Four essays on reactions to algorithmic decision-making and their boundary conditions in organizational contexts

Inaugural-Dissertation

to obtain the degree of Doktor der Wirtschaftswissenschaften

submitted to the Faculty of Business Administration and Economics

at the Heinrich Heine University Düsseldorf

Presented by

Josephine M. Moritz, M.Sc.

1st Supervisor: Prof. Dr. Marius Claus Wehner, Chair of Business Administration, especially Digital Management & Digital Work

2nd Supervisor: Prof. Dr. Andreas Engelen, Chair of Business Administration, especially Management

Date: 02.07.2025

Acknowledgements

Completing this dissertation has been an ongoing journey of learning and growth, and I am deeply grateful for the invaluable guidance, encouragement, and inspiration provided by many unique individuals throughout the last three years.

First, I sincerely thank my supervisor, Professor Dr. Marius Wehner, for his invaluable advice and steady support over the past three years. Your positivity, guidance, and honest feedback have been a constant source of motivation. I'm truly grateful to have had the chance to work with a supervisor so engaged and genuinely interested. I also want to thank Professor Dr. Andreas Engelen for taking on the role of the second supervisor.

I am grateful to my colleagues at the Chair of Digital Work & Digital Management and fellow doctoral students at Heinrich Heine University. Thank you for your support, helpful discussions, and good company. Further, I extend my sincere thanks to my co-authors, to Larissa Pomrehn, for our inspiring collaborations and to Dr. Holger Steinmetz, for answering every statistical question I could imagine and who I have learned from immensely.

Thanks to the journal editors, numerous anonymous reviewers, conference participants, and fellow researchers who have helped me refine my work and grow as a researcher. Especially, I would like to thank Leonie Freise, Dr. Mahei Li, Professor Dr. Sofia Schöbel, and Felix Hirsch for their support and our helpful discussions.

Of course, this doctoral journey did not only happen at work. I am honored to have received the untiring support of my family and friends. I would like to especially thank my parents for their confidence in me and for supporting me in everything I set my mind to. I would further like to thank my stepparents and Sylke, Andreas, and Alena for their ongoing support throughout my academic journey, and my sister Julia, Thomas, and of course Lilly. To Kirsten, Peter, and Andrew Fezer, thank you for your ongoing support. For being a source of joy and motivation and always supporting me, I would like to thank my dear friends Julia and Rike. Further, I would like to thank Kea, Jessi, Jannika, and Jana, as well as Paulin, Anna, Julia, Anja, Sara, Judith, Charlotte, Pia, and Fenna for their support and encouragement.

And finally, to Janis, thank you for being there for me as the person who has, without a doubt, most closely witnessed the highs and lows of this journey. Your support and constant encouragement have made every challenge feel manageable. Words cannot express how thankful I am for all you have done.

Table of contents

Table of contents .... **I**

List of figures .... **IV**

List of tables .... **V**

List of abbreviations .... **VI**

1 Introduction **1**

1.1 Motivation and relevance of the dissertation 1

1.2 Research objectives and development of research questions 11

2 A meta-analysis on reactions to algorithmic decision-making in human resource management (Essay I) **21**

2.1 Introduction 22

2.2 Current research, framework, and hypotheses development 29

2.2.1 Theoretical framework 29

2.2.2 System-related and organization-related reactions 30

2.2.3 Boundary conditions affecting reactions to ADM 36

2.3 Method 41

2.3.1 Literature search and data extraction 41

2.3.2 Included studies 43

2.3.3 Aggregation levels and coding 44

2.3.4 Analytical procedure 45

2.4 Results 47

2.4.1 Main relationships between ADM and reactions 47

2.4.2 Moderator Analysis 49

2.4.3 Robustness checks and assessment of study quality 53

2.5 Discussion 54

2.5.1 Theoretical implications 55

2.5.2 Practical implications 57

2.5.3 Future research avenues 58

2.5.4 Limitations 62

3 Rejected by an algorithm? A multigroup analysis of outcome-dependent trust in algorithmic decisions (Essay II) **64**

3.1 Introduction 65

3.2 Theoretical background and hypotheses development 68

3.2.1 Stimulus-organism-response model 68

3.2.2 Trust perceptions 69

3.2.3 Organizational attractiveness 71

3.2.4 Interventions to increase trust in algorithmic decision-making 72

3.3 Method 75
3.3.1 Experimental procedure and sample 75
3.3.2 Measurement Invariance 79
3.4 Results 80
3.5 Discussion and implications 83
3.5.1 General discussion and theoretical implications 83
3.5.2 Practical implications 86
3.5.3 Limitations and future research avenues 89
3.6 Conclusion 91

4 Justice evaluations of algorithmic management: The role of prior discrimination experience (Essay III) 92

4.1 Introduction 93
4.2 Theoretical background 96
4.2.1 Justice perceptions of algorithmic management 96
4.2.2 Prior human discrimination experience 98
4.2.3 Prior algorithmic discrimination experience 99
4.3 Overview of studies 100
4.4 Study 1 101
4.4.1 Methods 101
4.4.2 SEM results 104
4.4.3 Brief discussion of Study 1 107
4.5 Study 2 108
4.5.1 Methods 108
4.5.2 SEM results 109
4.5.3 Brief discussion of Study 2 111
4.6 Study 3a 112
4.6.1 Methods 112
4.6.2 SEM results 115
4.6.3 Brief discussion of Study 3a 118
4.7 Study 3b 119
4.7.1 Qualitative analysis of open-ended survey responses 119
4.8 Discussion 122
4.8.1 Main findings and theoretical contributions 122
4.8.2 Practical implications 124
4.8.3 Limitations and future research avenues 126
4.9 Conclusion 127

5 Understanding worker resistance to algorithmic management: The role of perceived algorithmic discrimination (Essay IV) 129

5.1 Introduction ... 130
5.2 Theoretical background and hypotheses development ... 134
5.2.1 Algoactivism ... 134
5.2.2 Discrimination experiences ... 135
5.2.3 Justice and anticipatory injustice ... 137
5.3 Method ... 140
5.3.1 Data collection and sample ... 140
5.3.2 Measures ... 141
5.4 Results ... 143
5.4.1 SEM results ... 143
5.5 Discussion ... 146
5.5.1 Main findings and theoretical contributions ... 146
5.5.2 Practical implications ... 148
5.5.3 Limitations and future research avenues ... 149
5.6 Conclusion ... 151
6 Conclusion of the dissertation ... 153
6.1 Summary of the research findings ... 153
6.2 Theoretical implications ... 155
6.2.1 Understanding stakeholder reactions through boundary conditions 155
6.2.2 Conceptualization of stakeholder reactions to ADM ... 157
6.2.3 Differences in reactions to human and algorithmic discrimination . 159
6.3 Practical implications ... 160
6.4 Limitations ... 164
6.5 Future research avenues ... 166
6.6 Summary ... 169
7 References ... 170
8 Appendix ... 213
8.1 Appendix A1 (Essay I) - Flow diagram ... 213
8.2 Appendix A2 (Essay I) - Overview of included studies ... 214
8.3 Appendix B1 (Essay II) - Exemplary survey flow for the HITL condition. 219
8.4 Appendix B2 (Essay II) - Exemplary survey flow for the ADM condition. 221
8.5 Appendix C1 (Essay III) - Scenario descriptions ... 223
8.6 Appendix C2 (Essay III) - Scenario descriptions for Study 1 and Study 2 . 223
8.7 Appendix C2 (Essay III) - Scenario descriptions for Study 3 ... 224

List of figures

Figure 1-1: Research model and overview of the four essays 12

Figure 2-1: Overarching theoretical framework and brief summary of results 29

Figure 3-1: Theoretical framework 69

Figure 3-2: General SEM results 81

Figure 3-3: Multigroup SEM results 82

Figure 4-1: Research model 101

Figure 4-2: SEM results, Study 1 105

Figure 4-3: Interaction effect of prior human discrimination on informational justice 106

Figure 4-4: Interaction effect of prior human discrimination on interpersonal justice 106

Figure 4-5: Interaction effect of prior human discrimination on procedural justice 107

Figure 4-6: SEM results, Study 2 111

Figure 4-7: SEM results, Study 3 117

Figure 4-8: Interaction effect of experimental discrimination on informational justice 117

Figure 4-9: Interaction effect of experimental discrimination on interpersonal justice 118

Figure 4-10: Interaction effect of experimental discrimination on procedural justice 118

Figure 4-11: Sample quotes per experimental group 120

Figure 5-1: Research model 140

Figure 5-2: SEM results 145

List of tables

Table 1-1: Characteristics of the four essays ... 18

Table 2-1: Overview of utilized theories to explain ADM reactions in HRM ... 25

Table 2-2: Results of the bivariate meta-analysis ... 48

Table 2-3: Results of the bivariate meta-analysis including type of stakeholder and type of interaction as moderators ... 50

Table 2-4: Results of the bivariate meta-analysis including type of task, extent of decision, and personal impact as moderators ... 52

Table 3-1: Means, standard deviations, and correlations ... 78

Table 3-2: Measurement invariance ... 80

Table 3-3: Multigroup SEM results ... 83

Table 4-1: Means, standard deviations, and correlations for variables in Study 1 ... 104

Table 4-2: SEM results, Study 1 ... 105

Table 4-3: Means, standard deviations, and correlations of variables in Study 2 ... 109

Table 4-4: SEM results, Study 2 ... 110

Table 4-5: Means, standard deviations, and correlations of variables in Study 3 ... 115

Table 4-6: SEM results, Study 3 ... 116

Table 5-1: Means, standard deviations, and correlations ... 143

Table 5-2: SEM results ... 145

List of abbreviations

ADM	Algorithmic decision-making
AI	Artificial intelligence
AM	Algorithmic management
ANOVA	Analysis of variance
AOM	Academy of Management
ATAI	Attitude towards artificial intelligence
B	Unstandardized regression coefficient
CASA	Computers-are-social-actors
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CI	Confidence interval
df	Degrees of freedom
e.g.	Exempli gratia (English: for example)
Et al.	Et alii (English: and others)
EU AI Act	European Union Artificial Intelligence Act
EURAM	European Academy of Management
F	*F*-test
HDM	Human decision-making
HICSS	Hawaii International Conference on System Sciences

HR	Human resources
HRM	Human resource management
i.e.	Id est (English: that is)
k	Number of independent samples
m	Number of effect sizes
M	Mean
M.Sc.	Master of Science
n	Sample size
p	Probability value
PRISMA	Preferred reporting items for systematic reviews
RMSEA	Root mean square error of approximation
SD	Standard deviation
SE	Standard error
SEM	Structural equation modeling
SOR	Stimulus-organism-response
SRMR	Standardized root mean square residual
t	*t*-test
t_1	Measurement time 1
t_2	Measurement time 2
t_3	Measurement time 3

t_4	Measurement time 4
U.S.	United States
vs.	versus
β	Standardized regression coefficient
I^2_w	Percentage of true heterogeneity in observed heterogeneity within studies
I^2_b	Percentage of true heterogeneity in observed heterogeneity between studies
$\hat{\rho}$	Weighted average effect size
τ_w	Estimate of the true heterogeneity within studies
τ_b	Estimate of the true heterogeneity between studies
χ^2	Chi-squared test

1 Introduction

1.1 Motivation and relevance of the dissertation

Algorithmic decision-making (ADM) changes the way we work (Benlian et al., 2022) and includes several application areas ranging from augmenting managerial decisions (Meijerink & Bondarouk, 2023; Meijerink et al., 2021) to automating the supervision of workers (Möhlmann et al., 2021). *ADM* refers to the use of advanced computational methods that enable machines to automate decisions (Breidbach, 2024) and falls under the umbrella term of artificial intelligence (AI; Haesevoets et al., 2024). *AI* describes technologies that are able to perform tasks that would conventionally require human cognition and decision-making (Tambe et al., 2019), such as learning and problem-solving (Breidbach, 2024). AI includes *machine learning* as a subfield, which refers to algorithms with the ability to learn without the need for task-specific programming (Koenig et al., 2023).

Organizations are increasingly using ADM for a wide range of tasks. Companies such as Uber or Upwork, which operate in the platform economy and connect stakeholders via digital platforms (Benlian et al., 2022), rely entirely on ADM systems to control workers and match them with clients (Möhlmann et al., 2021). In these contexts, scholars refer to the use of ADM for managerial tasks and decisions as *algorithmic management* (AM)[1]. Whereas AM has historically been used in the platform economy, it is increasingly adopted in traditional organizations, challenging reactions to managerial decisions (Hirsch et al., 2024). The increasing use of these novel technologies is reflected in a global survey by McKinsey, which reports that 78% of companies use AI- and ADM-based systems for at least one business function, compared to 72% in 2024 and 55% in 2023 (McKinsey & Company, 2025). Organizations use ADM for various tasks, for example, in the human resource

[1] This dissertation primarily uses the term ADM, with occasional references to AM or AI depending on the respective application area and literature stream.

management (HRM) domain, where algorithms augment or automate tasks in the recruiting process, such as resumé screening and interviews (Feldkamp et al., 2023; Köchling & Wehner, 2020).

Stakeholder reactions play a central role in the successful implementation of ADM (Langer & Landers, 2021). Here, industry reports reveal that 70% of difficulties that arise when adopting ADM or AI systems in organizations stem from people- or process-related issues (Boston Consulting Group, 2024). Consequently, while organizations implement ADM systems with the expectation of increasing accuracy, efficiency, and decreasing human bias (Budhwar et al., 2022; Newman et al., 2020), these benefits can only be fully realized if stakeholders perceive ADM as just and trustworthy, and ultimately accept ADM (Höddinghaus et al., 2021).

To understand stakeholder reactions to ADM, this dissertation draws on the stimulus-organism-response (SOR) model (Mehrabian & Russell, 1974), and its extension by Jacoby (2002). The SOR model proposes that external, environmental stimuli (S) affect internal, cognitive and affective processes or states of individuals, called organisms (O), which in turn impact their responses (R; Mehrabian & Russell, 1974). Within this model, the *stimulus* describes external factors, which, in the context of ADM, can refer to the system itself, system characteristics, as well as task and decision characteristics. The *organism* describes the internal processing of information induced by the stimulus, such as perceptions and cognitive evaluations. More broadly, the organism also refers to individual characteristics, such as prior experiences or knowledge (Jacoby, 2002). Finally, the *response* refers to the outcomes of these internal processes (Gazi et al., 2025), which may manifest as observable behaviors, unobservable behaviors, or behavioral intentions. In the context of ADM, responses can manifest, inter alia, as acceptance or (active) resistance to algorithmic systems. Following this classification, external stimuli, such as the ADM system itself or interaction characteristics, impact stakeholders' perceptions of ADM, which in turn impact subsequent responses[2].

In the research field of ADM, the SOR model has been widely applied to understand technology adoption and user reactions (e.g., Dalvi-Esfahani et al., 2023; Kordzadeh & Ghasemaghaei, 2021; Ochmann et al., 2024; Zheng & He, 2024). Here, stimuli are commonly conceptualized as system characteristics (Kordzadeh & Ghasemaghaei, 2021), whereas internal processes of organisms are often conceptualized as fairness and justice perceptions[3] (Ochmann et al., 2024). Furthermore, researchers assess subsequent responses, inter alia, as satisfaction with ADM systems (Ochmann et al., 2024) or decision adoption (Dalvi-Esfahani et al., 2023). Empirical findings reveal that stakeholders' perceptions of ADM are mixed (Langer & Landers, 2021). On the one hand, ADM is often negatively associated with fairness and justice (e.g., Köchling et al., 2025; Schlicker et al., 2021) as well as trustworthiness and trust (Höddinghaus et al., 2021). On the other hand, several findings reflect trust in algorithmic decisions, for example, in the context of forecasting (You et al., 2022) or when algorithmic decisions are modifiable (Dietvorst et al., 2018).

The literature describes these mixed results as algorithm aversion (Castelo et al., 2019; Dietvorst et al., 2015) and algorithm appreciation (Logg et al., 2019). While *algorithm aversion* describes the preference for human decision-making over ADM and often occurs when humans see algorithms err (Dietvorst et al., 2015), *algorithm appreciation* refers to the preference of ADM over human decision-making (Logg et al., 2019). Algorithm aversion occurs when (1) tasks or decisions conducted by algorithms are associated with "human"

[2] This dissertation uses the term *perception* to refer to internal processes of the organism (O) and *response* to refer to outcomes of these perceptions (R), as distinguished in the SOR model (Mehrabian & Russell, 1974). When referring to both perceptions and responses collectively, the term *reaction* is used.

[3] These perceptions are discussed in detail in Essay I.

skills such as social or emotional skills (Lee, 2018), (2) decisions are automated rather than augmented (De Cremer & McGuire, 2022), or (3) decisions made by algorithms have high stakes for individuals (Langer & Landers, 2021). Conversely, individuals show algorithm appreciation when (1) ADM is used for tasks associated with "mechanical" skills, for example, forecasting (Logg et al., 2019), (2) ADM systems are considered transparent, for example, by providing explanations (Langer et al., 2021; Schoeffer et al., 2022), and (3) stakeholders are familiar with the functions and use of ADM (Mahmud et al., 2024).

Algorithm aversion and appreciation may also depend on further boundary conditions, representing either external stimuli or internal processes of an organism, in alignment with the SOR model (Jacoby, 2002; Mehrabian & Russell, 1974). External stimuli include, for example, interaction characteristics, such as stakeholder types and their respective roles in the decision-making process (Mahmud et al., 2022). For instance, applicants confronted with ADM as part of an application process might perceive ADM inherently differently from human resource managers using ADM systems to augment their decision processes. In this example, algorithmic decisions can have substantial consequences for applicants, and this stakeholder group has little control over ADM use, whereas, for human resource managers, ADM use will typically have no direct personal consequences. This asymmetry can lead to differences in the reactions of stakeholders. Here, Langer and Landers (2021) distinguish between first, second, and third parties: *first parties* make decisions using ADM, *second parties* are affected by ADM decisions, and *third parties* observe ADM use but are not directly affected. Second parties may represent a particularly relevant target group, as they are often unable to opt out and must adhere to ADM imposed by organizational protocols (e.g., applicants subjected to automated resumé screening).

Furthermore, whether individuals experience algorithm aversion or appreciation may also depend on their prior experiences, which represent part of the organism variable in the

SOR model (Jacoby, 2002; Mehrabian & Russell, 1974). Following the model of justice expectations, individuals use prior experiences to infer how justly they feel treated in the workplace (Bell et al., 2004) and are more likely to confirm existing beliefs (Shapiro & Kirkman, 2001; Snyder & Swann, 1978). In the context of ADM, prior experiences of discrimination play an important role, as ADM is used with the intention of reducing human biases and discriminatory behavior (Newman et al., 2020). At the same time, ADM is also associated with perpetuating inequalities, for example due to underlying biased data (Kellogg et al., 2020; Kordzadeh & Ghasemaghaei, 2021).

Consequently, depending on prior experiences of discrimination, individuals can be more or less likely to experience algorithm aversion or appreciation. Here, research on discrimination experiences demonstrates mixed findings. On the one hand, individuals associate ADM systems with discriminatory decisions and bias (e.g., Köchling & Wehner, 2020; Kordzadeh & Ghasemaghaei, 2021). On the other hand, in the context of HRM, studies find individuals experience algorithm appreciation when they have experienced human discrimination (Bedemariam & Wessel, 2023; Koch-Bayram et al., 2023). This indicates that, although initially associated with discriminatory intent or the perpetuation of existing inequalities, ADM systems might also be perceived as more consistent, thus reducing the fear of experiencing discrimination again, which is observable in minority groups (Lima et al., 2025).

According to the SOR model, perceptions of ADM, such as experiencing algorithm aversion or appreciation, impact responses and thus have severe consequences for organizations. Specifically, negative fairness and justice perceptions impact turnover intentions positively (Köchling et al., 2025) and lead to decreased organizational attractiveness and job pursuit intentions (Acikgoz et al., 2020). Additionally, beyond experiencing algorithm aversion, individuals might take a further step by engaging in algoactivism. *Algoactivism* refers to individual or collective tactics aimed at influencing or resisting AM systems (Kellogg et al., 2020), and can range from passive forms of resistance, like voicing complaints about AM on online platforms (de Jong et al., 2025), to more active tactics, for example, disabling GPS or phones to manipulate work assignments in ride-hailing contexts (Jarrahi & Sutherland, 2019). Consequently, responses to ADM use can severely harm organizational reputation and impact workflows when algorithmic decisions are not accepted, underscoring the importance of understanding the boundary conditions that impact algorithm aversion and appreciation.

