



The Human Touch in Artificial Intelligence (AI) Decision-Making:

Investigating Employee Fairness Perceptions of Human-AI Collaboration in the Workplace

Anna Gescher

Dissertation written under the supervision of Professor Cristina Mendonça

Dissertation submitted in partial fulfilment of requirements for the International MSc in Management with a specialization in Strategy and Consulting, at the Universidade Católica

Portuguesa, July 25th, 2023.

Abstract

Title: The Human Touch in Artificial Intelligence (AI) Decision-Making: Investigating Employee Fairness Perceptions of Human-AI Collaboration in the Workplace

Author: Anna Pauline Gescher

The future of work is seen as a collaboration between humans and machines, but the use of AI for management decision-making raises concerns about employees' fairness perceptions. An experimental study investigated 797 employees' perceptions of procedural, interpersonal, and informational fairness towards three proposed hybrid decision-making approaches. The results are compared to perceptions of fully automated decision-making in three common management contexts. The study shows that not all hybrid decision-making approaches are generally perceived as fairer than fully automated decision-making. The aggregated human-AI decision-making approach, in which AI and humans collaborate by performing specific evaluation tasks based on their respective strengths, is perceived to be the fairest among employees. Fairness perceptions remained consistent across the decision contexts of the allocation of promotions, bonus payments, and career training, suggesting that the perceived fairness of hybrid approaches may not be context dependent. The qualitative responses shed light on employee satisfaction and concerns related to both AI and human involvement in hybrid decision-making. This thesis discusses the implications of these findings and future research directions. It contributes to understanding fairness perceptions in the context of human-machine collaboration and informs decision-makers on the most effective approach for implementing AI in the workplace.

Keywords: Artificial Intelligence, Hybrid Decision-Making, Algorithmic Fairness, Organizational Justice Theory, Employee Perception

Resumo

Título: O Toque Humano na Tomada de Decisão por Inteligência Artificial (IA): Investigando a Percepção de Justiça dos Trabalhadores sobre a Colaboração Humano-IA no Ambiente de Trabalho

Autor: Anna Pauline Gescher

O futuro do trabalho é visto como uma colaboração entre humanos e máquinas, mas o uso da Inteligência Artificial (IA) para a tomada de decisões de gestão levanta preocupações sobre as percepções de justiça dos trabalhadores. Um estudo experimental investiga as percepções de justiça processual, interpessoal e informacional de 797 trabalhadores face a três abordagens de tomada de decisão híbrida. Uma comparação é feita entre as percepções da tomada de decisão totalmente automatizada em três contextos de gestão comuns. O estudo mostra que nem todas as abordagens de tomada de decisão híbrida são geralmente percebidas como mais justas do que a tomada de decisão totalmente automatizada por IA. A abordagem de tomada de decisão agregada humano-IA, na qual IA e humanos colaboram realizando tarefas de avaliação específicas com base nas suas respectivas habilidades, é percebida como a mais justa pelos trabalhadores. As percepções de justiça permaneceram consistentes nos contextos de decisão referentes à alocação de promoções, pagamentos de bônus e treino profissional, sugerindo que a percepção de justiça das abordagens híbridas pode não depender do contexto. As respostas qualitativas lançaram luz sobre a satisfação e preocupações dos trabalhadores relacionadas tanto com o envolvimento da IA quanto ao dos humanos na tomada de decisão híbrida. Esta tese discute implicações dessas descobertas, bem como futuras direções de investigação. Ela contribui para a compreensão das percepções de justiça no contexto da colaboração entre humanos e máquinas e informa os decisores sobre a abordagem mais eficaz para implementar IA no local de trabalho.

Palavras-chave: Inteligência Artificial, Tomada de Decisão Híbrida, Justiça Algorítmica, Teoria da Justiça Organizacional, Percepção do Trabalhadores

Acknowledgments

This dissertation represents the completion of my International Master of Science degree and marks the end of an incredible yet challenging journey of continuous learning and development. I would like to acknowledge that this accomplishment would not have been possible without the unwavering support and guidance of those who have stood by me throughout this process.

I would like to thank my dedicated supervisor, Cristina Mendonça, for her support throughout this project. Her expertise, ideas, and experience have been invaluable in guiding me through this journey. I am grateful for her willingness to always make time to address my questions, clarify ambiguities, and provide valuable feedback despite her own busy schedule. With your support, my work has reached new heights, and I am truly grateful to have had you as my supervisor.

Besides this, I would like to thank everyone who participated in my online study. Without your support I would not have been able to gain such in depths insights and complete this research as exhaustive.

I would also like to thank my friends and colleagues who supported me and made time available for me whenever I needed a mental break. And finally, I could not have completed this academic journey and especially this dissertation without my loving family and Don. Thank you for standing by my side throughout this entire process of ups and downs; you are amazing.

Table of Contents

- Abstract..... I
- Resumo..... II
- Acknowledgments..... III
- Table of Contents..... IV
- List of Tables..... V
- List of Abbreviations..... VI
- 1. Introduction..... 1
 - 1.1. Opening Thought..... 1
 - 1.2. Relevance of the Topic..... 1
 - 1.3. Problem Statement and Research Objective..... 2
 - 1.4. Structure of the Dissertation..... 3
- 2. Literature Review..... 4
 - 2.1. AI in Decision-Making..... 4
 - 2.1.1. Potential of AI for Organizations..... 4
 - 2.1.2. Drawbacks of AI..... 5
 - 2.1.3. Algorithmic Fairness..... 5
 - 2.2. The Importance of People’s Perception..... 6
 - 2.3. Organizational Justice Theory and the Perception of AI Decision-Making..... 6
 - 2.4. Hybrid Decision-Making and the Role of Human Control..... 9
 - 2.5. Hybrid Decision-Making Approaches..... 10
 - 2.6. Decision-Making Contexts..... 12
- 3. Methodology..... 14
 - 3.1. Research Strategy and Design..... 14
 - 3.2. Participants..... 14
 - 3.3. Procedure..... 15
 - 3.4. Variable Measurement..... 16
 - 3.4.1. Main Variables..... 16
 - 3.4.2. Control Variables..... 17
- 4. Results..... 19
 - 4.1. Data Preparation..... 19
 - 4.2. Scale Reliability..... 19
 - 4.3. Demographic Analysis..... 21

4.4. Participants' General Fairness Perception.....	21
4.4.1. Qualitative Analysis.....	21
4.5. Hypotheses Testing.....	23
4.5.1. Differences in Perception across Decision Types – Overall Fairness.....	23
4.5.2. Differences in Perception across Decision Types – Fairness Dimensions.....	24
4.5.3. Controlling for Covariates – Overall Fairness.....	26
4.5.4. Controlling for Covariates – Fairness Dimension.....	27
4.5.5. Differences in Fairness Perception among Decision Contexts.....	29
5. Discussion.....	32
5.1. Academic and Practical Implications.....	35
5.2. Limitations and Future Research.....	36
5.3. Conclusion.....	37
References.....	38
Appendices.....	47
Appendix A: Qualtrics Survey.....	47
Appendix B: Vignettes per Decision Type and Context.....	59
Appendix C: Alternatives Factor Analysis.....	62
Appendix D: Demographic Analysis.....	63
Appendix E: Correlation among Variables.....	65
Appendix F: Quantitative Analysis of Participants' General Fairness Perception.....	66
Appendix G: Themes and Illustrative Quotes Qualitative Analysis.....	67
Appendix H: ANCOVA Fairness Dimensions – Covariates.....	73
Appendix I: Covariate Interaction with Independent Variables.....	74

List of Tables

Table 1. Average Overall Fairness Perception per Decision Type.....	24
Table 2. ANOVA for Fairness Dimensions.....	25
Table 3. Planned Contrasts for Fairness Perception on Dimensional Level.....	25
Table 4. ANCOVA Overall Fairness.....	26
Table 5. Covariate Adjusted Descriptive Statistics for Fairness Dimensions.....	28
Table 6. Repeated Contrasts.....	29
Table 7. Two-Way Factorial ANOVA for Fairness Dimensions.....	30
Table 8. Pairwise Comparison (Gabriel's Test) for Informational Fairness.....	31

List of Abbreviations

AI	Artificial Intelligence
ML	Machine Learning
OJ	Organizational Justice
HR	Human Resources
HRM	Human Resource Management
Ph.D.	Doctor of Philosophy
H1	Hypothesis 1 (2-3 respectively)
ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
df	Degrees of Freedom
F	F-Statistic
M	Sample Mean
N	Total Number of Cases
p	p-value
SD	Standard Deviation
SE	Standard Error
t	t-statistic

1 Introduction

“It’s not about machines replacing humans, but machines augmenting humans.”

- Robin Bordoli, former CEO of Figure Eight (2016)

1.1 Opening Thought

Artificial Intelligence (AI) is nothing new. In fact, AI has been around for more than six decades (Duan et al., 2019). What is new though are advances in computing power, the increasing availability of data, and new Machine Learning (ML) techniques (Haesevoets et al., 2021). These developments have fundamentally changed what AI is capable of, and, in turn, what it is used for today (Tolan, 2018). The rapid advances in AI technology have led to breakthrough successes in a variety of areas, ranging from improved fraud detection and forecasting in finance (Pallathadka et al., 2021), to more accurate medical diagnoses in healthcare (Shaheen, 2021), to autonomous vehicles and pathfinding in transportation (Medrano-Berumen & Akbaş, 2021). Each year, AI capabilities continue to expand, enabling systems to process larger volumes of data with remarkable speed and generate results with even greater accuracy (Sarker, 2021). According to McKinsey’s Global Survey on AI, AI adoption in 2022 has more than doubled compared to 2017 (Chui et al., 2022). Over 50% of responding companies report using AI in at least one business area (Chui et al., 2022). This year, AI seems to be everywhere – with generative AI such as ChatGPT making headlines – slowly but steadily establishing a relevant position for itself in people’s daily lives.

1.2 Relevance of the Topic

The fast development of AI comes with certain dilemmas, especially when it moves beyond its traditional role as a tool and into uncharted territory – such as the workplace. Traditionally, management decisions were just that: decisions made by managers based on experience, intuition, and an understanding of contextual factors (Agor, 1986). As companies leverage AI’s potential, there is a growing trend toward automating decision-making processes that were previously solely made by humans (Höddinghaus et al., 2021). Integrating AI into decision-making, such as hiring or performance evaluation, raises concerns, especially when it comes to the potential impact on employees’ perceptions and attitudes (Langer & Landers, 2021), which should be considered in any organizational setting (Greenberg, 1990).

Although AI has the potential to outperform humans' cognition with its greater computational information processing capacity (Jarrahi, 2018), most people believe that using AI for important decisions that affect their lives, such as decisions about their job, is unfair (Acikgoz et al., 2020; Binns et al., 2018; Starke et al., 2022). Some reasons for this aversion include the perception that AI lacks human intuition and emotion (Jussupow & Benbasat, 2020) and the potential for AI to inherit biases (Dietvorst et al., 2015). However, employees' attitudes and behaviors in the workplace are significantly influenced by their perceptions of fairness (Lind et al., 1997), which makes it a critical value that organizations should enhance and foster. Considering this, scholars call for a more human-centric approach to AI decision-making.

1.3 Problem Statement and Research Objective

Research suggests that people view AI decisions more positively when humans are also involved in the decision-making process (Kern et al., 2022; Nagtegaal, 2021), often referred to as hybrid decision-making. Moreover, experts agree that AI should not make decisions without human oversight (European Commission, 2019; European Union Parliament, 2020), and there is a growing agreement that the future of work relationships lies in co-creation between humans and machines rather than relying solely on one or the other (Haesevoets et al., 2021). While different approaches to hybrid decision-making have been proposed by the literature (European Commission, 2019; Methnani et al., 2021; Shrestha et al., 2019), uncertainties remain regarding individuals' preferred level of human involvement and the effective integration of human control versus system control in work-related AI-based decision-making (Langer & Landers, 2021; Schwemer, 2019). Additionally, studies indicate that people's fairness perceptions of AI decision-making vary depending on the decision context at hand (Starke et al., 2022). Given the growing use of AI in supporting managerial decisions, it is crucial to determine the optimal approach of human-AI collaboration in these contexts to prevent employee dissatisfaction (Clay-Warner et al., 2005), reduced company loyalty (De Boer et al., 2002) or loss of talented employees (Hom et al., 1984) – all potential consequences of perceiving organizational unfairness.

To fill this gap, this dissertation investigates different hybrid decision-making approaches in different decision contexts to identify the optimal role distribution between AI and humans, ensuring that employees feel fairly treated by their organization. Thus, this research intends to answer the following main research question: “How do different approaches to human oversight in AI decision-making affect employee fairness perceptions?”

This central research question was further divided into four related research questions:

- RQ1: To what extent do employees perceive hybrid decision-making as fair?
- RQ2: How does hybrid decision-making compare with employees' fairness perceptions of fully automated decision-making?
- RQ3: How do different hybrid decision-making approaches influence employees' fairness perceptions?
- RQ4: How does the decision context impact employees' fairness perceptions of different hybrid decision-making approaches?

To answer these questions, an extensive literature review and an experimental study were conducted. The experimental study examined participants' fairness perception regarding three distinct hybrid decision-making approaches proposed by the literature compared to fully AI-driven decision-making in three different managerial decision-making contexts. This research contributes to the existing literature on AI decision-making and algorithmic fairness, deepening the understanding of the role of the human in AI-involved decision-making. It sheds light on the most suitable human-AI collaboration perceived as relatively fairest by participants. Also, the study offers insights into participants' rationales for their fairness perception of the studied decision-making processes, providing practical considerations for companies that plan to implement AI or hybrid decision systems.

1.4 Structure of the Dissertation

This thesis is divided into five chapters. After this introduction, Chapter 2 follows with a literature review discussing 1) AI and its role in decision-making, along with its potential benefits and drawbacks; 2) the significance of employees' perception of AI and existing findings on their fairness perception; 3) the importance of human oversight in decision-making and three distinct approaches of hybrid decision-making; and 4) the relevance of decision contexts for fairness perception. Chapter 3 will outline the research strategy and design, describe the participant selection process and study procedure, and will define the measurement variables. Next, Chapter 4 will present the experimental study's results. Chapter 5 concludes this thesis with the research findings' discussion, related academic and practical implications, the study's limitations, suggestions for future research, and a conclusion.

2 Literature Review

2.1 AI in Decision-Making

In simple terms, AI refers to the use of digital technology to carry out tasks that would require human intelligence (Wamba-Taguimdje et al., 2020). The concept of AI initially emerged in the 1950s, and, since then, research in this field has been making consistent progress (Perifanis & Kitsios, 2023). Today, AI encompasses various theories and techniques for developing machines that can simulate intelligent behavior (Wamba-Taguimdje et al., 2020). Leveraging big data and ML, AI systems can now learn to autonomously perform tasks with human-like capabilities, without explicit human instructions (Zhu et al., 2021). These ML-based systems, trained on extensive datasets, can identify unknown patterns, and make optimal decisions within a specific context or framework (Zhu et al., 2021). These developments have started a debate about the potential replacement of domain experts by AI in the future (Hemmer et al., 2021), including decision-making traditionally carried out by leaders in the workplace.

2.1.1 Potential of AI for Organizations

In the workplace, AI developments have resulted in increasing automation of leadership functions (Höddinghaus et al., 2021). Computers can assign tasks to workers (M. K. Lee, 2018), evaluate work outcomes (Luo et al., 2021; Tong et al., 2021), determine compensation (Höddinghaus et al., 2021), or screen and recruit talent (Kaibel et al., 2019); all based on algorithms. For example, the consumer goods company Unilever uses AI in its hiring process (Marr, 2018). Their ML model filters respondents, invites potential prospects to participate in online games that assess skills, such as concentration under pressure, and top applicants proceed to complete a video interview where they respond to workplace challenges. Only afterwards are applicants forwarded to the HR department. Unilever claims that this approach makes their process more human, giving more applicants a chance (Marr, 2018). AI decision-making is believed to surpass human decision-making in terms of speed, information processing capacity, and cost reduction (Tolan, 2018), and practitioners argue that AI leads to improved decision outcomes (Angerschmid et al., 2022). The underlying idea is that machines “base decisions solely on facts, uninfluenced by human biases, discriminatory tendencies, or emotions” (Tolan, 2018, p. 1). Unilever, for instance, believes that, by removing human

judgment from a portion of the hiring process, they can attract both better and more diverse talent (Marr, 2018).

2.1.2 Drawbacks of AI

Despite its advantages, the downsides of relying solely on AI for decision-making must be acknowledged. There is evidence that algorithms can inherit or even perpetuate human biases in decision-making since ML models are trained on data provided by humans (Tolan, 2018). Besides biased training data, reasons for unfair AI include flawed algorithms and inadequate implementation (Tolan, 2018), or the initial transfer of decision authority from humans to AI, especially in sensitive matters (Starke et al., 2022). As an example, Amazon developed a computer program to automate talent searches in technical fields (Dastin, 2018). The ML-based algorithm assigned applicants scores from one to five stars. However, the system exhibited gender bias (Dastin, 2018). The bias originated from the model being trained with resumes primarily submitted by male applicants over the past decade, reflecting the industry's male dominance (Dastin, 2018). Thus, resumes containing terms like “women” or “female” were penalized by the algorithm (Dastin, 2018). This is one among several incidents that have raised concerns about the fairness of AI systems (i.e., the COMPAS algorithm or the PredPol algorithm). Consequently, fairness has become a core objective in developing AI to counteract such issues and enable the considerate use of AI in decision-making (Köchling & Wehner, 2020).

2.1.3 Algorithmic Fairness

According to Starke and colleagues (2022), algorithmic fairness refers to the principle that “decisions made by AI should not result in unjust, discriminatory, or disparate outcomes” (p. 2). So far, research on fairness in AI and ML covers technical aspects of fairness, including the evaluation of existing AI applications for fairness implications (X. Wang et al., 2022), establishing fairness metrics (Pagano et al., 2023), and developing approaches to correct unfair systems (Kern et al., 2022). An example of correcting unfair systems is fairness through awareness, which includes removing data that may introduce unfair considerations, such as ethnic origin (Tolan, 2018). Additionally, legal prerequisites for responsible AI use have been examined (e.g., Hagendorff, 2020; Sanderson et al., 2022). Yet, there is a difference between technical and legal research on non-discriminatory ML and the behavioral studies that investigate why people perceive AI as fair or unfair (Starke et al., 2022). As humans are ultimately impacted by AI decisions, there is a call for human-centric approaches

to studying algorithmic fairness to ensure the AI system's legitimacy. A body of literature has emerged, examining how individuals assess AI usage empirically.

2.2 The Importance of People's Perception

People's perceptions and attitudes significantly impact technology adoption (Lichtenthaler, 2020). For AI to be embraced in decision-making, acceptance is important from two groups: a) users who rely on AI to automate or augment their decision-making processes, and b) individuals that are directly impacted by these decisions in their lives, work, and well-being (Langer & Landers, 2021). This research focuses on the latter's perception, of employees. Decision-makers can typically challenge or override AI recommendations in favor of their judgment. However, employees have limited agency to escape or influence decisions made by AI or based on its results, unless they remove themselves entirely from the decision-making context, such as by quitting their job (Langer & Landers, 2021).

