (Bies & Shapiro, 1987; Colquitt, 2001). While the outcome of the decision-making process is a fundamental factor in judgments of distributive fairness, the outcome of the decision-making process can also be an essential factor in individuals' perceptions of procedural fairness. The outcome of a decision-making process can interact with procedural fairness (Lind, 2001). For instance, Brockner et al. (1994) demonstrated that lower perceptions of procedural fairness negatively affect organizational trust, particularly when outcomes are negative. Furthermore, Brockner et al. (2007) showed that perceived unfairness of a PE process can evoke stronger feelings of anger, especially when the outcome of the evaluation is unfavorable.

Prior research on decision-making has established individuals' perception of fairness as an important factor in explaining individual responses to decisions made by or in cooperation with AI systems (Gonzalez et al., 2022; Köchling et al., 2024). For instance, Köchling et al. (2024) found that employees who received a PE from an AI system or an AI-assisted decision-making process involving a human resource (HR) manager and an AI system perceived the decision as less fair than when the decision was made by the HR manager alone, leading to a higher employees' turnover intention. Similarly, Gonzalez et al. (2022) showed that receiving a hiring decision from an AI system rather than from a human manager can lead to lower fairness perceptions and reduced willingness to accept a job offer, especially when the individuals are unfamiliar with the AI system. However, as AI systems' capabilities advance, the landscape of AI-involved decision-making continues to evolve, introducing new types such as HAEDM. These emerging decision-making processes require a more nuanced understanding of how different types of AI-involved decision-making affect individuals' fairness perceptions and what consequences they entail.

3 Hypothesis Development

Fairness heuristic theory proposes that individuals rely on their fairness judgments as a heuristic that affects their perceptions and behaviors (Lind, 2001). When assessing a PE process, procedural fairness judgments have been proven to be particularly relevant (Köchling et al., 2024; Sholihin & Pike, 2009), which are formed by the individuals' perception of the fairness of the PE process (Lind et al., 1993). Recent research has demonstrated that the involvement of AI systems in decision-making can significantly impact individuals' perceived fairness of the decision-making process (Gonzalez et al., 2022; Köchling et al., 2024). For instance, Köchling et al. (2024) found that PEs conducted by AI systems or through human-AI augmentation were perceived as less fair compared to evaluations conducted by human managers (Köchling et al., 2024). Notably, in the context of PE, employees seem to prefer decisions made solely by humans, as they prefer intuitive and subjective decision-making rather than the objectivity of algorithms (Lee, 2018). Furthermore, while AI systems can make, in some cases, more accurate decisions than humans (Fuegener et al., 2022), they are also often inscrutable and known for having biases (Berente et al., 2021; Choudhary et al., 2023). In DAIDM and HUAEDM, AI systems are either solely or partly responsible in making the final decision. Therefore, we argue that when an individual receives a PE resulting from DAIDM or HAEDM, it will be perceived as less fair than receiving a PE solely from a human manager. Accordingly, we derive our first hypotheses:

H1: *Receiving a PE resulting from (a) DAIDM or (b) HAEDM will lead to a lower decision recipient's perceived procedural fairness than receiving a PE made exclusively by a human manager.*

Furthermore, according to the fairness heuristic theory, when individuals perceive a decision-making process as unfair, they tend to react adversely toward the authority involved (Lind et al., 1993). Specifically, the perception of unfairness can evoke mistrust (Lind, 2001). In the context of PE, this mistrust is often directed toward the associated manager. Sholihin and Pike (2009) demonstrated that perceiving a PE process as unfair can lead to lower trust in the manager making the decision. Such mistrust reflects a negative perception of the managers' trustworthiness, defined as a perceived lack of ability, benevolence, and integrity (Colquitt et al., 2007). Therefore, and in line with

H1, we assume that lower perceived procedural fairness of a DAIDM or HAEDM process leads to a lower perception of the managers' trustworthiness involved. Accordingly, we derive our second hypothesis.