Despite these initial findings on reactions to ADM and their boundary conditions, four critical gaps remain unaddressed. First, the literature on reactions to ADM use is scattered and findings are inconclusive (e.g., Jussupow et al., 2024; Köchling et al., 2025; Logg et al., 2019). Furthermore, to understand ADM reactions, previous research uses a wide range of theoretical frameworks and models, such as the organizational justice theory (Colquitt, 2001), the affective response model (Zhang, 2013), or the integrative model of organizational trust (Mayer et al., 1995). This leads to a scattered research landscape that further complicates a common understanding of algorithm aversion and appreciation. Additionally, research on boundary conditions impacting ADM reactions is mixed and, while several qualitative reviews have examined ADM use in organizations (e.g., Bankins et al., 2024; Basu et al., 2023; Budhwar et al., 2023; Bujold et al., 2024; Cheng & Hackett, 2021; Langer & Landers, 2021; Pan & Froese, 2023), there is currently no quantitative synthesis or study-level comparison of boundary conditions. Thus, existing literature lacks empirical and theoretical integration of boundary conditions, and it remains unclear how stakeholders make sense of ADM systems across different settings and decision types.

From a theoretical perspective, the inconclusiveness of results and the use of different theories and models lead to difficulties in generalizing findings, which ultimately hinders a thorough understanding of ADM reactions. From a practical perspective, given labor shortages and the ongoing war for talent, organizations must understand how individuals react to ADM systems, and whether these reactions can be positively impacted to manage human capital effectively and to gain a competitive advantage (Black & van Esch, 2021). Additionally, employees are more likely to accept and follow decisions they perceive as fair (Bankins et al., 2022; Lind, 2001). Therefore, it is important to understand when algorithm aversion may occur so that organizations can intervene accordingly.

Second, limited understanding and generalizability of research on reactions to ADM systems result in limited understanding of interventions to address algorithm aversion. Specifically, while research has found explainability and transparency of decision-making processes to be of importance (e.g., Binns et al., 2018; Schoeffer et al., 2024; Schoeffer et al., 2022), there is little understanding of how interventions need to be tailored depending on decision favorability. Given that ADM reactions are task- and context-dependent (Lee, 2018; Mahmud et al., 2022), it is important to understand the circumstances in which interventions are effective.

Third, although theoretical assumptions of justice perceptions suggest that individuals' prior experiences shape their justice perceptions (Bell et al., 2004; Shapiro & Kirkman, 2001; Snyder & Swann, 1978), empirical research on how prior experiences impact these perceptions in the context of ADM remains scarce. This is particularly important given the tension inherent in ADM systems: while they are often expected to deliver objective and consistent decisions (Bonezzi & Ostinelli, 2021), they also have the potential to reproduce or even aggravate existing social inequalities (Kellogg et al., 2020). Importantly, an individual may perceive discrimination from humans or ADM systems differently. This could impact initial ADM perceptions, making prior discrimination experiences a critical boundary condition. Additionally, it is particularly important to understand how justice evaluations are shaped as these determine decision acceptance (Lind, 2001) and relate to broader organizational consequences such as turnover intention and job performance (Moon, 2017).

Fourth, although worker resistance is not a new phenomenon in management literature (Cohen & Diamant, 2019; Hodson, 1995; Kelloway et al., 2010), the shift of managerial decisions traditionally made by humans to ADM systems offers new means to resist management practices (e.g., by engaging in algoactivism). However, the current understanding of the circumstances in which algoactivism occurs, and of whether assumptions drawn from human–human contexts are applicable, is limited. Nevertheless, a theoretical understanding is particularly relevant as worker resistance is negatively associated with both individual (e.g., worker well-being; Searle, 2022) and organizational consequences (e.g., customer satisfaction or productivity; Carpenter et al., 2021), representing a major risk for individuals and organizations.

To address these gaps, it is crucial to gain a deeper understanding of the perceptions of ADM and their boundary conditions, including (1) interaction characteristics, (2) interventions to address algorithm aversion, and (3) prior discrimination experiences, as well as (4) subsequent responses, including algoactivism. The outlined gaps and associated concerns lead to the overarching research question of this dissertation:

How do interaction characteristics, interventions, and prior discrimination experiences shape stakeholders' perceptions of and responses to algorithmic decision-making in organizations?

Following the SOR model, this dissertation examines each of the three components (i.e., external stimuli, internal processes of an organism, and responses; Mehrabian & Russell, 1974) separately, along with relevant boundary conditions, to gain a deeper understanding of how the reactions to ADM use in organizations are shaped. Depending on their characteristics, boundary conditions can represent both external stimuli (S) and internal processes of an organism (O). External stimuli include (1) interaction characteristics, such as the type of stakeholder, the type of interaction, and task and decision characteristics, and (2) interventions, specifically transparency and a human-in-the-loop design. Internal processes of an organism (O) include prior experiences, with a specific focus on prior human and algorithmic discrimination experience. Further, this dissertation examines justice, fairness, trustworthiness, trust, and affective reactions as ADM perceptions of organisms (O), as well as algoactivism, pursuit intention, and organizational attractiveness as subsequent responses to ADM (R). By examining the relationship between ADM use, boundary conditions, perceptions, and responses, this dissertation makes four contributions to the literature.

First, it advances theoretical understanding of ADM perceptions by addressing the fragmented and inconclusive literature related to algorithm aversion and appreciation. By systematically examining stakeholders' perceptions of ADM use and the boundary conditions impacting this relationship, this dissertation offers robust findings on which reactions occur under which conditions. Prior research has largely considered these conditions in isolation, while this dissertation thus provides a more integrated perspective by examining boundary conditions across primary studies. Additionally, this dissertation offers a theoretical understanding by proposing an overarching framework that links specific boundary conditions to stakeholder reactions toward both the system itself and the organization implementing it.

Second, this dissertation addresses the circumstances under which interventions mitigate algorithm aversion. Specifically, it examines how interventions that address algorithm aversion (i.e., transparency and a human-in-the-loop design) are more or less effective depending on outcome favorability. This contributes to a theoretical understanding by challenging the assumption that interventions work universally and provides a direction for organizations to implement ADM more effectively by aligning interventions with outcome favorability.

Third, by examining the role of human and algorithmic discrimination experiences, this dissertation contributes to research that has mainly focused on the impact of human discrimination on reactions to ADM (e.g., Koch-Bayram et al., 2023; Schulte Steinberg & Hohenberger, 2023). Comparing human and algorithmic discrimination experiences offers an initial understanding of the tension between the expectation of ADM systems to reduce human bias (Bonezzi & Ostinelli, 2021) and the potential of perpetuating inequalities (Kellogg et al., 2020). Thus, this dissertation offers an understanding of how justice evaluations of ADM are shaped.

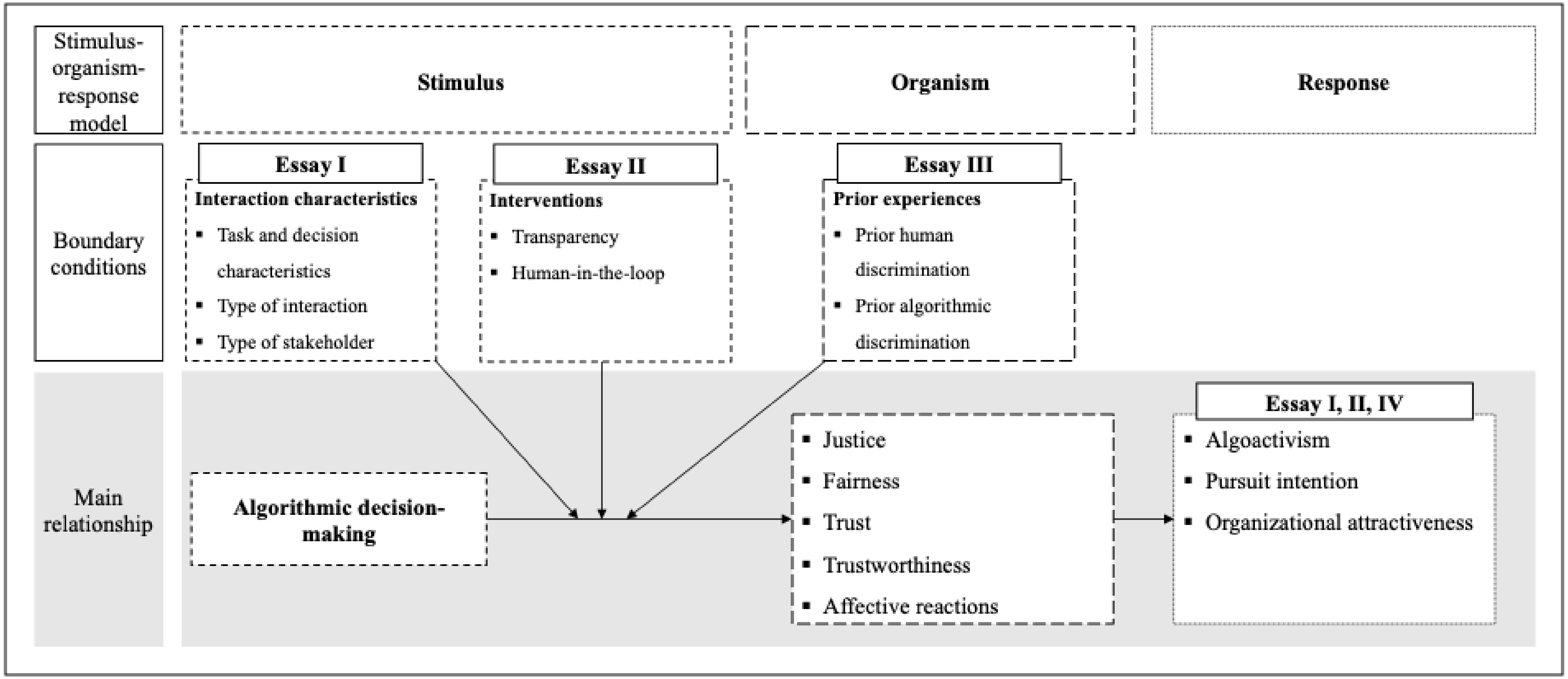
Fourth, this dissertation examines the mechanisms underlying algoactivism and thus offers a theoretical understanding of when workers engage in this form of resistance. Drawing on the justice model of counterproductive work behavior (Cohen & Diamant, 2019), which states that individuals who feel treated unjustly are more likely to engage in resistance behaviors, this dissertation investigates how perceived algorithmic discrimination, justice expectations, and justice evaluations drive resistance behaviors. By extending established theories of worker resistance to the context of ADM, this work contributes to a more nuanced and context-sensitive understanding of algoactivism as a response to ADM use.

Together, these contributions respond to theoretical and practical challenges arising from the shift from human decision-making to ADM in organizational contexts. By addressing these challenges, this dissertation advances theoretical understanding of stakeholder reactions and provides this emergent research field with an overarching theoretical framework. Further, it underlines the practical implications of interventions that increase trust in ADM and mitigate the consequences of algorithm aversion.

1.2 Research objectives and development of research questions

To address the overarching research question, this dissertation consists of four essays, each represented in a separate chapter. Figure 1-1 illustrates the essays based on the meta-theoretical model used in this dissertation, i.e., the SOR model (Mehrabian & Russell, 1974).

Figure 1-1: Research model and overview of the four essays



Note. Algorithmic decision-making is compared to human decision-making in all four essays. Interaction characteristics and interventions represent external stimuli. Prior experiences are subsumed under the organism variable in alignment with the stimulus-organism-response model (Mehrabian & Russell, 1974).

Essay I (chapter 2) presents a meta-analysis that synthesizes primary studies on reactions to ADM use within the HRM domain, an area where ADM is frequently used (Budhwar et al., 2022). To gain a deeper understanding of initial reactions to ADM in this domain, Essay I differentiates between system-related reactions (those directly associated with ADM systems, such as fairness, justice, trust, trustworthiness, and affective reactions) and organization-related reactions (those that refer to reactions to organizations using ADM systems, including organizational attractiveness and pursuit intention). Furthermore, the meta-analysis examines the impact of boundary conditions on reactions to ADM. Here, the meta-analysis includes the type of stakeholder, the type of interaction, the type of task, the extent of decisions, and the personal impact of decisions. Accordingly, Essay I addresses the following question:

Research question 1: How do individuals react to the use of ADM in the HRM domain and which boundary conditions impact these reactions?

By addressing this research question, Essay I examines the relationship between ADM as an external stimulus (S) and internal cognitive and affective perceptions of organisms (O), as well as responses (R), as outlined in the SOR model (Mehrabian & Russell, 1974), and thus makes several contributions to the current literature. First, by analyzing 53 studies, published between 2019 and 2025, comprising 73 individual samples (n = 24,578), the meta-analysis provides a quantitative synthesis of initial findings regarding reactions to ADM and creates a consensus in this field. Second, the meta-analysis includes different boundary conditions that impact system- and organization-related reactions to ADM use. These are tested across primary studies and thus help develop a thorough understanding of ADM reactions. Third, Essay I uses the computers-are-social-actors framework (CASA; Nass & Moon, 2000; Nass et al., 1994) to reason why theories from human–human interaction can also be utilized to understand human–algorithm interactions. The CASA framework postulates that humans interact with ADM systems like they do with other humans, employing stereotypes, interactional rules, and norms. Consequently, we propose that theories used to understand human–human interaction are applicable in the human–algorithm context. Based on this argument, the meta-analysis provides an overarching theoretical framework that includes reactions to ADM and their boundary conditions.

Essay II focuses on interventions (i.e., transparency and a human-in-the-loop design) to increase trust in ADM in the recruiting context. Building on prior research on interventions that increase trust (e.g., through explainability; Shulner-Tal et al., 2022), Essay II proposes that these interventions depend on outcome favorability (i.e., whether applicants are rejected or accepted). According to attribution theory (Weiner, 1985) and self-serving bias (Zuckerman, 1979), individuals tend to attribute unfavorable decisions externally (i.e., to the decision entity), while taking personal credit for favorable outcomes. This attribution bias can lead to discrepancies in the perception of trust, so that trust-building interventions are impacted by outcome favorability. Thus, Essay II aims to answer the following research question:

Research question 2: How does the outcome of a selection procedure (i.e., acceptance or rejection) impact trust in ADM and how effective are trust-building interventions (i.e., transparency and human-in-the-loop) depending on outcome favorability?

To address this research question, Essay II employs an online between-subjects vignette experiment (n = 483). Here, individuals are confronted with a favorable (i.e., acceptance) or unfavorable (i.e., rejection) decision outcome as a stimulus (S), which impacts trust as a perception of an organism (O). This, in turn, impacts the individual's response (R) to the organization, namely organizational attractiveness. By employing multigroup structural equation modeling, Essay II demonstrates that ADM use is negatively associated with trust when participants are rejected, which is not apparent when individuals are accepted. Thus, Essay II contributes to prior research by demonstrating that transparency and a human-in-the-loop design have different effects on trust perceptions depending on outcome favorability. This challenges the current understanding that interventions addressing algorithm aversion have similar effects across contexts.

Essay III examines the role of prior human and algorithmic discrimination experiences as a boundary condition that impacts justice perceptions of ADM. To offer a theoretical understanding, Essay III draws on the model of justice expectations by Bell et al. (2004) and the idea of anticipatory injustice (Shapiro & Kirkman, 2001), which suggest that prior experiences impact justice expectations and evaluations. Additionally, Essay III uses qualitative open-ended questions to gain deeper insights into potential mechanisms underlying the proposed relationships. Examining experiences of discrimination by both humans and ADM systems allows a better understanding of how justice evaluations are shaped in this context. Consequently, Essay III addresses the following research question:

Research question 3: How do prior experiences of human and algorithmic discrimination impact the relationship between ADM use and justice perceptions?

To answer this research question, Essay III employs three vignette-based online experiments (n_1 = 82, n_2 = 83, n_3 = 216), which demonstrate that human and algorithmic discrimination are perceived differently. Essay III extends prior research by examining the mechanisms underlying these different perceptions. Qualitative open-ended responses suggest that individuals associate algorithmic discrimination with indirect sources, such as biased data, whereas human discrimination is linked directly to the decision-maker. Consequently, Essay III proposes that the decontextualization often associated with ADM systems (Newman et al., 2020) may act as a double-edged sword, lowering overall justice evaluations while leading individuals to believe that algorithms lack discriminatory intent.

Furthermore, Essay III contributes to research by extending the model of justice expectations (Bell et al., 2004) from human–human to human–algorithm contexts and thus extends the theoretical understanding of how justice perceptions are shaped. This also challenges the prevailing focus in ADM research, which emphasizes situational factors (e.g., decision transparency or stakes of a decision; Langer, König, & Papathanasiou, 2019; Schoeffer et al., 2022), while largely neglecting the role of individuals' prior experiences.

Finally, Essay IV examines counterproductive work behavior (i.e., algoactivism) as a response to negative perceptions of ADM, specifically through experiences of algorithmic discrimination. According to the SOR model (Mehrabian & Russell, 1974), perceptions of an organism lead to (behavioral) responses. Apparent in Essay I, much research focused on the perceptions of employees subject to ADM, such as trust (e.g., Höddinghaus et al., 2021), fairness (e.g., Lavanchy et al., 2023), and justice (e.g., Feldkamp et al., 2023), reflecting the need for a better understanding of the subsequent responses. Here, Essay IV proposes that discrimination experience lowers justice expectations and justice perceptions, which in turn lead to algoactivism to counteract negative justice perceptions. Accordingly, Essay IV addresses the following research question:

Research question 4: How do perceived algorithmic discrimination and justice evaluations impact algoactivism within organizations using ADM?

To answer this research question, Essay IV includes a four-wave survey (n_{t1} = 276, n_{t2} = 200, n_{t3} = 155, n_{t4} = 124) among workers who interact with ADM systems in their daily work. By doing so, this essay offers several contributions to the existing literature. First, it extends the justice model of counterproductive work behavior (Cohen & Diamant, 2019) from a human–human context to a human–algorithm context, offering a theoretical understanding of the mechanisms underlying algoactivism. Second, Essay IV uses a four-wave survey design, extending prior research that primarily relies on cross-sectional or qualitative data, thus offering a more nuanced understanding of the mechanisms underlying algoactivism. Third, Essay IV presents practical implications, such as establishing feedback mechanisms when employees feel discriminated against, and avenues for future research.

In summary, the four essays contribute to an understanding of ADM perceptions in organizational settings, the responses that result from these perceptions, and how these relationships are impacted by boundary conditions. Table 1-1 provides an overview of the research objectives, methods and samples, main results, and the current status of the respective essay.