Research on AI decision-making shows that employees exhibit resistance toward AI, primarily driven by a lack of trust in such systems and perceived unfairness (Lichtenthaler, 2020; Starke et al., 2022). However, wanting to be treated fairly is a fundamental human desire (Konradt et al., 2020), and fairness perception strongly influences people's beliefs, attitudes, and behaviors (Greenberg, 1990). Upholding fairness principles is vital for organizational effectiveness and employee satisfaction (Greenberg, 1990). Perceiving the organization as fair can increase employees' organizational commitment and elicit positive behaviors, benefitting organizations' performance (Clay-Warner et al., 2005; Colquitt & Zipay, 2015). In contrast, if employees perceive the actions of the organization or supervisor as unfair, they are likely to act against their organization (Zoghbi-Manrique-de-Lara & Suárez-Acosta, 2014). Thus, to prevent negative consequences, such as increased employee turnover (Hom et al., 1984), managing employees' fairness perceptions is crucial for determining the appropriate use of AI in the workplace (Kern et al., 2022).

2.3 Organizational Justice Theory and the Perception of AI Decision-Making

For effective implementation of AI systems, it is important to understand why employees perceive AI decisions as unfair (Starke et al., 2022). The study of fairness in the workplace is known as Organizational Justice (OJ) (Greenberg, 1990). OJ theory assesses employees' fairness perceptions across four dimensions: distributive, procedural, interpersonal, and informational fairness (Colquitt, 2001).

Distributive fairness captures the extent to which the distribution of outcomes such as rewards and other resources is perceived as fair. It assumes that individuals perceive decision outcomes as fair if they are allocated equally and in proportion to their efforts (Leventhal, 1976). Distributive fairness can impact employee attitudes and behaviors, including job satisfaction and motivation (Greenberg, 1990). AI can promote equality in decision outcomes in situations where human bias and error are likely, owing to the system being fact-based, data-driven, and objective (Morse et al., 2022). Additionally, people tend to view algorithms as fairer when they produce positive outcomes for themselves (Shulner-Tal et al., 2022; R. Wang et al., 2022). Wang and colleagues (2020) show that this outcome favorability has a greater impact on people's fairness perception than disclosing biased predictions across demographic groups. Thus, perceived distributive fairness of AI decisions significantly hinges on people's self-interest. This aligns with earlier findings from OJ research that emphasize the importance for people to have control over processes and outcomes (Greenberg, 1990). To eliminate potential biases stemming from outcome favorability and to focus solely on the overall impact of using AI-based decision-making on fairness perception, distributive fairness has been excluded from further research.

Procedural fairness refers to people's fairness perception of the decision-making process (Greenberg & Tyler, 1987). It has seven components: process control (i.e., ability to voice one's views during the process), decision control (i.e., ability to influence the outcome), consistency (i.e., everyone is treated equally), accuracy (i.e., process is based on valid, high-quality information), ethicality (i.e., process does not violate personal standards of ethics and morality), bias suppression (i.e., process prevents favoritism), and correctability (i.e., people can challenge or correct decisions) (Colquitt, 2001). AI may increase procedural fairness perception by performing more consistently than humans. Human decision-making can be influenced by factors such as stereotypes and subjective judgment (Tversky & Kahneman, 1974), and AI may improve procedural fairness perception by following standardized procedures and being unaffected by biases and emotions (Starke et al., 2022). For instance, Kaibel and colleagues (2019) show that AI enhances standardization in selection processes, leading participants to perceive AI-driven application processes as more consistent. Similarly, Lee (2018) reveals that AI-involved decision processes are perceived as more accurate, objective, and thus, less biased. However, AI decision-making has limitations when it comes to allowing people to express their thoughts or emotions about the procedure. (Acikgoz et al., 2020; Langer & Landers, 2021). The absence of human interaction hinders people to appeal

final decisions, contributing to higher perceptions of procedural unfairness (Langer & Landers, 2021).

Interpersonal fairness involves the extent to which employees are treated with respect, dignity, and consideration by the decision-making authority (Colquitt, 2001). It plays a crucial role in shaping employee behavior because of the frequent interpersonal encounters in organizations (Blodgett et al., 1997). People's interpersonal fairness perception seems to be harmed by AI (Langer & Landers, 2021), with the notion that AI is dehumanizing (Dietvorst et al., 2015) and impersonal (Binns et al., 2018). People believe that AI suffers from reductionism, oversimplifying things and breaking them down into simpler parts (Langer & Landers, 2021). Hence, they think that AI may miss important details (i.e., unique situations or qualitative information) and reduces decision-making to a number (Langer & Landers, 2021). For example, Kaibel and colleagues (2019) found that applicants with a non-standard professional background feel that AI cannot evaluate them fairly based on predefined standards in selection processes. Also, Bankins and colleagues (2022), who studied employees' perception in the Human Resource Management (HRM) context, observed that interpersonal fairness perception among employees is lower when decision outcomes were generated by AI compared to humans. The absence of human interaction in AI decision-making is seen as a drawback, making AI less appropriate as a decision-maker (Bankins et al., 2022).

Lastly, informational fairness relates to the fairness of communicated information, in terms of its completeness, truthfulness, and justification (Cropanzano & Ambrose, 2015). Moreover, transparency of AI systems has been identified as a crucial indicator for the fulfillment of this dimension (Starke et al., 2022). Informational fairness of AI decision-making is generally perceived as low (Langer & Landers, 2021; M. K. Lee et al., 2015). As noted by Holzinger and colleagues (2022), the methods used in AI are often "so complex that it is difficult for a human to re-trace, to understand, and to interpret how a certain result was achieved" (p. 13). Thus, transparency, interpretability, and explainability are important to ensure that algorithms are understandable to those affected by them (Renjith et al., 2020; Shin, 2020). Research suggests that technical design decisions prioritizing these aspects can influence the perceived fairness of AI. For example, Vallejos and colleagues (2017) found that young people perceive AI as fairer if they receive more information about it and the data it uses. Other studies (e.g., Angerschmid et al., 2022; Binns et al., 2018; Shin, 2020) support this, emphasizing that providing decision explanations improve perceived fairness. Also, Wang and colleagues

(2020) note that transparency about the AI developer impact people's fairness perceptions, with outsourced teams being perceived as less fair than in-house developers.

Concluding, individuals impacted by the adoption of AI in decision-making often meet it with skepticism. Consequently, organizations should consider all OJ dimensions when considering implementing and using AI, ensuring it is perceived as fair by both users and those affected by AI. Research proposes hybrid approaches that incorporate human control within AI decision-making processes (Starke et al., 2022).

2.4 Hybrid Decision-Making and the Role of Human Control

Hybrid decision-making combines human intelligence and AI, solving problems using the thoughtful allocation of tasks among algorithmic and human agents (Dellermann et al., 2019). Hybrid decision-making is considered more realistic because, in real-life tasks, humans are involved in the decision-making process and AI does not decide autonomously; yet (Kern et al., 2022). Indeed, studies that account for some sort of hybrid decision-making show that it is perceived as fairer than pure AI decisions. For example, Nagtegaal and colleagues (2021) found that in complex tasks, decision-making involving both humans and AI is considered more procedurally fair than algorithmic decision-making. Similarly, Kern and colleagues (2022) discovered that combining automated recommendations with a human's final decision is perceived as fairer than decisions made entirely by an AI. Also, Newman and colleagues (2020) revealed that decisions allowing humans to only consult AI are seen as fairer than fully automated decisions. Thus, by including human oversight and intervention in AI decision-making, fairness perception may be improved while still leveraging AI's capabilities. Based on this and considering individuals' concerns regarding AI-driven decisions mentioned in section 2.3, I hypothesize that employees will view hybrid decision-making processes as fairer than those based solely on AI owing to the additional human element.

H1: Employees who are exposed to hybrid decision-making procedures will perceive higher levels of procedural, interpersonal, and informational fairness compared to those who are exposed to full AI decision-making procedures.

Regulatory bodies and guidelines stress the importance of human control in AI decision-making (European Commission, 2019; European Union Parliament, 2020; UNIDIR, 2014). According to the European Commission, AI systems should enhance and complement human abilities – not replace them (European Commission, 2019). The General Data Protection

Regulation states the right not to be subject to completely automated decisions significantly affecting a data subject (European Union Parliament, 2020), and the United Nations Institute for Disarmament Research highlights the need for human oversight to prevent algorithmic harm (UNIDIR, 2014).

However, existing studies on human-AI collaboration in decision-making have not yet provided details about the specific roles of humans and AI in the process (Schwemer, 2019; Starke et al., 2022), and evidence concerning the control exerted by humans versus the system in work-related decisions is lacking (Langer & Landers, 2021). An exception is Jiang and colleagues' research (2022), which examined the impact of the order of decision-makers (i.e., AI-human or human-AI) on procedural fairness perceptions in multistage decision-making. The human-AI order resulted in lower procedural fairness perception (Jiang et al., 2022), indicating that the specific placement of humans in collaboration with AI influences fairness perception.

2.5 Hybrid Decision-Making Approaches

Shrestha and colleagues (2019) present three approaches to combining human and AI decision-making in organizations: full human-to-AI delegation, hybrid sequential decision structures, and aggregated human-AI decision-making (Shrestha et al., 2019). The first approach involves autonomous decision-making by AI. The other approaches are explained below.

Hybrid sequential decision-making offers two approaches where humans and AI make decisions sequentially (Shrestha et al., 2019), similar to the order of decision-makers investigated by Jiang and colleagues (2020). In the first approach, AI output serves as input for human decision-making. AI filters out inappropriate alternatives and passes suitable alternatives to the human decision-maker who makes a final decision based on these alternatives (Shrestha et al., 2019). This structure is comparable to the judge-advisor system, where the advisor gives the judge recommendations for decisions, which the judge then evaluates and considers in the final decision (Sniezek & Buckley, 1995). Here, the AI acts as the advisor, allowing the final decision-maker to effectively handle situations with many alternatives, such as reviewing job applications (Shrestha et al., 2019). Using AI to narrow down alternatives saves resources, and the algorithm can provide confidence levels for its suggestions, enabling the human to weigh the presented alternatives (Shrestha et al., 2019). However, AI decisions could lead to omission bias, potentially missing important information

or reporting it incompletely (Chanda & Banerjee, 2022). This can hide viable alternatives from human decision-makers unknowingly. Moreover, AI is trained with historical data and can reproduce human biases, impacting the choices in alternatives forwarded to the human decision-maker (Kordzadeh & Ghasemaghaei, 2022), as seen in the Amazon incident.

In the second approach, AI and humans switch roles. This structure allows for human control in the first phase, in which the human decision-maker can include personal expertise and intuition (Shrestha et al., 2019). AI then uses its predictive power and data analysis capabilities to make an informed final decision (Shrestha et al., 2019). An example where such a decision-making approach is already used is sports analytics (Millington & Millington, 2015). However, AI-based decision-making in the second phase requires careful selection of potential alternatives by the human in the first phase (Shrestha et al., 2019). As mentioned before, research on human decision-making indicates that humans are prone to biases. Meaning, irrational choices by humans can result in unfair final decisions made by AI. Also, the AI decision may lack comprehensibility for people due to the complexity of the algorithms used (Schoeffer et al., 2022). This in turn can hinder explanations of outcomes and prevent human decision-makers from learning from past decisions (Shrestha et al., 2019).

In aggregated human-AI decision-making, AI and humans get assigned specific aspects of the decision based on their strengths. Their outcomes are then combined into a collective decision (Shrestha et al., 2019). This approach allows the simultaneous, but independent, consideration of input and expertise from both sources (Oksana et al., 2022). Thus, this approach can prevent the reinforcement of human error or AI bias. Also, it could aid in avoiding irrational decisions, as the AI could cover decision-making aspects that humans may be incapable of noticing (Shrestha et al., 2019). For example, in the recruitment process, human decision-makers can use their intuition and experience to assess subjective factors that are important for a successful hire, while AI can evaluate the objective or numeric factors (Shrestha et al., 2019). However, this approach faces challenges in interpreting the AI component and replicating the human element too, impacting explainability, and consistency (Shrestha et al., 2019).

Concluding, these hybrid decision-making approaches have human intervention at different stages of the decision process, demonstrating advantages and disadvantages. Haesevoets and colleagues (2021) investigated managers' acceptance of human-machine partnerships in managerial decisions. Although their study focused on managers, it has implications for

employees' fairness perceptions. Their findings revealed that 5% of managers preferred machines to have the upper hand, 15% preferred an equal partnership, 50% preferred humans to have the upper hand, and 30% preferred humans to have complete control in decision-making (Haesevoets et al., 2021). Based on this, I hypothesize that employees' perception of hybrid decision-making may align with the observed preferences among managers, resulting in a similar distribution of fairness perception.

H2: Employees' fairness perceptions in hybrid decision-making will vary depending on the specific stage of human intervention. Specifically, employees will perceive decision-making procedures where humans have the last say as the fairest, followed by procedures with an equal partnership between humans and machines, and finally, procedures where AI has the last say will be perceived as the least fair.

2.6 Decision-Making Contexts

Fairness perception of AI also seems to be context-specific (Starke et al., 2022). When comparing people's perceptions and reactions to AI in different decision contexts, research suggests that task characteristics (human skills vs. mechanical skills) (Lee, 2018), task stakes (high-stakes vs. low-stakes) (Araujo et al., 2020; Langer et al., 2019), and task complexity (low complexity vs. high complexity) (Nagtegaal, 2021) may cause differences.

Previous research has explored various ways in which employers use AI in decision-making. Kellogg (2021) identified three main groups of AI use: directing employees, evaluating employees, and disciplining employees. Bankins and colleagues (2022) further associated the groups with specific HRM functions, including task allocation, training recommendation, promotion, firing, and performance management measures such as bonus payment. They examined employees' interpersonal fairness perceptions of AI decision-making compared to human decision-making in these six functions, along with other factors. However, their findings indicate consistent fairness perceptions, irrespective of the decision context involved (Bankins et al., 2022).

While these results could be explained by arguing that all HRM decisions may carry high stakes due to their potential consequences for employees, other studies emphasize distinctions among certain decision contexts in people's reactions. For example, Langer and colleagues (2019) compared people's reactions regarding AI decision-making in personnel selection (considered high-stakes) versus training recommendation (considered low-stakes) and found

that people reacted more negatively to the first. Höddinghaus and colleagues (2021) considered both promotions and vocational training as high-stake decisions and found that the automation of the first was perceived more positively than the latter. In contrast, Ötting and Maier (2018) found no significant difference in employees' procedural fairness perceptions between task allocation (considered low-stakes) and vocational training decisions (considered high-stakes). Thus, how decision contexts are defined and perceived seems to differ greatly across studies.

Due to these inconsistent findings, exploring potential discrepancies in fairness perception across different managerial decisions is important to account for any impact that different contexts may have on employees' perceptions of hybrid decision-making.

H3: Employees' fairness perceptions will vary depending on the decision context, with some contexts potentially leading to higher fairness perception of hybrid decision-making compared to others.

3 Methodology

3.1 Research Strategy and Design

This study investigated how different approaches to human oversight in AI decision-making affect employee fairness perception. For this, an experimental survey study using vignette methodology was designed (Malhotra et al., 2017, p. 80). Experimental survey studies enable researchers to assess causality in hypothetical scenarios by manipulating factors believed to impact individuals' judgments (Malhotra et al., 2017, p. 81). Building upon prior findings regarding the influence of AI decision-making on fairness perception (Bankins et al., 2022; Lee, 2018; Ötting & Maier, 2018), I used a quantitative approach with an online study implemented in the survey platform Qualtrics.

The study employed a two-way 4 (decision-making type: full AI vs. sequential AI-human vs. sequential human-AI, vs. aggregated human-AI) x 3 (decision context: promotion vs. bonus payments vs. career training) between-subjects design, with fairness perception as the dependent variable. This design resulted in 12 scenarios, to which participants were randomly and evenly assigned, with each participant only being exposed to one level of each factor. The between-subjects design was chosen to compare participants that were exposed to different scenario conditions, avoiding knowledge spillover and order effects (Aguinis & Bradley, 2014).

3.2 Participants

The sample size was determined beforehand by running a power analysis in G*Power (Faul et al., 2007). G*Power indicates that at least 225 participants are required for a two-way factorial ANOVA with 80% power, $\alpha = .05$, and a medium effect size ($f = 0.25$) for analyzing interaction effects. However, as this study had 12 conditions, this would result in each condition having only approximately 18 participants. With such a sample size, the power to detect medium-sized differences ($d = 0.5$) in planned contrasts would be limited to 32%. When comparing two groups, with 80% power, $\alpha = .05$, two-tailed, with a medium-sized effect ($d = 0.50$), 64 participants per condition are recommended. Thus, with 12 conditions, the desired sample size amounted to 768 participants.

The recruitment of participants was done in several steps. First, personal and professional networks were used. Second, participants were invited via social media platforms such as LinkedIn, Instagram, and Facebook. Additionally, two online platforms - SurveyCircle and

SurveySwap - were used to reach participants outside of my personal network. Also, to reach people who are inactive on social media, flyers including a QR-code leading to the survey were distributed in Münster, Germany. Finally, a gender-balanced sample of 65 participants was recruited from the crowdsourcing platform Profilic, with participants being paid 1.30€ for taking part in the survey.

Between May 2nd and May 22nd, a total of 1,095 survey responses were collected. Out of these, 8 responses were from participants who declined to consent to participate and 272 participants did not complete the survey. After excluding those, 810 responses remained. However, 18 of these failed the attention check and were eliminated from the analysis. The valid sample consisted of 797 participants, with 47.9% being male and 51.2% female. Participants' age ranged from 18 to 81 years old ($M = 31.00$, $SD = 11.67$), and most possessed a bachelor's or master's degree ($N = 581$). Also, most participants were full-time employed ($N = 299$), part-time employed ($N = 188$), or students ($N = 253$), with the majority working in the computer or IT ($N = 97$), consulting or management ($N = 84$), marketing or PR ($N = 58$), and engineering or manufacturing ($N = 51$) sectors. Most participants were European ($N = 715$), of which most were from Germany ($N = 544$), followed by Portugal ($N = 31$) and Austria ($N = 31$). Besides, participants originated from Asia ($N = 9$), America ($N = 13$), and Africa ($N = 15$). 83% of the participants somewhat or strongly agreed to be familiar with the term AI, and 66% reported that completing the survey was slightly, moderately, or extremely easy.

3.3 Procedure

Participants voluntarily and anonymously started the study after giving informed consent. Next, they had to indicate their familiarity with the term AI. To ensure a common understanding among participants, a standardized definition of the term was provided afterwards.

Then, participants were randomly assigned to one of the 12 hypothetical vignettes (see Appendix B). These depicted them as employees in a fictitious company, providing information about their roles, and the company's decision plans. All vignettes followed a consistent structure, were of similar length, and included the same information aspects: 1) the specific decision to be made (i.e., allocation of promotion, bonus payments, or career training; depending on the condition), 2) the employee data gathered (i.e., one subjective and one objective data point to suit both the human and AI decision maker), 3) the decision-making type applied (i.e., full AI, sequential human-AI, sequential AI-human, or aggregated human-

AI; depending on the condition), and 4) a detailed explanation of the decision-making process. A process flow diagram following business process model notation was provided to enhance comprehension of the process.

After reading the vignette, participants rated the decision-making procedure and provided a brief explanation for their rating. Two manipulation check questions followed, assessing the decision-making procedure and context. Next, participants were presented with statements concerning their procedural, informational, and interpersonal fairness perception. Then, trust in technology, trust in automation, and trust in people were measured as covariates. An attention check question was included (“I have never used a computer”), based on Curran and Hauser’s proposal (2019). Demographic questions were asked at the end. See Appendix A for the entire survey.