H2: *Receiving a PE resulting from (a) DAIDM or (b) HAEDM compared to a PE exclusively made by a human manager will lead to lower perceived trustworthiness of the human manager due to lower perceived procedural fairness.*

Moreover, according to the fairness heuristic theory, the effect of individuals' perception of procedural fairness on individuals' reactions can interact with the outcome of the decision-making process (Lindt, 2001). For instance, Brockner et al. (1994) demonstrated that individuals reacted less favorably to perceived procedural fairness when the process outcome was perceived as negative. Furthermore, according to Brockner et al. (2007), negative reactions from decision recipients' caused by their procedural fairness judgment when receiving a performance evaluation can be mitigated when the evaluation outcome is favorable. Accordingly, we derive our third hypothesis:

H3: *The effect of receiving a PE resulting from (a) DAIDM or (b) HAEDM compared to a PE made exclusively by a human manager on the perceived trustworthiness of the manager will be moderated by the outcome of the PE in that a higher PE will mitigate the negative effect of individuals perceived procedural fairness.*

4 Method

4.1 Experimental Design

To investigate how different types of AI-involved decision-making affect decision recipients, we conducted an online vignette-based experiment using a 3 (Type of Decision-Making: HUDM, DAIDM, HAEDM) x 2 (PE outcome: Higher vs. Lower) mixed within- and between-subjects design. For assessing the Type of Decision-Making condition, we chose a within-subjects design approach to control for individual differences. Following established standards for vignette-based experiments (Aguinis & Bradley, 2014), we structured our online experiment as follows: (1) First, we welcomed the participants to the experiment, introduced them to the data policies, and informed them that they would participate in an experiment on PE processes. (2) Next, we used a questionnaire to collect data on our control variables, such as participants' AI literacy and trust in AI. (3) In the third step, we informed all participants that they would encounter a series of scenarios about different PE processes and that they would need to immerse themselves in these scenarios. (4) In the next step, they were asked to put themselves in the shoes of a fictitious person named Alex, who was introduced as a project manager working at a mid-sized company for five years. They also read that Alex would receive an annual employee bonus linked to their PE. (5) In the fifth step, each participant encountered the three different decision-making scenarios. In these scenarios, Alex's PE was determined by one of the following methods: solely by the manager (HUDM), by JcEvalAI, an AI system, after the manager delegated the PE task to it (DAIDM), or by an ensemble decision of the manager and JcEvalAI (HAEDM). Additionally, participants in the higher (vs. lower) PE outcome condition were informed in each scenario that they would receive a bonus of 80% (vs. 20%) of the maximum possible amount based on their PE. All scenarios are listed in Table A.1 in the appendix. To increase the scenarios' comprehensiveness, all participants saw an image for each scenario, visualizing the PE process described. Each time the participants were presented with one of the three decision-making scenarios, they had to answer questions to measure our mediator and dependent variable. To avoid unintended effects due to the positioning of the scenarios in the experiment, we randomized the order of the three scenarios for each participant. At the end of the experiment, all participants completed a post-experimental questionnaire to capture their demographics. To ensure that our PE outcome manipulation worked as intended, we conducted a manipulation check using three items (e.g., "The bonus payment I received was high", 7-point Likert scale) to measure if participants perceived the bonus determined in the higher PE outcome condition substantially higher than in the lower condition. The results of an ANOVA showed that participants in the higher condition perceived the bonus significantly higher than in the lower condition ($M = 5.61$ vs. 2.82; $F = 69.18$, $p < 0.001$). To check the within-subject manipulation, we asked participants after each decision manipulation who made the decision ("Both the manager and the AI, averaged recommendations"/ "Only the manager, no AI involvement"/ "Manager delegated to AI"). Several binomial tests confirmed most participants correctly identified the related decision-making process ($p < 0.001$).

4.2 Measured Variables

We measured our mediator and dependent variable for each decision-making scenario separately. We adapted three items from Skarlicki et al. (1998) to measure our mediator variable, *Procedural Fairness*. For the measurement of our dependent variable, *Trustworthiness of the Manager*, we used five items adapted from Mayer & Davis (1999), covering all three facets of trustworthiness (ability, benevolence, and integrity of the trustee) (Colquitt et al., 2007; Mayer & Davis, 1999). In addition, we measured *AI Literacy* with three items from Pinski & Benlian (2023), *Trust in AI* with one item adapted from Gonzalez et al. (2022), and demographics such as age and gender as control variables. We controlled for AI literacy, assuming that participants with higher AI literacy may be more aware of the capabilities of AI systems in decision-making, which may affect their general attitude towards AI systems in decision-making processes. We controlled for *Trust in AI*, assuming that participants with higher trust in AI may perceive a PE made by an AI system as fairer compared to participants who do not trust AI to make important decisions at work.