Table 1-1: Characteristics of the four essays

Essay	Research objectives	Method and sample	Main results	Current status	Conferences	Share of contributions
Essay I (chapter 2) A meta-analysis on reactions to algorithmic decision-making in human resource management	This study synthesizes reactions to ADM use and their boundary conditions in the context of HRM. It integrates theoretical perspectives through the CASA framework and proposes an overarching theoretical model.	Meta-analysis including 73 samples in 53 studies with n = 24,578.	ADM is negatively associated with system-related (i.e., justice, fairness, trust, trustworthiness) and organization-related (i.e., organizational attractiveness, job pursuit intention) reactions, which depend on boundary conditions (i.e., type of interaction, type of decision).	2nd Revision (moderate) in *Human Resource Management Review*	Presented at: 23rd European Academy of Management Conference, Dublin, Ireland, 2023. 83rd Annual Meeting of the Academy of Management, Boston, USA, 2023.	J. M. Moritz 50% L. Pomrehn 30% H. Steinmetz 10% M. C. Wehner 10%
Essay II (chapter 3) Rejected by an algorithm? A multigroup analysis of outcome-dependent trust	This study investigates how outcome favorability impacts trust in ADM and evaluates the impact of two trust-building interventions (i.e., transparency and	One vignette-based between-subject experiment with n = 483.	Multigroup structural equation modeling demonstrated that trust in ADM is lower for rejected applicants, and that the impact of interventions differs by outcome. A human-in-the-loop	2nd Revision (major) in *Communications of the Association for Information Systems*	Presented at: 84th Annual Meeting of the Academy of Management, Chicago, USA, 2024.	J. M. Moritz 70% J. Witte 10% M. C. Wehner 10% N. R. Gier-Reinartz 10%

Essay	Research objectives	Method and sample	Main results	Current status	Conferences	Share of contributions
in algorithmic decisions	human-in-the-loop) in the recruiting context.		design increases trust after acceptance, while transparency increases trust after rejection.			
Essay III (chapter 4) Justice evaluations of algorithmic management: The role of prior discrimination experiences	This study investigates prior discrimination experience as a moderator of the relationship between ADM use and justice perceptions. Additionally, it compares human and algorithmic discrimination experiences and their impact on justice perceptions.	Three vignette-based between-subject experiments with $n_1 = 82$, $n_2 = 83$, $n_3 = 216$.	Prior human discrimination experience decreases algorithm aversion, while prior algorithmic discrimination has no impact on the relationship between ADM and justice. Qualitative responses indicate that attribution processes might be an explanatory mechanism underlying these perceptions.	Under review at *Business & Information Systems Engineering*	<u>Presented at:</u> 32nd European Conference on Information Systems, Paphos, Cyprus, 2024. 84th Annual Meeting of the Academy of Management, Chicago, USA, 2024. <u>Awards:</u> Best paper award runner-up, European Conference on Information Systems 2024.	J. M. Moritz 90% M. C. Wehner 10%
Essay IV (chapter 5) Understanding worker	This study investigates how perceived algorithmic discrimination impacts	Four-wave survey with $n_{t1} = 276$, $n_{t2} = 200$,	Perceived algorithmic discrimination was positively associated with algoactivism	Submitted to *Journal of the Association for*	<u>Accepted at:</u> 33rd European Conference on	J. M. Moritz 100%

Essay	Research objectives	Method and sample	Main results	Current status	Conferences	Share of contributions
resistance to algorithmic management: The role of perceived algorithmic discrimination	workers' engagement in algoactivism by examining the mediating roles of justice expectations and justice evaluations.	n_{t3} = 155, n_{t4} = 124.	through the proposed mediators justice expectations and justice evaluations.	*Information Systems*	Information Systems, Amman, Jordan, 2025. <u>Presented at:</u> 22nd European Congress of Work and Organizational Psychology, Prague, Czech Republic, 2025.	
Total contribution by Josephine Moritz						**310%**

2 A meta-analysis on reactions to algorithmic decision-making in human resource management (Essay I)[4]

Abstract

In this meta-analysis, we assess reactions to algorithmic decision-making (ADM) in human resource management. Drawing on the computers-are-social-actors (CASA) paradigm as a meta-theoretical framework, we offer an overarching theoretical framework to explain the relationships between ADM and (1) system-related reactions (e.g., perceptions of justice, trust) as well as between ADM and (2) organization-related reactions (e.g., organizational attractiveness, turnover intention). Based on our overarching theoretical framework, we examine boundary conditions impacting these reactions. Our meta-analysis includes 365 effect sizes from 73 samples in 53 studies (n = 24,578) and reveals that ADM is negatively associated with system-related (i.e., justice, fairness, trust, trustworthiness) and organization-related (i.e., organizational attractiveness, job pursuit intention) reactions, which depend on boundary conditions (i.e., type of interaction, extent of decision). Based on our overarching theoretical model, we provide an empirical and theoretical synthesis of primary studies and outline avenues for future research.

Keywords:

Algorithmic Decision-Making; Employee Reactions; Applicant Reactions; Human Resource Management; Meta-Analysis

[4] **Authors:** Moritz, J. M., Pomrehn, L., Steinmetz, H., & Wehner, M. C.
Second revision in *Human Resource Management Review*
A similar version of this manuscript was presented at the following conferences:
- 23rd European Academy of Management Conference, Dublin, Ireland, 2023
- 83rd Annual Meeting of the Academy of Management, Boston, USA, 2023

2.1 Introduction

Organizations increasingly use algorithmic decision-making (ADM) to optimize and accelerate the tasks of their human resource management (HRM; Chowdhury et al., 2023; Meijerink & Bondarouk, 2023), specifically in recruitment and selection processes (Bankins et al., 2022; Campion et al., 2024; Koenig et al., 2023), career development and training (Köchling et al., 2025), and performance management of employees (Zhang & Amos, 2023). ADM refers to systems that use advanced computational techniques and falls under the umbrella term of artificial intelligence (AI; Langer & Landers, 2021; Meijerink et al., 2021). AI refers to the ability of machines to analyze large datasets to make predictions and solve problems within complex environments and includes machine learning (Koenig et al., 2023; Prikshat et al., 2023). In this context, machine learning is a method through which systems analyze large sets of data, learn patterns from it, and apply these patterns to make decisions or predictions without being explicitly programmed for each task (Koenig et al., 2023). ADM systems can benefit managers by augmenting or automating HRM tasks, such as screening resumes or predicting employee turnover, thereby improving efficiency and potentially reducing human bias in decision-making (Daugherty & Wilson, 2018).

While the rapid growth of ADM holds significant promise for HRM, its benefits can only be fully realized if all stakeholders in the decision-making process accept and effectively engage with the ADM technology (Langer & Landers, 2021). However, if individuals respond negatively to ADM systems (e.g., with feelings of distrust or unfairness), these individual reactions may negatively affect the entire organization, for example, by diminishing organizational attractiveness or increasing turnover intentions (Koch-Bayram et al., 2023; Köchling et al., 2025). In times of labor shortages and the ongoing war for talent, it is crucial for organizations to understand how individual reactions to ADM emerge and whether they are able to positively influence these reactions to effectively manage their human capital and achieve competitive advantages. Prior literature has subsumed negative ADM reactions under the term *algorithm aversion* that is often applied when algorithms make mistakes (Castelo et al., 2019; Dietvorst et al., 2015) and positive ADM reactions under the term *algorithm appreciation*, which is often demonstrated for tasks such as forecasting (Logg et al., 2019). In addition, these reactions depend on boundary conditions, such as type of task (Lee, 2018) or included stakeholders (Langer & Landers, 2021). Furthermore, although several empirical papers (e.g., Bedemariam & Wessel, 2023; Kleinlogel et al., 2023) and qualitative reviews (e.g., Bankins et al., 2024; Basu et al., 2023; Budhwar et al., 2023; Bujold et al., 2024; Cheng & Hackett, 2021; Langer & Landers, 2021; Pan & Froese, 2023), depicted the landscape of ADM and AI use in organizations, the theoretical basis to explain reactions to ADM is rather scattered.

Following a thorough literature review of studies on ADM use in the HRM domain, we identified three gaps in this field. First, although research often highlights negative reactions to ADM, findings from previous studies are controversial and inconclusive (e.g., Acikgoz, 2019; Feldkamp et al., 2023). On the one hand, prior research demonstrated algorithm aversion when (1) the decision or task is expected to require social or emotional skills (Castelo et al., 2019; Lee, 2018), (2) decisions are made by an algorithm with little or no human interaction or human-in-the-loop (De Cremer & McGuire, 2022), and (3) decisions involve high stakes for individuals (Langer, König, & Papathanasiou, 2019). On the other hand, studies report algorithm appreciation when (1) ADM is used for tasks considered mechanical (i.e., forecasting; Logg et al., 2019), (2) ADM is considered transparent (Suen & Hung, 2023), or (3) participants experienced prior discrimination by humans and hence prefer ADM (Koch-Bayram et al., 2023; Schulte Steinberg & Hohenberger, 2023). From a theoretical point of view, this polarization hinders the development of a coherent framework for understanding reactions to ADM, as inconsistent results make it difficult to establish generalizable mechanisms. Practically, this ambiguity can complicate ADM implementation in organizations, especially when stakeholder perceptions are not considered, thereby risking employee disengagement and reducing trust in ADM and the organization itself. Addressing this ambiguity is essential for advancing the research field and supporting organizations to implement ADM in a manner that is perceived as fair and trustworthy by stakeholders.

Second, research on boundary conditions is mixed and there is no quantitative synthesis to date, although several systematic and narrative reviews do examine the consequences of AI and ADM use (e.g., Bankins et al., 2024; Basu et al., 2023; Budhwar et al., 2023; Bujold et al., 2024; Cheng & Hackett, 2021; Langer & Landers, 2021; Pan & Froese, 2023). Additionally, previous research on boundary conditions often focuses on individual moderators (e.g., type of task, cultural differences; Lee, 2018; Kleinlogel et al., 2023), but findings on study-specific moderators, such as type of interaction, are lacking across primary studies. The lack of focus on study-specific moderators and absence of a quantitative synthesis is concerning as it leaves critical boundary conditions underexplored, yet these factors are essential for understanding how ADM reactions vary across contexts. This gap hinders the ability to adapt ADM systems to specific stakeholder needs, which is crucial for ensuring successful adoption.

Third, theoretical approaches used in previous literature lack a coherent foundation, as prior studies focused on specific aspects of ADM reactions rather than making use of one overarching theoretical framework (see Table 2-1). While existing theories such as the computers-are-social-actors (CASA) framework (Nass & Moon, 2000; Reeves & Nass, 1996) address human–technology interactions, other theories emphasize specific variables, for example, justice theory (Colquitt, 2001; Greenberg, 1987; Leventhal, 1980), Gilliland’s (1993) model of applicant reactions, or Mayer et al.’s (1995) integrative model of organizational trust. These different theoretical approaches could lead to various challenges, such as difficulty in developing a unified understanding of reactions to ADM. Without such understanding, researchers may struggle to identify patterns across studies, as inconsistent use of theoretical constructs can undermine the validity of empirical research and complicate efforts to build cumulative knowledge in this field.

Table 2-1: Overview of utilized theories to explain ADM reactions in HRM

Theories	Description/main argument	Articles
	Theories to explain system-related reactions	
Advice Response Theory (ART; Feng & MacGeorge, 2010)	People react to advice depending on the message, the advisor, and the receiver.	Chacon et al. (2024)
Affective Response Model (Zhang, 2013)	(Unconscious) evaluation processes lead to affective responses, which guide cognition and action.	Köchling et al. (2023), Pomrehn and Wehner (2025)
Applicant Attribution-Reaction Theory (Ployhart & Herold, 2004)	Job applicants interpret/attribute reasons behind recruitment decisions and these interpretations influence their reactions.	Oostrom et al. (2023)
Automation Model (Parasuraman et al., 2000)	Guides automation design by matching system functions with appropriate automation types and levels.	Langer, König, and Papathanasiou (2019), Langer, König, Sanchez, et al. (2019)
Attachment Theory (Mikulincer & Shaver, 2005)	Assumes that individuals internalize early interactions with primary caregivers and that resulting emotions guide their social behavior.	Deriu et al. (2024)
CASA (Computers Are Social Actors; Nass & Moon, 2000) / Social Response Theory (Nass & Moon, 2000)	Suggests that people tend to treat computers and other technology as if they were human, applying social norms to their interactions with machines.	Moritz and Schmidt (2024), Moritz et al. (2024), Moritz et al. (2025), Suen and Hung (2023), Suen et al. (2019)
Explanation-for-Trust-Theory (Pieters, 2011)	Transparent explanations for system behavior can build trust in automated systems.	Suen and Hung (2023)
Exposure Theory (Zajonc, 1968)	Repeated exposure to a stimulus increases an individual's preference for it.	Koch-Bayram et al. (2023)
Fairness Heuristics Theory (Lind, 2001)	People rely on fairness heuristics to decide how to respond to organizational decisions.	Bedemariam and Wessel (2023), Choi and Chao (2024), Luo and Zhang (2023)
Integrative Model of Organizational Trust (Mayer et al., 1995)	Defines trust as a willingness to be vulnerable to another party, whereas trustworthiness includes ability, benevolence, and integrity.	Höddinghaus et al. (2021), Kares et al. (2023), Lacroux and Martin-Lacroux (2022), Langer et al. (2023), Moritz and Schmidt (2024), Moritz et al. (2024), Sondern et al. (2025)
Media Richness Theory (Daft & Lengel, 1986)	Explains how the effectiveness of communication depends on the medium's ability to convey rich information.	Köchling and Wehner (2023), Langer, König and Hemsing (2019), Suen et al. (2019)
Model of Applicants' Reactions to Employment Selection Systems (Gilliland, 1993)	Procedural and distributive justice impact applicants' fairness perceptions during selection.	Acikgoz et al. (2024), Acikgoz et al. (2020), Langer, König, and Papathanasiou (2019), Langer, König, Sanchez et al. (2019), Lavanchy et al. (2023), Noble et

Theories	Description/main argument	Articles
		al. (2021), Oostrom et al. (2023), Suen and Hung (2023), Suen et al. (2019)
Model of Justice Expectation (Bell et al., 2004)	Individuals form expectations about justice based on prior experiences and adjust their behavior according to whether these expectations are met.	Koch-Bayram et al. (2023), Moritz and Schmidt (2024)
Model of Social Validity (Schuler & Stehle, 1983)	Information, participation, transparency, and feedback are used to evaluate whether a selection procedure is socially acceptable.	Koch-Bayram et al. (2023)
Organizational Justice Theory (Colquitt, 2001)	Justice comprises distributive, procedural, interpersonal, and informational justice, which impact workplace behavior.	Bankins et al. (2022), Bedemariam and Wessel (2023), Gonzalez et al. (2022), Koch-Bayram et al. (2023), Köchling et al. (2024), Köchling and Wehner (2023), Lavanchy et al. (2023), Moritz and Schmidt (2024), Moritz et al. (2024), Moritz et al. (2025), Nagtegaal (2021), Noble et al. (2021), Oostrom et al. (2023), Schlicker et al. (2021), Zhang and Amos (2023)
Self-Determination Theory (Deci & Ryan, 1985)	Suggests that motivation is driven by the fulfillment of three basic needs: autonomy, competence, and relatedness.	Gonzalez et al. (2022)
Social Exchange Theory (Blau, 1964)	People assess relationships by weighing costs and benefits in exchanges.	Dutta and Mishra (2024), Oostrom et al. (2023)
Social Information Processing Theory (Walther, 2011)	Individuals infer characteristics of other individuals based on implicit and explicit cues.	Suen et al. (2019)
Social Interface Theory (Long, 2003)	Humans interact with computer interfaces similarly to human–human interaction.	Suen and Hung (2023), Suen et al. (2019)
Social Monitoring System (Pickett & Gardner, 2005)	Individuals have a heightened sensitivity to social cues when their belonging is threatened.	Lavanchy et al. (2023)
Stereotype Content Model (Fiske et al., 2002)	Individuals distinguish others according to warmth and competence.	Sondern et al. (2025)
Stereotype Theory (Oakes & Turner, 1986)	Stereotypes are cognitive shortcuts that help individuals categorize people or things.	Pomrehn et al. (2024)
Task-Technology Fit Theory (Furneaux, 2012)	Technology is used effectively when its capabilities match the requirements of the task.	Acikgoz et al. (2024)
Technology Acceptance Model (Davis, 1989)	Perceived ease of use and usefulness are the primary factors influencing technology adoption.	Choung et al. (2023), Suen and Hung (2023)
Theory of Planned Behavior (Ajzen, 1991)	Behavioral intention is shaped by an individual's attitude towards the behavior, the influence of social norms, and perceived control.	Mirowska (2020)
Three-Factor Theory of Anthropomorphism (Epley et al., 2007)	Anthropomorphism depends on anthropocentric knowledge, motivation to explain and understand the other agent, and desire for social motivation.	Liu et al. (2023)
Trust in Automation (Lee & See, 2004)	Trust in automation develops over time and impacts human reliance on automated systems.	Lacroux and Martin-Lacroux (2022)

Theories	Description/main argument	Articles
Uncanny Valley Theory (Mori, 1970)	People feel discomfort when a digital representation appears almost human, creating a sense of eeriness.	Suen and Hung (2023)
Theories to explain organization-related reactions		
HR Attribution Framework (Nishii et al., 2008)	Suggests that employees' perceptions of HR practices are influenced by the attributions they make about the organization's intentions.	Koch-Bayram and Kaibel (2023)
Signaling Theory (Spence, 1973)	Explains how signals (such as qualifications or organizational practices) convey information in situations of uncertainty.	Acikgoz et al. (2024), Acikgoz et al. (2020), Bedemariam and Wessel (2023), Gonzalez et al. (2022), Keppeler (2023), Koch-Bayram et al. (2023), Koch-Bayram and Kaibel (2023), Köchling et al. (2023), Luo and Zhang (2023), Mirowska (2020), Moritz et al. (2025), Oostrom et al. (2023), Suen and Hung (2023), Yan et al. (2024)
Social Identity Theory (Highhouse et al., 2007)	Proposes that people derive part of their self-concept from their membership in social groups.	Keppeler (2023)

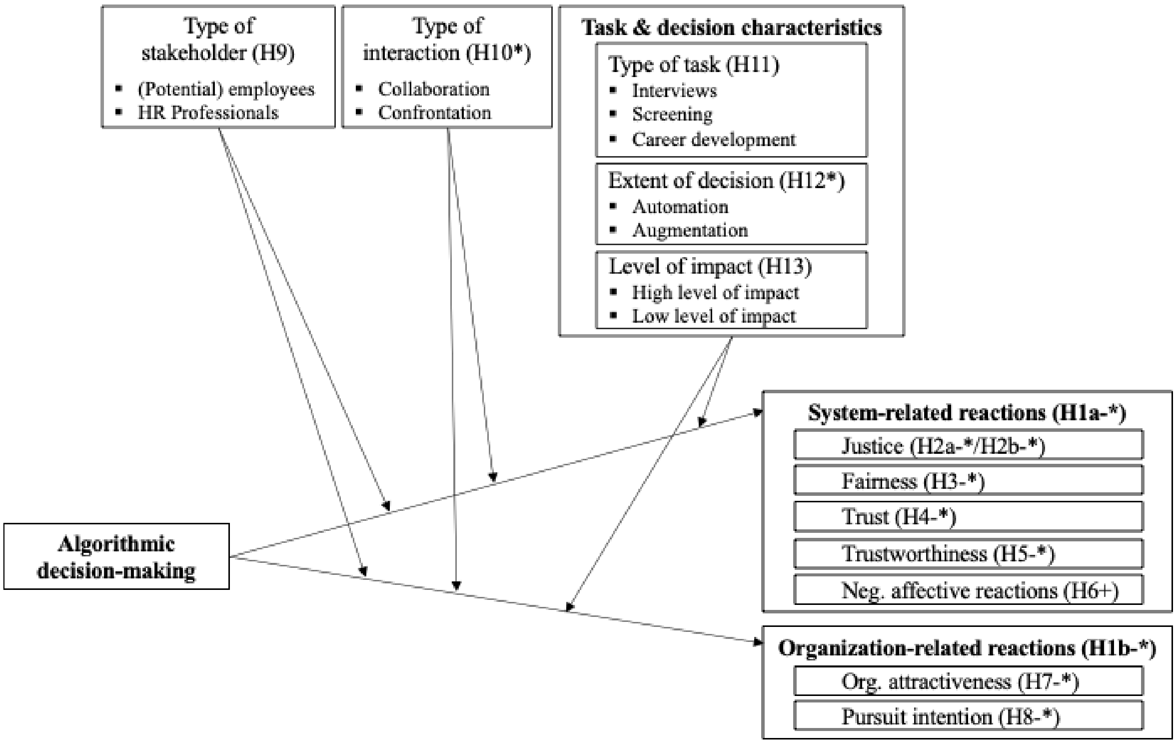
To address the aforementioned gaps, we conducted a meta-analysis of reactions to the use of ADM in the HRM domain. We analyzed 53 studies consisting of 73 individual samples (n = 24,578) published between 2019 and 2025. We offer three main contributions:

First, given the mixed results of previous research, this meta-analysis offers a synthesis of initial findings. We differentiate two reactions to ADM: *System-related reactions* include those reactions directly related to the system used for decision-making (e.g., justice, fairness, trust, trustworthiness, affective reactions). *Organization-related reactions* are linked to the respective organization which uses ADM systems (e.g., organizational attractiveness, job pursuit intention). By quantitatively synthesizing prior research on ADM in HRM, we provide the most robust evidence to date on reactions to ADM, thereby contributing to the creation of a new consensus in the field. This is a crucial step towards understanding reactions to ADM due to the inevitable adoption of these systems in organizations, as highlighted by Del Giudice et al. (2023).