3.4 Variable Measurement

3.4.1 Main Variables

Employee fairness perception: This study’s dependent variable was employees’ fairness perception, measured using an adaptation of Colquitt’s organizational justice scale (Colquitt, 2001). It has been used in parts or complete by researchers to measure fairness perception in organizational contexts (Blodgett et al., 1997; Cropanzano & Ambrose, 2015), including AI-related fairness perception (Lee et al., 2019; Morse et al., 2022). The scale originally includes 20 items, of which the 4 items on distributive justice were neglected, to avoid outcome favorability bias. Thus, 16 items were included: 7 items for procedural fairness (e.g., “Have those procedures been free of bias?”), 4 items for interpersonal fairness (e.g., “Has (he/she) treated you with dignity?”), and 5 items for informational fairness (e.g., “Has (he/she) explained the procedures thoroughly?”). Participants responded on a 5-point Likert scale ranging from 1 (*to a small extent*) to 5 (*to a large extent*). Changes were made to align the scale with the study’s hypothetical situation. First, the verb tense was adapted (e.g., “Will this procedure be free of bias?”). Second, the interpersonal fairness items were modified to fit the study context (e.g., “Will the decision-making procedure make you feel treated with respect?”). Lastly, the informational fairness items were adapted to fit the control and treatment groups (e.g., “Will your manager and the AI system be able to explain the procedures thoroughly?”).

Decision-making type: The study used a first categorical independent variable consisting of four different types of decision procedures, including one full AI (control) and three hybrid

decision-making procedures (treatment): sequential AI-human, sequential human-AI, and aggregated human-AI (Shrestha et al., 2019). In the scenarios, participants were randomly assigned to one of these procedures to ensure equal distribution across treatment conditions.

Decision context: The second independent variable was the decision context. Similar to Bankins and colleagues' (2022) approach, I focused on the three AI functions defined by Kellogg and colleagues (2021). The contexts included training (representing directing), bonus payment (representing disciplining), and promotion (representing evaluating). These contexts were selected to maintain a manageable scope and provide a relatively positive context for analysis. Negative decision contexts such as firing or work allocation, as used by Bankins and colleagues (2022), were not included. Again, participants were randomly assigned to one of these contexts to ensure equal distribution.

3.4.2 Control Variables

Trust in technology: People's trust in technology may influence the fairness perception of AI (Pierson, 2017; R. Wang et al., 2022). Two scales by Mcknight and colleagues (2011) were used. The first scale included four items (e.g., "A large majority of technologies are excellent."), measuring faith in general technology on a 5-point Likert scale from 1 (*Strongly disagree*) to 5 (*Strongly agree*). The second scale included three items (e.g., "I usually trust a technology until it gives me a reason not to trust it."), and measured trusting stance in general technology on the same Likert scale. These scales were examined individually.

Trust in automation: Trust in automation influences individuals' reliance on or rejection of automated systems (J. D. Lee & See, 2009), and has been found to impact the fairness perception of AI (Bankins et al., 2022). Thus, I included participants' propensity to trust automation as a covariate, measured using Körber's (2018) 3-item scale (e.g., "I rather trust a system than I mistrust it.") on a 5-point Likert scale from 1 (*Strongly agree*) to 5 (*Strongly disagree*).

Trust in people: Hybrid decision-making could cause differing perceptions, depending on people's trust in humans. Research has demonstrated that people who have experienced discrimination from a human authority are more likely to perceive AI as fair (Kaibel et al., 2019). Thus, I included participants' propensity to trust people as a covariate and measured it with Schoorman and colleagues' (1996) 8-item scale (e.g., "Most people can be counted on to

do what they say they will do.”), using a 5-point Likert scale from 1 (*Strongly disagree*) to 5 (*Strongly agree*).

Demographics: Several participant characteristics were analyzed, including age, gender, nationality, educational background, current employment status, and industry sector. Gender options were male, female, or other. Age was obtained numerically in years through an open-ended question. Nationality, educational background, and industry sector were collected using a single-choice question format.

4 Results

4.1 Data Preparation

Manipulation checks assessed participants' accuracy in identifying the decision type and context presented in the scenario. Among all participants, 7.4% incorrectly recognized the decision context, 14.7% incorrectly recognized the decision type, and 18.4% got both the decision type and context wrong. Still, I decided not to exclude these participants to maintain the desired minimum sample size per condition, but the manipulation check was included as a control variable in subsequent analyses. Also, dummy variables were created for participants' origin (i.e., Asian, American, African, and European), educational background, employee status, and industry sector. A binary variable was created for gender (0 = female, 1 = male).

4.2 Scale Reliability

First, reverse-scored items were recoded to align their scoring direction with the other scale items. Also, items within a scale should exhibit internal consistency, indicating that they are measuring the same underlying construct (Bland & Altman, 1997). Since the scales used in this study consisted of more than two items, a factor analysis, as recommended by Field (2009), was conducted. Also, Cronbach's α coefficient was calculated to assess each scale's reliability. When comparing groups, Cronbach's α values between .70 and .80 are considered satisfactory (Bland & Altman, 1997).

A principal component analysis was conducted on the 16 items from the OJ scale using orthogonal rotation (varimax). The adequacy of the sample for the analysis was verified using the Kaiser-Meyer-Olkin (KMO) measure, resulting in a KMO value of .89, which is considered great according to Field (2009). Additionally, all item KMO values exceeded .81, well above the acceptable limit of .5 (Field, 2009). Bartlett's test of sphericity, $\chi^2(120) = 5099.72$, $p < .001$, indicated that the correlations between items were sufficiently large for conducting the principal component analysis. The initial analysis yielded Eigenvalues for each component, with four components having Eigenvalues above Kaiser's criterion of 1. Together, these components explained 62.16% of the variance. Component 1 represented interpersonal and Component 2 informational fairness. However, the items related to procedural fairness were split across two components. Also, two items loaded on incorrect components ("Will this decision-making procedure uphold ethical and moral standards?" loaded on interpersonal fairness and "Will this decision-making procedure refrain from

improper remarks or comments?” on one of the procedural fairness components). I explored alternative approaches (see Appendix C for more detail), to find a better-fitting model. First, I reran the analysis, specifying the extraction of only three components. This resulted in a worse model fit compared to the four-component model (55.69% variance explained), with five items loading incorrectly. I then turned to an approach following Maharee-Laweler and colleagues (2010) who faced a similar split of procedural fairness: Using oblimin rotation, this extraction yielded the same component structure and item loading pattern as the initial approach. Consequently, I chose the four-factor model as the best-fitting model for several reasons. First, the two items that loaded incorrectly had the second-highest loading on the correct component, indicating some degree of association. Second, the four-factor model only had two items failing to load correctly. Third, when calculating Cronbach’s α for the subscales, the two procedural fairness components showed higher or similar values ($\alpha = .76$ and $.71$) than merging them into a single factor ($\alpha = .71$). Thus, procedural fairness was treated as two separate subscales in the subsequent analyses: procedural-voice referring to items associated with individuals’ opportunities to influence, appeal, or express opinions during the decision process, and procedural-core referring to items related to process characteristics. The entire OJ scale yielded a satisfactory Cronbach’s α (.89), as did the interpersonal and informational fairness subscales (.82 and .78 respectively).

The factor analyses of the items used in faith in general technology, trusting stance in general technology, and propensity to trust automation scales each revealed a single underlying factor, indicating that the items measured the same construct. Cronbach’s α for faith in general technology and trusting stance in general technology were satisfactory at .76 and .73, respectively. However, propensity to trust automation yielded a Cronbach’s α of .61, which is not considered satisfactory according to Bland and Altman (1997). Although this subscale, taken from Körber’s questionnaire, originally yielded a Cronbach’s α of .75, Körber (2018) explained that smaller subscales may exhibit sensitivity to different situations and contexts. Also, participants may hold a tendency to trust automated systems and perceive them as generally functioning well while also believing that one must be cautious with unfamiliar automated systems, especially considering they were presented with the AI-workplace scenario. Thus, the obtained Cronbach’s α below .70 can be justified but should be interpreted cautiously.

The factor analysis of the propensity to trust people scale extracted two components. The first consisted of the two negatively worded items (after reverse coding) and the second consisted

of the remaining positively worded items. As the scale yielded a satisfactory Cronbach's α of .72, I concluded that this was an impact of wording and that the items still correspond to the same underlying concept.

After ensuring the reliability of the different scales, scores were obtained by averaging each scale's items.

4.3 Demographic Analysis

I analyzed the impact of participants' demographic characteristics on overall fairness perception (see Appendix D). No significant differences were observed based on gender, nationality, and employment status. However, age negatively correlated with overall fairness perception. As age increased, fairness perception decreased, $r(794) = -.11, p = .002$. Also, a negative correlation was found between overall fairness perception and participants with a Ph.D., $r(795) = -.09, p < .05$. Considering participants' lines of work, significant negative correlations were found between overall fairness perception and the science/pharmaceutical sector, $r(795) = -.10, p < .01$, and the environment/agriculture sector, $r(795) = -.08, p < .05$. Appendix E provides a correlation table between the dependent variables and non-demographic covariates.

4.4 Participants' General Fairness Perception

To address the exploratory RQ1, I reviewed participants' quantitative and qualitative answers to the question of how fair they generally believed the decision-making procedure in the presented scenario to be. Due to space constraints, noteworthy findings from the quantitative analysis are mentioned in Section 4.5.1. For the entire quantitative analysis see Appendix F.

4.4.1 Qualitative Analysis

The open-text responses were thematically analyzed according to the step-by-step guide provided by Braun and Clarke (2006) to uncover meaningful patterns within the data. German responses were translated into English, and a data-driven, inductive approach was used for theme identification (Braun & Clarke, 2006). In this approach, the data was coded without attempting to fit it into pre-existing categories, with themes emerging naturally directly from the data itself. (Braun & Clarke, 2006). After becoming familiar with the data, common themes were identified based on explicit or surface meanings (Braun & Clarke, 2006), and the

responses were organized according to the identified themes and then summarized for interpretation.

Before analysis, the responses were categorized based on sentiment: positive, mixed, negative, and neutral (see Table G1, Appendix G). Mixed responses occurred when participants provided contradictory statements, mentioning both positive and negative aspects of the decision-making process. For instance, participants often mentioned positive aspects and ended their statement with a “but”, expressing concerns or doubts. In such cases, I coded the response with all relevant positive and negative themes. All relevant data were coded under at least one minor theme. Minor themes were grouped into three major themes: positive, negative, or neutral. Appendix G provides the frequency of theme usage, including example quotes representing the minor themes. The following paragraphs summarize the qualitative analysis results.

The aggregated human-AI condition showed the largest percentage of themes that were positive (70%). The most frequent positive themes were 1) the combination of subjective and objective evaluation (20%; e.g., “The AI can review the objective data just fine. I think it could have had problems with reviewing the subjective data, so that’s a good thing that a human is doing that.” [Promotion]), 2) that each decision-making actor was responsible for something they are good in (13%; e.g., “Qualities of both agents are fairly distributed” [Bonus]), and 3) that the human was making the final decision (9%; e.g., “The use of AI is useful for the analysis of objective indicators, as long as it is integrated with qualitative data analyzed by humans. I think it is important that the final decision is human made.” [Bonus]). Also, this condition had the lowest percentage of negative themes (23%).

Interestingly, the sequential human-AI condition (not the full AI) had the largest percentage of themes that were negative (54%). The most frequent negative themes were 1) that subjective factors, interpersonal components, or soft skills were neglected (14%; e.g., “I can see how the process makes sense, but I think that there are also other important characteristics of potential employees that go beyond hard skills etc.” [Training]), and 2) that the decision could have been subject of human bias or partial human judgment (14%; e.g., “The selection of the shortlist can already have a strong bias, so that the AI can no longer evaluate neutrally.” [Promotion]). Thus, it seems that in addition to concerns regarding AI bias in decision-making (Kordzadeh & Ghasemaghaei, 2022), people also identify human bias as an issue.

Similar to sequential human-AI, the full AI condition had more negative (50%) than positive themes (35%). The most frequent positive theme was that 1) the fully automated decision was objective or neutral (10%; e.g., “Kind of fair as it is objectively chosen by an AI and no human.” [Promotion]), followed by 2) the decision was performance-oriented, data-driven or fact-based (8%; e.g., “The system refers to actual data on the candidate’s performance which is fair in my opinion.” [Training]). The most frequent negative themes again included 1) that subjective factors were neglected (16%; e.g., “This process can’t take into account soft skills such as interpersonal connection, empathy, etc.” [Promotion]), and that 2) the human aspect was missing (9%; e.g., “AI should not solely make decisions about such things, the system lacks human behavior and emotion and might not be able to see other important factors that are clear to the human manager.” [Training]). Also, the full AI condition had the largest percentage of neutral themes (14%), including the minor theme that fairness depends on the AI and the data it is trained with. Here, this condition yielded the largest percentage (11%).

Finally, the sequential AI-human condition had more positive (51%) than negative (41%) themes. The most common positive theme was that the human made the final decision (12%; e.g., “It’s okay because the final decision is still made by the manager.” [Promotion]), and the most common negative theme was that subjective factors were neglected by the AI (10%; e.g., “The personality and the interpersonal factors are not taken into account.” [Training]). The second most common negative theme was that potential prospects might have been omitted during the preselection of the AI (7%; e.g., “Some candidates might not be on the AI-generated shortlist although the manager might think they would be appropriate.” [Promotion]). Interestingly, human bias or partial human judgment was the third most common negative theme in the sequential AI-human condition (6%).

4.5 Hypotheses Testing

4.5.1 Differences in Perception across Decision Types – Overall Fairness

H1 predicted that hybrid decision-making would be perceived as fairer by participants than full AI decision-making, while H2 predicted that participants’ fairness perceptions would vary depending on the specific stage of human intervention. To test these hypotheses, I conducted an independent one-way ANOVA, with overall fairness perception as the dependent variable and decision type as the independent variable (see Table 1 for *N*, *M*, and *SD* per condition). Both the normality (Shapiro-Wilk test $p > .05$) and the homogeneity of variances (Levene’s test $p > .05$) assumptions were met.

The ANOVA yielded a significant effect of decision type on the average overall fairness perception, $F(3, 793) = 18.78, p < .05, \eta^2 = .07$. The small, medium, and large effect size cutoffs for η^2 are .01, .06, and .14, respectively (Richardson, 2011). Thus, a significant difference was observed among the decision types regarding participants' average overall fairness perception¹.

Table 1. *Average Overall Fairness Perception per Decision Type*

Condition	<i>N</i>	<i>M</i>	<i>SD</i>
Full AI	203	2.83	0.67
Sequential Human-AI	196	2.91	0.60
Aggregated Human-AI	196	3.29	0.67
Sequential AI-Human	202	3.00	0.63

To identify group differences, I conducted three planned contrasts, which allowed the testing of specific hypotheses by comparing conditions based on predefined assumptions (Field, 2009). Contrast 1 tested H1 by comparing full AI to hybrid decision-making. The results supported H1, showing greater fairness perception in the hybrid compared to full AI decision-making, $t(793) = 4.55, p < .05, d = 1.11$. Contrast 2 and 3 compared the hybrid decision-making types to each other, testing H2. In Contrast 2, sequential AI-human and aggregated human-AI were compared to sequential human-AI decision-making, revealing a statistically significant difference in overall fairness perception, supporting H2, $t(793) = 4.11, p < .05, d = 0.72$. However, Contrast 3, comparing sequential AI-human to aggregated human-AI, found a statistically significant difference in the opposite direction of H2, $t(793) = -4.38, p < .05, d = -0.44$. Thus, H2 was partially supported, indicating differences in overall fairness perception among hybrid decision-making procedures, but with aggregated human-AI, not sequential AI-human, being perceived as the fairest.

4.5.2 Differences in Perception across Decision Types – Fairness Dimensions

I repeated the procedure of the previous subsection for each fairness dimension. Before this, I tested the dependent variables for normality and homogeneity. All variables were normally

¹ The study included one item measuring general fairness (see Chapter 4.4) before the overall fairness scale. The analysis of that item was consistent with the overall fairness results. Notably, the analysis of the general fairness rating revealed no significant difference between full AI ($M = 4.32, SD = 1.71$) and sequential human-AI decision-making ($M = 4.38, SD = 1.59$), $Z = -0.19, p = .849$. For the detailed analysis, see Appendix F.

distributed, and variances were homogeneous for the procedural and the informational fairness variables, $p > .05$, but not for interpersonal fairness, $p < .05$. Thus, I used Welch’s robust test for equality of means to analyze interpersonal fairness, accounting for the unequal variances between groups (Delacre et al., 2017). The ANOVAs showed significant effects of decision type on fairness perception in each dimension, see Table 2. Thus, significant differences exist in participants’ procedural-voice, procedural-core, interpersonal, and informational fairness perception across the different decision-making procedures.

Table 2. ANOVA for Fairness Dimensions

	Full AI		Sequ. Human-AI		Aggregated Human-AI		Sequ. AI-Human		F (3, 793)	p	η^2
	M	SD	M	SD	M	SD	M	SD			
Procedural-Voice	1.85	0.85	2.17	0.84	2.6	0.94	2.26	0.95	23.43	<.001	.08
Procedural-Core	3.3	0.82	3.2	0.79	3.44	0.82	3.26	0.75	3.40	.017	.01
Interpersonal	2.86	1.00	2.93	0.84	3.49	0.84	3.15	0.87	19.58	<.001	.07
Informational	3.02	0.88	3.12	0.75	3.41	0.83	3.12	0.77	8.68	<.001	.03

Note. Interpersonal fairness results are based on Welch’s robust test for equality of means.

I conducted planned contrasts to identify specific group differences within the fairness dimensions (see Table 3). Contrast 1 showed significant differences between full AI and the three hybrid conditions, except for procedural-core. Contrast 2 revealed significant results, indicating that participants perceived sequential AI-human and aggregated human-AI as fairer than sequential human-AI decision-making across all dimensions. Also, Contrast 3 showed a significant difference, suggesting that aggregated human-AI was perceived as fairer than sequential human-AI across all dimensions.

Table 3. Planned Contrasts for Fairness Perception on Dimensional Level

	Contrast	Value of Contrast	SE	$t(793) \dagger$	d
Procedural-Voice	1 – Full AI vs. Hybrid	1.47	0.22	6.74*	1.64
	2 – Sequ. Human-AI vs. Other Hybrid	0.53	0.16	3.38*	0.59
	3 – Sequ. AI-human vs. Aggregated	-0.33	0.09	-3.72*	-0.37
Procedural-Core	1 – Full AI vs. Hybrid	0.01	0.19	0.04	0.01
	2 – Sequ. Human-AI vs. Other Hybrid	0.31	0.14	2.25*	0.39
	3 – Sequ. AI-human vs. Aggregated	-0.18	0.08	-2.29*	-0.23

	Contrast	Value of Contrast	SE	$t(793) \dagger$	d
Interpersonal	1 – Full AI vs. Hybrid	0.99	0.23	4.21*	1.11
	2 – Sequ. Human-AI vs. Other Hybrid	0.77	0.15	5.18*	0.86
	3 – Sequ. AI-human vs. Aggregated	-0.34	0.09	-3.92*	-0.38
Informational	1 – Full AI vs. Hybrid	0.6	0.20	3.07*	0.75
	2 – Sequ. Human-AI vs. Other Hybrid	0.29	0.12	2.08*	0.36
	3 – Sequ. AI-human vs. Aggregated	-0.29	0.08	-3.54*	-0.36

Note. * $p = <.05$, Cohen's d categorizes effect sizes as small (0.2), medium (0.5), and large (0.8) (Cohen, 1992). \dagger degrees of freedom for interpersonal fairness differ as equal variances were not assumed for Contrast 1 (308.90), Contrast 2 (393.63), and Contrast 3 (396.00).