We ensured our measures' internal consistency, discriminant validity, and convergent validity by assessing item loadings, average variance extracted (AVE), and Cronbach's alpha. Cronbach's alpha (0.90-0.97) exceeded the threshold of 0.70, indicating the internal consistency of our measures (Nunnally, 1978). In addition, the item loadings exceeded the 0.70 threshold (0.74-0.98), and the AVEs (0.76-0.92) exceeded the 0.50 threshold for our measures, indicating convergent validity (Hair, 2019). Furthermore, the square roots of the AVEs of our measures exceeded the inter-construct correlations, indicating the discriminant validity of our measures (Fornell & Larcker, 1981). Table A.2 in the appendix lists more details about our measured variables.

4.3 Data Collection and Sample Description

We recruited participants for our online experiment through our personal environment and Prolific.com. Prolific.com is a crowdsourcing platform specialized for research purposes that is widely used for IS research (e.g., Bauer & Gill, 2024; Schmidt et al., 2025) and is known for providing high-quality data (Palan & Schitter, 2018). We recruited a total of 111 participants who completed our online experiment. To ensure high-quality data for our analysis, we excluded all participants who failed at least two of our three attention checks (32 participants) based on the embedding of nonsensical items (e.g., "I swim across the Atlantic Ocean to get to work every day," 7-point Likert scale) (Paolacci et al., 2010) leading to a final data set of 79 participants and 237 responses for our analysis. Table 1 provides the descriptive statistics of our dataset. We conducted several analyses of variance and chi-squared tests to ensure that our participants were successfully randomly assigned across our two conditions. The results demonstrate that there was no significant difference between *AI Literacy* (F = 0.13, p > 0.05), *Trust in AI* (F = 0.30, p > 0.05), *Age* ($\chi 2$ = 5.56, p > 0.05), and *Gender* ($\chi 2$ = 0.00, p > 0.05), indicating that our conditions were successfully randomized.

Table 1. Descriptive Statistics

Variable	Type	Total (N = 79)	Lower PE Outcome (N = 39)	Higher PE Outcome (N = 40)
		Mean (SD)	Mean (SD)	Mean (SD)
Procedural Fairness	HUDM	5.17 (1.27)	4.98 (1.30)	5.36 (1.22)
	DAIDM	3.46 (1.99)	3.42 (1.94)	3.50 (2.06)
	HAEDM	5.11 (1.46)	4.85 (1.47)	5.36 (1.43)
Trust in the Manager	HUDM	5.02 (1.39)	4.83 (1.39)	5.21 (1.37)
	DAIDM	3.26 (1.74)	3.44 (1.66)	3.08 (1.82)
	HAEDM	4.78 (1.49)	4.54 (1.42)	5.02 (1.54)
AI Literacy		4.23 (1.44)	4.29 (1.38)	4.18 (1.51)
Trust in AI		3.68 (1.77)	3.79 (1.58)	3.58 (1.95)
Age[1]		2.37 (1.25)	2.66 (1.43)	2.08 (0.96)
Gender[2]		0.49 (0.50)	0.49 (0.51)	0.50 (0.51)

Note: [1] 1 = "18–24"; 2 = "25–34"; 3 = "35–44"; 4 = "45–54"; 5 = "55–64", 6 = 65+";
[2] 1= "Female"; 0 = "Male"/ "Other"

5 Results

To test our hypotheses, we conducted a series of linear mixed-effect regressions on our mediator variable, *Procedural Fairness*, and our dependent variable, *Trustworthiness of the Manager*, resulting in 4 models. Our moderator variable, *Performance Evaluation,* was binary coded with one equal to higher PE outcome and zero to lower PE outcome. In all models, we included our control variables. Table 2 displays the results of the analyses.