Second, we follow the call by Colquitt et al. (2023) to provide evidence about relevant boundary conditions of the effects of ADM in HRM. Thus, the present meta-analysis examines study-specific moderators as boundary conditions, which include type of stakeholder (i.e., applicants, employees, and human resource [HR] professionals), type of interaction (i.e., confrontation vs. collaboration), type of task (i.e., interviews vs. screening applications vs. career development tasks), extent of decisions (i.e., automation vs. augmentation) as well as personal impact concerning the decision (i.e., being personally affected by the decision vs. others being affected by it). Figure 2-1 depicts our theoretical framework.

Third, we adopt CASA (Nass & Moon, 2000; Reeves & Nass, 1996) as a meta-theoretical framework to integrate and extend the diverse theories that have been applied in ADM research within HRM. CASA posits that individuals apply social norms and expectations to interactions with technology, including ADM systems, particularly when these systems simulate human characteristics (Gambino et al., 2020; Nass & Moon, 2000). This makes CASA well-suited to serve as a meta-theoretical framework: It allows us to build bridges between theories developed in human–human contexts—such as justice theory (Colquitt, 2001) or the integrative model of organizational trust (Mayer et al., 1995)—and their application in human–algorithm contexts. Table 2-1 illustrates the theoretical heterogeneity in prior ADM research and underscores the lack of synthesis across studies. By situating these constructs within CASA as a meta-theoretical framework, our overarching theoretical framework advances the field by clarifying which types of reactions are most relevant in ADM contexts and by specifying key boundary conditions, such as type of stakeholder or type of task. In doing so, our framework helps reconcile previous theoretical inconsistencies and outlines avenues for future research.

Figure 2-1: Overarching theoretical framework and brief summary of results



Note. Neg. = negative, org. = organizational. Algorithmic decision-making is always compared with human decision-making. The corresponding hypotheses are given in parentheses with an asterisk (*) if $p < .05$. For direct effects on system-related and organization-related outcomes, we included the proposed positive or negative direction of the hypotheses. H9–H13 include sub-hypotheses for (a) system-related and (b) organization-related reactions; however, for brevity, only the main hypotheses are presented. (Potential) employees include current employees and applicants. A high level of impact means that individuals are directly affected by the decision, whereas a low level of impact means that they are not directly affected.

2.2 Current research, framework, and hypotheses development

2.2.1 Theoretical framework

Current literature on human–algorithm interaction is scattered, with studies drawing on various theories to explain reactions to ADM use. Table 2-1 illustrates the fragmented use of theories, which gives rise to several challenges. A central challenge lies in the fact that most of the theories and frameworks were originally developed to understand human–human interactions and are now being adapted to the context of ADM. Theories, such as justice theory (Colquitt, 2001; Greenberg, 1987), social exchange theory (Blau, 1964), the integrative model of organizational trust (Mayer et al., 1995), and signaling theory (Spence, 1973), assume interpersonal dynamics, yet are frequently used to understand reactions to

ADM (e.g., Feldkamp et al., 2023; Schlicker et al., 2021). Even models intended for human–computer interaction, such as Zhang’s (2013) affective response model, are rooted in assumptions about emotional and social responses that originate from human–human contexts.

We propose that the CASA framework (Nass & Moon, 2000; Reeves & Nass, 1996) can be used as a meta-theoretical framework to explain the application of theories from human–human contexts to human–algorithm contexts. Here, CASA suggests that individuals perceive a computer as an independent entity, distinct from its programmer, and as possessing its own source of information (Sundar & Nass, 2000). Following this, people respond to technologies as they would to human beings, attributing human-like characteristics to these systems (Gambino et al., 2020), which allows the application of human–human theories in this context. By adopting CASA, we build a bridge between theories from human–human contexts, extend these theories to the ADM field, and offer an overarching theoretical framework for understanding how stakeholders react to ADM in HRM settings (see Figure 2-1). This integration offers a comprehensive theoretical foundation for analyzing how stakeholders perceive and engage with ADM systems in HRM.

2.2.2 System-related and organization-related reactions

System-related reactions reflect how individuals respond to ADM itself, based on their expectations of human-like behaviors, such as accountability, transparency, and responsiveness. We use system-related reactions as an umbrella term for all reactions relating to the ADM system including justice, fairness, trust, trustworthiness, and affective reactions. All of these specific variables are introduced separately as they offer a theoretical lens (i.e., justice theory, model of applicant reactions to employment selection systems, organizational trust, affective reactions; Colquitt, 2001; Gilliland, 1993; Greenberg, 1987; Leventhal, 1980; Mayer et al., 1995; Zhang, 2013) through which they can be investigated.

As ADM systems grow more interactive, expectations for human-like behaviors increase (Chandra et al., 2022). However, key differences between human–human and human–algorithm interactions remain. While ADM may offer advantages, such as increased objectivity, availability, efficiency, and feedback (Campion & Campion, 2023), it can also lead to unfulfilled social expectations (Glikson & Woolley, 2020). When individuals apply inappropriate norms from human interactions, they may experience negative system-related reactions to ADM, stemming from the system's inability to fully replicate human-like accountability and social responsiveness (Castelo et al., 2019; Zerilli et al., 2019). This gap reflects current limitations in the capacity of ADM systems to meet these expectations.

Beyond system-related reactions, we propose that transferring HRM tasks from humans to ADM can shape *organization-related reactions*. Drawing on signaling theory (Spence, 1973), we argue that ADM use sends observable cues, "signals," about the organization's values, priorities, and work culture to stakeholders such as applicants, employees, and HR professionals. For example, using ADM may signal innovation and efficiency (van Esch & Black, 2019) or a lack of personal engagement (Bedemariam & Wessel, 2023), impacting how stakeholders view the organization itself. In HRM, humans typically signal aspects of the organization through their behavior and communication. Based on the CASA framework (Nass & Moon, 2000), ADM becomes a "social actor" and thus a representative of the organization. However, algorithms might fail to meet expectations associated with human interaction, such as offering explanations or personal interactions (Gambino et al., 2020; Glikson & Woolley, 2020; Gursoy et al., 2019). This can result in a negative view on the organization, which might manifest in higher turnover intention or decreased organizational attractiveness. These reactions extend beyond the system itself, because ADM is perceived as a proxy for the organization (Acikgoz et al., 2020; Köchling et al., 2024). Thus, signaling theory and CASA jointly explain why ADM may negatively affect both system-related and organization-related reactions. We thus hypothesize:

Hypothesis 1: *Compared to human decision-making (HDM), the use of ADM is negatively associated with (a) system-related reactions (e.g., justice, trust) and (b) organization-related reactions (e.g., organizational attractiveness, perceived organizational support).*

2.2.2.1 Justice and fairness

Within the organizational justice literature, justice is characterized as “the degree to which one’s company […] is perceived to act consistently, equitably, respectfully, and truthfully in decision contexts” (Colquitt & Rodell, 2015, p. 188). Whereas scholars oftentimes use the terms fairness and justice interchangeably (Beugré, 2009), Goldman and Cropanzano (2015) distinguish between both terms by stating that justice rules precede fairness perceptions. Hence, individuals draw on procedural, interactional (i.e., consisting of interpersonal and informational justice), and distributive justice to determine whether they feel fairly treated (Colquitt, 2001; Cropanzano et al., 2015; Goldman & Cropanzano, 2015).

Justice and fairness play a pivotal role in organizational decision-making because higher levels of justice and fairness are associated with higher job performance (Cohen-Charash & Spector, 2001). Additionally, individuals are more likely to follow a decision when it is perceived to be fair (Lind, 2001), whereas low levels of justice and fairness are associated with counterproductive work behaviors (Cohen-Charash & Spector, 2001). Unlike HDM, ADM systems lack human intuition and the ability to assess an individual’s social context, a critical factor in shaping perceptions of justice (Köchling et al., 2024; Lee, 2018; Suen et al., 2019). Wiblen and Marler (2021) highlighted that ADM systems may be less effective when addressing subjective factors that cannot be objectively quantified. Thus, drawing on organizational justice theory (Colquitt, 2001), we propose that ADM use in HRM is associated with lower justice evaluations and hence lower fairness perceptions. Due to inappropriate social expectations (Nass & Moon, 2000), ADM is often viewed as lacking transparency (Suen & Hung, 2023) or being less responsive to individual circumstances (Glikson & Woolley, 2020). When ADM does not meet expectations for transparency (i.e., informational justice), personal and adequate treatment (i.e., interpersonal justice), and consistent procedures (i.e., procedural justice), these limitations together lower perceptions of justice and fairness. We thus hypothesize:

Hypothesis 2*: Compared to HDM, the use of ADM is negatively associated with (a) overall justice evaluations, (b) interactional justice, and (c) procedural justice.*[5]

Hypothesis 3*: Compared to HDM, the use of ADM is negatively associated with perceived fairness.*

2.2.2.2 Trust and trustworthiness

Trust is defined as the willingness to expose oneself to the actions of another party, driven by the belief that "the other will carry out a specific action of importance to the trusting party, regardless of the ability to observe or control that party" (Mayer et al., 1995, p. 712). Mayer and colleagues (1995) suggest that trustworthiness precedes trust. Trustworthiness refers to the qualities or attributes of the trustee that make them worthy of trust and comprises three factors: ability, benevolence, and integrity (Mayer et al., 1995). Cabiddu et al. (2022) propose that ADM may be perceived as more trustworthy due to higher precision and objectivity, which has been supported by empirical evidence (Logg et al., 2019; Wang & Benbasat, 2016). In the HRM context, however, the use of ADM might be associated with lower perceived trustworthiness and trust compared to HDM due to the nature of tasks and decisions, which are often associated with emotional or social skills, and the perceived lack of high morals and values (i.e., integrity and benevolence; Lacroux &

[5] Due to the limited availability of primary studies, we were unable to test for distributive justice.

Martin-Lacroux, 2022; Langer et al., 2023). Additionally, Cabiddu et al. (2022) propose that trust in algorithms will develop over time. However, the use of ADM systems in HRM is still rather uncommon, which indicates that individuals might generally have lower trust perceptions. Furthermore, prior research demonstrates ADM is associated with a lack of human warmth (Lee, 2018), responsibility (Binns et al., 2018), and familiarity (Cabiddu et al., 2022), resulting in lower perceived trustworthiness and trust by individuals. We thus hypothesize:

Hypothesis 4: *Compared to HDM, the use of ADM is negatively associated with perceived trust.*

Hypothesis 5: *Compared to HDM, the use of ADM is negatively associated with perceived trustworthiness.*

2.2.2.3 Affective reactions

Extant research focused on emotional or affective responses as consequences of ADM (e.g., Köchling et al., 2023; Langer, König, & Papathanasiou, 2019). Based on Zhang's (2013) affective response model, affective responses to technology, such as anxiety or emotional creepiness, arise from (unconscious) evaluation processes. These evaluation processes impact how individuals think about technology and shape their subsequent behavior by prompting specific actions, such as engaging with or avoiding technology, depending on whether the emotions are positive or negative (Zhang, 2013). One of the most common measurements assessing affective responses to ADM is emotional creepiness (Köchling & Wehner, 2023; Langer & König, 2018; Oostrom et al., 2023), referring to the "potentially negative and uncomfortable emotional response paired with perceptions of ambiguity toward a person, technology or even during a situation" (Langer & König, 2018, p. 2). Emotional creepiness oftentimes occurs when individuals experience uncertainty or unpredictability (Langer & König, 2018). Because ADM systems are relatively new to the

HRM field, stakeholders might experience uncertainty and, hence, exhibit stronger negative affective reactions towards ADM use. We thus hypothesize:

Hypothesis 6*: Compared to HDM, the use of ADM is positively associated with negative affective reactions (e.g., emotional creepiness, creepy ambiguity).*

2.2.2.4 Organizational attractiveness and job pursuit intention

Drawing on signaling theory (Spence, 1973), ADM also impacts (potential) employees' and HR professionals' perceptions of the organization, which we operationalize with the most common measurements for organizational reactions within the ADM–HRM field, namely organizational attractiveness and job pursuit intention (Qu et al., 2023). While organizational attractiveness describes the general positive attitude or affect towards an organization reflecting a desire to establish a relationship (Gomes & Neves, 2011; Highhouse et al., 2003), job pursuit intention describes the "intention to pursue a job or to remain in the applicant pool" (Chapman et al., 2005, p. 929). When organizations use ADM systems for recruitment and selection processes, such as using asynchronous video interviews evaluated by ADM systems, applicants may interpret this as a signal of what it is like to work for the potential employer or how the organization will treat them in the future. Here, individuals may perceive the company as modern, data-driven, and efficient. Conversely, individuals could perceive this as an indication of being treated impersonally (Hunkenschroer & Luetge, 2022; Lee, 2018), a violation of reciprocity norms (Oostrom et al., 2023), and a signal of an organizational culture that devalues personal contact(Mirowska & Mesnet, 2022). Both types of perceptions—positive and negative—ultimately impact organizational attractiveness and job pursuit intention, as individuals use observable signals (i.e., such as ADM use) to infer unobservable information about the potential employer.

Drawing on CASA, we suggest that reactions to ADM are more negative than those of HDM, as individuals may sense an absence of interpersonal interaction and reasoning

(Binns et al., 2018), potentially leading to feelings of dehumanization (Bankins et al., 2022). For current employees and HR professionals, the use of ADM for career development (Köchling et al., 2024) or performance measurements (Zhang & Amos, 2023) introduces novel work dynamics. This shift may evoke a sense of continuous electronic surveillance, dehumanization, or the assumption that evaluations by algorithms might be disadvantageous (Bankins et al., 2022). We thus hypothesize:

***Hypothesis** 7: Compared to HDM, the use of ADM is negatively associated with organizational attractiveness.*

***Hypothesis 8:** Compared to HDM, the use of ADM is negatively associated with job pursuit intention.*

2.2.3 Boundary conditions affecting reactions to ADM

Following our meta-theoretical framework based on CASA (Nass & Moon, 2000), we propose that the relationship between ADM usage and reactions is impacted by boundary conditions. Based on theories and findings from human–human and human–algorithm contexts, we differentiate between type of stakeholder (Langer & Landers, 2021), type of interaction (Burton et al., 2020), as well as task and decision characteristics (i.e., type of task, extent of decision, personal impact; Deci & Ryan, 1985; Höddinghaus et al., 2021; Lee, 2018).[6]

2.2.3.1 Type of stakeholder

We propose that ADM usage involves different *types of stakeholders* (i.e., HR professionals, line managers, and [potential and/or current] employees; adapted from the HR triad; Jackson & Schuler, 2003) in the HRM context. The HR triad emphasizes the shared responsibility of the involved stakeholders: HR professionals offer HR expertise (e.g.,

[6] We acknowledge that it was not possible to include all boundary conditions identified in the literature. Instead, we focused on those conditions that could be empirically tested based on the available primary studies.

training, bonus payments, and promotions; Budhwar et al., 2022), line managers implement HR practices, and current employees actively participate in shaping their own development. We also include potential employees (i.e., applicants from the labor market and current employees applying for internal positions) because they represent key stakeholders who are directly affected by ADM decisions but have limited influence over them.

In accordance with self-determination theory (Deci & Ryan, 1985), which posits that individuals are motivated to grow when their basic psychological needs for autonomy, competence, and relatedness are fulfilled, we argue that competence is an important factor in differentiating stakeholder reactions to ADM. In particular, HR professionals may fulfill their need for competence by learning new skills and overcoming challenges when using ADM systems over time compared to (potential) employees who are more interested in shaping their own development (Deepa et al., 2024). This aligns with the differentiation of stakeholders related to the HR triad and the classification into *first, second,* and *third parties* by Langer and Landers (2021), which highlights that stakeholders have distinct roles and levels of involvement in HRM. Here, first parties use ADM to make decisions affecting others (i.e., HR professionals), second parties are directly impacted by ADM (i.e., applicants, employees), and third parties observe ADM use and may be indirectly affected in the near future (Langer & Landers, 2021). Therefore, HR professionals can use ADM to enhance decision-making and support daily tasks, whereas employees and applicants, focused on personal outcomes, may feel more threatened by ADM, as its communication may fail to meet expectations compared to human interaction. Specifically, in hiring, HR professionals may value ADM for its efficiency in screening applicants, whereas applicants might perceive automated decisions as impersonal or unfair. We thus hypothesize:

Hypothesis 9: *Type of stakeholder moderates the relationship between the use of ADM and (a) system-related reactions and (b) organization-related reactions, such that* *negative reactions to ADM usage are stronger for (potential) employees than for HR professionals.*[7]

2.2.3.2 Type of interaction

Our framework distinguishes stakeholders' *type of interaction*, and we categorize this as confrontation or collaboration. Confrontation refers to a type of interaction in which individuals are subjected to decisions made by ADM systems without any opportunity to provide input (e.g., applicants receiving a hiring decision). Collaboration means that stakeholders are actively involved in decision-making (e.g., HR professionals using ADM insights; Daugherty & Wilson, 2018). Following our distinction, the same task, for example, automated CV screening, can involve both collaboration and confrontation, depending on the individuals' perspective. In our example, HR professionals could use automated CV screening to decide who to invite for an interview (i.e., collaboration), whereas applicants may be rejected by ADM based on CV screening (i.e., confrontation). When individuals are confronted with ADM, they may perceive a lack of agency or control over the process, which can increase negative reactions such as lower trust, frustration, or feelings of unfairness (Burton et al., 2020). In contrast, collaborating with ADM systems allows individuals to retain a sense of control and agency, as they are actively involved in the decision-making process (Langer et al., 2025). This interaction can foster trust and acceptance, as the ADM system is perceived as a tool that enhances, rather than replaces, human judgment (Burton et al., 2020). We thus hypothesize:

Hypothesis 10: *Type of interaction moderates the relationship between the use of ADM and (a) system-related reactions and (b) organization-related reactions, such that*

[7] We were not able to include line managers here as none of the primary studies included line managers as a stakeholder group.

negative reactions are stronger when individuals are confronted with ADM than when they collaborate with it.