4.5.3 Controlling for Covariates – Overall Fairness

People's trust in technology, automation, and people have been identified as factors that may influence fairness perception of AI decision-making, possibly confounding the relationship between decision type and the dependent variables. I conducted an ANCOVA, assessing the unique impact of decision type on participants' overall fairness perception while controlling for covariates. All covariates, except the propensity to trust people and the manipulation check, were statistically significant (see Table 4), suggesting that overall fairness perception was impacted by them. However, after including the covariates, decision type remained significantly related to overall fairness, $F(3, 780) = 21.08, p < .001, \eta^2 = .07$, suggesting a robust effect.

Table 4. ANCOVA Overall Fairness

	β	$F(1, 780)$	p	η^2
Faith in General Technology	.23	41.58	<.001	.05
Trusting Stance in General Technology	.11	13.36	<.001	.02
Propensity to Trust Automation	.20	32.29	<.001	.04
Propensity to Trust People	.05	2.17	.141	.00
Manipulation Check	-.09	2.32	.128	.00

Furthermore, to assess the fairness perception of hybrid decision-making compared to full AI decision-making, testing H1, while controlling for covariates, I used planned contrasts. A new dummy variable was created to represent decision type, with full AI as the reference group. The contrasts indicated that full AI procedures significantly reduced overall fairness

perception compared to aggregated human-AI, $t(788) = 7.65, p < .001, \eta^2 = .07$, and sequential AI-human, $t(788) = 3.09, p = .002, \eta^2 = .01$, but not compared to sequential human-AI decision-making, $t(788) = 1.95, p = .052, \eta^2 = .01$. Thus, controlling for covariates, H1 was only partially supported regarding overall fairness, suggesting that the influence attributed to the decision type in the ANOVA may have been confounded by the covariates.

To assess the robustness of H2 while considering the covariates, I performed repeated contrasts. I wanted to validate the previous ANOVA results, specifically with aggregated human-AI being perceived as the fairest and sequential human-AI as the least fair. A new dummy variable representing decision type was created, following the order suggested by the ANOVA. The repeated contrasts allowed comparing the groups in this specific order. The order of results was in line with the previous findings, from most fair to least: aggregated human-AI ($M_{\text{adjusted}} = 3.26, SE = 0.04$), sequential AI-human ($M_{\text{adjusted}} = 3.00, SE = 0.04$), sequential human-AI ($M_{\text{adjusted}} = 2.94, SE = 0.04$), and full AI ($M_{\text{adjusted}} = 2.84, SE = 0.04$). However, only the differences between aggregated human-AI and sequential AI-human, as well as between sequential human-AI and full AI, were statistically significant at $p < .05$. Thus, H2 was not supported in terms of overall fairness.

4.5.4 Controlling for Covariates – Fairness Dimensions

I examined the unique impact of different decision types on participants' procedural-voice, procedural-core, interpersonal, and informational fairness perception while controlling for the covariates. Levene's test yielded non-significant results for all ANCOVAs, except for the procedural-voice dimension. However, the significant Levene's test only emerged when the manipulation check was included, suggesting that participants' fairness perception in the procedural-voice dimension varied based on their understanding of the assigned decision type. To address this violation, I performed a rank transformation before the parametric calculation, following the approach recommended by Conover and Iman (1982). The rank transformation involved replacing the procedural-voice data with corresponding ranks, and the F-statistics were calculated based on the ranks rather than the raw data (Conover & Iman, 1982).

Most covariates showed a significant relationship with fairness perception across the individual dimensions (see Appendix H). Propensity to trust people had a significant association with interpersonal fairness perception only. Also, participants' performance in the manipulation checks strongly influenced procedural-voice fairness, with significance

observed only in this dimension. Regardless of the covariates, decision type had a significant effect on fairness across all dimensions.

Next, I performed planned and repeated contrasts to examine whether the differences in perception across decision types aligned with the ANOVA findings at the dimensional level while controlling for the covariates. Covariate adjusted descriptives are presented in Table 5. Planned contrasts revealed that full AI decision-making significantly decreased interpersonal fairness perception compared to aggregated human-AI, $t(788) = 7.21, p < .001, \eta^2 = .06$, and sequential AI-human decision-making, $t(788) = 3.53, p < .001, \eta^2 = .02$, but not compared to sequential human-AI decision-making, $t(788) = 1.24, p = .214, \eta^2 = .002$. Moreover, full AI decision-making significantly decreased the informational fairness perception compared to aggregated human-AI, $t(788) = 4.99, p < .001, \eta^2 = .03$, but not compared to sequential AI-human, $t(788) = 1.33, p = .186, \eta^2 = .002$, or sequential human-AI decision-making, $t(788) = 1.75, p = .080, \eta^2 = .004$. Additionally, full AI decision-making did not significantly decrease the procedural-core fairness perception compared to aggregated human-AI, $t(788) = 1.34, p = .179, \eta^2 = .002$, sequential AI-human, $t(788) = -.91, p = .365, \eta^2 = .001$, or sequential human-AI decision making, $t(788) = -1.32, p = .187, \eta^2 = .00$. Lastly, full AI decision-making significantly decreased procedural-voice fairness perception compared to aggregated human-AI, $t(788) = 9.19, p < .001, \eta^2 = .08$, sequential AI-human, $t(788) = 5.73, p < .001, \eta^2 = .04$, and sequential human-AI decision-making, $t(788) = 4.74, p < .001, \eta^2 = .03$.

Table 5. *Covariate Adjusted Descriptive Statistics for Fairness Dimensions*

Dimension	Decision Type	M^a	SE
Procedural-Voice†	Aggregated Human-AI	492.19	15.22
	Sequential AI-Human	416.45	14.97
	Sequential Human-AI	369.24	15.21
	Full AI	294.33	15.07
Procedural-Core	Aggregated Human-AI	3.42	0.05
	Sequential AI-Human	3.25	0.05
	Sequential Human-AI	3.22	0.05
	Full AI	3.32	0.05
Interpersonal	Aggregated Human-AI	3.45	0.06
	Sequential AI-Human	3.15	0.06
	Sequential Human-AI	2.96	0.06
	Full AI	2.87	0.06

Dimension	Decision Type	M^a	SE
Informational	Aggregated Human-AI	3.40	0.05
	Sequential AI-Human	3.20	0.05
	Sequential Human-AI	3.15	0.05
	Full AI	3.02	0.05

Note. ^a = adjusted, † Adjusted mean values for procedural-voice fall out of line because of the rank transformation.

Continuing with the repeated contrasts (see Table 6), only aggregated human-AI and sequential AI-human decision-making showed significant differences across all dimensions. Aggregated human-AI was consistently perceived as the fairest across all fairness dimensions. A significant difference was shown between sequential AI-human and sequential human-AI, indicating that the former was perceived as fairer than the latter in terms of interpersonal fairness. Sequential human-AI only differed significantly from full AI decision-making in the procedural-voice dimension.

Table 6. *Repeated Contrasts*

	Interpersonal		Informational		Procedural-Voice		Procedural-Core	
	Contrast Estimate	p	Contrast Estimate	p	Contrast Estimate	p	Contrast Estimate	p
1	0.30	<.001	0.27	<.001	75.74	<.001	0.16	.024
2	0.18	.023	-0.03	.650	20.21	.344	0.03	.655
3	0.10	.227	0.13	.078	101.91	<.001	-0.10	.182

Note. 1 = Aggregated vs. Sequential AI-human, 2 = Sequential AI-human vs. Sequential human-AI, 3 = Sequential human-AI vs. Full AI. Contrast estimates for procedural-voice fall out of line because of the rank transformation.

4.5.5 Differences in Fairness Perception among Decision Contexts

The final hypothesis, H3, examined employees' fairness perceptions in different decision contexts: promotion, career training, and bonus payment. Two-way factorial ANOVAs were performed to analyze the main effects of decision context on the dependent variables and explore potential interactions between decision type and context.

First, I examined the effects and potential interactions regarding overall fairness perception, followed by analyzing each fairness dimension. However, Levene’s test for the two-way factorial ANOVA on overall fairness, informational fairness, and procedural-voice fairness perception violated the assumption of equal variances ($p < .05$). To address this, the Aligned Rank Transform (ART) ANOVA, a non-parametric approach, was employed in these cases. The ART method involves aligning and ranking the data for each main and interaction effect, providing a more robust analysis when equal variances are violated (Wobbrock et al., 2011). This transformation ensures appropriate Type I error rates and suitable power for main effects and interactions (Wobbrock et al., 2011).

Regarding overall fairness perception, the analysis revealed a significant main effect of decision type, $F = 16.40$, $p < .05$, $\eta^2 = .06$, indicating that the specific decision type significantly influenced participants’ overall fairness perception. However, there was no significant main effect of decision context, $F = 2.91$, $p > .05$, $\eta^2 = .01$, meaning the decision context did not significantly influence participants’ overall fairness perception. Also, no statistically significant interaction between the effect of decision type and decision context was found, $F = 0.85$, $p > .05$, $\eta^2 = .01$, suggesting that the impact of different decision types on overall fairness perception did not differ across the decision contexts, not supporting H3.

Furthermore, despite the absence of a main effect of decision context and an interaction effect in overall fairness, the procedure was repeated for each fairness dimension, see Table 7. The results showed consistent significant main effects of decision type across all dimensions. However, decision context did not show significant effects in the fairness dimension, except in the case of informational fairness. Here, both decision type, $F = 8.29$, $p < .05$, $\eta^2 = .03$, and context, $F = 3.52$, $p < .05$, $\eta^2 = .01$, had significant main effects. However, no significant interaction effect was found in any fairness dimension. Thus, the decision types and contexts seem to not interact in their effect on fairness perceptions in the dimensions.

Table 7. *Two-Way Factorial ANOVA for Fairness Dimensions*

	Effect	F	df	p	η^2
	Type	19.22	3	<.001	.07
Interpersonal	Context	1.73	2	.178	.00
	Type * Context	0.63	6	.706	.00

	Effect	<i>F</i>	<i>df</i>	<i>p</i>	η^2
Informational	Type	8.29	3	<.001	.03
	Context	3.52	2	.030	.01
	Type * Context	1.47	6	.184	.01
Procedural-Core	Type	3.37	3	.018	.01
	Context	2.67	2	.070	.01
	Type * Context	0.59	6	.743	.00
Procedural-Voice	Type	24.68	3	<.001	.09
	Context	0.86	2	.425	.00
	Type * Context	1.76	6	.104	.01

Lastly, I conducted follow-up tests for the decision context in the informational fairness dimensions, to identify how contexts differed. As I had no specific priori prediction about the differences between contexts, pairwise comparisons were employed to explore any between-group differences. Pairwise comparisons are designed to compare all different combinations of the treatment groups (Field, 2009). I performed Gabriel's pairwise test as my sample sizes across the three decision context groups slightly differed but had equal variances. Here, Gabriel's test provides good control over the Type I error rate and ensures great statistical power (Field, 2009). Table 8 shows that the decision contexts that differed significantly from each other in terms of informational fairness were the bonus payment ($M = 3.28$, $SD = 0.79$) and career training ($M = 3.08$, $SD = 0.86$) contexts. Thus, H3 was not supported, except for the informational dimension, where participants perceived bonus payment decisions to be, overall, fairer than career training decisions.

Table 8. *Pairwise Comparison (Gabriel's Test) for Informational Fairness*

Decision Context (I)	Decision Context (J)	Mean Difference (I-J)	<i>SE</i>	<i>p</i>
Promotion	Bonus Payment	-0.14	0.07	.131
	Career Training	0.07	0.07	.722
Bonus Payment	Promotion	0.14	0.07	.131
	Career Training	0.21	0.07	.011
Career Training	Promotion	-0.07	0.07	.722
	Bonus Payment	-0.21	0.07	.011

5 Discussion

The use of AI in organizations is changing how managers make decisions (Höddinghaus et al., 2021). However, employees often feel that AI-based decisions affecting them are unfair (Starke et al., 2022), harming organizations (Greenberg, 1990). To tackle this issue, the concept of hybrid decision-making, combining AI algorithms with human oversight, has emerged. Nevertheless, the optimal level of human versus system control and the specific roles of both humans and AI in the decision process remains uncertain. Thus, this paper aimed to answer: “How do different approaches to human oversight in AI decision-making affect employee fairness perceptions?”. I compared three hybrid decision-making approaches to full AI decision-making and assessed employees’ fairness perceptions across three distinct decision contexts.

Initially, I examined participants’ responses, quantitative and qualitative, regarding their general fairness perceptions in the presented decision-making scenarios. Quantitative analysis showed that participants rated full AI as less fair than aggregated human-AI and sequential AI-human, but not than sequential human-AI decision-making. Among the hybrid decision-making approaches, sequential human-AI received the lowest and aggregated human-AI the highest general fairness score. This aligned with the qualitative analysis, that categorized participants’ responses into themes. For sequential human-AI, participants provided a similar number of positively and more negatively themed justifications as for full AI. Regarding full AI decision-making, participants recognized AI’s objectivity, and fact-based nature but expressed concerns about neglecting subjective and interpersonal components. They felt that the human element, including emotions, sympathy, and intuition, was missing from the process. Similar concerns were expressed for sequential human-AI decision-making. This is consistent with existing literature, observing that people perceive automated decision-making as dehumanizing and impersonal (Binns et al., 2018; Dietvorst et al., 2015). Interestingly, in hybrid decision-making, participants expressed concerns about potential partial judgment or bias from the human counterpart, particularly in sequential human-AI decision-making. So far, only Bankins and colleagues (2022) reported a similar tendency.

As for hypotheses, H1 tested whether the hybrid is perceived as fairer than full AI decision-making in terms of overall, procedural-voice, procedural-core, interpersonal, and informational fairness. H1 was partially supported. The results indicate that overall fairness perception differed between full AI and hybrid decision-making, with the latter being

perceived as fairer. However, these findings were not entirely robust when considering the covariates. Trust in technology and automation significantly increased participants' overall fairness perception. Controlling for covariates, no significant difference was found between full AI and sequential human-AI decision-making. This suggests that the impact attributed to the decision type may have been confounded by one or more covariates. Literature suggests that algorithm literacy and computer knowledge positively influence the fairness perception of AI decision-making (Pierson, 2017; R. Wang et al., 2022).

Considering the fairness dimensions, full AI significantly differed from hybrid decision-making across all dimensions, except for procedural-core fairness. When accounting for covariates, this pattern remained consistent, but only aggregated and sequential AI-human decision-making were found to significantly increase interpersonal fairness perception, and only aggregated human-AI decision-making significantly increased informational fairness perception, compared to full AI.

Procedural-core fairness assessed accuracy, consistency, absence of bias, and adherence to ethical standards in decision processes. The results indicate that individuals may not perceive significant differences in these characteristics between full AI and hybrid decision-making. This aligns with previous research highlighting the objectivity, accuracy, and standardization of AI compared to human decision-making (Kaibel et al., 2019; Lee, 2018). Also, it corresponds with participants' qualitative responses regarding full AI decision-making, emphasizing its objective nature. Participants believed that fully automated processes would minimize human prejudice, promote equal opportunities, and ensure consistency.

Interpersonal fairness relates to individuals' interpersonal treatment during the decision-making process. The findings suggest that participants perceived a similar degree of dignity, respect, and truthfulness in fully automated and sequential human-AI decision-making. This similarity may arise because these approaches ultimately have the AI make the final decision, not humans. Participants' qualitative responses in these conditions criticized the lack of human involvement in the final decision-making or at least double-checking AI's decision. To address feelings of disrespect or indignity, including further human review could be beneficial in implementing either approach.

Informational fairness concerns the truthfulness and justification of information provided by decision-makers. AI's opaque nature can result in perceptions of unfairness (Holzinger et al., 2020). However, research has shown that explanations for decisions can mitigate perceptions

of unfairness (e.g., Angerschmid et al., 2022). The perception of aggregated human-AI being more informationally fair than the other approaches may stem from AI handling objective while humans address subjective data. Participants noted that subjective factors were overlooked in full AI and both sequential decision-making approaches and expressed a preference for dividing responsibilities based on strengths in aggregated human-AI decision-making. Thus, the latter approach may allow better explanations of decisions, increasing employees' informational fairness perception.

H2 examined how different types of human involvement in hybrid decision-making impact employees' fairness perceptions. Again, I considered overall fairness and the individual fairness dimensions. When not accounting for covariates, differences in perceptions among hybrid decision-making were observed, although not entirely aligning with the predicted outcomes. Consistent with H2 and Haesevoets and colleagues' (2021) findings, overall fairness perception was lowest in sequential human-AI decision-making, where AI had the final decision-making authority. However, the highest fairness was perceived in aggregated human-AI decision-making, where AI and humans collaborate with distinct and independent roles, and not in sequential AI-human. This observation remained consistent across all dimensions and contradicted H2. However, when accounting for covariates, only aggregated human-AI decision-making showed a significant difference from sequential AI-human regarding overall fairness. No significant difference was found between the two sequential approaches. In the dimensions, only aggregated human-AI and sequential AI-human significantly differed. Thus, considering the covariates, H2 was not supported. However, it can be concluded that aggregated human-AI decision-making yields the highest fairness perception among the hybrid approaches.

H3 explored employees' fairness perceptions across different decision-making contexts and predicted variations based on these contexts. H3 was not supported, as neither significant differences in fairness perception among the studied contexts (i.e., promotion, bonus payment, and career training allocation), nor significant interactions between contexts and decision types were observed, except for informational fairness, where participants perceived bonus payment decisions to be fairer than career training decisions. This contradicts most existing research (e.g., Langer et al., 2019; Nagtegaal, 2021), but aligns with Bankins and colleagues (2022), who focused on interpersonal fairness perception in HRM decisions, and Ötting and Maier (2018), who examined procedural fairness perception in task allocation versus vocational training. Thus, participants may not assign different levels of importance to these

contexts and may have rated them as similarly high- or low-stakes. Considering informational fairness, employees may perceive bonus payment allocation as a task requiring more mechanical than human skills, leading to differing fairness perceptions (Lee, 2018). Also, bonus payment decisions might be perceived as less complex, more objective, and performance-oriented compared to career training decisions, which might also impact fairness perception (Nagtegaal, 2021).

5.1 Academic and Practical Implications

This study holds academic and practical implications. Academically, it adds fresh insights to the literature on AI decision-making and algorithmic fairness. By examining distinct hybrid decision-making approaches and investigating employees' perceptions, it reinforces previous findings on the impact of hybrid decision-making compared to fully automated approaches (e.g., Kern et al., 2022; Nagtegaal, 2021), and highlights that the inclusion of a human element alone, as mentioned by the literature, may not always enhance fairness perception compared to fully automated decision-making (i.e., sequential human-AI decision-making). Instead, it confirms Jiang and colleagues' (2020) suggestion and reveals that the specific placement of humans in collaboration with AI may play a valuable role in shaping employees' fairness perception. This research reveals the most promising hybrid decision-making approach among those studied. It guides organizations that seek to use AI in decision-making while ensuring fairness and employee acceptance. Managers could minimize risks by adopting aggregated human-AI decision-making, which was perceived as the fairest regarding all OJ dimensions.

Moreover, this research contradicts some previous findings (e.g., Nagtegaal, 2021; Ötting & Maier, 2018) and showed that fairness perception did not depend on the studied decision contexts. Thus, the use of AI in this study's decision-making contexts should be treated with the same degree of caution by organizations.

Additionally, the qualitative analysis revealed employee concerns not only about AI but also about bias stemming from the human element in hybrid decision-making. This aligns with Bankin and colleagues' (2022) observations, who compared AI and human decision-making, and noted concerns regarding human bias in human decision-making. The thesis demonstrates that human bias might also be an issue in the context of human-AI collaboration. Once the human was integrated into the process, concerns about partial judgment, personal preferences, or favoritism influencing decisions and pre-selection were expressed. While prior research has

emphasized that AI cannot be a fair decision-maker due to replicated biases in training data (Kordzadeh & Ghasemaghaei, 2022; Roselli et al., 2019), this study highlights that biases introduced by the human component are of equal concern to employees.