Table 1. Results of Linear Mixed-Effect Regressions

	Procedural Fairness	Trustworthiness of the Manager		
	Model 1	Model 2	Model 3	Model 4
Intercept	3.52 (0.53)***	3.72 (0.59)***	1.35 (0.45)**	1.27 (0.51)*
DAIDM	-1.71 (0.23)***	-1.76 (0.21)***	-0.61 (0.17)***	-
HAEDM	-0.07 (0.22)	-0.24 (0.21)	-0.20 (0.15)	
Procedural Fairness	-	-	0.67 (0.05)***	0.70 (0.06)***
PE Outcome	-	-	-	-0.61 (0.40)
Procedural Fairness x PE Outcome	-	-	-	0.11 (0.08)
Age	-0.02 (0.09)	-0.04 (0.10)	-0.03 (0.07)	-0.04 (0.07)
Gender	0.27 (0.21)	0.22 (0.24)	0.04 (0.17)	0.00 (0.17)
Trust in AI	0.33 (0.07)***	0.26 (0.08)***	0.04 (0.06)	0.01 (0.05)
AI Literacy	0.09 (0.90)	0.09 (0.10)	0.03 (0.07)	0.01 (0.07)
Marginal R^2	0.32	0.29	0.63	0.62
Conditional R^2	0.37	0.45	0.71	0.69

Note: beta coefficients; () = standard error; *p < 0.05; **p < 0.01; ***p < 0.001.

Our results of Model 1 show only a significant effect of *DAIDM* (β = -1.71, p < 0.001, CI = [-2.16, -1.27]) on *Procedural Fairness* but not for *HAEDM* (β = -0.07, p > 0.05, CI = [-0.51, 0.38]), thus providing **support for H1a but not for H1b**. Receiving a PE resulting from a DAIDM process can lead to a lower decision recipient's perceived procedural fairness than receiving a PE made exclusively by a human manager. However, receiving a PE resulting from a HAEDM process did **not** lead to a significantly lower decision recipient's perceived procedural fairness in comparison to receiving a PE made exclusively by a human manager. Furthermore, the insignificant effect of *HAEDM* on *Procedural Fairness* (β = -0.07, p > 0.05, CI = [-0.51, 0.38]) demonstrates **no support for H2b and H3b**. However, Model 3 demonstrates that *Procedural Fairness* partially mediates the effect of *DAIDM* on *Trustworthiness of the Manager* (β = 0.67, p < 0.001, CI = [0.58, 0.76]), indicating **support for H2a**. In addition, our results of Model 4 demonstrate that the negative effect of *Procedural Fairness* on *Trustworthiness of the Manager* is not moderated by the *PE Outcome* (β = 0.11, p > 0.05, CI = [-0.04, 0.26]), providing **no support for H3**. The PE outcome does not significantly influence the negative effect of perceived procedural fairness on the perceived trustworthiness of the manager when receiving a decision made via DAIDM or HAEDM.

To strengthen our mediation analysis, we conducted several non-parametric bootstrap analyses based on ordinary least squares regressions of Models 1 and 3 with 5,000 bootstrap samples and 95% confidence intervals. We ran the bootstrap analyses for our complete data set as well as for one subset, including participants in the lower PE outcome condition and one in the higher PE outcome condition. The results of the bootstrap analyses show that *Perceived Fairness* mediates the effect of *DAIDM* on *Trustworthiness of the Manager* in all three analyses, providing **support for H2a**. We found a significant indirect effect for both conditions (indirect effect = -1.19, CI = [-1.55, -0.84]), for the lower PE outcome condition (indirect effect = -1.03, CI = [-1.59, -0.54]), and for the higher PE outcome condition (indirect effect = -1.35, CI = [-1.95, -0.82]).

6 Discussion

6.1 Theoretical Contributions and Practical Implications

As AI systems become more relevant in organizational decision-making, this study explores how DAIDM and HAEDM PE processes affect decision recipients' perceived procedural fairness and the trustworthiness of the human manager. While recent studies have shown that automated AI and AI-assisted decision-making are often perceived as less fair compared to decision-making processes with only a human manager involved (e.g., Gonzalez et al., 2022; Köchling et al., 2024), DAIDM and HAEDM have hardly been given any attention in this relationship so far, despite their potential to increase the accuracy and efficiency of decision-making (Choudhary et al., 2023; Fuegener et al., 2022). Therefore, we contribute to the literature about AI-involved decision-making by demonstrating that receiving a PE resulting from DAIDM is perceived as less fair than receiving a PE made exclusively by a human manager (i.e., HUDM). Furthermore, we show that this effect does not emerge when receiving the PE resulting from HAEDM. Humans often perceive decisions made by AI systems as less fair than those made by humans, largely due to the opacity of AI systems, which makes it difficult to obtain detailed information about the decision-making process (Langer & Landers, 2021). Moreover, while AI-based decisions can be more accurate than human decisions (Fuegener et al., 2022), individuals may feel dehumanized when evaluated by an algorithm (Lee, 2018). The HAEDM process leverages the benefits of AI decision-making, such as potentially increased accuracy, while mitigating the feeling of dehumanization and ensuring a human point of contact for follow-up inquiries. Overall, our findings provide evidence that different types of AI-involved decision-making need to be investigated in a nuanced manner.