2.2.3.3 Type of task

We argue that tasks that require tacit knowledge or social skills, such as interviews or career development, could be perceived as less suitable for ADM and are likely to evoke more negative reactions compared to mechanical tasks, such as resume screening or work scheduling (Castelo et al., 2019; Lee, 2018). Following the cognitive psychology and human factors literature, certain skills or types of knowledge, such as procedural and tacit knowledge, are not suitable for programming in computer programming languages (Lee, 2018; Reber, 1989). These types of knowledge and respective skills are typically acquired through experiences and practices, encompassing intuitive judgments and the understanding of emotions (Lee, 2018). Consequently, ADM is often considered more appropriate for tasks that are structured, quantifiable, and less dependent on human judgment, such as resume screening or data analysis, where standardized processes are prioritized over interpersonal interaction (Lee, 2018). We thus hypothesize:

Hypothesis 11: *Type of task (interviews, screening, and career development) moderates the relationship between the use of ADM and (a) system-related reactions and (b) organization-related reactions, such that negative reactions to ADM usage are stronger when it is used for interviews and career development compared to screening.*

2.2.3.4 Extent of decision

Extent of decision comprises whether ADM automates or augments human decisions (Langer & Landers, 2021). Here, researchers differentiate between ADM systems that autonomously make decisions without human involvement (i.e., automation), and those that serve as decision support, namely augmentation (Candrian & Scherer, 2022). ADM systems that fully automate decisions may evoke more negative responses (e.g., reduced justice or trust) due to a perceived loss of autonomy (Deci & Ryan, 1985). In contrast, using ADM for decision support—where human intervention is possible—may result in less negative reactions (e.g., higher perceptions of justice or trust), because individuals perceive that a human retains control over final decisions (Burton et al., 2020). We thus hypothesize:

Hypothesis 12: *Extent of decision moderates the relationship between the use of ADM and (a) system-related reactions and (b) organization-related reactions, such that negative reactions to ADM usage are stronger when it is used for automation compared to augmentation.*

2.2.3.5 Level of personal impact

We argue that the *level of personal impact* can affect stakeholders' reactions to ADM. When individuals are directly affected by ADM decisions (e.g., applicants, second parties), they may experience stronger negative reactions due to a perceived loss of control by themselves or humans (Langer & Landers, 2021). These individuals might feel psychologically closer to the situation as the outcomes of the decisions affect their lives directly in the long run (e.g., getting a job or being laid off; Langer et al., 2019; Trope & Liberman, 2010). In such contexts, individuals may evaluate the introduction of ADM more critically as the decision process becomes more salient due to its potential impact on their lives. In contrast, individuals who use ADM in the decision-making process (e.g., HR professionals; first parties) may respond less negatively, because they feel less personally connected to ADM and are less likely to be significantly impacted by its outcomes (Deci & Ryan, 1985; Trope & Liberman, 2010). In turn, reactions from less affected stakeholders could be more positive as they include a broader, rather distant decision rationale. We thus hypothesize:

Hypothesis 13: *Level of personal impact moderates the relationship between the use of ADM and (a) system-related reactions and (b) organization-related reactions, such that negative reactions are stronger when personal impact is high compared to when it is low.*

2.3 Method

2.3.1 Literature search and data extraction

We conducted a literature search of articles published between January 2010 and March 2025. The search process consisted of four steps drawn from the PRISMA protocol (preferred reporting items for systematic reviews and meta-analyses; Liberati et al., 2009; Page et al., 2021): identification, screening, eligibility, and inclusion. A flow diagram of the search process is provided in Appendix A1 and an overview of included studies is provided in Appendix A2.

As the first step (i.e., identification), the sample collection was based on two complementary search strategies (i.e., database search and manual search). We used two databases for our initial search, specifically, EBSCO Business Source Premier and Web of Science (Social Sciences Citation Index). We applied search terms used in previous studies (Hunkenschroer & Luetge, 2022; Köchling & Wehner, 2020) and extended these by adding additional synonyms. We searched for articles by applying the following combination of search terms: "algorithm," "artificial intelligence," or "algorithmic decision," potential reactions, such as "fairness" or "trust," and our context "human resources" or "HRM," whereas we extended all search words with further terms and synonyms (i.e., fairness was extended by justice terms). The initial search resulted in 8,790 studies. After eliminating duplicates, we kept 8,107 unique studies for further evaluation. Additionally, following the approach by Ravid et al. (2023), we applied an issue-by-issue search of journals and conference proceedings,[8] resulting in 34 additional articles for initial screening. We identified relevant journals by using the FT50 ranking and by using those journals most frequently represented in the initial database search. Further, we compared our findings with samples from recently published reviews (Bankins et al., 2024; Basu et al., 2023; Budhwar et al., 2023; Cheng & Hackett, 2021; Giermindl et al., 2022; Kaushal et al., 2023; Köchling & Wehner, 2020; Langer & Landers, 2021; Pan & Froese, 2023; Votto et al., 2021; Vrontis et al., 2022) on the topic of ADM use in organizations. We identified one article from this step of the search process. Discrepancies between our meta-analysis and the reviews were often due to the methods used in the primary studies (e.g., qualitative approaches), but more importantly, they resulted from differences in the context, as our analysis focused specifically on the HRM context. As an additional step, we used AI-based tools, such as Connectedpapers, ResearchRabbit, and Litmaps, to identify other articles that would fit our inclusion criteria. However, no additional articles were identified through this process. We contacted authors of papers already included for additional published and unpublished articles, resulting in 15 additional articles. To complement our manual search, we conducted a forward search for articles (i.e., screening articles that cited the articles included in the analysis), resulting in three additional articles.

In the second step (i.e., screening), two of the authors independently examined titles, key words, and abstracts of the identified studies and excluded all studies that did not fit into the scope of the meta-analysis (i.e., studies not focused on HRM). Here, we used the tool Rayyan (Ouzzani et al., 2016). While we did not use Rayyan's automation features, we did use it to conduct blinded and independent screening. Consistent with previous meta-

[8] Academy of Management Journal, Academy of Management Proceedings, Big Data and Society, Computers in Human Behavior, Human Resource Management, Information Systems Frontiers, Information Systems Research, International Journal of Selection and Assessment, Journal of Applied Psychology, Journal of Business Ethics, Journal of Management, Journal of Management Information Systems, Journal of Management Studies, Management Information Systems Quarterly, Management Science, Organizational Behavior and Human Decision Processes, Personnel Psychology, Research Policy, Technovation.

analyses, we calculated intercoder agreement as a measurement for reliability (Ravid et al., 2023). Initial intercoder agreement was 99.42%, corresponding to 47 disagreements among 8,107 articles. The authors discussed and resolved all disagreements collaboratively.

For the third and fourth step (i.e., eligibility and inclusion), two authors examined the full text of the studies according to our a priori inclusion criteria: Included publications should report (a) at least one quantitative study, (b) the reaction to ADM in comparison to HDM, and (c) at least one correlation coefficient between the use of ADM and system-related reactions (e.g., fairness, emotional creepiness) or organization-related reactions (e.g., organizational attractiveness, litigation intention, turnover intention). For this step, we read 84 full-text articles stemming from the initial database search and 53 articles stemming from our manual search. The most common reason for article exclusion was that the studies did not compare HDM to ADM but rather compared different ADM systems or assessed general ADM reactions (e.g., Edwards et al., 2024). When studies fit our inclusion criteria but did not report the data needed, we contacted authors to ask for further information. Studies were excluded if the authors did not respond or were unable to provide the requested data. Our final dataset consists of 53 studies and 73 samples (n = 24,578; 52.97% female, mean age = 34.32 years).

2.3.2 Included studies

As we required studies to compare ADM to HDM for the analysis, most of the included studies used an experimental approach (except for Dutta & Mishra, 2024; Keppeler, 2023; Suen & Hung, 2023, who all employed field experiments). Typically, participants in the experimental group were presented with a scenario—such as a job application—in which an ADM-based system made a decision (e.g., regarding their hiring). Participants in the control group were presented the same scenario with one difference: The decision was made by a human entity (e.g., a HR professional). Consequently, the studies measured differences between the two groups in their reactions to the scenario. The effect sizes coded in each study reflected the mean differences. We coded the correlation coefficient such that the treatment (i.e., ADM) group was coded as 1 and the control (i.e., HDM) group was coded as 0 (HDM = 0, ADM = 1). A positive correlation coefficient indicated a higher mean in the outcome variable for the ADM group versus the HDM group. We were not able to apply Cohen's *d*, an effect size measure traditionally used in meta-analyses, as most of the primary studies did not provide sufficient information (e.g., missing information on standard deviations).

2.3.3 Aggregation levels and coding

To address the heterogeneity of primary studies (Steel et al., 2021), we analyzed different aggregation levels of the dependent variables. First, we aggregated all variables subsumed under the *main relationship* (i.e., including both system-related and organization-related constructs). Second, we grouped the effect sizes in two broad aggregates of system-related and organization-related variables to understand and to address differences between both types of reactions. Third, we grouped the effect sizes according to the specific constructs, namely *justice, fairness, trust, trustworthiness, affective reactions, organizational attractiveness,* and *job pursuit intention*. For justice, we were able to further differentiate between *procedural justice* and *interactional justice*. Two of the authors independently coded the 365 effect sizes of the first model according to the guidelines by Steel et al. (2021). We calculated the intercoder agreement, which was 96.44% (13 disagreements) for all coded relationships. All disagreements were resolved through discussions.

2.3.3.1 Coding of system-related and organization-related reactions and moderators

System-related reactions comprise all constructs that refer to reactions directly linked to the system used for decision-making, such as perceived fairness or justice. For organization-related reactions, we grouped the effect sizes of variables that were related to organizational parameters. To identify the respective correlations, we examined the research model and the descriptions of the variables in the studies. In addition to the effect sizes reported in the primary studies, we included study-level moderators. We used information on *stakeholders* targeted in the respective primary study, *type of interaction* and *type of task*, *extent of decision*, and *personal impact*. Here, we report on the first two levels of aggregation (i.e., main relationships, system-related reactions, and organization-related reactions) because the lowest aggregation level separated into specific constructs (e.g., justice, fairness) did not allow meaningful investigation due to the low number of available samples.

2.3.4 Analytical procedure

To address multiple correlation coefficients per study and the inherent dependability of effect sizes, we used three-level random and mixed effects models (Assink & Wibbelink, 2016; Cheung, 2019). A three-level model analysis has the advantage over a two-level random effect model to add a third variance component to understand the differences in correlation coefficients within the respective studies (Assink & Wibbelink, 2016; Hox et al., 2017). Furthermore, this model acknowledges that the multiple correlation coefficients are nested (within each primary study) and not statistically independent. Three-level models can be used to create averages per study and—in contrast—enable the investigation of the extent of heterogeneity within (and between) studies, while addressing nonindependence. The random effects model is a simplified version of a regression model without predictors:

$$y_{ij} = \beta_0 + u(2)_{ij} + u(3)_j + \varepsilon_{ij}$$

with y_{ij} representing the correlation coefficient i in study j that, in our case, is the relationship between the use of ADM and system-related or organization-related reactions. The intercept β_0 reflects the weighted correlation coefficient, $Var(u(2)_{ij}) = \tau^2(2)$ reflects the true within-study variance, $Var(u(3)_j) = \tau^2(3)$ represents the true between-study variance, and $Var(\varepsilon_{ij})$

reflects random sampling error. Following Aguinis et al.'s (2011) recommendations, we report the square root of the two tau-squares, representing an estimate of the "true" (i.e., nonrandom) heterogeneity in the form of an easily interpretable standard deviation of the effect size. We refer to the within-study heterogeneity as τ_w and to the between-study heterogeneity as τ_b.

When analyzing moderators (i.e., type of stakeholder, type of interaction, type of task, extent of decision, personal impact), the model is expanded to a mixed effects regression model by including these as predictors, potentially allowing the prediction of portions of the between-study or within-study variance. In this model, one or several predictors, x_j, are added. When analyzing categorical moderators (e.g., type of task), we used dummy coding. For these categorical moderators, we estimated the model as a mixed effects model, excluding the intercept β_0. As a result, the reported coefficients of the model represent the weighted average in each category of the moderator depicted by $\hat{\rho}$. To investigate differences across levels of the moderator, we report the result of an omnibus χ^2-test (i.e., the Q_M test), which assesses whether significant differences exist in the average effect sizes across the levels of the moderator, thereby indicating whether the moderator explains variability in the respective effect size. A significant Q_M value suggests that the moderator contributes to differences in the observed relationships. For this test, the selection of the reference category (determined by the first alphabetical occurrence) is inconsequential, as the test examines differences across all included categories collectively. The analyses were conducted with the open-source software R (R Core Team, 2020). Packages used were tidyverse (Wickham et al., 2019) for data management, and metafor (Viechtbauer, 2010) and metaSEM (Cheung, 2019) for the meta-analytical procedures.

2.4 Results

2.4.1 Main relationships between ADM and reactions

Table 2-2 presents the results for the bivariate correlation analysis of the main effects. Regarding the main relationship, findings reveal that ADM was perceived negatively, including system-related and organization-related reactions ($\hat{\rho}$ = –.13, p < .001). In addition, ADM was negatively related to system-related reactions ($\hat{\rho}$ = –.12, p < .001), including justice ($\hat{\rho}$ = –.17, p < .001), and more specifically, also to procedural ($\hat{\rho}$ = –.14, p < .001) and interactional justice ($\hat{\rho}$ = –.21, p < .001), fairness ($\hat{\rho}$ = –.13, p = .002), trust ($\hat{\rho}$ = –.20, p < .001), and trustworthiness ($\hat{\rho}$ = –.13, p = .033). We found no significant relationship between ADM and negative affective reactions ($\hat{\rho}$ = .08, p = .099). Overall, our results supported hypotheses 1a and 2 to 5, demonstrating a negative relationship between ADM and system-related reactions, justice, fairness, trust, and trustworthiness, but did not support hypothesis 6 related to negative affective reactions. Further, we found that ADM was negatively related to organization-related reactions ($\hat{\rho}$ = –.15, p < .001), including organizational attractiveness ($\hat{\rho}$ = –.16, p < .001) and job pursuit intention ($\hat{\rho}$ = –.12, p = .003), supporting hypotheses 1b, 7, and 8.

Table 2-2: Results of the bivariate meta-analysis

Dependent variable	k	m	N	$\hat{\rho}$	SE	p-value	95% CI		τ_w (I^2_w)	τ_b (I^2_b)	Publication bias	Outlier
Main relationship	73	365	24,578	$-.13^{***}$	.02	< .001	–.16	–.09	.19 (.66)	.13 (.30)	No	Yes (Corr. $\hat{\rho}$ = –.14, p < .001)
System-related reactions	69	286	23,601	$-.12^{***}$	.02	< .001	–.16	–.08	.22 (.75)	.12 (.21)	No	Yes (Corr. $\hat{\rho}$ = –.13, p < .001)
Justice	29	88	7,015	$-.17^{***}$	.03	< .001	–.23	–.10	.16 (.57)	.13 (.36)	Yes (Corr. $\hat{\rho}$ = –.15, p < .001)	Yes (Corr. $\hat{\rho}$ = –.14, p < .001)
Procedural justice	26	61	6,191	$-.14^{***}$	.03	< .001	–.20	–.07	.16 (.58)	.12 (.34)	Yes (Corr. $\hat{\rho}$ = –.11, p < .001)	No
Interactional justice	12	23	2,993	$-.21^{***}$	.05	< .001	–.32	–.11	.15 (.51)	.14 (.42)	No	No
Fairness	33	40	14,109	$-.13^{**}$	.04	.002	–.21	–.05	.16 (.44)	.17 (.52)	No	No
Trust	21	28	5,619	$-.20^{***}$	.06	< .001	–.31	–.09	.15 (.35)	.20 (.60)	No	No
Trustworthiness	5	16	1,069	$-.13^{*}$	.06	.033	–.25	–.01	.15 (.68)	.08 (.21)	No	No
Negative affective reactions	16	38	6,811	.08	.05	.099	–.02	.18	.06 (.09)	.20 (.86)	No	Yes (Corr. $\hat{\rho}$ = .12, p < .001)
Organization-related reactions	34	79	11,230	$-.15^{***}$	.02	< .001	–.19	–.10	.04 (.06)	.13 (.79)	No	No
Organizational attractiveness	22	39	5,572	$-.16^{***}$	.03	< .001	–.22	–.11	.06 (.20)	.11 (.59)	No	No
Job pursuit intention	10	20	2,813	$-.12^{**}$	.04	.003	–.20	–.04	.07 (.26)	.10 (.56)	No	No

Note. *** $p < .001$, ** $p < .01$, * $p < .05$; k = number of independent samples; m = number of effect sizes; n = number of analyzed individuals; $\hat{\rho}$ = weighted average effect size; SE = standard error; CI = confidence interval (lower and upper bound of a 95% CI given); τ_w and τ_b = estimate (standard deviation) of the true heterogeneity within and between studies; I^2_w and I^2_b = percentage of true heterogeneity in observed heterogeneity; Corr. = corrected. Whenever the estimate is zero, the random effects model was reduced to a fixed effects model. Independent variable is HDM versus ADM for all relationships (HDM = 0, ADM = 1).

2.4.2 Moderator Analysis

2.4.2.1 Type of stakeholder and type of interaction

Results of the moderator analysis for type of stakeholder and type of interaction are reported in Table 2-3. For type of stakeholder, we found a significant relationship between ADM on the main relationship for applicants ($\hat{\rho}$ = –.14, p < .001) and employees ($\hat{\rho}$ = –.13, p < .001), but not for HR professionals ($\hat{\rho}$ = –.07, p = .222). Similarly, on the level of system-related reactions, we found significant effects for applicants ($\hat{\rho}$ = –.12, p < .001) and employees ($\hat{\rho}$ = –.13, p < .001), but not for HR professionals ($\hat{\rho}$ = –.08, p = .207). Results showed a significant relationship between ADM and organization-related reactions for applicants ($\hat{\rho}$ = –.17, p < .001) and for employees ($\hat{\rho}$ = –.10, p = .012), but not for HR professionals ($\hat{\rho}$ = –.19, p = .196). The difference tests showed no significant differences for all aggregation levels; hence, we cannot confirm our hypothesis 9.

For type of interaction, results indicated that on the main relationship, confrontation has significant effects ($\hat{\rho}$ = –.14, p < .001) compared to collaboration ($\hat{\rho}$ = –.04, p = .417). A similar pattern emerges for system-related reactions (collaboration: $\hat{\rho}$ = –.04, p = .360; confrontation: $\hat{\rho}$ = –.13, p < .001) and organization-related reactions (collaboration: $\hat{\rho}$ = –.02, p = .881; confrontation: $\hat{\rho}$ = –.16, p < .001). The difference tests indicated significant differences on the main relationship ($\chi^2(1) = 4.64$, p = .030), but not on the level of system-related or organization-related reactions, leading us to reject our hypothesis 10.

Table 2-3: Results of the bivariate meta-analysis including type of stakeholder and type of interaction as moderators

Dependent variable	k	m	N	$\hat{\rho}$	SE	p-value	95% CI		$Q_M(df)$
Main relationship									
Type of stakeholder									$\chi^2(2) = 0.84$, $p = .660$
Applicants	42	254	11,521	$-.14^{***}$	.03	< .001	−.18	−.09	
Employees	22	91	7,736	$-.13^{***}$	.04	< .001	−.20	−.05	
HR professionals	9	29	5,321	−.07	.06	.222	−.19	.04	
Type of interaction									$\chi^2(1) = 4.64^{*}$, $p = .030$
Collaboration	14	63	7,405	−.04	.04	.417	−.12	.05	
Confrontation	55	294	16,432	$-.14^{***}$	.02	< .001	−.18	−.10	
System-related reactions									
Type of stakeholder									$\chi^2(2) = 0.50$, $p = .780$
Applicants	39	183	10,736	$-.12^{***}$	.03	< .001	−.17	−.07	
Employees	21	75	7,544	$-.13^{***}$	.04	< .001	−.20	−.05	
HR professionals	9	28	5,321	−.08	.06	.207	−.20	.04	
Type of interaction									$\chi^2(1) = 3.28$, $p = .070$
Collaboration	14	61	7,405	−.04	.04	.360	−.13	.05	
Confrontation	51	219	15,455	$-.13^{***}$	.02	< .001	−.17	−.08	
Organization-related reactions									
Type of stakeholder									$\chi^2(2) = 2.18$, $p = .340$
Applicants	21	62	5,495	$-.17^{***}$	.03	< .001	−.23	−.12	
Employees	12	16	5,530	$-.10^{*}$	.04	.012	−.18	−.02	
HR professionals	1	1	205	−.19	.15	.196	−.48	.10	
Type of interaction									$\chi^2(1) = 2.10$, $p = .150$
Collaboration	2	2	1,384	−.02	.10	.881	−.21	.18	
Confrontation	31	75	9,430	$-.16^{***}$	.02	< .001	−.21	−.11	

Note. *** $p < .001$, ** $p < .01$, * $p < .05$; k = number of independent samples, m = number of effect sizes, N = number of analyzed individuals; $\hat{\rho}$ = weighted average effect size; SE = standard error; CI = confidence interval (lower and upper bound of a 95% CI given). Differences were tested with a Q_M test, which is a χ^2-test with the depicted degrees of freedom. Independent variable is HDM versus ADM for all relationships (HDM = 0, ADM = 1).