Finally, current research on algorithmic fairness has been criticized for its heterogeneity and inconsistent use of theoretical concepts (Starke et al., 2022). Previous studies have primarily focused on individual dimensions of OJ theory, making it difficult to compare perceptions across dimensions. Therefore, this study focused on different dimensions to allow for a coherent analysis and to facilitate the comparison of perceptions across dimensions. It showed that covering different fairness dimensions seems worthwhile, as certain results differed depending on the dimension, such as procedural-core fairness.

5.2 Limitations and Future Research

This research has its limitations, which must be considered when interpreting its results. These limitations offer opportunities for future research. First, this study examined three specific hybrid decision-making approaches and three decision contexts, potentially limiting the generalizability of the findings to other hybrid models or decision contexts and domains. Hybrid Intelligence (HI) is an increasingly prominent subject among scientists and researchers (i.e., The HI Centre at VU Amsterdam). Future research could explore a broader range of hybrid decision-making approaches or examine alternative decision contexts for a comprehensive topic understanding.

Next, due to space constraints, this research did not include a comparison to human decision-making. Prior studies indicate that full AI decision-making is often perceived as less fair than human decision-making, while some suggest that hybrid decision-making is relatively fairer than full AI decision-making. However, these studies also suggest that hybrid decision-making does not surpass human decision-making regarding fairness (Kern et al., 2022; Nagtegaal, 2021). Future research should explore different human-AI collaborations compared to human decision-making, especially considering this study's findings on varying fairness perceptions among employees based on the specific way of human involvement in AI decision-making.

Third, this research used a non-probability sampling technique, potentially diminishing the generalizability of findings. Moreover, experimental designs with vignette studies are often criticized for prioritizing internal over external validity (Aguinis & Bradley, 2014). While this

method facilitated hypothesis testing in a controlled setting, the results' generalizability may be compromised as the presented scenarios may differ from real-world experiences and generate different responses in actual situations. As AI-driven decision-making becomes more prevalent, I recommend replicating this study with real-world experiments. Additionally, the technical feasibility of the processes used in this study for the organizational contexts has been neglected, warranting further research into their practical usefulness.

Lastly, limitations exist concerning this study's measurement instruments. The factor analysis of Colquitt's OJ scale revealed the presence of a new latent variable, procedural-voice. This has been found in a few other studies (i.e., Jespen, 2009, Maharee-Laweler, 2010), suggesting that it may be caused by the procedural fairness items' wording. Another perspective suggests that employees distinguish between the decision-making procedure's characteristics and their desire for active participation in the process. Given the dynamic and decentralized nature of modern workplaces, it can be assumed that employees today have higher expectations for participation in decision-making. However, this implies that the obtained results regarding the procedural dimensions cannot be directly compared to existing findings of procedural fairness in AI decision-making. Moreover, procedural-voice was influenced by the failure of the manipulation check, indicating the need for future research to verify results using more reliable samples.

5.3 Conclusion

As AI evolves, organizations must strike the right balance between human and AI capabilities when implementing it in decision-making. This balance is essential for establishing a fair, efficient, and supportive workplace. This research confirms that hybrid decision-making is more promising than relying solely on AI, considering employees' fairness perception. However, not all human-AI collaborations are equally successful in achieving this balance. As Robin Bordoli (2016) emphasized, the objective is not to replace but to augment humans and their abilities. To thrive in the future of work, organizations must consider diverse stakeholder perceptions and should carefully evaluate different hybrid decision-making options. By doing so, they can effectively leverage the unique strengths that may arise from human-machine partnerships.

References

- Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, 28(4), 399–416. <https://doi.org/10.1111/ijsa.12306>
- Agor, W. H. (1986). The Logic of Intuition: How Top Executives Make Important Decisions. *Organizational Dynamics*, 14(3), 5–18. [https://doi.org/https://doi.org/10.1016/0090-2616\(86\)90028-8](https://doi.org/https://doi.org/10.1016/0090-2616(86)90028-8)
- Aguinis, H., & Bradley, K. J. (2014). Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods*, 17(4), 351–371. <https://doi.org/10.1177/1094428114547952>
- Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., & Holzinger, A. (2022). Fairness and Explanation in AI-Informed Decision Making. *Machine Learning and Knowledge Extraction*, 4(2), 556–579. <https://doi.org/10.3390/make4020026>
- Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI and Society*, 35(3), 611–623. <https://doi.org/10.1007/s00146-019-00931-w>
- Bankins, S., Formosa, P., Griep, Y., & Richards, D. (2022). AI Decision Making with Dignity? Contrasting Workers' Justice Perceptions of Human and AI Decision Making in a Human Resource Management Context. *Information Systems Frontiers*, 24(3), 857–875. <https://doi.org/10.1007/s10796-021-10223-8>
- Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). “It’s reducing a human being to a percentage”; perceptions of justice in algorithmic decisions. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*. <https://doi.org/10.1145/3173574.3173951>
- Bland, J. M., & Altman, D. G. (1997). *Statistics notes: Cronbach's alpha*.
- Blodgett, J. G., Hill, D. J., & Tax, S. S. (1997). *The Effects of Distributive, Procedural, and Interactional Justice on Postcomplaint Behavior*. [https://doi.org/https://doi.org/10.1016/S0022-4359\(97\)90003-8](https://doi.org/https://doi.org/10.1016/S0022-4359(97)90003-8)
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Chanda, S. S., & Banerjee, D. N. (2022). Omission and commission errors underlying AI failures. *AI and Society*. <https://doi.org/10.1007/s00146-022-01585-x>
- Chui, M., Hall, B., Singla, A., Mayhew, H., & Sukharvesky, A. (2022). *The state of AI in 2022-and a half decade in review*. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review#/>

- Clay-Warner, J., Reynolds, J., & Roman, P. (2005). Organizational justice and job satisfaction: A test of three competing models. *Social Justice Research, 18*(4), 391–409. <https://doi.org/10.1007/s11211-005-8567-5>
- Cohen, J. (1992). QUANTITATIVE METHODS IN PSYCHOLOGY A Power Primer. *Psychological Bulletin, 112*(1), 155–159. <https://doi.org/10.1037//0033-2909.112.1.155>
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*(3), 386–400. <https://doi.org/10.1037/0021-9010.86.3.386>
- Colquitt, J. A., & Zipay, K. P. (2015). Justice, Fairness, and Employee Reactions. In *Annual Review of Organizational Psychology and Organizational Behavior* (Vol. 2, pp. 75–99). Annual Reviews Inc. <https://doi.org/10.1146/annurev-orgpsych-032414-111457>
- Conover, W. J., & Iman, R. L. (1982). Analysis of covariance using the rank transformation. *Biometrics, 38*, 715–724.
- Cropanzano, R., & Ambrose, M. L. (2015). Chapter 8: Measuring Justice and Fairness. In *The Oxford handbook of justice in the workplace* (pp. 187–202). Oxford University Press. <https://doi.org/https://doi.org/10.1093/oxfordhb/9780199981410.001.0001>
- Dastin, J. (2018, October 11). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- De Boer, E. M., Bakker, A. B., Syroit, J. E., & Schaufeli, W. B. (2002). Unfairness at work as a predictor of absenteeism. *Journal of Organizational Behavior, 23*(2), 181–197. <https://doi.org/10.1002/job.135>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use welch’s t-Test instead of student’s t-Test. *International Review of Social Psychology, 30*(1), 92–101. <https://doi.org/10.5334/irsp.82>
- Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business and Information Systems Engineering, 61*(5), 637–643. <https://doi.org/10.1007/s12599-019-00595-2>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- European Commission. (2019). *HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE SET UP BY THE EUROPEAN COMMISSION ETHICS GUIDELINES*

FOR TRUSTWORTHY AI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

- European Union Parliament. (2020). *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. In *Behavior Research Methods* (Vol. 39, pp. 175–191).
<https://doi.org/https://doi.org/10.3758/BF03193146>
- Field, A. P. (2009). *Discovering statistics using SPSS : (and sex and drugs and rock “n” roll)*. SAGE Publications.
- Greenberg, J. (1990). Organizational Justice: Yesterday, Today, and Tomorrow. *Journal of Management*, 16(2), 399–432. <https://doi.org/10.1177/014920639001600208>
- Greenberg, J., & Tyler, R. (1987). Why Procedural Justice in Organizations? *Social Justice Research*, 1, 127–142. <https://doi.org/https://doi.org/10.1007/BF01048012>
- Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior*, 119. <https://doi.org/10.1016/j.chb.2021.106730>
- 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>
- Hemmer, P., Schemmer, M., Vössing, M., & Kühl, N. (2021). Human-AI Complementarity in Hybrid Intelligence Systems: A Structured Literature Review. *Conference: Pacific Asia Conference on Information Systems (PACIS)*.
https://www.researchgate.net/publication/352882174_Human-AI_Complementarity_in_Hybrid_Intelligence_Systems_A_Structured_Literature_Review
- Höddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, 116. <https://doi.org/10.1016/j.chb.2020.106635>
- Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, · Klaus-Robert, & Samek, W. (2020). *xxAI-Beyond Explainable AI*. <https://link.springer.com/bookseries/1244>
- Hom, P. W., Griffeth, R. W., & Sellaro, C. L. (1984). The Validity of Mobley’s (1977) Model of Employee Turnover. *ORGANIZATIONAL BEHAVIOR AND HUMAN PERFORMANCE*, 34, 141–174. [https://doi.org/https://doi.org/10.1016/0030-5073\(84\)90001-1](https://doi.org/https://doi.org/10.1016/0030-5073(84)90001-1)

- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586.
<https://doi.org/10.1016/j.bushor.2018.03.007>
- Jiang, L., Qin, X., Dong, X., Chen, C., & Liao, W. (2022). How and When AI-Human Order Influences Procedural Justice in a Multistage Decision-Making Process. *Academy of Management Proceedings*, 2022(1).
<https://doi.org/https://doi.org/10.5465/AMBPP.2022.29>
- Jussupow, E., & Benbasat, I. (2020, June 15). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *Proceedings of the 28th European Conference on Information Systems (ECIS)*.
<https://www.researchgate.net/publication/344401293>
- Kaibel, C., Koch-Bayram, I., Biemann, T., & Mühlenbock, M. (2019). Applicant Perceptions of Hiring Algorithms - Uniqueness and Discrimination Experience as Moderators. *Academy of Management Proceedings*, 2019(1), 18172.
<https://doi.org/0.5465/AMBPP.2019.210>
- Kern, C., Gerdon, F., Bach, R. L., Keusch, F., & Kreuter, F. (2022). Humans versus machines: Who is perceived to decide fairer? Experimental evidence on attitudes toward automated decision-making. *Patterns*, 3(10).
<https://doi.org/10.1016/j.patter.2022.100591>
- Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795–848.
<https://doi.org/10.1007/s40685-020-00134-w>
- Konradt, U., Oldeweme, M., Krys, S., & Otte, K. P. (2020). A meta-analysis of change in applicants' perceptions of fairness. *International Journal of Selection and Assessment*, 28(4), 365–382. <https://doi.org/10.1111/ijsa.12305>
- Kordzadeh, N., & Ghasemaghahi, M. (2022). Algorithmic bias: review, synthesis, and future research directions. In *European Journal of Information Systems* (Vol. 31, Issue 3, pp. 388–409). Taylor and Francis Ltd. <https://doi.org/10.1080/0960085X.2021.1927212>
- Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, 27(3), 217–234. <https://doi.org/10.1111/ijsa.12246>
- Langer, M., & 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.
<https://doi.org/10.1016/j.chb.2021.106878>

- Lee, J. D., & See, K. A. (2009). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data and Society*, 5(1). <https://doi.org/10.1177/2053951718756684>
- Lee, M. K., Jain, A., Cha, H. J. I. N., Ojha, S., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW). <https://doi.org/10.1145/3359284>
- Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. *Conference on Human Factors in Computing Systems - Proceedings, 2015-April*, 1603–1612. <https://doi.org/10.1145/2702123.2702548>
- Leventhal, G. S. (1976). What Should Be Done with Equity Theory? New Approaches to the Study of Fairness in Social Relationships. In *Social Exchange: Advances in Theory and Research* (pp. 27–55). John Wiley.
- Lichtenthaler, U. (2020). Extremes of acceptance: employee attitudes toward artificial intelligence. *Journal of Business Strategy*, 41(5), 39–45. <https://doi.org/10.1108/JBS-12-2018-0204>
- Lind, E. A., Tyler, T. R., Huo, Y. J., Tyler, T. R., & Lind, E. A. (1997). Procedural Context and Culture: Variation in the Antecedents of Procedural Justice Judgments. *Journal of Personality and Social Psychology*, 73(4), 767–780. <https://doi.org/10.1037/0022-3514.73.4.767>
- Luo, X., Qin, M. S., Fang, Z., & Qu, Z. (2021). Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions. *Journal of Marketing*, 85(2), 14–32. <https://doi.org/10.1177/0022242920956676>
- Malhotra, N. K., Nunan, D., & Birks, D. F. (2017). *MARKETING RESEARCH: AN APPLIED APPROACH* (5th ed.). Pearson .
- Marr, B. (2018, December 14). *The Amazing Ways How Unilever Uses Artificial Intelligence To Recruit & Train Thousands Of Employees*. Forbes. <https://www.forbes.com/sites/bernardmarr/2018/12/14/the-amazing-ways-how-unilever-uses-artificial-intelligence-to-recruit-train-thousands-of-employees/?sh=f960d836274d>
- Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, 2(2). <https://doi.org/10.1145/1985347.1985353>

- Medrano-Berumen, C., & Akbaş, M. İ. (2021). Validation of decision-making in artificial intelligence-based autonomous vehicles. *Journal of Information and Telecommunication*, 5(1), 83–103. <https://doi.org/10.1080/24751839.2020.1824154>
- Methnani, L., Aler Tubella, A., Dignum, V., & Theodorou, A. (2021). Let Me Take Over: Variable Autonomy for Meaningful Human Control. *Frontiers in Artificial Intelligence*, 4. <https://doi.org/10.3389/frai.2021.737072>
- Millington, B., & Millington, R. (2015). “The datafication of everything”: Toward a sociology of sport and big data. *Sociology of Sport Journal*, 32(2), 140–160. <https://doi.org/10.1123/ssj.2014-0069>
- Morse, L., Teodorescu, M. H. M., Awwad, Y., & Kane, G. C. (2022). Do the Ends Justify the Means? Variation in the Distributive and Procedural Fairness of Machine Learning Algorithms. *Journal of Business Ethics*, 181(4), 1083–1095. <https://doi.org/10.1007/s10551-021-04939-5>
- Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees’ procedural justice. *Government Information Quarterly*, 38(1). <https://doi.org/10.1016/j.giq.2020.101536>
- 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>
- Oksana, M., Kotsipak, M., Dolgikh, S., Bilak, Y., Radivilova, T., & Baranovskyi, O. (2022). Collaborative Human-AI Decision-Making Systems with Numerical Channels. *2022 12th International Conference on Advanced Computer Information Technologies, ACIT 2022*, 5–8. <https://doi.org/10.1109/ACIT54803.2022.9913201>
- Ötting, S. K., & Maier, G. W. (2018). The importance of procedural justice in Human–Machine Interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior*, 89, 27–39. <https://doi.org/10.1016/j.chb.2018.07.022>
- Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (2023). Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. In *Big Data and Cognitive Computing* (Vol. 7, Issue 1). MDPI. <https://doi.org/10.3390/bdcc7010015>
- Pallathadka, H., Ramirez-Asis, E. H., Loli-Poma, T. P., Kaliyaperumal, K., Ventayen, R. J. M., & Naved, M. (2021). Applications of artificial intelligence in business management, e-commerce and finance. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2021.06.419>

- Perifanis, N. A., & Kitsios, F. (2023). Investigating the Influence of Artificial Intelligence on Business Value in the Digital Era of Strategy: A Literature Review. In *Information (Switzerland)* (Vol. 14, Issue 2). MDPI. <https://doi.org/10.3390/info14020085>
- Pierson, E. (2017). Demographics and discussion influence views on algorithmic fairness. *ArXiv: Computers and Society. Computer Science*. <https://doi.org/https://doi.org/10.48550/arXiv.1712.09124>
- Renjith, S., Sreekumar, A., & Jathavedan, M. (2020). An extensive study on the evolution of context-aware personalized travel recommender systems. *Information Processing and Management*, 57(1). <https://doi.org/10.1016/j.ipm.2019.102078>
- Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. In *Educational Research Review* (Vol. 6, Issue 2, pp. 135–147). <https://doi.org/10.1016/j.edurev.2010.12.001>
- Roselli, D., Matthews, J., & Talagala, N. (2019). Managing bias in AI. *The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019*, 539–544. <https://doi.org/10.1145/3308560.3317590>
- Sanderson, C., Lu, Q., Douglas, D., Xu, X., Zhu, L., & Whittle, J. (2022). Towards Implementing Responsible AI. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 5076–5081. <https://doi.org/10.1109/BigData55660.2022.10021121>
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3). <https://doi.org/10.1007/s42979-021-00592-x>
- Schoeffer, J., Machowski, Y., & Kuehl, N. (2022, September 13). Perceptions of Fairness and Trustworthiness Based on Explanations in Human vs. Automated Decision-Making. *Hawaii International Conference on System Sciences 2022*. <https://doi.org/https://doi.org/10.48550/arXiv.2109.05792>
- Schwemer, S. (2019, November 9). Humans-in-the-"Loop" in EU. *XXXVI Nordic Conference*. https://www.jus.uio.no/ifp/om/organisasjon/seri/arrangementer/2021/schwemer_2021_os_lo_nordic_conference%281%29.pdf
- Shaheen, Y. M. (2021). Applications of Artificial Intelligence (AI) in healthcare: A review. *ScienceOpen Preprints*. <https://doi.org/10.14293/S2199-1006.1.SOR-.PPVRY8K.v1>
- Shin, D. (2020). User Perceptions of Algorithmic Decisions in the Personalized AI System: Perceptual Evaluation of Fairness, Accountability, Transparency, and Explainability. *Journal of Broadcasting and Electronic Media*, 64(4), 541–565. <https://doi.org/10.1080/08838151.2020.1843357>
- Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review*. <https://doi.org/10.1177/0008125619862257>

- Shulner-Tal, A., Kuflik, T., & Kliger, D. (2022). Enhancing Fairness Perception—Towards Human-Centred AI and Personalized Explanations Understanding the Factors Influencing Laypeople’s Fairness Perceptions of Algorithmic Decisions. *International Journal of Human-Computer Interaction*.
<https://doi.org/10.1080/10447318.2022.2095705>
- Snieszek, J. A., & Buckley, T. (1995). Cueing and Cognitive Conflict in Judge-Advisor Decision Making. *Organizational Behavior and Human Decision Processes*, 62(2), 159–174. <https://doi.org/10.1006/obhd.1995.1040>
- Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022a). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data and Society*, 9(2). <https://doi.org/10.1177/20539517221115189>
- Tolan, S. (2018). *Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges*. <https://doi.org/10.48550/arXiv.1901.04730>
- Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 42(9), 1600–1631. <https://doi.org/10.1002/smj.3322>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. In *New Series* (Vol. 185, Issue 4157).
- UNIDIR. (2014). *The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward*.
<https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-considering-how-meaningful-human>
- Wamba-Taguimdje, S. L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893–1924. <https://doi.org/10.1108/BPMJ-10-2019-0411>
- Wang, R., Harper, F. M., & Zhu, H. (2022). *Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences*. <https://doi.org/10.1145/313831.3376813>
- Wang, X., Zhang, Y., & Zhu, R. (2022). A brief review on algorithmic fairness. *Management System Engineering*, 1(1). <https://doi.org/10.1007/s44176-022-00006-z>
- Wobbrock, J. O., Findlater, L., & Higgins, J. J. (2011). The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)*, 143–146.
<https://doi.org/10.1145/1978942.1978963>

Zhu, Y. Q., Corbett, J. ueline, & Chiu, Y. Te. (2021). Understanding employees' responses to artificial intelligence. *Organizational Dynamics*, 50(2).