Furthermore, our study extends our knowledge of the unintended consequences of receiving a PE based on different types of AI-involved decision-making, which goes beyond previous findings in the literature. While previous studies demonstrated that the perception of unfairness of a decision-making process can lead to distrust of the manager, those studies focused solely on the relationship between employee and manager without considering the involvement of AI systems in the decision-making process. Just a few studies demonstrated that decisions made by algorithms can be perceived as less trustworthy (Lee, 2018); however, we know little about how employees perceive the trustworthiness of the manager when receiving a PE resulting from AI-involved decision-making processes. Our findings address this research gap by showing that receiving a PE resulting from DAIDM compared to a PE solely made by a human manager can lead to lower perceived trustworthiness of the manager due to a lower perceived procedural fairness. Moreover, we demonstrate that this effect is not moderated by the level of PE outcome, indicating that delegating a task to an AI system can decrease the perception of the managers' trustworthiness due to perceptions of unfairness, even if the employee perceives a high PE. However, while some studies argue that the outcome of a decision can affect how fair the recipients perceive it to be (e.g., Brockner et al., 2007), Lind (2001) argues that this effect is particularly pronounced when the outcome is unexpected. This could explain the limited role of PE outcome level in our study since we deliberately avoided generating outcome expectancies in our vignette study.

Finally, our findings show that managers planning to involve AI systems in decision-making processes should be aware that employees may perceive a decision as unfair when the task of making the decision is delegated to an AI system, even though AI systems can often make more objective decisions than humans. In the context of PE, we recommend that these managers build PE processes on HAEDM instead of DAIDM to minimize the perception of unfairness in the PE process and its potential negative effect on employees' perception of managers' trustworthiness. However, when managers prefer to rely on DAIDM to utilize its potential efficiency, they should implement AI systems that follow the concept of explainable AI because these systems are more transparent in their decision-making. Managers should provide employees with as much information as possible about the logic behind the decision-making processes of the AI systems involved to increase procedural transparency. Furthermore, managers should take deliberate measures to mitigate potential feelings of dehumanization, such as establishing channels for employee feedback and explicitly positioning the AI system as a decision-support tool rather than a fully autonomous system. These measures may mitigate the negative effects of DAIDM on employees' perceived fairness of the PE process, thus enabling managers to utilize DAIDM with minimized consequences.

6.2 Limitations and Directions for Future Research

There are several limitations to our study that may provide interesting directions for future research. First, our study focuses on two specific types of AI-involved decision-making processes in a specific context – DAIDM and HAEDM in the context of PE. However, while other types of decision-making processes, such as automated AI decision-making, have already been extensively considered in the literature, exploring these types in comparison to DAIDM and HAEDM may lead to a more nuanced view of the effects of different types of decision-making processes on decision recipients' perceptions. Second, our study is based on a vignette-based online experiment and a one-time interaction. Vignette-based online experiments are already a well-established method to investigate how different types of decision-making processes affect the perceptions of the decision recipient' (Brockner et al., 2007; Gonzalez et al., 2022), and we believe that our method is an appropriate method to begin to investigate the effects of involving AI systems in the decision-making process. However, vignette-based experiments lack real-world interaction over a longer period, which can limit external validity (Eifler & Petzold, 2019). Future research could help to address this limitation and strengthen the robustness of our findings by conducting related studies based on experiments with multiple real-world interactions. Despite these limitations, our study provides important theoretical and practical implications and helps to understand how involving AI systems in decision-making processes affects the perceptions of the decision recipients.

References

Adams, J. S. (1965), 'Inequity In Social Exchange', Advances in Experimental Social Psychology (Vol. 2, pp. 267–299). Academic Press.

Aguinis, H., & Bradley, K. J. (2014), 'Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies', Organizational Research Methods, 17(4), 351–371.