2.4.2.2 Type of task, extent of decision, and personal impact

Table 2-4 summarizes the findings of the moderator analysis for type of task, extent of decision, and personal impact. Regarding the type of task on the main relationship, ADM was perceived negatively in interviews ($\hat{\rho} = -.17$, $p < .001$), screening ($\hat{\rho} = -.15$, $p < .001$), and career development ($\hat{\rho} = -.18$, $p < .001$). Similarly, for system-related reactions, results indicated that ADM in interviews ($\hat{\rho} = -.15$, $p < .001$), screening ($\hat{\rho} = -.14$, $p = .002$), and career development ($\hat{\rho} = -.18$, $p < .001$) was perceived negatively. In terms of organization-related reactions, we could not confirm this relationship for career development ($\hat{\rho}$ = –.11, p = .116), compared to interviews ($\hat{\rho}$ = –.26, p < .001) and screening ($\hat{\rho}$ = –.16, p = .004). The difference tests were not significant; therefore, we reject our hypothesis 11.

For the extent of the decision, we found a significant relationship between ADM on the main relationship when decisions were automated ($\hat{\rho}$ = –.16, p < .001) compared to when decisions where augmented ($\hat{\rho}$ = –.04, p = .220). On the level of system-related reactions, we found associations between ADM and automation ($\hat{\rho}$ = –.16, p < .001), but not between ADM and augmentation ($\hat{\rho}$ = –.04, p = .217). Regarding organization-related reactions, the relationships between ADM and automation ($\hat{\rho}$ = –.13, p < .001) as well as between ADM and augmentation ($\hat{\rho}$ = –.16, p < .001) were significant. Our results revealed significant differences for the main relationship and system-related reactions, but not for organization-related reactions, thereby confirming H12a and rejecting H12b.

For the level of personal impact, our results indicated that ADM was perceived negatively in high personal impact ($\hat{\rho}$ = –.14, p < .001) and low personal impact scenarios ($\hat{\rho}$ = –.09, p = .029) on the main relationship. Regarding system-related reactions, ADM was significantly related to high personal impact ($\hat{\rho}$ = –.13, p < .001), but not to low personal impact ($\hat{\rho}$ = –.06, p = .182). In terms of organization-related reactions, we found significant associations between ADM and high personal impact ($\hat{\rho}$ = –.13, p < .001) as well as between ADM and low personal impact ($\hat{\rho}$ = –.26, p < .001). The difference tests suggested that there are no significant differences between low or high levels of impact, leading us to reject hypothesis 13.

Table 2-4: Results of the bivariate meta-analysis including type of task, extent of decision, and personal impact as moderators

Dependent variable	*k*	*m*	*N*	$\hat{\rho}$	*SE*	*p*-value	95% *CI*		$Q_M(df)$
Main relationship									
Type of task									$\chi^2(3) = 3.78, p = .290$
Interviews	18	86	3,317	–.17***	.04	< .001	–.24	–.09	
Screening	18	64	8,006	–.15***	.04	< .001	–.23	–.07	
Career development	8	34	3,111	–.18***	.05	< .001	–.28	–.07	
Extent of decision									$\chi^2(2) = 7.13^*, p = .030$
Automation	48	201	17,276	–.16***	.02	< .001	–.21	–.11	
Augmentation	21	137	6,171	–.04	.04	.220	–.11	.03	
Personal impact									$\chi^2(1) = 0.85, p = .360$
High impact	55	299	17,473	–.14***	.02	< .001	–.18	–.09	
Low impact	18	66	7,105	–.10*	.04	.029	–.17	–.01	
System-related reactions									
Type of task									$\chi^2(3) = 4.00, p = .260$
Interviews	17	71	3,133	–.15***	.04	< .001	–.22	–.07	
Screening	16	46	7,405	–.14**	.04	< .001	–.22	–.05	
Career development	8	28	3,111	–.18***	.06	< .001	–.29	–.07	
Extent of decision									$\chi^2(2) = 6.88^*, p = .030$
Automation	44	164	16,299	–.16***	.03	< .001	–.21	–.11	
Augmentation	20	104	6,270	–.04	.03	.217	–.11	.03	
Personal impact									$\chi^2(1) = 2.30, p = .130$
High impact	53	228	16,947	–.13***	.02	< .001	–.18	–.09	
Low impact	16	58	6,654	–.06	.04	.182	–.14	.03	
Organization-related reactions									
Type of task									$\chi^2(3) = 4.46, p = .220$
Interviews	7	15	1,735	–.26***	.05	< .001	–.36	–.16	
Screening	6	18	1,532	–.16**	.06	.004	–.27	–.05	
Career development	4	6	2,146	–.11	.07	.116	–.24	.03	
Extent of decision									$\chi^2(2) = 0.53, p = .770$
Automation	19	37	6,579	–.13***	.03	< .001	–.20	–.07	
Augmentation	11	33	3,742	–.16***	.04	< .001	–.24	–.08	
Personal impact									$\chi^2(1) = 2.69, p = .100$
High impact	30	71	10,368	–.13***	.02	< .001	–.18	–.09	
Low impact	4	8	862	–.26***	.07	< .001	–.39	–.12	

Note. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$; k = number of independent samples, m = number of effect sizes, N = number of analyzed individuals; $\hat{\rho}$ = weighted average effect size; *SE* = standard error; *CI* = confidence interval (lower and upper bound of a 95% CI given). Differences were tested with a Q_M test, which is a χ^2-test with the depicted degrees of freedom. Independent variable is HDM versus ADM for all relationships (HDM = 0, ADM = 1).

2.4.3 Robustness checks and assessment of study quality

To ensure robustness of our results, we tested for outliers and publication bias. For the outlier analysis, we combined the studentized residuals and Cook's distance to analyze if outliers with a high leverage were present in the sample (Viechtbauer & Cheung, 2010). When this was the case, the three-level model was re-estimated without the outlier, and the resulting correlation coefficient and significance level are reported in Table 2-2. Additionally, we analyzed publication bias by testing for funnel plot asymmetry by Egger et al. (1997) and by conducting the rank test by Begg and Mazumdar (1994). If both tests showed significance, we acknowledged the presence of publication bias. Similarly to the outlier test, we used the trim-and-fill procedure and estimated a corrected correlation coefficient (see Table 2-2). In summary, we found several instances of publication bias and outliers, but these had no impact on the results' significance or the study's implications.

Furthermore, as a robustness check, we incorporated mean age (conceptualized by the mean age of the participants in the study) and gender (conceptualized by the percentage of female participants in the studies) in our bivariate meta-analysis to check for potential effects. For the main relationship and the levels of system-related as well as organization-related reactions, our results revealed no significant difference effects for mean age and gender. In addition, we controlled for the terminology used in the scenarios regarding the automated decision entity (i.e., AI, algorithm, automation). These results indicate that there were no significant differences in reactions to the three terms across all aggregation levels. We introduced the experimental approach (between vs. within designs) as a robustness check, and again found no differences at any level of aggregation. Additionally, we used country of origin as a robustness check. We operationalized country of origin based on the country specified in the sample description. We found no significant differences for the main relationship and system-related reactions. However, we did find significant differences between countries for organization-related reactions ($\chi^2(2) = 8.98$, $p = .010$) with positive reactions in India ($\hat{\rho} = .13$, $p = .276$) compared to Germany ($\hat{\rho} = -.14$, $p < .001$) and the U.S. ($\hat{\rho} = -.21$, $p < .001$).

In addition, we controlled for study quality by analyzing both the journal quality as well as the differences between study approaches (true experiments vs. quasi-experiments) according to PRISMA guidelines (Page et al., 2021). Journal quality was assessed using the Scientific Journal Ranking (SJR) from 2023. No significant differences were found between higher ranked versus lower ranked journals neither for the main relationship nor for the division into system-related and organization-related reactions. In terms of the differences of the study approaches, two authors engaged in coding whether the included studies randomized (or reported on randomizing) the allocation of participants to scenarios. Results indicated that randomized or nonrandomized approaches did not differ significantly in terms of the main relationship and on the level of system-related reactions. However, a significant difference emerged at the level of organization-related reactions, as confirmed by the difference test, indicating that careful interpretation of organization-related reactions is needed.

2.5 Discussion

To advance the understanding of reactions to ADM usage in the HRM domain, we conducted a meta-analysis on system-related and organization-related reactions and their boundary conditions. Overall, we found significant negative reactions towards ADM use in general, and particularly regarding system-related reactions (i.e., procedural, interactional, and overall justice, fairness, trust, trustworthiness) and organization-related reactions (i.e., organizational attractiveness, job pursuit intention). In addition, we found differences in the proposed relationships depending on certain boundary conditions, namely, the type of interaction and the extent of decision. Specifically, negative effects occurred when individuals were confronted with ADM compared to HDM, but not when they collaborated with the system. Similarly, ADM use led to more negative reactions when it automated rather than augmented decisions. Although not significant in the difference test, our results indicate that employees and applicants responded more negatively to ADM than HR professionals, and reactions were more negative when individuals were personally affected by decisions. We did not find differences regarding type of task, indicating that context sensitivity might not be as pronounced as initially assumed (e.g., Lee, 2018). However, all included tasks focused on the HRM domain due to the scope of this meta-analysis, so it is possible that the tasks might have been too similar.

2.5.1 Theoretical implications

We expand the growing literature on reactions to ADM systems by (1) synthesizing existing heterogeneous findings on ADM usage and related reactions, (2) explaining these findings through an examination of boundary conditions that impact them, and (3) proposing an overarching theoretical framework to account for these findings. Our initial synthesis revealed that ADM is generally perceived negatively, confirming the presence of algorithm aversion in the existing literature. However, our findings indicate that algorithm aversion is not a static phenomenon, but rather depends on boundary conditions (e.g., extent of decision).

First, our findings revealed negative associations between ADM usage and perceptions of both fairness and justice. Notably, the negative relationship was stronger for justice than for fairness, with the largest differences between interactional justice and fairness perceptions, supporting organizational justice theory (Colquitt et al., 2001; Greenberg, 1987). This posits that individuals apply justice rules to form their fairness perceptions, emphasizing that justice and fairness are distinct concepts (Goldman & Cropanzano, 2015). However, fairness and justice are often used interchangeably in primary studies; we thus urge researchers to clarify whether they are assessing rule-based justice or subjective fairness perceptions. Notably, interactional justice showed the strongest negative correlation with ADM usage, suggesting that such systems may particularly undermine expectations of informational and interpersonal treatment (Binns et al., 2018). This highlights the importance of analyzing justice dimensions separately to better understand reactions to ADM.

Second, these main results vary according to specific boundary conditions as proposed in our overarching theoretical framework. Even though our difference test did not reveal significant differences across stakeholder types, our results suggested that applicants and employees evaluated the introduction of ADM more negatively than HDM—a trend that was not evident among HR professionals. This finding aligns with the differentiation proposed by Langer and Landers (2021), who argue that HR professionals often occupy the position of first parties (i.e., actively using ADM systems) whereas applicants and employees are more likely to be positioned as second parties (i.e., individuals directly affected by decisions) or third parties (i.e., external observers). As first parties, HR professionals may possess greater familiarity with the system's functionalities and potential benefits, which could reduce skepticism and promote a more neutral or even positive evaluation of ADM (Binns et al., 2018).

Third, drawing on CASA (Nass & Moon, 2000; Reeves & Nass, 1996), we offer an overarching theoretical framework to clarify why theories from human–human contexts are relevant in human–algorithm contexts, building a bridge between both contexts. Since people often engage with ADM systems as if interacting with humans, we can apply theories, such as organizational justice theory (Colquitt, 2001; Greenberg, 1987), to explain reactions to ADM. However, CASA also suggests that ADM may prompt negative reactions when human-like expectations remain unmet, which may be particularly strong in HR contexts that inherently imply human interactions. Focusing solely on the HRM domain, this could also explain why we observed no task-specific effects. Additionally, we suggest that considering the various boundary conditions affecting ADM reactions may help account for prior inconsistent findings.

2.5.2 Practical implications

The results of this study provide several practical implications for organizations. First, ADM might offer advantages from a business perspective (e.g., increased efficiency, reduction of human bias). However, our study indicates that individuals tend to react negatively to ADM usage. Here, the proposed boundary conditions of our overarching theoretical framework might help to guide organizations. Specifically, individuals' negative perceptions were strengthened in cases where ADM was used as automation and when they were confronted with ADM decisions and had no opportunity to interact with the system. Taking this into consideration, organizations might introduce human oversight to facilitate trust as well as fairness and justice perceptions. Additionally, organizations might increase interactivity when incorporating ADM systems. However, the responsible decision-making entity should be clearly defined in advance to prevent HR professionals from placing blame on the algorithm for morally difficult decisions (Maasland & Weissmueller, 2022). Furthermore, it is important to note that human-in-the-loop designs are not a panacea for all concerns of ADM usage in this context, and several studies point out the difficulties of combining HDM and ADM. Langer et al. (2025), for instance, suggest that humans need to learn how to detect errors in ADM before intervening. Consequently, organizations are challenged to train their employees to work with ADM systems efficiently. Additionally, HRM should create clear guidelines to ensure that the introduction and application of ADM work as intended.

Moreover, we found no significant negative relationships associated with HR professionals interacting with ADM systems. This suggests that HR professionals may view ADM systems as tools to enhance efficiency and decision quality. To support this potential benefit, organizations could offer training to strengthen HR professionals' competencies in ADM usage.

Lastly, when individuals were not affected by a decision, they reacted more positively to ADM use. Hence, current employees might not experience lower levels of organizational attractiveness simply because their employer uses ADM. However, when potential employees are targeted, organizations should introduce interventions to increase positive reactions specifically for affected individuals, for example, by providing more detailed information and transparency about the algorithm in use (Suen & Hung, 2023).

2.5.3 Future research avenues

We encourage scholars to explore the following research avenues:

2.5.3.1 Main Relationship

We found that ADM is negatively associated with system-related and organization-related reactions. These negative reactions are typically measured by assessing fairness, trust, or justice. However, some studies report that they found perceptions of unfairness or distrust to be apparent (despite measuring fairness and trust), although previous literature established that low levels of fairness or trust do not necessarily equate to unfairness or distrust (Lewicki et al., 1998). Instead, previous studies suggest that, for example, trust and distrust are independent constructs which need to be viewed separately (Kramer, 1999). Future research could examine whether fairness, justice, and trust reactions are distinguishable from constructs such as unfairness, injustice, or distrust, thus leading to our first research avenue:

Research avenue 1: *What roles do constructs such as distrust, injustice, and unfairness play in shaping responses to ADM? How are these constructs distinct from negative perceptions of trust, justice, and/or fairness?*

2.5.3.2 Type of Stakeholder

Primary studies have included HR professionals, current employees, and applicants as participants, while generally overlooking other potential stakeholders. Line managers' reactions to ADM remain underexplored, although they play a key role in implementing HR practices as established by the HR triad (Jackson & Schuler, 2003). We argue that line managers may act as both first and second parties in ADM-related HR decisions: they are involved in making decisions, while also being affected by HR decisions in which ADM is used, such as performance evaluations or promotions.

In addition, most scenarios for HR professionals have been framed from the perspective of HR managers. We define HR professionals as all staff involved in developing formal HR policies and processes (Jackson & Schuler, 2003), noting that while HR managers are included, they also hold additional management responsibilities. Following this, we propose that HR managers are more likely to occupy the role of first parties (i.e., individuals using ADM and making decisions), whereas HR professionals may simultaneously serve as both first and second parties (Langer & Landers, 2021) similar to line managers. Due to the potentially overlapping roles, HR professionals and line managers may respond differently than other stakeholders. Furthermore, we urge future research to empirically examine the underlying mechanisms differentiating stakeholders (i.e., differing tasks according to the HR triad, Jackson & Schuler, 2003; competence based on self-determination theory, Deci & Ryan, 1985), as this area of research is currently underdeveloped. In addition, the dynamic between HR managers and their subordinates introduces an important yet underexplored influencing factor. Investigating how these groups jointly evaluate ADM implementation (e.g., dealing with conflicting interests) offers a promising avenue for future research. We thus propose our second research avenue:

Research avenue 2: *How do line managers/HR professionals (not on the management level) react to ADM systems in comparison to other stakeholder groups (e.g., employees, applicants)? Which underlying mechanisms play a role in the differences in stakeholder reactions?*

2.5.3.3 Type of Interaction

We showed that the type of interaction (i.e., either collaboration or confrontation) impacts ADM reactions. Building on this, researchers could examine how ADM reactions might be improved when individuals are confronted with these decisions (i.e., second parties; Langer & Landers, 2021). Therefore, we suggest our third research avenue:

Research avenue 3: *How can organizations improve perceptions of ADM for individuals that are confronted with algorithmic decisions? How can decision processes become more interactive for second parties?*

2.5.3.4 Type of Task

In terms of our moderator type of task, we were able to include a differentiation of interviews, screening, and career development in ADM. However, other tasks, such as payroll, remain underexplored. Therefore, we propose our fourth research avenue:

Research avenue 4: *Which HR tasks beyond interviews, screening, and career development may be suitable for ADM implementation? How do reactions differ between these tasks?*

2.5.3.5 Extent of Decision

We found differences between ADM used for automation versus augmentation. However, it remains questionable whether human and ADM in collaboration yield better results than HDM or ADM alone (Langer et al., 2025), as previous literature indicates that results could become more or less accurate depending on situational aspects (Vaccaro et al., 2024). Furthermore, the factors that underpin successful collaboration remain unclear, leading to our next research avenue:

Research avenue 5: *How can successful human–ADM collaboration be achieved? How can human–ADM collaboration create better results compared to HDM or ADM alone?*

2.5.3.6 Personal Impact

The level of personal impact was coded into two categories: low or high. However, we argue that future studies could differentiate this further to understand the specific differences arising from the proximity of the decision to the respective stakeholder. In addition, we stress the importance of examining the subjective severity of decisions for individuals (i.e., important vs. unimportant decisions).

Research avenue 6: *How do different levels of personal impact affect reactions to ADM? Which role does the subjective importance of the decision play for affected stakeholders?*

2.5.3.7 Country of Origin

While our meta-analysis included country of origin solely as a robustness check, we see potential for future research in this area. Country of origin can act as a proxy for formal and informal institutions (North, 1990) that shape how individuals perceive ADM. Formal institutions (i.e., laws, regulations) such as the EU AI Act (European Commission, 2021) or the U.S. Algorithmic Accountability Act (Mökander et al., 2022), may foster trust and perceived fairness or, when absent, raise concerns about bias and lack of control (Fessenko & Jasperse, 2024; North, 1990). Informal institutions, including cultural norms and values (North, 1990), might also impact ADM reactions. As prior research on informal institutions is mixed (e.g., Kleinlogel et al., 2023; Mahmud et al., 2022), this might be a particularly relevant future research avenue. Given that most studies in our meta-analysis were conducted in the U.S. and Germany, we encourage future research to explore ADM reactions in diverse institutional contexts to broaden theoretical insights beyond Western perspectives. This leads us to our final research avenue:

Research avenue 7: *How do informal (e.g., norms, traditions) and formal (e.g., political standards, legislation, regulation) institutions in different countries shape reactions to ADM?*

2.5.4 Limitations

This meta-analysis has limitations in terms of study inclusion and methodology. First, meta-analyses are always limited by the availability of primary studies. Due to the absence of primary studies (in our case: studies including line managers' reactions to ADM and studies including severity of the decision), some effects could not be tested adequately ($k < 3$). Hence, the results need to be interpreted carefully. However, as Schmidt et al. (1985) argued, even with a small number of studies and sample sizes, meta-analyses remain the optimal method for synthesizing findings.

Second, most of the included primary studies tested their research models in hypothetical experimental scenarios. This creates a dependency on the participants to be able to realistically imagine themselves being evaluated by a human or an algorithm in the HRM domain. Furthermore, it might be difficult to compare the studies and consequently the reported effect sizes, as some scenarios were more realistic and sophisticated than others. We therefore invite future researchers to investigate real-life scenarios or conduct field experiments that analyze actual introductions of ADM in organizations. Further, one of our inclusion criteria was to only include studies comparing HDM and ADM directly, a strategy that contained the possibility of leaving out any studies that investigated differences between two algorithms that differed in certain functionalities. Therefore, scholars could analyze differing reactions towards ADM in the future to clarify and synthesize the current research landscape even further. This might be particularly important due to the proposed boundary conditions that could be explored by comparing different ADM systems.