<https://doi.org/10.1016/j.orgdyn.2020.100786>

Zoghbi-Manrique-de-Lara, P., & Suárez-Acosta, M. A. (2014). Employees' reactions to peers' unfair treatment by supervisors: The role of ethical leadership. *Journal of Business Ethics*, 122(4), 537–549.

<https://doi.org/10.1007/s10551-013-1778-z>

Appendices

Appendix A: Qualtrics Survey

Informed Consent Dear Participant, thank you for considering participating in this experimental study. This study is part of my master's Thesis in International Management with Specialization in Strategy and Consulting at Universidade Católica Portuguesa.

This experimental study involves reading a scenario and answering a few questions, which should take no longer than 8-10 minutes. Since all questions are based on what you personally think or believe, there is no right or wrong answer. Your participation is voluntary, and you may withdraw at any time without any consequences. All data collected will be treated anonymously and is used only for research purposes. If you have any questions or are interested in the results of this study, please feel free to contact me at s-agescher@ucp.pt.

By consenting you confirm that you have read and understood the above information and that you agree to participate in this study. Thank you, your contribution is greatly appreciated!

Best, Anna Gescher

- I consent, begin the study (1)
 - I do not consent, I do not wish to participate (2)
-

Q2 Thank you for taking part in my study! First, please indicate your level of agreement with the following statements (1 = strongly disagree to 5 = strongly agree).

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I am familiar with the term "Artificial Intelligence" (AI)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

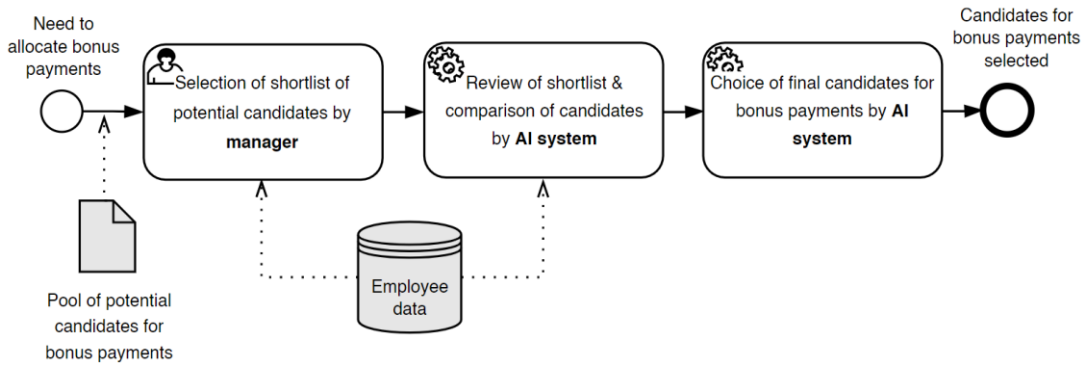
Q3 The term "AI" used in this study's context refers to computer systems that can perform tasks that normally need human intelligence, such as recognizing images, understanding language, or making decisions. For this, the method of Machine Learning (ML) is leveraged. ML uses algorithms to enable computers to learn from data and improve over time. In decision-making, ML algorithms can analyze large amounts of data to identify patterns and relationships. This enhances decision-making by learning from historical data and forecasting future events. ML-based AI systems may help people make better decisions by learning from data. AI and ML can increase the accuracy and efficiency of decision-making.

Q4 On the next page, you will find a description of a situation that could occur in everyday work life. Please **read the description carefully and imagine yourself in that situation**. Later, you will be asked questions about the situation.

Q5 Imagine you're an employee at BrightPath, a fast-growing technology company. You've been with the company for several years. Recently, **BrightPath announced that they're allocating bonus payments** to deserving employees, and you're excited about the opportunity to receive a bonus.

BrightPath recognizes the importance of using both subjective and objective data to make a well-informed decision about the final candidates. For this, BrightPath has been gathering data on employees' qualifications, experiences, and contributions to the company. The following decision-making procedure is applied.

Your manager is responsible for reviewing the gathered data and selecting a shortlist of potential candidates for the bonus payments. Your manager will be using **the data at hand, expertise, and knowledge** of the team members to identify the most promising candidates from the pool of potential candidates. After the initial selection, the shortlist will be sent to an **AI system**. **The AI system** will carefully compare and analyze the selected candidates based on various criteria (e.g., past performance, alignment with the company's strategic goals and values). Based on this evaluation, **the AI system** will choose the employees who will receive the bonus payment. * Condition: Bonus Payment x Sequential Human-AI. See Appendix B for the remaining vignettes.



Q6 Please answer the question to the best of your ability, under careful consideration of the information presented above.

	(1) Very unfair	(2) Moderately unfair	(3) Slightly unfair	(4) Neither fair nor unfair	(5) Slightly fair	(6) Moderately fair	(7) Very fair
How fair is the decision-making process in this context in your opinion? (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q7 Please provide a short explanation for your answer. (You can answer in German or English)

Q8 Read the presented scenario **again** to ensure that you understood all the information **before** proceeding to the next page. You will not be able to go back.

Q53 **Manipulation Check** How was the decision in the presented scenario made?

- First AI system selects a shortlist, then manager reviews this list and makes final decision.
 - First manager selects a shortlist, then AI system reviews this list and makes final decision.
 - Manager and AI system simultaneously analyze different data according to their capabilities/strengths.
 - The AI system decided without any human intervention.
-

Q54 **Manipulation Check** Please select the decision context that was subject in the scenario presented to you.

- Promotion to new team lead
 - Allocation of bonus payments
 - Allocation of career training
-

Q55 Thank you for the responses! Now we are turning to a set of questions on **your perception** of the scenario presented to you. Please keep in mind the details of the presented situation when answering the questions. As a reminder, you are an employee at BrightPath who is directly affected by the decision.

Q56: **OJ Scale (Hybrid Condition)** Following items relate to the **decision-making procedure** used to arrive at the final decision. In your opinion, to what extent:

	To a small extent (1)	(2)	(3)	(4)	To a large extent (5)
Will you be able to express your views and feelings during this decision-making procedure?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will you have influence over the decision made with this procedure?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will this procedure be applied consistently?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will this procedure be free of bias?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will this procedure be based on accurate information?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will you be able to appeal (oppose, veto) the decision under the given decision-making procedure?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will this decision-making procedure uphold ethical and moral standards?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q57 Following items refer to the decision-making procedure, **specifically considering the decision-making actor(s)** within the procedure. In your opinion, to what extent:

	To a small extent (1)	(2)	(3)	(4)	To a large extent (5)
Will the decision-making procedure make you feel treated in a polite manner?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the decision-making procedure make you feel treated with dignity?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the decision-making procedure make you feel treated with respect?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the decision-making procedure refrain from improper remarks or comments?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q58 Following items refer to the decision-making procedure, **specifically considering the decision-making actor(s)** within the procedure. **Imagine a decision was made**, in your opinion, to what extent:

	To a small extent (1)	(2)	(3)	(4)	To a large extent (5)
Will your manager and the AI system be candid (honest, open) in their communications with you?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will your manager and the AI system be able to explain the procedures thoroughly?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the explanations regarding the procedures be reasonable?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will your manager and the AI system communicate details in a timely manner?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will your manager and the AI system tailor their communications to your specific needs?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q60 **OJ Scale (AI Condition)** Following items relate to the **decision-making procedure** used to arrive at the final decision. In your opinion, to what extent:

	To a small extent (1)	(2)	(3)	(4)	To a large extent (5)
Will you be able to express your views and feelings during this decision-making procedure?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will you have influence over the decision made with this procedure?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will this procedure be applied consistently?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will this procedure be free of bias?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will this procedure be based on accurate information?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will you be able to appeal (oppose, veto) the decision under the given decision-making procedure?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will this decision-making procedure uphold ethical and moral standards?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q61 Following items refer to the decision-making procedure, **specifically considering the decision-making actor(s)** within the procedure. In your opinion, to what extent:

	To a small extent (1)	(2)	(3)	(4)	To a large extent (5)
Will the decision-making procedure make you feel treated in a polite manner?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the decision-making procedure make you feel treated with dignity?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the decision-making procedure make you feel treated with respect?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the decision-making procedure refrain from improper remarks or comments?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q62 Following items refer to the decision-making procedure, **specifically considering the decision-making actor(s)** within the procedure. Imagine a decision was made, in your opinion, to what extent:

	To a small extent (1)	(2)	(3)	(4)	To a large extent (5)
Will the AI system be candid (honest, open) in the communications with you?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the AI system be able to explain the procedures thoroughly?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the explanations regarding the procedures be reasonable?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the AI system communicate details in a timely manner?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will the AI system tailor their communications to your specific needs?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q63 **Covariate Questions** You are almost done. The next set of questions are related to your trust in technology, automated systems, and people.

Q64 Please indicate your level of agreement with the following statements:

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I believe that most technologies are effective at what they are designed to do.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A large majority of technologies are excellent.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Most technologies have the functions required for their respective fields.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think most technologies enable me to do what I need to do.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My typical approach is to trust new technologies until they prove to me that I shouldn't trust them.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I usually trust a technology until it gives me a reason not to trust it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I generally give a technology the benefit of the doubt when I first use it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q65 Please indicate your level of agreement with the following statements:

** Automated systems are a combination of both software and hardware that are designed and programmed to work automatically with only little or no need for a human operator to provide inputs and instructions. Examples of automated systems are robots, self-driving vehicles, or machine learning algorithms.*

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
One should be careful with unfamiliar automated systems.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I rather trust a system than I mistrust it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Automated systems generally work well.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q66 Please indicate your level of agreement with the following statements:

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
One should be very cautious with strangers.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Most experts tell the truth about the limits of their knowledge.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Most people can be counted on to do what they say they will do.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
These days, you must be alert or someone is likely to take advantage of you.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have never used a computer. Attention Check	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Most repair people will not overcharge people who are ignorant of their specialty.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Most people answer public opinion polls honestly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Most adults are competent at their jobs.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Most salespeople are honest in describing their products.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q67: **Demographic Questions** Finally, the last set of questions! Please answer the following demographic questions.

Q68 How old are you?

Q69 Gender: How do you identify?

- Male
 - Female
 - Other (please specify) _____
 - Prefer not to say
-

Q70 Where do you come from?

▼ Afghanistan (1) ... Prefer not to say (194)

Q71 What is your highest level of education?

- High school graduate
 - Bachelor's degree
 - Master's degree
 - Ph.D.
 - Other (please specify) _____
 - Prefer not to say
-

Q72 What is your current occupation?

- Employed full time
 - Employed part time (e.g., working student)
 - Student
 - Unemployed
 - Retired
 - Prefer not to say
-

Q73 In which sector do you currently work?

▼ Consultancy or management (1) ... Prefer not to say (28)

Q74 Please answer the following statement (1 = Extremely difficult to 5 = Extremely easy).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
How difficult was it for you to imagine the scenario and answer the relating questions?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Thank you for taking the time to complete this survey. Your response has been recorded and will be incredibly valuable to my research. I really appreciate your contribution.

Appendix B: Vignettes per Decision Type and Context

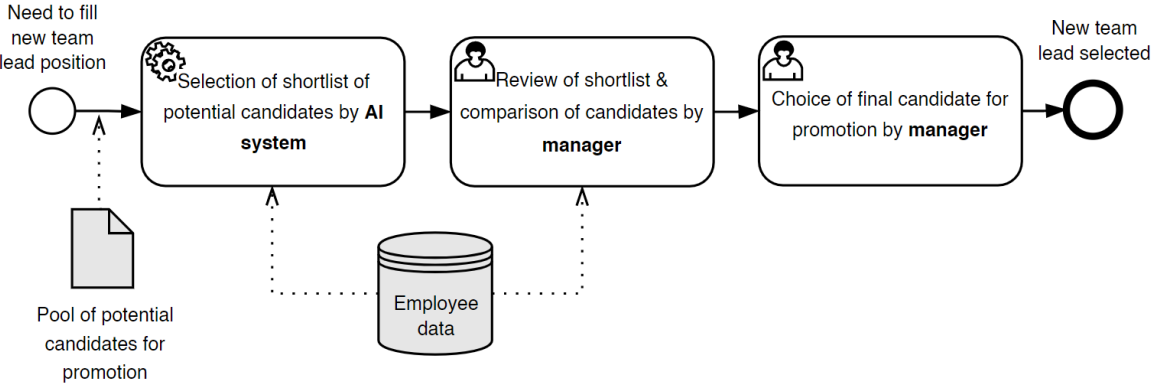
In the following, four vignettes will be presented, to show the structure of the description of the decision-procedure as well as the decision context. Sequential Human-AI has already been presented in the previous Appendix.

Promotion Decision Context x Sequential AI-Human

Imagine you're an employee at BrightPath, a fast-growing technology company. You've been with the company for several years. Recently, BrightPath announced that **they're looking to fill a new team lead position**, and you're excited about the opportunity to apply for the role.

BrightPath recognizes the importance of using both subjective and objective data to make a well-informed decision about the final candidate. For this, BrightPath has been gathering data on employees' qualifications, experiences, and **potential to lead effectively**. The following decision-making procedure is applied.

An **AI system** is responsible for reviewing the gathered data and selecting a shortlist of potential candidates for the new team lead position. The **AI system** will use its data processing capability to identify the most promising candidates from the pool of candidates based on various criteria (e.g., past performance, alignment with the company's strategic goals and values). After the initial selection, the shortlist will be sent to **your manager**. **Your manager** will carefully compare and analyze the selected candidates using the data at hand, expertise, and knowledge of the team member. Based on this evaluation, **your manager** will choose the new team lead.



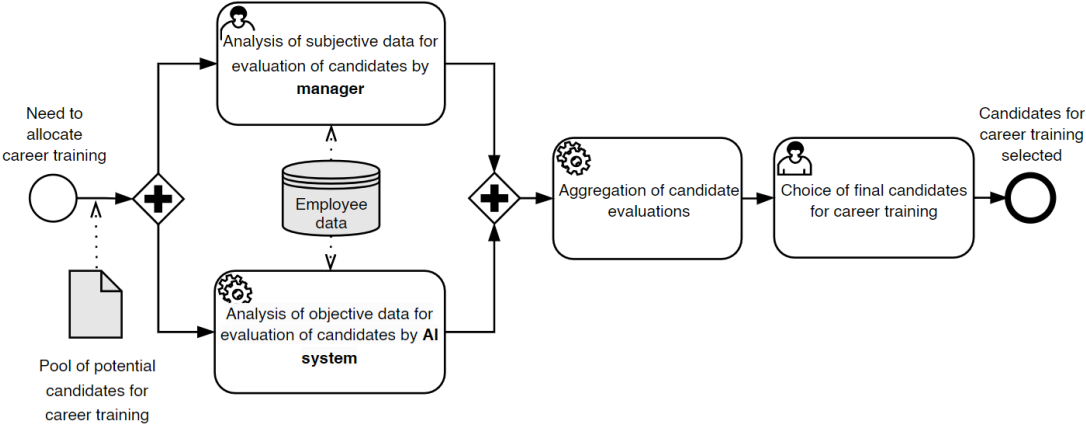
Career Training Decision Context x Aggregated Human-AI

Imagine you're an employee at BrightPath, a fast-growing technology company. You've been with the company for several years. Recently, BrightPath announced that **they're allocating a**

career training opportunity to deserving employees, and you're excited about the opportunity to apply for the training program.

BrightPath recognizes the importance of using both subjective and objective data to make a well-informed decision about the final candidates. For this, BrightPath has been gathering data on employees' qualifications, experiences, and **potential to benefit from the training program**. The following decision-making procedure is applied.

Your manager and an AI system are cooperatively reviewing and analyzing the gathered data according to their respective capabilities/strengths and selecting the candidate for the career training program. The **AI system** will analyze the **objective data points** (e.g., past performance), while **your manager** will analyze the **subjective data points** (e.g., alignment with the company's strategic goals and values). Based on the combination of their evaluations, the employees who will receive the career training opportunity will be chosen.

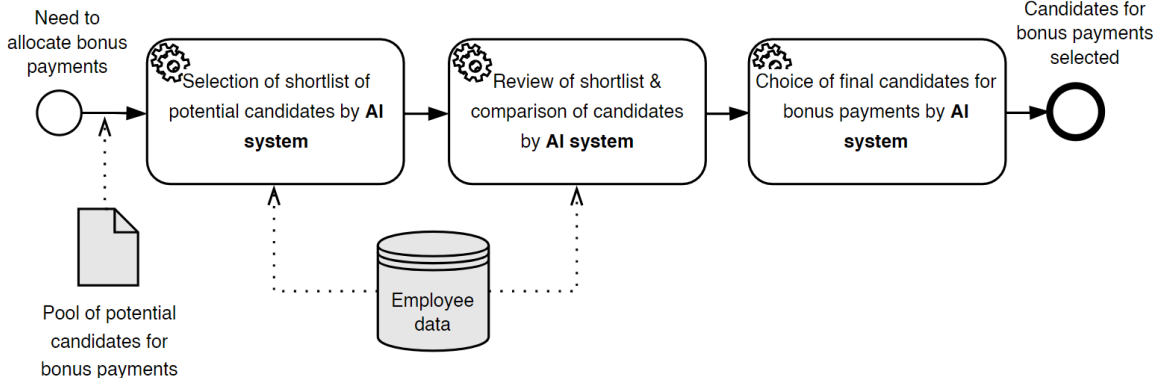


Bonus payment Decision Context x Full AI

Imagine you're an employee at BrightPath, a fast-growing technology company. You've been with the company for several years. Recently, BrightPath announced that **they're allocating bonus payments to deserving employees**, and you're excited about the opportunity to receive a bonus.

BrightPath recognizes the importance of using both subjective and objective data to make a well-informed decision about the final candidates. For this, BrightPath has been gathering data on employees' qualifications, experiences, and **contributions to the company**. The following decision-making procedure is applied.

An **AI system** is responsible for reviewing the gathered data and selecting a shortlist of potential candidates for the bonus payments. The **AI system** will identify the most promising candidates from the pool of candidates based on various criteria (e.g., past performance, alignment with the company’s strategic goals and values). After the initial selection, the **AI system** will carefully compare and analyze the selected candidates on the shortlist using the data at hand, expertise, and knowledge of the team members. Based on this evaluation, the **AI system** will choose the employees who will receive the bonus payment.



Appendix C: Alternatives Factor Analysis

Due to the split of procedural fairness and the incorrectly loading items in the initial factor analysis, I explored alternative approaches for a better fitting model, see Table C1. First, I performed a principal component analysis (varimax), extracting only three components. The KMO measure resulted in a great value of .89, with all item KMO values exceeding .81. Bartlett's test of sphericity confirmed the suitability of the data for the analysis, $\chi^2(120) = 5099.72$, $p < .001$. However, the three components with Eigenvalues above Kaiser's criterion of 1 only accounted for 55.69% of the variance and five items loaded incorrectly. Thus, principal axis factoring following Maharee-Laweler and colleagues (2010) who also experienced the split of the procedural fairness dimension across two components was performed. Using oblimin rotation, it yielded the same component structure and item loading pattern as the initial approach presented in Chapter 4.2, explaining 62.16% of the variance.