Anthropic. (2024), 'Introducing the next generation of Claude', https://www.anthropic.com/news/claude-3-family

Baird, A., & Maruping, L. M. (2021), 'The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts', MIS Quarterly, 45(1), 315–341.

Bauer, K., & Gill, A. (2024), Mirror, Mirror on the Wall: Algorithmic Assessments, Transparency, and Self-Fulfilling Prophecies. Information Systems Research, 35(1), 226–248.

Berente, N., Bin Gu, Recker, J., & Santhanam, R. (2021), 'Managing Artificial Intelligence', MIS Quarterly, 45(3), 1433–1450.

Bies, R. J., & Shapiro, D. L. (1987), 'Interactional fairness judgments: The influence of causal accounts', Social Justice Research, 1(2), 199–218.

Brockner, J., Fishman, A. Y., Reb, J., Goldman, B., Spiegel, S., & Garden, C. (2007), 'Procedural fairness, outcome favorability, and judgments of an authority's responsibility', Journal of Applied Psychology, 92(6), 1657–1671.

Brockner, J., Konovsky, M., Cooper-Schneider, R., Folger, R., Martin, C., & Bies, R. J. (1994), 'Interactive Effects of Procedural Justice and Outcome Negativity on Victims and Survivors of Job Loss', Academy of Management Journal, 37(2), 397–409.

Brower, H. H., Lester, S. W., Korsgaard, M. A., & Dineen, B. R. (2009), 'A Closer Look at Trust Between Managers and Subordinates: Understanding the Effects of Both Trusting and Being Trusted on Subordinate Outcomes', Journal of Management, 35(2), 327–347.

Choudhary, V., Marchetti, A., Shrestha, Y. R., & Puranam, P. (2023), 'Human-AI Ensembles: When Can They Work?', Journal of Management, 01492063231194968.

Colquitt, J. A. (2001), 'On the dimensionality of organizational justice: A construct validation of a measure', Journal of Applied Psychology, 86(3), 386–400.

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007), 'Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance', Journal of Applied Psychology, 92(4), 909–927.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023), '"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy', International Journal of Information Management, 71, 102642.

Eifler, S., & Petzold, K. (2019), 'Validity Aspects of Vignette Experiments: Expected "What-If" Differences Between Reports of Behavioral Intentions and Actual Behavior', Experimental Methods in Survey Research (pp. 393–416), John Wiley & Sons, Ltd.

Fornell, C., & Larcker, D. F. (1981), 'Evaluating Structural Equation Models with Unobservable Variables and Measurement Error', Journal of Marketing Research, 18(1), 39–50. JSTOR.

Fuegener, A., Grahl, J., Gupta, A., & Ketter, W. (2022), 'Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation', Information Systems Research, 33(2), 678–696.

Gonzalez, M. F., Liu, W., Shirase, L., Tomczak, D. L., Lobbe, C. E., Justenhoven, R., & Martin, N. R. (2022), 'Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes', Computers in Human Behavior, 130, 107179.

Hair, J. F. (2019), 'Multivariate data analysis', Cengage.

Keding, C., & Meissner, P. (2021), 'Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions', Technological Forecasting and Social Change, 171, 120970.

Köchling, A., Wehner, M. C., & Ruhle, S. A. (2024), 'This (AI)n't fair? Employee reactions to artificial intelligence (AI) in career development systems', Review of Managerial Science.

Langer, M., & Landers, R. N. (2021). 'The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers', Computers in Human Behavior, 123, 106878.

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

Lind, E. A. (2001), 'Fairness heuristic theory: Justice judgments as pivotal cognitions in organizational relations', In Advances in organization justice (pp. 56–88). Stanford University Press.

Lind, E. A., Kulik, C. T., Ambrose, M., & de Vera Park, M. V. (1993), 'Individual and corporate dispute resolution: Using procedural fairness as a decision heuristic', Administrative Science Quarterly, 224–251.

Lind, E. A., & Tyler, T. R. (1988), 'The Social Psychology of Procedural Justice', Springer Science & Business Media.

Lunenburg, F. C. (2010), 'The decision making process', National Forum of Educational Administration & Supervision Journal, 27(4).

Macorva. (2025), 'Why Choose Macorva? Discover the Power of AI-driven Experience Management', https://www.macorva.com/why-macorva

Mayer, R. C., & Davis, J. H. (1999), 'The effect of the performance appraisal system on trust for management: A field quasi-experiment', Journal of Applied Psychology, 84(1), 123–136.