3 Rejected by an algorithm? A multigroup analysis of outcome-dependent trust in algorithmic decisions (Essay II)[9]

Abstract

Algorithmic decision-making (ADM) is widely implemented in organizations and is associated with higher efficiency, accuracy, and objectivity. However, research suggests that stakeholders often perceive algorithmic decisions as untrustworthy. Using personnel selection as a sample context, we show that reactions to ADM systems and interventions aimed at building trust depend on decision outcomes (i.e., whether applicants are accepted or rejected for a job). We examined transparency and a human-in-the-loop (HITL) design as interventions to increase applicants' trust. To test our hypotheses, we conducted an online experiment (n = 483) and analyzed the data using multigroup structural equation modeling. While our results show that ADM was trusted less in cases of negative decision outcomes than in positive ones, this difference was not statistically significant. However, we observed significant differences in the acceptance and rejection group regarding both interventions. HITL was perceived with greater trust in positive decision outcomes compared to both ADM and a human counterpart, whereas transparency enhanced trust only in unfavorable decision outcome scenarios (i.e., being rejected). Consequently, these results demonstrate that the effectiveness of trust-building interventions varies by decision outcomes.

Keywords:

Algorithmic Decision-Making, Trust, Transparency, Human-in-the-Loop

---

[9] **Authors:** Moritz, J. M., Witte, J., Wehner, M. C., & Gier-Reinartz, N. R.
Second revision in *Communications of the Association for Information Systems*
A similar version of this manuscript was presented at the following conference:

- 84th Annual Meeting of the Academy of Management, Chicago, USA, 2024

3.1 Introduction

Algorithmic decision-making (ADM) is one of the most influential technologies of our time and reshapes the way we work (Kushwaha et al., 2022; Rigotti & Fosch-Villaronga, 2024). In organizations, the use of ADM systems is associated with higher efficiency (Cabiddu et al., 2022), accuracy (Dietvorst et al., 2015), and objectivity (Kordzadeh & Ghasemaghaei, 2021). Thus, ADM is being increasingly adopted by organizations, with 75% expected to implement it by 2027 (Di Battista et al., 2023). Among other organizational processes, ADM will increasingly play a central role in recruitment and selection, which impact applicant reactions to digital assessments and selection decisions (Gonzalez et al., 2019; Hunkenschroer & Luetge, 2022). These applicant reactions are the main drivers of organizational attractiveness or job acceptance intentions (Chapman et al., 2005; Hausknecht et al., 2004), and electronic word-of-mouth about potential employers (Evertz et al., 2019). In turn, this has important implications for organizations aiming to attract and select the best candidates for their job vacancies.

However, the integration of ADM challenges stakeholders' trust in these systems and the organizations deploying them (Glikson & Woolley, 2020; Lee, 2018). Trust in ADM varies depending on the task (Lee, 2018) and the stakeholders involved (Langer & Landers, 2021; Park et al., 2022). For instance, ADM usage for tasks associated with "human" skills, such as social or emotional skills, is perceived as less trustworthy, whereas this is not the case if tasks are associated with mechanical skills (Lee, 2018; Mahmud et al., 2022). In addition, Langer and Landers (2021) identified different reactions according to the stakeholder types, which they referred to as first and second parties. First parties include stakeholders who directly interact with ADM systems (e.g., recruiters using an ADM system), while second parties include those stakeholders (e.g., applicants) who are affected by ADM without control over the decision (Langer & Landers, 2021). As second parties are neither able to interact with the decision entities nor change the decisions made, they are more likely to react negatively to ADM compared to first parties (Köchling et al., 2023; Langer & Landers, 2021; Ochmann et al., 2024). Consequently, if algorithms take over tasks typically associated with human skills (Lee, 2018), such as evaluating applicants' performance during interviews or assessments, trust in both ADM and the organization may be undermined. As organizations increasingly adopt ADM systems for selection, it is essential to examine perceptions of trust among affected stakeholders and identify potential interventions to address the needs of this stakeholder group.

In addition to trust, we propose that individuals attribute negative decisions externally (i.e., to the decision entity), while taking personal credit for positive outcomes, described as self-serving bias (Zuckerman, 1979). This attribution bias can lead to discrepancies in the perception of trust, so that trust-building interventions are more effective than others depending on the outcome. While unfavorable outcomes (e.g., rejection) are associated with a decrease in perceived fairness (Wang et al., 2020) and procedural justice (Bedemariam & Wessel, 2023), little is known about whether trust-building interventions are equally effective across different outcome conditions. This is problematic because existing intervention approaches, such as explanations (Schlicker et al., 2021; Schoeffer et al., 2022) or human oversight (Langer et al., 2025), may not work universally. Instead, the effectiveness of interventions may depend on whether individuals receive favorable or unfavorable decisions. If this is the case, organizations risk implementing trust-building strategies that are ineffective or even counterproductive.

To address this issue, we provide new insights into outcome-dependent trust interventions and examine their effectiveness based on outcome conditions (i.e., acceptance or rejection). Drawing on attribution theory and self-serving bias, we assume that individuals attribute their experiences internally or externally. We examine this within the context of personnel selection, where individuals face either favorable (i.e., job offer) or unfavorable (i.e., rejection) decisions. This context is particularly relevant for two reasons. First, applicants have limited information about organizations and therefore might make assumptions about the organization based on the selection process and its outcome, as suggested by signaling theory (Spence, 1973). Consequently, the use of technology, and ADM in particular, can either enhance or diminish organizational attractiveness, ultimately influencing the quality of the applicant pool. Second, applicants do not voluntarily choose to be evaluated by an algorithm but are instead forced to follow corporate hiring processes. Thus, unlike consumers, they are second parties and have no control over algorithmic decisions (Czernietzki et al., 2023; Johnson et al., 2016), making it even more important to increase their trust in these systems.

Building on this background and grounded in the stimulus-organism-response (SOR) model (Mehrabian & Russell, 1974), we examine two widely recommended interventions to increase trust in ADM: (1) a human-in-the-loop (HITL) design (De Cremer & McGuire, 2022; Middleton et al., 2022) and (2) transparency (Glikson & Woolley, 2020; Middleton et al., 2022; Ochmann et al., 2024). When ADM is fully applied in a selection process, applicants may perceive it as untrustworthy due to the associations with human skills during the evaluation (Lee, 2018). Therefore, we assume that an HITL design could help mitigate these negative perceptions because a human component is still part of the decision-making process (Fügener et al., 2021). Furthermore, we propose that the amount of information individuals receive about algorithmic decisions should have a positive effect on their trust perception because individuals should better understand decisions made by algorithms if their decision criteria are transparent to them (Bitzer et al., 2023; Ochmann et al., 2024). We thus address the following two research questions:

Research question 1: *How does the outcome of a selection procedure (i.e., acceptance or rejection) impact trust in ADM?*

Research question 2: *How effective are trust-building interventions (i.e., HITL and transparency) depending on the outcome of a selection procedure?*

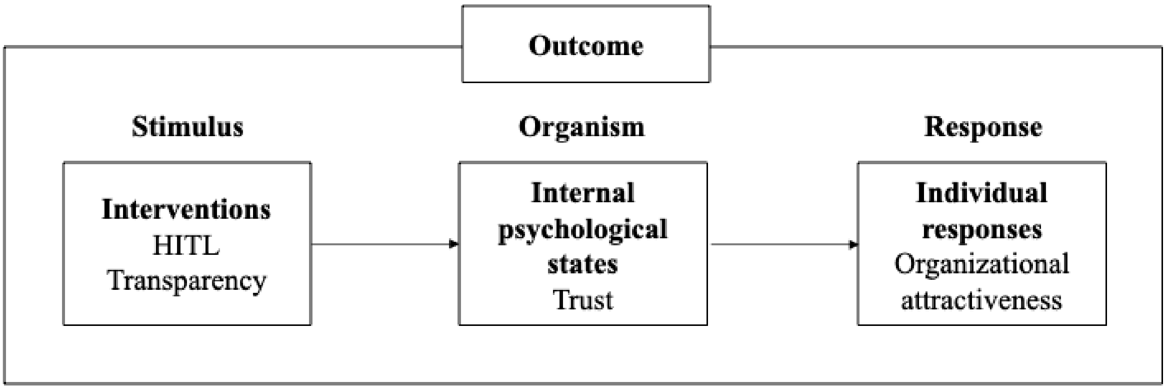
To address these questions, our paper is structured as follows: Section 2 reviews relevant literature from information systems, psychology, and human-computer interaction, followed by the theoretical background and hypotheses development. Section 3 details the experimental design (n = 483). Section 4 presents the main results, which are discussed in Section 5, where we critically reflect on the (over)reliance on ADM systems. In addition, Section 5 highlights the study's theoretical and practical implications, along with its limitations. Finally, Section 6 summarize our key findings.

3.2 Theoretical background and hypotheses development

3.2.1 Stimulus-organism-response model

To examine the effect of interventions on trust perceptions, we draw on the SOR model (Mehrabian & Russell, 1974). This model posits that environmental, external stimuli (S) impact internal psychological states of an organism (O), which, in turn, impact an individual's response (R; Mehrabian & Russell, 1974). Here, environmental stimuli refer to external cues that impact individuals, while internal psychological states encompass cognitive and affective processes, such as perceptions and beliefs. The resulting responses can manifest as either internal (non-observable) or external responses. The SOR model has been widely applied to understand technology adoption and user reactions (e.g., Kordzadeh & Ghasemaghaei, 2021; Ochmann et al., 2024; Zheng & He, 2024). In the context of ADM systems, stimuli have often been referred to as system characteristics (Kordzadeh & Ghasemaghaei, 2021), whereas justice evaluations and fairness perceptions have been assessed as internal psychological states (Ochmann et al., 2024). Building on this, we examine HITL design and transparency as environmental stimuli that impact trust perceptions. Mediated by trust, we propose that the individuals' responses in personnel selection manifest in their reactions to organizations using ADM systems and related interventions. Further, we extend the SOR model by arguing that outcome favorability impacts both the individual's perception of ADM usage and effectiveness of interventions. The theoretical framework is depicted in Figure 3-1.

Figure 3-1: Theoretical framework



To extend this theoretical framework to our research model, we first focus on individuals' trust perceptions and reactions, followed by an examination of the interventions aimed at increasing trust.

3.2.2 Trust perceptions

Trust is defined as the willingness to be vulnerable to another party's actions, based on the expectation that "the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al., 1995, p. 712). You et al. (2022) extended this definition to algorithm-based systems, framing trust as "trust in an advisor's ability to make good predictions" (p. 339), such as selecting the best candidate during recruitment. Given that trust is embedded in all kinds of social relationships (Munnukka et al., 2022), it is a critical factor in determining user acceptance of emerging technologies (Glikson & Woolley, 2020; Suen & Hung, 2023), including ADM.

While some research suggests that ADM may be perceived as more trustworthy than human decision-making (HDM) in terms of precision and objectivity (Cabiddu et al., 2022; Logg et al., 2019; Wang & Benbasat, 2016; You et al., 2022), empirical findings in personnel selection indicate the opposite, with ADM generally perceived as less trustworthy than HDM (Feldkamp et al., 2023; Langer et al., 2023). This observation may be attributed to the nature of this context, which is more likely to involve tasks and decisions associated with human skills (Lacroux & Martin-Lacroux, 2022; Lee, 2018). Theoretically, these findings might be explained by the computers are social actors (CASA) framework (Nass & Moon, 2000). Drawing on the principles of media equation (Reeves & Nass, 1996), the framework assumes that humans interact with computers and other technologies similarly to how they interact with other humans (Xu et al., 2022). This includes applying social norms and stereotypes—that is, interactional rules common to human–human interaction (Gambino et al., 2020). However, the application of these interactional rules might be inappropriate for human–algorithm interaction as interactions with computers or other technologies differ greatly from human–human interaction (Jain et al., 2023; Reicherts et al., 2022). Following the CASA framework, individuals may have certain expectations of interacting with technology that may be unrealistic due to the differences between HDM and ADM. Failure to meet these expectations, solely due to the nature of the decision-making entity, could lead to negative perceptions of trust. For example, while humans are generally able to provide reasoning for their decisions, algorithms do not inherently offer such rationales (Binns et al., 2018), making ADM appear opaque and reducing trust perceptions. Based on this reasoning, we hypothesize:

Hypothesis 1: *Compared to HDM, ADM is negatively associated with trust.*

3.2.3 Organizational attractiveness

Organizational attractiveness is defined as an expressed general positive affect or attitude towards an organization with the desire to initiate or strengthen a relationship (Gomes & Neves, 2011). In personnel selection, perceptions of organizational attractiveness are essential for effective candidate recruitment (Highhouse et al., 2009). Empirical findings on the relationship between ADM usage and organizational attractiveness reveal negative associations between ADM use in selection procedures and organizational attractiveness (Acikgoz et al., 2020; Roulin et al., 2023), especially when used in later stages of the selection process (Köchling & Wehner, 2023). This might be explained by signaling theory (Spence, 1973), which suggests that when one party (i.e., an applicant) lacks information about another party (i.e., the organization), the first party draws conclusions on missing information from observable cues. Therefore, applicants might derive missing information on the organization from the selection procedure. Hence, following signaling theory (Spence, 1973), we assume that the use of ADM in personnel selection impacts trust perceptions. These perceptions, in turn, impact how the organization using ADM is viewed.

Furthermore, according to the CASA framework, individuals may develop unrealistic expectations regarding human–algorithm interactions. Since these interactions differ from human-to-human interactions, ADM-based systems might fail to meet these expectations, potentially leading to decreased trust. While perceptions of greater trust are linked to the belief that the organization and its representatives act in a reliable and ethical manner (Mayer et al., 1995; You et al., 2022), a lack of trust is likely to convey a negative image of the organization. We assume that candidates base their perception of an organization's attractiveness on trust in the decision entity, as this entity represents the organization during the selection process. We thus hypothesize:

Hypothesis 2: *Compared to HDM, ADM is negatively associated with organizational attractiveness.*

Hypothesis 3: *Trust perceptions mediate the relationship between the decision entity and organizational attractiveness.*

3.2.4 Interventions to increase trust in algorithmic decision-making

3.2.4.1 Human-in-the-Loop

HITL as a term evolved around simulation and modeling, indicating the involvement of humans in algorithmic systems (Grønsund & Aanestad, 2020). This involvement can include human feedback, exception handling, and taking over control of and responsibility for a final decision (Sheridan, 1995). Fügener et al. (2021) suggest that humans and ADM systems complement each other, potentially leading to superior outcomes. This might be the case due to the combination of the precision and objectivity of algorithms with the accountability and transparency of humans (Middleton et al., 2022). Previous studies indicate that humans prefer human–algorithm collaboration when humans (compared to algorithms) lead the decision-making process (De Cremer & McGuire, 2022). This preference might be explained by a lack of accountability and responsibility in systems that are solely based on algorithms, resulting in perceptions of unfairness and untrustworthiness (Binns et al., 2018; Lee, 2018). To mitigate negative ADM perceptions, a human component might be able to account for responsibility and accountability. Consequently, we assume that an HITL design is positively associated with perceptions of trust compared to the use of ADM alone. Additionally, we assume that the combination of a human component (which provides accountability and responsibility) and algorithms (associated with increased objectivity and consistency) will lead to stronger trust perceptions compared to a human counterpart. Thus, we hypothesize that:

Hypothesis 4: *Compared to (a) HDM and (b) ADM, an HITL design is positively associated with trust perceptions.*

3.2.4.2 Transparency

Algorithmic transparency refers to the disclosure of information about these systems, enabling stakeholders to monitor, verify, and intervene as needed (Diakopoulos & Koliska, 2017). In general, algorithmic transparency can serve three purposes: (1) debugging models, (2) detecting bias, and (3) building trust (Brennen, 2020), of which we will focus on the latter. Bitzer et al. (2023) identify two distinct types of transparency. Accordingly, "transparency as action" occurs when developers reveal information about the algorithm, enabling transparency. In contrast, "transparency as perception" refers to the information or level of transparency that stakeholders perceive, which is influenced not only by information provided by developers but also by factors such as individuals' knowledge of ADM. Following the CASA framework (Nass & Moon, 2000) and the interactional rules of algorithm interactions, humans often expect to understand the decisions and underlying processes of ADM systems—a concept known as the interpretability of algorithmic decisions (Weber et al., 2023). This interpretability can be achieved through transparent decision-making processes (Weber et al., 2023). If transparency is lacking, algorithms may be perceived as a "black box," leading to a decrease in trust perceptions (von Eschenbach, 2021), whereas the presence of transparency might foster trust by providing stakeholders with necessary information to make informed judgments (Liao & Sundar, 2021). Consequently, humans might experience lower trust levels due to their inability to comprehend decisions made with or by algorithms in cases of absent (perceived) transparency (Bitzer et al., 2023; Ochmann et al., 2024). We therefore propose:

Hypothesis 5: *Transparency is positively associated with trust perceptions.*

3.2.4.3 Decision outcome as a moderator

We propose that the stated hypotheses are moderated by the decision outcome of the selection process. When applicants are rejected, they may perceive the entire selection process—including ADM—differently than those who receive a job offer, a phenomenon explained by attribution theory. According to Weiner (1985), individuals seek causes for their successes and failures, attributing them to either internal factors (e.g., inability to solve a work sample test) or external factors (e.g., an overly difficult selection process or an unfair decision-making process). Self-serving bias describes that individuals tend to overestimate themselves in such a way that they tend to attribute failure to external factors and success to internal factors (Zuckerman, 1979). Thus, rejected applicants are likely to attribute their rejection to the process rather than their own performance (Bedemariam & Wessel, 2023), resulting in decreased trust in the representatives or the organization. Moreover, we assume that the effects of the interventions— HITL design and transparency—are more pronounced for rejected individuals. Such candidates may scrutinize the procedure more critically than those who are offered a job. Regarding the HITL design, this could mitigate negative trust perceptions as candidates might trust a dual-entity system (ADM and a human) more than ADM only. As for transparency, knowing why they were rejected can be crucial for candidates continuing their job search. Feedback on their performance during the selection process can clarify the reasons for their rejection and indicate areas for improvement, potentially leading to enhanced trust perceptions. Based on this reasoning, we propose:

Hypothesis 6: *The effect of (a) ADM and (b) HITL on trust is stronger for the rejection group compared to the acceptance group.*

Hypothesis 7: *The effect of transparency on trust is stronger for the rejection group compared to the acceptance group.*

3.3 Method

To test the hypotheses and theoretical research model, we designed a 3 (decision entity) x 2 (decision outcome) x 2 (transparency) between-subject online experiment. To test the research model, we applied covariance-based multigroup structural equation modeling (SEM) in RStudio (version: 2023.12.0+369) with the R-packages lavaan (Rosseel, 2012) and semTools (Jorgensen et al., 2018).

3.3.1 Experimental procedure and sample

The online experiment simulated a fictitious digital assessment center with real-time interaction, where participants applied for a trainee position at the fictitious company "Marzeo AG." We used this fictitious company, along with a corresponding URL (www.marzeo.de), originally developed by Evertz et al. (2019), to ensure that participants would not find contradictory information about the company online. Similar to other experiments in this context (e.g., Köchling et al., 2023; Ochmann et al., 2024), we did not use an actual algorithm in this study but rather designed the study so participants believed they were being evaluated by an algorithm (depending on their assigned condition).

The online experiment started with a general introduction to the study, followed by a scenario description. In the scenario description, participants were told that they would complete a digital assessment center online, with an interacting agent to guide them through the process, providing them with feedback on their performance after each task. They were also informed that their performance would be evaluated at the end to determine their acceptance or rejection. Resulting from the decision entity (HDM vs. ADM vs. HITL), decision outcome (acceptance vs. rejection), and whether participants received an explanation of the decision parameters (no transparency vs. transparency), twelve experimental conditions were created, to which participants were randomly assigned. In the HDM scenarios, participants believed that a human resource manager, as interacting agent, evaluated candidates' performance and made the hiring decision. Similarly, in the ADM scenarios, they were led to believe that an ADM system, as interacting agent, evaluated candidates' performance and determined the outcome. In the HITL condition, an ADM system was introduced as the interacting agent during the assessment center, but participants were led to believe that a human resource manager made the final decision.