Table C1. *Factor Loading of OJ Items*

Items	3-factor			4-factor			
	1	2	3	1	2	3	4
Procedural Fairness							
1	<u>0.62</u>	0.36	0.22	0.14	0.13	<u>-0.33</u>	0.32
2	0.45	<u>0.50</u>	0.02	0.22	0.01	-0.12	<u>0.42</u>
3	0.37	<u>0.45</u>	-0.10	0.03	0.08	0.05	<u>0.62</u>
4	0.41	<u>0.44</u>	-0.16	0.07	-0.03	-0.06	<u>0.52</u>
5	0.18	0.09	<u>0.78</u>	-0.06	0.84	0.00	0.10
6	0.09	0.06	<u>0.75</u>	-0.03	0.73	0.02	0.04
7	0.08	0.05	<u>0.74</u>	0.08	0.57	-0.06	-0.10
Interpersonal Fairness							
8	0.83	0.20	0.28	0.03	-0.02	<u>-0.88</u>	0.03
9	0.83	0.15	0.29	-0.03	0.02	<u>-0.88</u>	0.03
10	0.80	0.13	0.29	0.01	0.02	<u>-0.81</u>	0.00
11	0.46	0.28	-0.15	-0.02	-0.02	-0.13	<u>0.42</u>
Informational Fairness							
12	0.22	0.71	0.18	0.57	0.02	0.00	0.20
13	0.14	0.69	0.26	0.75	-0.03	-0.01	-0.04
14	0.09	0.69	0.08	0.46	-0.01	0.05	0.19
15	0.29	0.64	0.28	0.66	-0.01	-0.14	0.02
16	0.12	0.42	<u>0.59</u>	0.50	0.28	-0.11	-0.18

Note. Underlined values indicate the highest loading for each item and bold values indicate where the items should have loaded.

Appendix D: Demographic Analysis

Pearson correlation coefficients examined relationships between participants' overall fairness perception and demographic factors, including age, gender, origin, educational background, employee status, and the industry sector they worked in. To prevent the correlation table from becoming too large, it was divided into individual tables. Table D1 shows a significant negative correlation between overall fairness perception and age, with fairness perception decreasing as age increases, $r(794) = -.11, p = .002$. Table D2 shows a negative correlation between overall fairness perception and participants with a Ph.D., $r(795) = -.09, p < .05$.

Table D1. *Correlation - Participants' Age, Gender, Origin and Fairness Perception*

	1	2	3	4	5	6
1 Overall Fairness						
2 Age	-.11**					
3 Gender	-.04	.00				
<u>Origin</u>						
4 America	-.01	-.02	-.06			
5 Asia	-.01	.00	-.01	-.01		
6 Africa	.06	-.04	.09*	-.02	-.01	
7 Europe	-.04	.04	-.02	-.38**	-.32**	-.41**

Note. * Correlation is significant at .05 level; ** Correlation is significant at .01 level

Table D2. *Correlation - Participants' Education, Employment Status Fairness Perception*

	1	2	3	4	5	6	7	8	9
1 Overall Fairness									
<u>Educational Background</u>									
2 Highschool	-.03								
3 Bachelor's	.06	-.53**							
4 Master's	.01	-.28**	-.54**						
5 PhD	-.09*	-.06	-.12**	-.07					
<u>Employment Status</u>									
6 Full-Time Employed	-.02	-.14**	-.08*	.20**	.05				
7 Part-Time Employed	-.02	-.03	.04	-.01	.00	-.43**			
8 Student	0.04	.15**	.11**	-.19**	-.08*	-.53**	-.38**		
9 Unemployed	0.01	.08*	-.09*	0.04	-0.02	-.14**	-.10**	-.13**	
10 Retired	0.02	0.01	-.07*	-0.03	0.06	-.11**	-.08*	-.10**	-0.03

Note. * Correlation significant at .05 level; ** Correlation significant at .01 level

Also, considering participants' lines of work, significant negative correlations were only found between overall fairness perception and the science/pharmaceutical sector, $r(795) = -.10, p < .01$, as well as the environment/agriculture industry sector, $r(795) = -.08, p < .05$.

Appendix E: Correlation among Variables

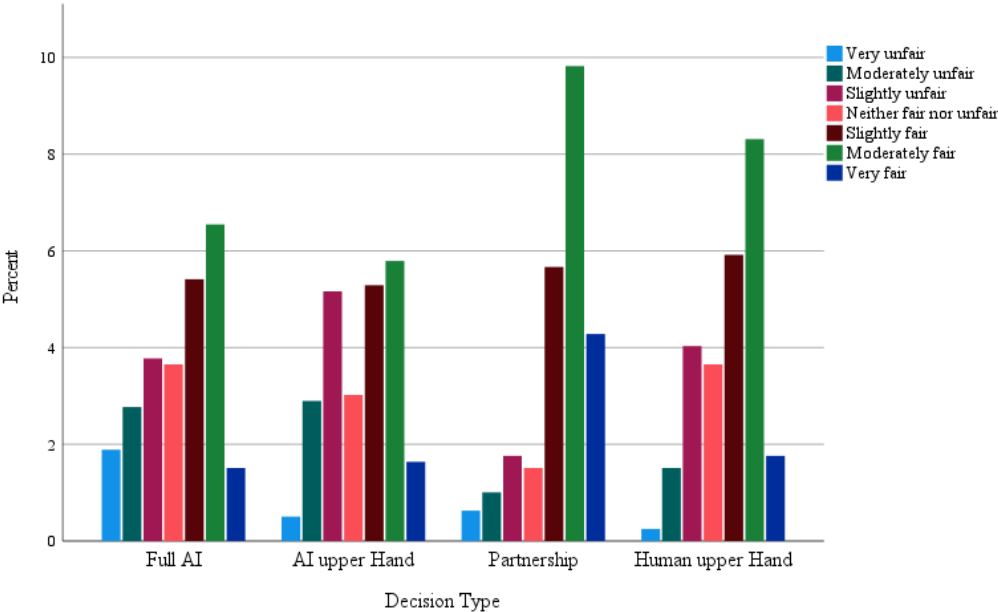
	1	2	3	4	5	6	7	8	9	10
1 Overall Fairness	-									
2 Interpersonal	.82**	-								
3 Informational	.84**	.54**	-							
4 Procedural-Voice	.59**	.33**	.37**	-						
5 Procedural-Core	.78**	.60**	.55**	.23**	-					
6 Faith in GT	.40**	.35**	.38**	.10**	.37**	-				
7 Trust Stance GT	.38**	.33**	.31**	.17**	.33**	.51**	-			
8 Trust Automation	.42**	.38**	.35**	.17**	.34**	.46**	.52**	-		
9 Trust People	.12**	.14**	.07	.02	.13**	.05	.12**	.20**	-	
10 Familiarity AI	.04	.06	.05	.01	.00	.17**	.05	.05	-.11**	-
11 Difficulty	.23**	.19**	.21**	.07*	.22**	.24**	.14**	.14**	.04	.14**

Note. $N = 797$. ** $p < .01$. To accommodate the limited table space, abbreviated variable names were used (e.g., GT = General Technology).

Appendix F: Quantitative Analysis of Participants’ General Fairness Perception

Figure F1 shows the frequency of participants’ fairness ratings across the different decision-making types. I performed Mann-Whitney U tests to evaluate whether general fairness differed by decision type, because the Shapiro-Wilk test yielded significant results for fairness across all groups ($p < .001$), indicating that the data for general fairness was not normally distributed. I compared full AI with each of the hybrid conditions and the hybrid decision-making conditions with each other. The results indicate that there was no significant difference between the general fairness score of the full AI ($M = 4.32, SD = 1.71$) and the sequential human-AI decision type ($M = 4.38, SD = 1.59$), $Z = - 0.190, p = .849$. However, the full AI decision type had a significantly lower general fairness score than the aggregated human-AI decision type ($M = 5.32, SD = 1.46$), $Z = - 6.158, p < .001$, and the sequential AI-human decision type ($M = 4.79, SD = 1.43$), $Z = - 2.623, p = .009$. Moreover, the sequential human-AI decision type had a lower general fairness score than the aggregated human-AI decision type, $Z = - 6.029, p < .001$, as well as the sequential AI-human decision type, $Z = - 2.544, p = .011$. Finally, the aggregated human-AI decision type had a significantly greater general fairness score than sequential AI-human, $Z = -4.119, p < .001$.

Figure F1. *Participants’ Fairness Scores According to Decision-Making Type*



Note. Full AI received the highest proportion of ratings categorized as “very unfair”, aggregated human-AI (=partnership) the largest proportion of ratings categorized as “moderately” and “very fair”.

Appendix G: Themes and Illustrative Quotes Qualitative Analysis

Table G1. *Sentiment Categorization*

	Full AI			Sequential Human-AI			Sequential AI-Human			Aggregated Human-AI		
	P	B	T	P	B	T	P	B	T	P	B	T
Positive	16	21	14	15	20	15	25	29	25	34	41	38
Mixed	20	15	19	17	8	10	13	14	11	12	10	4
Negative	25	19	28	26	21	26	19	13	16	7	8	8
Neutral	4	3	1	2	1	5	2	1	6	1	0	4
Total	65	58	62	60	50	56	59	57	58	54	59	54

Note. Responses were categorized based on sentiment before classifying them into minor and major themes. Mixed sentiments occurred when participants provided contradictory statements. Aggregated Human-AI decision-making received the highest number of positive and the fewest negative responses. Conversely, full AI decision-making elicited the most mixed responses, with participants acknowledging some advantages of AI while also expressing concerns about aspects they disliked. P = Promotion, B = Bonus Payment, T = Career Training.

Table G2. *Major and Minor Themes with related Frequency (%) per Decision Type*

Theme	1	2	3	4
Positive				
No (or reduces) human prejudice or preferences	6%	3%	6%	2%
Objective, neutral	10%	9%	7%	3%
Performance-oriented, data-driven, or fact-based	8%	2%	3%	2%
Good that AI Manager makes first selection	0%	5%	0%	0%
Equal chances, standardization, consistency, same parameters	2%	1%	2%	1%
No human bias, reduces human bias	4%	7%	5%	7%
AI only analysis objective data points	0%	0%	0%	5%
AI is a useful tool, standard process in which AI can be used	2%	1%	1%	3%
Two-way approach (e.g., allows redundancy, two pairs of eyes)	0%	0%	0%	2%
Human makes final decision	0%	0%	12%	9%
Efficient (e.g., time saving, manager can focus more on his analysis)	0%	0%	4%	2%
Evaluation criteria seems fair	3%	1%	1%	1%
Each actor is responsible for something they are good in	0%	0%	0%	13%
Combination of subjective and objective, factual and emotional etc.	0%	6%	9%	20%
Total Positive	35%	35%	51%	70%
Negative				
Human decision should be double checked by another Human	0%	1%	0%	1%
Missing human aspect (e.g., emotion, sympathy, intuition)	9%	0%	2%	0%

Neglects personal circumstances	4%	1%	2%	1%
AI should not make moral decisions; AI should not make decision (alone)	3%	5%	3%	2%
AI might err, AI not good enough yet	2%	1%	1%	2%
Explainability, controllability, transparency of AI	4%	2%	1%	2%
No subjective factors, interpersonal components, soft skills	16%	14%	10%	1%
No non-standard treatment possible	2%	0%	0%	0%
No feedback, no opportunity for voice	0%	0%	1%	0%
Suitable candidates could have been omitted in the preselection	0%	3%	7%	0%
Human should double check AI decision	1%	6%	2%	1%
Human should make final decision	3%	6%	0%	0%
AI bias	2%	1%	4%	2%
Not legal	0%	0%	0%	0%
AI can be manipulated	2%	0%	1%	1%
Human bias, partial judgement	0%	14%	6%	7%
Total Negative	50%	54%	41%	23%
Neutral				
Neutral decision	1%	1%	2%	1%
Missing information to rate fairness	2%	1%	1%	2%
Should be other way around	0%	4%	0%	0%
Depends on data, AI, training data, historical data	11%	5%	4%	5%
Total Neutral	14%	10%	8%	8%

Note. The cumulative percentages for the three major themes are highlighted in bold, while the remaining numbers represent the percentage of times each minor theme was used across the four different decision type conditions. 1 = Full AI, 2 = Sequential. Human-AI, 3 = Sequential. AI-Human, 4 = Aggregated Human-AI.

Table G3. Illustrative Quotes

Theme Description	Illustrative Quotes
Positive	
AI only analysis objective data points	<i>"Having the AI preselecting candidates based on performance and numbers and backing this up with human expertise, human making final decision seems like a pretty good process. Would not mind this procedure." [Sequential AI-Human, Promotion]</i>
	<i>"Seems fair because the AI is only analyzing the objective data which shouldn't cause any issues." [Aggregated Human-AI, Bonus]</i>
Human makes final decision	<i>"As long as the direct supervisor (here "manager") has the final say and uses his judgment and knowledge of human nature to use the AI results only as a decision-making aid and data basis, I think the use of AI based on this example is good." [Sequential AI-Human, Bonus]</i>
	<i>"AI should only be used to get "inspiration" or an idea, it should not be used in such a way that the full decision is made by a machine. No matter how intelligent that machine is. The final decision should still be made by a human." [Sequential AI-Human, Promotion]</i>

Two-way approach (e.g., allows redundancy, two pairs of eyes)	<p>"I think it is more fair because there is a combination of the manager's decision and the AL system. By having the AL system make the final decision, subjective factors are eliminated because the system does not know the candidates." [Sequential Human-AI, Promotion]</p> <p>"The two-way approach allows redundancy (if one is in a bad position with the manager, one's own factual performance still counts, and in other circumstances, which the computer cannot evaluate (e.g., failure due to burnout), the manager can pull up the rating) and is more acceptable, since the manager is accused of being too subjective/preferential, and vice versa with AI." [Aggregated Human-AI, Bonus]</p>
No (or reduces) human prejudice or preferences	<p>"Unlike humans, the AI should not have any (unconscious) prejudices." [Full AI, Promotion]</p> <p>"I think, the bonus payment is fair to be evaluated like this. The manager can preselect eligible people and then the machine can objectively pick one. For the manager the final decision could be harder - maybe he would favor someone intentionally or unintentionally. Like this no one gets in trouble or unfairly treated based on personal preferences." [Sequential Human-AI, Bonus]</p>
Objective, neutral	<p>"I think it is fair because the AI system can make an objective decision according to the respective qualifications. The 'brownie point factor' or vitamin B (= relationships) therefore no longer plays a role." [Full AI, Promotion]</p> <p>"The AI system probably evaluates the input data more objectively than the manager who, for example, has more contact, friendships, etc. with some employees. In the second step, the manager may then bring in knowledge about the employees that the AI did not have." [Sequential AI-Human, Bonus]</p>
Performance oriented, data-driven, or fact-based	<p>"The system refers to actual data on the candidate's performance which is fair in my opinion." [Full AI, Training]</p> <p>"I feel it is fair because the reward is based on facts and therefore less likely to be distorted by the employer's attitude." [Sequential AI-Human, Bonus]</p>
Good that Manager makes first selection	<p>"I think this is pretty fair considering that there is a human element prior to the AI. If it was all AI, I would not think it's fair since it might disregard elements such as commitment to the team, motivation, drive, etc. Like this my manager can make a pre-selection and the AI goes in more detail on the data aspects." [Sequential Human-AI, Bonus]</p> <p>"As the shortlist is selected by the manager, this decision-making procedure seems fair." [Sequential Human-AI, Promotion]</p>
Each actor is responsible for something they are good in	<p>"Since the subjective data is still evaluated by humans, which could be problematic for an AI, it is even fairer that an impartial authority is used for the rest. In the end, a human still has the final say and can monitor the results." [Aggregated Human-AI, Bonus]</p> <p>"Qualities of both agents is fairly distributed." [Aggregated Human-AI, Bonus]</p>
Combination of subjective and objective, factual and emotional etc.	<p>"The supervisor makes a subjective preselection and then uses the AI. Thus, the interpersonal criteria may be usefully supplemented with the objective criteria of the AI." [Sequential AI-Human, Bonus]</p> <p>"The use of an AI system to analyze objective data helps in eliminating potential biases and ensures a more objective evaluation of the candidates. Additionally, the involvement of the manager, who can provide a subjective assessment based on their familiarity with the team members, adds a human element to the decision-making process. This combination of objective and subjective evaluation attempts to create a fair and comprehensive selection process." [Aggregated Human-AI, Training]</p>
No human bias, reduces human bias	<p>"The AI will not have any racial or social biases. This makes the process way fairer. But it also cannot assess character traits and social qualities as well as humans. This is why I only gave it the overall grade "slightly" fair." [Full AI, Promotion]</p> <p>"It's fairer than when the human decides alone, I think. Like this the first selection will be without potential bias of the human which gives also others that might have no good relationship with the boss a fair chance." [Sequential AI-Human,</p>

	Promotion]
	<i>"In my opinion using AI as a support in this way would be relatively fair. The final decision would still be made by the human and the AI can calculate the hard facts much more objectively and accurately than the human - helping out. For a team lead position, it can be avoided that someone gets the position solely because of the managers judgement, which can always be biased as I think they do have some people they find particular suitable and favor." [Aggregated Human-AI, Promotion]</i>
AI is a useful tool, standard process in which AI can be used	<i>"The decision is the company's and AI is their tool." [Full AI, Bonus]</i> <i>"I think it's part of the century's performance review procedure. Indeed, the manager should consider performance and team fit but can use the AI tool as a baseline for that." [Aggregated Human-AI, Promotion]</i>
Evaluation criteria seems fair	<i>"It seems someone has put some thought into the process. The criteria evaluated by AI also seem fair to me." [Full AI, Bonus]</i>
Efficient (e.g., time saving, manager can focus more on his analysis)	<i>"It's fair that the AI does all the sorting and selection process as this might be a lengthy task and it's best that the final decision is done by the manager who will have an in-depth knowledge of the selected individuals and their achievements" [Sequential AI-Human, Bonus]</i> <i>"I think it's legitimate to use something like that. It saves a lot of work. However, it should be ensured that the AI makes decisions based on objective data that is as neutral as possible." [Sequential AI-Human, Training]</i>
Equal chances, standardization, consistency, same parameters	<i>"As it's based on parameters such as expertise and level of qualification it can be somewhat compared between individuals. However, I believe that every individual should be personally assessed e.g., in an interview, as this gives the individual the chance to personally introduce themselves." [Full AI, Training]</i> <i>"Preselection of AI seems more unbiased than from human and I then like that the human has to work with this preselection and still makes the final decision. Everyone has same chance to get into the pool for further selection." [Sequential AI-Human, Training]</i>
Negative	
Missing human aspect (e.g., emotion, sympathy, intuition)	<i>"There is no human element anymore in this decision-making process. I find it inhuman and not personal enough." [Full AI, Bonus]</i> <i>"Can be very superficial because of AI evaluation." [Sequential AI-Human, Training]</i>
Neglects personal circumstances	<i>"Unlikely to take personal circumstances and challenges into consideration, only based on hard facts, creates pressure on employees to meet some quotas." [Full AI, Bonus]</i> <i>"The human factor is completely eliminated in the pre-selection, so some candidates may not be considered. For example, a person who is busy training new employees. Spends time and perhaps has a poorer performance due to this." [Sequential AI-Human, Bonus]</i>
AI bias	<i>"I think that many of the human sources of error for unfair/inequitable selection can be eliminated by this. However, the data with which AI systems are trained and "fed" can also be biased and contain biases, which can lead to the AI system merely reproducing and applying human biases." [Full AI, Training]</i> <i>"The metrics used by AI can be biased to a certain degree. Preventing some members to even be considered." [Sequential AI-Human, Training]</i>
Human decision should be double checked by another Human	<i>"With the subjective selection of the shortlist by the manager, the quality of the decision or fairness could be further enhanced by the opinions of other managers (if any). The fact that in the end the AI as an objective part of the selection process decides on the final result, I feel is very fair." [Sequential Human-AI, Promotion]</i>