Nunnally, J. C. (1978), 'An Overview of Psychological Measurement', Clinical Diagnosis of Mental Disorders (pp. 97–146). Springer US.

Palan, S., & Schitter, C. (2018), 'Prolific.ac—A subject pool for online experiments', Journal of Behavioral and Experimental Finance, 17, 22–27.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010), 'Running experiments on Amazon Mechanical Turk', Judgment and Decision Making, 5(5), 411–419.

Pinski, M., & Benlian, A. (2023), 'AI Literacy—Towards Measuring Human Competency in Artificial Intelligence', Hawaii International Conference on System Sciences.

Schmidt, J.-H., Bartsch, S. C., Adam, M., & Benlian, A. (2025), 'Elevating Developers' Accountability Awareness in AI Systems Development', Business & Information Systems Engineering, 67(1), 109–135.

Sholihin, M., & Pike, R. (2009), 'Fairness in performance evaluation and its behavioural consequences', Accounting and Business Research, 39(4), 397–413.

Skarlicki, D. P., Ellard, J. H., & Kelln, B. R. C. (1998), 'Third-party perceptions of a layoff: Procedural, derogation, and retributive aspects of justice', Journal of Applied Psychology, 83(1), 119–127.

Strich, F., Mayer, A.-S., & Fiedler, M. (2021), 'What Do I Do in a World of Artificial Intelligence? Investigating the Impact of Substitutive Decision-Making AI Systems on Employees' Professional Role Identity', Journal of the Association for Information Systems, 22(2), 304–324.

Van Den Bos, K., Vermunt, R., & Wilke, H. a. M. (1996), ‘The consistency rule and the voice effect: The influence of expectations on procedural fairness judgements and performance’, European Journal of Social Psychology, 26(3), 411–428.

Workable. (2024), ‘Recruiting and HR features powered by AI – Workable’, https://www.workable.com/workable-ai

Appendix A

Table A.1. Text-Based Scenarios

Human-AI Ensemble Decision-Making (HAEDM) Your performance evaluation was conducted by a combination of your manager, who has 10 years of experience, and an artificial intelligence (AI) based algorithm named JcEvalAI with access to 10 years of historical performance data. Both the manager and JcEvalAI independently provided a performance evaluation, which were then averaged to determine your final performance evaluation.	**Delegated AI Decision-Making (DAIDM)** Your performance evaluation was conducted by a combination of your manager, who has 10 years of experience, and an artificial intelligence (AI) based algorithm named JcEvalAI system with access to 10 years of historical performance data. Your manager delegates your performance evaluation to JcEvalAI. JcEvalAI then provided a performance evaluation, which is your final performance evaluation.
Human Only Decision-Making (HUDM) Your performance evaluation was conducted entirely by your manager, who has 10 years of experience. Your manager provided a performance evaluation, which is your final performance evaluation.	**Additional Information Presented at the End of Each Scenario** As a result of your performance evaluation, you have been awarded a bonus equal to 20%/80% of the maximum possible amount. Note: JcEvalAI and the manager used the same dataset for your performance evaluation.

Table A.2. Measured Variables

Constructs	Items
Procedural Fairness (Skarlicki et al., 1998)	Generally, the procedures used for the performance evaluation were fair. The way the performance was evaluated was fair. The procedures used for the performance evaluation were acceptable.
Trustworthiness of the Manager (Mayer & Davis, 1999)	The manager is well-qualified to conduct performance evaluations. The manager is very concerned about my welfare. The manager really looks out for what is important to me. The manager tries hard to be fair in dealings with employees. Sound principles seem to guide the manager's behavior.
AI Literacy (Pinski & Benlian, 2023)	In general, I know the unique facets of artificial intelligence (AI) and humans and their potential roles in human-AI collaboration. I am knowledgeable about the steps involved in AI decision-making. Considering all my experience, I am relatively proficient in the field of AI.
Trust in AI (Gonzalez et al., 2022)	In general, do you trust artificial intelligence to make important work decisions (e.g., hiring, performance evaluation)? (Not at all (1) to Completely (7))

All items used 7-point Likert scales (1 = Strongly disagree, 7 = Strongly agree) unless noted.