Suitable candidates could have been omitted in the preselection (due to bias, faulty data, wrong criteria etc.)	<p><i>"The AI will probably make a great decision based on provided criteria, however, to select a good leader you need also interpersonal competencies and a lot of soft skills. This is also hard to track and gather data on that. So, these valuable skills will not be considered by the AI in the first place. A good performing employee is not necessary a good leader, so some great candidates might not be preselected."</i> [Sequential AI-Human, Promotion]</p> <p><i>"Because the first step is the subjective selection of possible candidates. So, the AI is only able to select from this list."</i> [Sequential Human-AI, Promotion]</p>
Human bias, partial judgement	<p><i>"Performance ratings can be unfair in general as managers are humans and therefore their decisions are easily biased."</i> [Sequential AI-Human, Bonus]</p> <p><i>"Since AI and manager weigh hand in hand who is eligible for the position, the process seems to be quite fair. However, the manager still has the final say, so in the end the decision is still made by his subjective preferences and thus can no longer be seen as 100% fair."</i> [Aggregated Human-AI, Promotion]</p>
AI should not make moral decisions; AI should not make such a decision (alone)	<p><i>"The choice of the final candidates should not be made by the AI in my opinion, because AI can't make moral decisions."</i> [Full AI, Bonus]</p> <p><i>"No AI should make real-life decision that critically and financially influenced a real person's life."</i> [Sequential Human-AI, Bonus]</p>
AI might err, AI not good enough yet	<p><i>"It may be more fair, but on the other hand the AI is not ideal and may make wrong decisions."</i> [Full AI, Bonus]</p> <p><i>"AI could misinterpret facts."</i> [Aggregated Human-AI, Promotion]</p>
Explainability, controllability, intransparency of AI	<p><i>"Data is still biased because of provided shortlist by the manager; difficulty of explaining AI generated decision to candidates."</i> [Sequential Human-AI, Promotion]</p> <p><i>"The system is fair at first, since all employees go through it equally. The boss could be biased, of course. My biggest problem, however, is the lack of transparency in terms of criteria and weighting of the "objective" data in the black box called AI."</i> [Aggregated Human-AI, Training]</p>
No subjective factors, no interpersonal component, no soft skills, no qualitative data	<p><i>"This process can't take into account soft skills such as interpersonal connection, empathy, etc."</i> [Full AI, Promotion]</p> <p><i>"Of course, you can list and evaluate the most important qualifications and skills required. But whether a person should really be promoted and then stay on in the team or even move up depends also on emotional criteria, whether "the chemistry is right"."</i> [Full AI, Training]</p>
Human should double check AI decision	<p><i>"The procedure is rather fair than unfair, as the ML Algorithm will make a decision based on objective data, so no personal preference will be included in the final decision first. But still, I think a person should review the decision in the final step."</i> [Full AI, Bonus]</p> <p><i>"I would say, the idea to implement AI to select candidates is an interesting one. Regardless, I think that at the moment AI is not sufficient enough to actually make such decisions and therefore, a human should review the past performance of the applicants as well."</i> [Aggregated Human-AI, Promotion]</p>
Human should make final decision	<p><i>"It seems a bit unfair that AI is not only helping in shortlisting candidates, but also selects the candidates in the end, as its decision-making process is highly untransparent and incomprehensible. The final decision and a cross-check should always be performed by a human being, to prevent a biased AI (data taken from the internet that is used to train AI often leads to a bias based on race, gender or religious believe)."</i> [Full AI, Training]</p> <p><i>"I do not think it's fair that the human has no say in the end with the final decision. A person should be selected by a person."</i> [Sequential Human-AI, Promotion]</p> <p><i>"Personnel decisions should be made by people."</i> [Aggregated Human-AI, Bonus]</p>

Not legal	<i>"The main problem is that the entire process is carried out by the AI system and therefore an automatic case-by-case decision is made (I don't think this is legally allowed at all/only to a limited extent). Since it is not transparent how the AI system works, which data is relevant and if there is a bias, it is not fair and therefore attackable from my point of view - especially if I don't get the job (but this is the case for every rejected candidate)."</i> [Full AI, Promotion]
AI can be manipulated	<i>"In order to trick the system, you can write down things that maybe are not true. The AI will not detect this."</i> [Full AI, Promotion] <i>"I would say it's moderately fair, since it gives every person to prove themselves at the objective interview with the AI and show that they know the goals and the values at the company. Also, it helps if there is any personal matter between the manager and the person applying. On the other hand, if someone figured out a way to trick the AI, then it's not all done with tech by having the manager there as a safety net."</i> [Aggregated Human-AI, Promotion]
No non-standard treatment possible	<i>"I don't believe AI is able to identify nonstandard personal strength."</i> [Full AI, Bonus] <i>"It is fair because all potential candidates are evaluated with the same parameters and without bias. However, one problem with this method, in my opinion, is that employees with "abnormal" strengths may not be well assessed by the system. Example: the ability to motivate others in the team or to increase their performance."</i> [Full AI, Training]
No feedback, no opportunity for voice	<i>"It is very rational and doesn't include real feedback. In my opinion the human side is missing as well as some interaction or consideration of other valuables. It should include more personal variables as well as an argumentation for some performance indicators."</i> [Full AI, Bonus]
Neutral	
Depends on data, AI, training data, historical data	<i>"I think the fairness depends on the algorithm. Assuming the AI includes every possible variable, this should lead to a fair result related to its parameters. Human intuition, which would deliver unfair results, would be excluded."</i> [Full AI, Promotion] <i>"Data based decision inherently are more objective than purely human based decisions. However, depending on the training data that the ML algorithm received the decision by the AI system will align with the human who trained it, hence it's no longer providing a opposition solution or different view point to the manager as the AI will probably be training on past decisions to allocate bonus payments."</i> [Aggregated Human-AI, Bonus]
Neutral decision	<i>"It doesn't really matter, if the selection is made by a person or by AI, it can always be unfair or fair."</i> [Full AI, Promotion] <i>"All procedures have advantages and disadvantages in terms of fairness."</i> [Sequential AI-Human, Training]
Missing information to rate fairness	<i>"Not assessable under the information given. Assessments of candidates are very complex decisions, and it needs to be clear what kind of objective data is used by the AI to judge a candidate."</i> [Aggregated Human-AI, Promotion]
Should be other way around	<i>"I don't like that the AI system makes the pre-selection, as it might already eliminate me. My manager knows me best and should make a pre-selection and THEN the AI system could evaluate each person's individual data for allocation of bonus payments."</i> [Sequential AI-Human, Bonus] <i>"Other way around would be fairer. preselection by manager could be biased, final decision by AI should be checked by human."</i> [Sequential Human-AI, Bonus]

Appendix H: ANCOVA Fairness Dimensions – Covariates

		<i>F</i>	<i>df</i>	<i>p</i>	η^2
Interpersonal	Faith in General Technology [†]	24.53	1, 788	<.001	.03
	Trusting Stance General Technology	8.26	1, 788	.004	.01
	Propensity to Trust Automation [†]	29.54	1, 788	<.001	.04
	Propensity to Trust People	5.13	1, 788	.024	.01
	Manipulation Check	0.265	1, 788	.607	.00
	Decision Type	20.07	3, 788	<.001	.01
Informational	Faith in General Technology [†]	46.71	1, 788	<.001	.06
	Trusting Stance General Technology [†]	4.04	1, 788	.045	.01
	Propensity to Trust Automation	18.74	1, 788	<.001	.02
	Propensity to Trust People [†]	0.13	1, 788	.722	.00
	Manipulation Check	0.46	1, 788	.497	.00
	Decision Type	8.98	3, 788	<.001	.03
Procedural- Core	Faith in General Technology	33.98	1, 788	<.001	.04
	Trusting Stance General Technology	10.12	1, 788	.002	.01
	Propensity to Trust Automation	15.19	1, 788	<.001	.02
	Propensity to Trust People	3.71	1, 788	.054	.01
	Manipulation Check	1.67	1, 788	.197	.00
	Decision Type	2.79	3, 788	.040	.01
Procedural- Voice	Faith in General Technology	0.154	1, 788	.695	.00
	Trusting Stance General Technology	4.319	1, 788	.038	.01
	Propensity to Trust Automation	12.281	1, 788	<.001	.02
	Propensity to Trust People	0.137	1, 788	.711	.00
	Manipulation Check	25.289	1, 788	<.001	.03
	Decision Type	28.742	3, 788	<.001	.10

Note. [†] Interaction with the independent variable. Simply removing these covariates from the analysis could have resulted in an overestimation of the effect of decision type on the dependent variables. Since the focus was on controlling for the covariates and they were not directly related to the main hypothesis, I proceeded without excluding them. Regardless, the interaction can be inspected in Appendix I.

Appendix I: Covariate Interaction with Independent Variables

There was an interaction of some covariates with the independent variable.

Figure I1. Relationship between Propensity to Trust People and Decision Type in the Informational Fairness Dimension

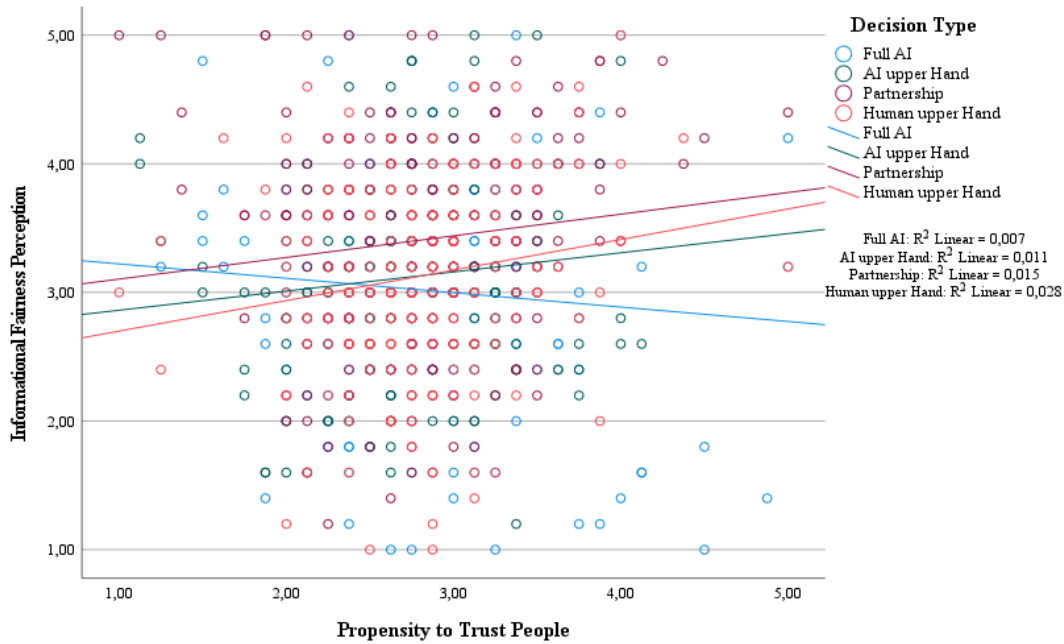


Figure I2. Relationship between Faith in General Technology and Decision Type in the Informational Fairness Dimension

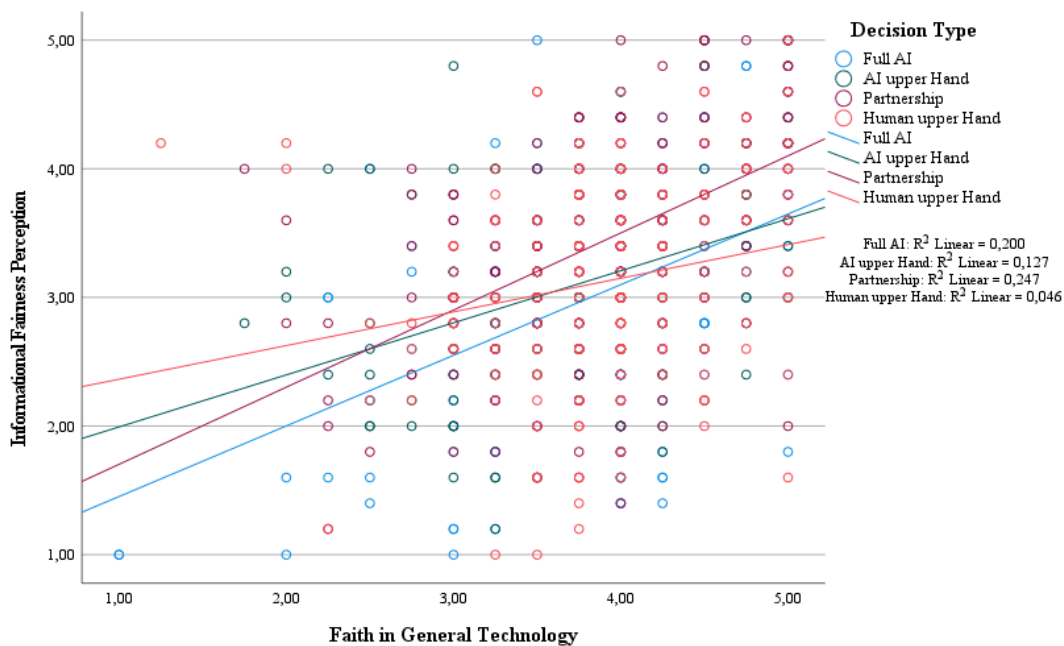


Figure I3. Relationship between Trusting Stance in General Technology and Decision Type in the Informational Fairness Dimension

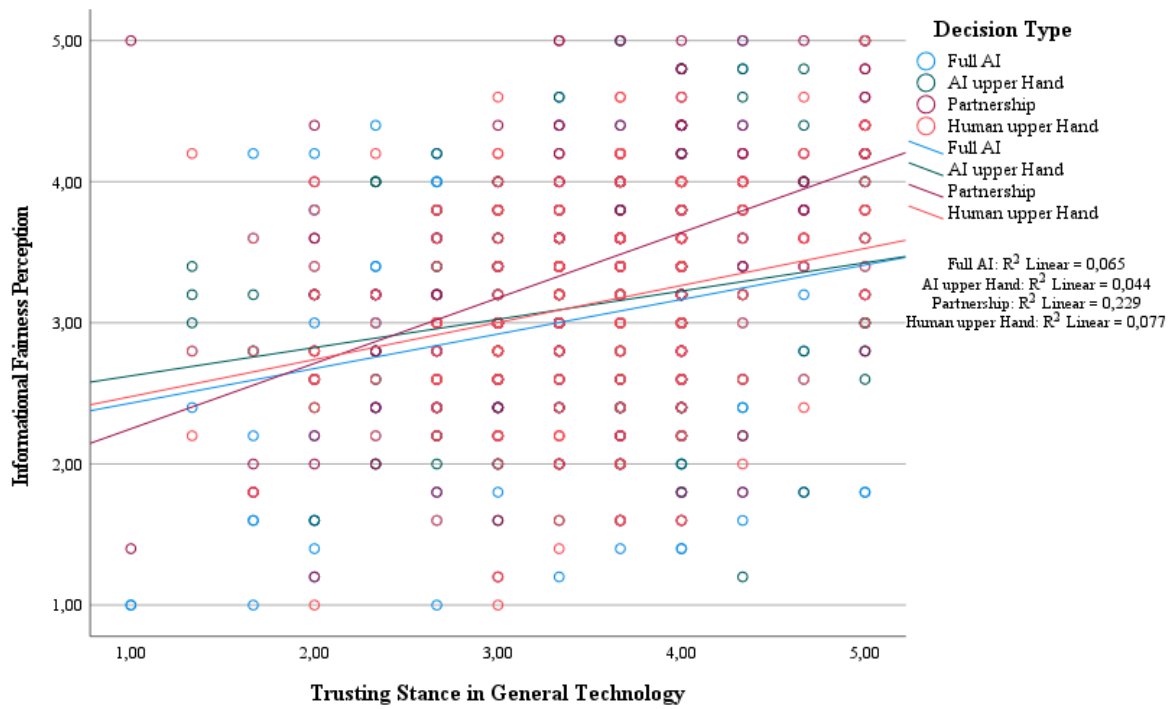


Figure I4. Relationship between Faith in General Technology and Decision Type in the Interpersonal Fairness Dimension

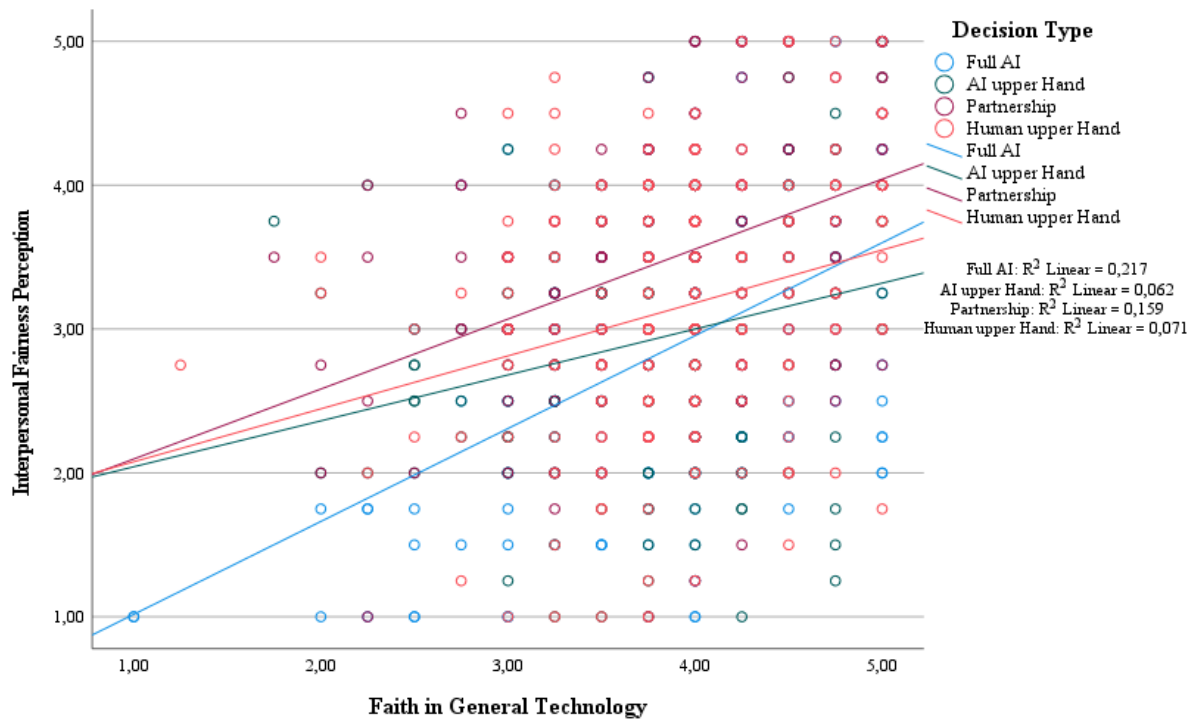


Figure 15. Relationship between Propensity to Trust Automation and Decision Type in the Interpersonal Fairness Dimension