After participants confirmed that they had read the scenario description and were ready to start the digital assessment center, a loading bar appeared, followed by a chat introducing the interacting agent. The chat content was identical across all scenarios, with only the name of the interacting agent and decision agent adjusted according to the respective scenario. Designed as an interactive chat, this introduction aimed to mimic an online interaction and replicate the real-time experience of a digital assessment center. However, the chat was simulated and one-sided, meaning that participants could not respond to the interacting agent. Thereafter, participants began working on 24 randomized cognitive tasks (completing sequences of numbers and identifying words that do not fit into a series of words), taken from a website offering training tasks for assessment centers (Ausbildungspark Verlag, 2023). The tasks were pretested on a separate sample ($n = 151$, $M_{age} = 39.86$; $SD_{age} = 11.43$; male 62.25%) to ensure that they did not significantly differ in difficulty and response times. After completion of each task, participants received randomized feedback from the interacting agent in the form of chat messages, which was either positive or negative (equally distributed across the 24 tasks). Examples of positive feedback included "This was easy for you" or "You completed the task very quickly," while negative feedback included statements such as "This could have been done quicker" or "This was not very easy for you."

At the end of the assessment center, participants received the final decision outcome from the decision entity, presented as a pre-programmed chat. Transparency was manipulated in the chat message by providing an additional explanation on how the evaluation was executed, with participants in the "no transparency" group only informed about whether they had passed this round of the selection process. The transparency group did not receive individual performance information but rather information on the specific evaluation criteria. The following message was displayed: "I have evaluated your tasks. For the evaluation, I considered the following aspects: speed of response, accuracy of response, number of positions available in the company, and your performance in relation to all other applicants." Two exemplary survey flows are depicted in Appendix B1 and Appendix B2.

Using a panel provider (ISO 20252:19 certified), we recruited an online, quota-based sample of $n = 531$ participants, representative of the German working population. To ensure a realistic simulation of the assessment scenario, we required that participants had prior experience with a selection process and had used a personal computer to take part in the study. We used boxplots to eliminate $n = 48$ outliers regarding the duration of the study, leaving us with a final sample size of $n = 483$ ($M_{\text{age}} = 46.08$; $SD_{\text{age}} = 10.79$; male = 50.52 %).

To test our treatments, we built dichotomous dummy variables for the ADM and HITL groups, with HDM as the reference group. We then applied a multigroup SEM to compare participants who were accepted (group A) with participants who were rejected (group R). We used established multi-item measures to assess our constructs. We used the trust scale by Höddinghaus et al. (2021) and used an adapted version of the organizational attractiveness scale by Highhouse et al. (2003). As a control variable, we included attitude towards artificial intelligence (ATAI), which we assessed using the scale by Sindermann et al. (2021). Cronbach's alphas of all constructs exceed the .7 threshold in both groups (Cortina, 1993), indicating the reliability of our scales (trust$_{\text{GroupA}}$: $\alpha = .97$; trust$_{\text{GroupR}}$: $\alpha = .98$; organizational attractiveness$_{\text{GroupA}}$: $\alpha = .95$; organizational attractiveness$_{\text{GroupR}}$: $\alpha = .94$; ATAI $_{\text{GroupA}}$: $\alpha = .82$; ATAI $_{\text{GroupR}}$: $\alpha = .77$). Additionally, we controlled for participants' performance by calculating the average percentage of correct answers across all tasks. Table 3-1 presents the means, standard deviations, and correlations of all our variables.

Table 3-1: Means, standard deviations, and correlations

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. Trust	3.32	1.80							
2. Organizational attract.	2.56	1.17	.69**						
3. Transparency[a]	.52	.50	.07	.06					
4. Outcome[b]	.50	.50	.25**	.34**	–.01				
5. ATAI	4.09	1.33	.26**	.28**	–.05	.11*			
6. Performance	0.89	0.11	–.19**	–.26**	–.07	–.07	–.01		
7. Age	46.08	10.79	–.12**	–.10*	.03	–.02	–.08	.08	
8. Gender[c]	0.49	0.50	–.21**	–.19**	–.02	–.13**	–.15**	.14**	.03

Note. [a] 0 = no transparency, 1 = transparency. [b] 0 = rejection, 1 = acceptance. [c] 0 = male and 1 = female. *M* = mean, *SD* = standard deviation, Organizational Attract. = organizational attractiveness. * Indicates $p < .05$. ** indicates $p < .01$.

Furthermore, we conducted a confirmatory factor analysis (CFA) to verify several validity criteria. The CFA fits the data well ($\chi^2 = 46.55$; $df = 26$; $\chi^2/df = 1.79$; $p < .001$, CFI = .99; RMSEA = .06; SRMR = .01; Hair et al., 2014). All factor loadings were significant and above the cut-off criteria of .70 (Hair et al., 2014). The average variance extracted for all constructs is above the .50 threshold for both groups (trust$_{\text{GroupA}}$ = .91; trust$_{\text{GroupR}}$ = .93; organizational attractiveness$_{\text{GroupA}}$ = .84; organizational attractiveness$_{\text{GroupR}}$ = .81), indicating convergent validity (Fornell & Larcker, 1981; Hair et al., 2014). Composite reliabilities also exceeded the threshold of .6 (Bagozzi & Yi, 1988) for all our constructs (trust$_{\text{GroupA}}$ = .97; trust$_{\text{GroupR}}$ = .98; organizational attractiveness$_{\text{GroupA}}$ = .95; organizational attractiveness$_{\text{GroupR}}$ = .94). Additionally, discriminant validity was confirmed for our constructs as the average variance extracted exceeded the squared correlations with other constructs (Fornell & Larcker, 1981). We randomized the rotation of the items of each scale to reduce common method bias. We further addressed common method bias by implementing two approaches. First, Harman's one-factor test showed that no single factor accounted for more than 50% of the total variance. Nonetheless, Harman's one-factor test has been criticized (Podsakoff et al., 2024). Consequently, we also incorporated a theoretically unrelated marker variable (Lindell & Whitney, 2001). After adjusting the correlation matrix based on this variable, the significance of all correlations between the main constructs was maintained, suggesting that the risk of common method bias in our data was effectively mitigated. Moreover, to ensure comparability of both groups, we performed an analysis of variance to examine whether our control variables ATAI as well as a performance score (i.e., the average percentage of correct answers across all tasks) differed between the twelve groups. No significant differences were found, allowing us to proceed with further analyses.

3.3.2 Measurement Invariance

To ensure accurate model specification and verify the prerequisites for group comparisons, we first assessed measurement invariance before testing our multigroup SEM. Based on our grouping variable outcome, we divided the dataset into one acceptance and one rejection group utilizing the lavaan (Rosseel, 2012) and the semTools (Jorgensen et al., 2018) packages in R. We evaluated configural, metric, scalar, and strict invariance, using the comparative fit index (CFI) to determine if measurement invariance was achieved. According to Cheung and Rensvold (2002), the CFIs of two nested models should not differ by more than .01 when assessing measurement invariance. The results of our measurement invariance analysis are presented in Table 3-2. The CFI of all models did not differ by more than .01, indicating that measurement invariance was established.

Table 3-2: Measurement invariance

Model	χ^2/df	CFI	RMSEA	SRMR	ΔCFI
1. Configural invariance	1.337	.992	.037	.018	
2. Metric invariance	1.314	.992	.036	.019	.000
3. Scalar invariance	1.316	.991	.036	.020	–.001
4. Strict invariance	1.541	.983	.047	.020	–.008

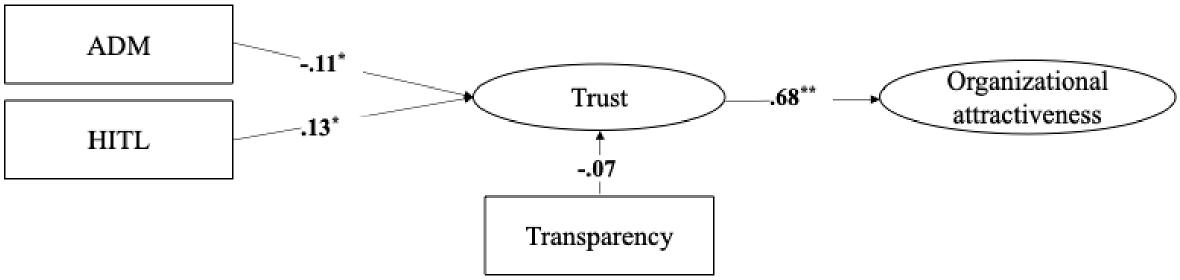
Note. χ^2 = chi-square test, *df* = degrees of freedom, CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual.

3.4 Results

To test our hypotheses, we calculated one SEM including all participants independently of their assigned outcomes and one multigroup SEM using outcome (acceptance vs. rejection) as the grouping variable. We tested all indirect effects of the decision entities on our mediator trust using 500 bootstrapping samples. Both the general SEM (χ^2 = 71.70; *df* = 39; χ^2/df = 1.84; *p* <. 001; CFI = .99; RMSEA = .04; SRMR = .01), and the multigroup SEM (χ^2 = 124.32; *df* = 83; χ^2/df = 1.50; *p* <. 002; CFI = .99; RMSEA = .05; SRMR = .02) demonstrated a good model fit (Hair et al., 2014). Figure 3-2 illustrates the results of the general SEM, while the results of the multigroup SEM are depicted in Figure 3-3.

Regarding the general model, we found trust was negatively associated with ADM (β = –.11; *p* = .026) and positively associated with HITL (β = .13; *p* = .015). The effect of transparency on trust was not significant (β = .07; *p* = .097). Furthermore, organizational attractiveness was positively associated with trust (β = .68; *p* < .001). However, we did not find significant effects for both ADM and HITL regarding organizational attractiveness. Both indirect effects of ADM (β = –.07; *p* = .028) and HITL (β = .09; *p* = .014) on organizational attractiveness via our mediator trust were significant. Consequently, we support hypotheses H1, H3, H4, while rejecting H2 and H5. Regarding our control variables, we found a positive relationship between ATAI and trust (β = .27; *p* < .001) as well as organizational attractiveness (β = .09; $p = .013$). Performance was negatively associated with trust (β = –.20; $p < .001$) and organizational attractiveness (β = –.13; $p < .001$).

Figure 3-2: General SEM results

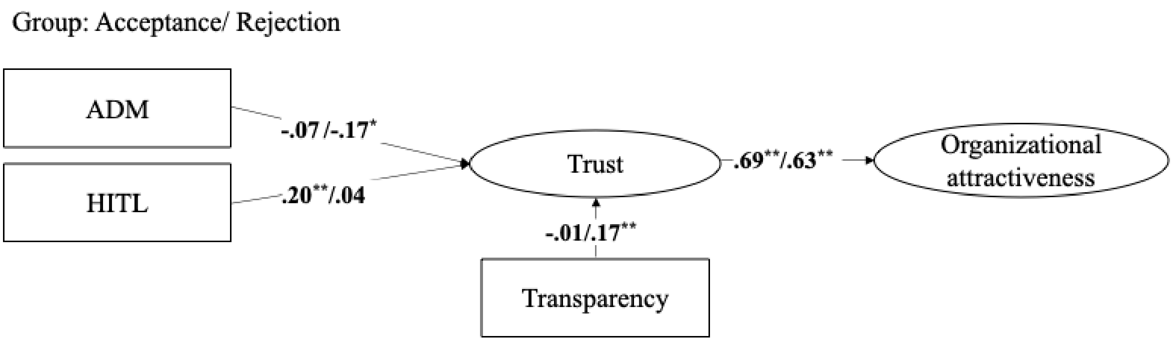


Note. Reference group ADM and HITL: human decision–making (= 0), Reference group transparency: no transparency (= 0). Control variables: attitude towards AI, performance.
* Indicates $p < .05$. ** indicates $p < .01$.

Regarding the results of our multigroup SEM, we found that ADM usage was negatively associated with trust in both groups; however, this was only statistically significant in the rejection group (β = –.17; $p = .018$). An HITL design was positively associated with trust in the acceptance group (β = .20; $p = .007$), with no significant effects for the rejection group. In the acceptance group, transparency did not have a significant effect on trust, whereas it positively impacted trust in the rejection group (β = .17; $p = .011$). We did not find significant effects of the decision entities on organizational attractiveness in either group. Additionally, we found a significant relationship between trust and organizational attractiveness in both groups (acceptance group: β = .69; $p < .001$; rejection group: β = .63; $p < .001$). Furthermore, in the acceptance group, we found a significant indirect effect of HITL on organizational attractiveness via trust (β = .14; $p = .007$), and in the rejection group, ADM had a significant indirect effect on organizational attractiveness via our mediator trust (β = –.10; $p = .032$).

To determine statistically significant differences between groups, we compared the coefficients using chi-square difference tests. We compared a model with freely estimated coefficients (our multigroup SEM) to a model where the regression weights were constrained to be equal across groups. The chi-square difference test allowed us to examine the difference between these two models and determine whether the constraint of equal regression coefficients significantly worsens model fit, indicating a moderation by our grouping variable. For the relationship between ADM and trust, the chi-square difference test revealed no significant difference between the constrained model and our multigroup SEM ($\Delta\chi^2 = 0.86$, $\Delta df = 1$, $p = 0.354$); leading us to reject H6a. Regarding HITL, the chi-square difference test revealed a significant ($p < .10$) difference ($\Delta\chi^2 = 3.80$, $\Delta df = 1$, $p = 0.051$), indicating that the effect in the acceptance group was stronger than in the rejection group. Given that we hypothesized the opposite, we also had to reject H6b. Finally, when testing the relationship of transparency and trust between the acceptance and rejection group, we found a significant ($p < .10$) difference ($\Delta\chi^2 = 2.72$, $\Delta df = 1$, $p = 0.099$), supporting H7. The results are depicted in Figure 3-3 and Table 3-3.

Figure 3-3: Multigroup SEM results



Note. Reference group ADM and HITL: human decision-making (= 0), Reference group transparency: no transparency (= 0). Control variables: attitude towards AI, performance.
* Indicates $p < .05$. ** indicates $p < .01$.

Table 3-3: Multigroup SEM results

Path	Acceptance				Rejection			
Hypothesized effects	***B***	***SE***	**β**	***p***	***B***	***SE***	**β**	***p***
ADM → Trust	–.16	.16	–.07	.328	–.37	.17	–.17*	.018
HITL → Trust	.47	.18	.20**	.007	.10	.16	.04	.551
Transparency → Trust	–.01	.14	–.01	.925	.35	.13	.17**	.011
Trust → Organizational attractiveness	.93	.10	.69**	<.001	.87	.11	.63**	<.001
ADM → Organizational attractiveness	.02	.18	.01	.904	.30	.18	.10	.122
HITL→ Organizational attractiveness	–.17	.17	–.06	.312	–.26	.19	–.09	.174
ADM → Trust → Organizational attractiveness	–.15	.15	–.05	.333	–.32	.15	–.10*	.032
HITL → Trust → Organizational attractiveness	.44	.17	.14*	.007	.08	.14	.03	.557
Control effects								
ATAI → Trust	.23	.06	.29**	<.001	.18	.06	.22**	.001
ATAI → Organizational attractiveness	.11	.06	.10	.064	.10	.06	.08	.137
Performance → Trust	–2.04	.75	–.21**	.005	–1.75	.79	–.17*	.041
Performance → Organizational attractiveness	–1.61	.50	–.12**	.003	–1.94	.78	–.13*	.008

Note. The ADM and HITL groups were not evaluated by actual algorithms; rather, participants perceived they were being evaluated by ADM or HITL. *B* = unstandardized effect, *SE* = standard error, β = standardized effect, ATAI = attitude towards AI. * indicates p < .05. ** indicates p < .01.

3.5 Discussion and implications

3.5.1 General discussion and theoretical implications

Our research provides new and distinctive insights into how stakeholders directly affected by ADM perceive its use and trust-building interventions depending on the outcome. Overall, the results align with SOR model (Mehrabian & Russell, 1974), supporting the notion that external (system) characteristics can be classified as stimuli impacting individuals' psychological states (i.e., their trust perceptions), which, in turn, shape their reactions—such as their assessment of organizational attractiveness. Furthermore, the findings extend prior applications of SOR model (e.g., Ochmann et al., 2024) by demonstrating that outcome favorability moderates internal psychological states. Here, we found that individuals tend to perceive trust-building interventions differently depending on whether the outcome is positive (i.e., acceptance) or negative (i.e., rejection).

First, we found that when individuals are confronted with a positive outcome (i.e., acceptance), they are less likely to question the selection process and organizational representatives compared to individuals who are rejected. Following, in the acceptance group, ADM did not negatively impact trust perceptions, which appears contrary to prior research findings at first glance (Feldkamp et al., 2023). In the rejection group, ADM was negatively associated with trust perceptions, aligning with attribution theory (Weiner, 1985) and self-serving bias (Zuckerman, 1979), which suggest that failures are more likely to be attributed externally (e.g., to the selection procedure or the organizational representative). Thus, negative ADM perceptions might also be explained by external attribution. However, given that we did not find statistically significant differences among our two outcome groups concerning the ADM–trust relationship, we are cautious to conclude that applicant reactions differ substantially based on the decision outcome. Instead, our results suggest that ADM without human oversight is generally trusted less than human recruiters and evaluators alone.

Second, an HITL design increased trust compared to both ADM and HDM in the acceptance group, and this effect was significantly stronger than in the rejection group. Here, the HITL design might reflect the perceived benefits of both ADM and HDM, indicating high relevance in this specific context. Participants' expectations about the objectivity and presumed consistency of an ADM-based system during task performance, in combination with the perceived higher levels of social skills, accountability, and responsibility of a human, might have led to this positive applicant reaction. Moreover, individuals who were accepted might have felt more confident in this decision as it was made by both ADM and a human decision entity. Regarding transparency, we found no effect on trust in the acceptance group, indicating that accepted individuals are less likely to question the application process. According to attribution theory, they might attribute the success to their own abilities, with no need for an external explanation of why they were offered a job. This finding is also supported by existing research indicating that—in addition to transparency and interpretability—trust in ADM might be influenced by the outcome of the decision process rather than the process itself (Ahn et al., 2024; Hidalgo et al., 2021).

Third, we found positive effects of transparency on trust in the rejection group, and this effect was significantly stronger than in the acceptance group. Thus, transparent decisions seem to be considered more trustworthy, a notion that is supported empirically (Grimmelikhuijsen, 2023), but explanations seem to be more important to individuals who are confronted with negative outcomes (i.e., rejection). This might be because rejected individuals might question the selection procedure by attributing the cause externally (Weiner, 1985), which increases their need for more transparency. Thus, the effect of transparency on trust seems to depend on conditional criteria, a finding that is in line with previous research (Kizilcec, 2016). Moreover, this finding could be important for rejected applicants and their (electronic) word-of-mouth behavior (Evertz et al., 2019) as transparent communication with rejected applicants might reduce the possibility that they provide negative employer reviews. In turn, these rejected candidates may be less likely to harm the organization’s future applicant pool, despite their rejection.

Fourth, we found that an HITL design had positive indirect effects on organizational attractiveness in the acceptance group, while ADM had negative indirect effects on organizational attractiveness in the rejection group both mediated by trust. Consequently, the use of HITL and ADM impacts organizational attractiveness, mediated by trust and dependent on the decision outcome, which is aligned with signaling theory (Spence, 1973).

Interestingly, the control variable performance was negatively associated with both trust and organizational attractiveness. This suggests that high-performing applicants are more likely to question (AI-based) selection procedures compared to low-performing applicants. This finding aligns with prior research suggesting that the use of AI in personnel selection may reduce individuals' perceived opportunity to perform (Köchling et al., 2023), which might explain why specifically high-performing individuals showed more negative reactions here.

3.5.2 Practical implications

The findings of this study provide valuable insights for organizations implementing ADM-based systems in their personnel selection procedures. First, our general model, which includes all participants, indicates that solely using ADM can lead to negative trust perceptions and negatively impact organizational attractiveness through the mediator trust. This effect may be attributed to the context of personnel selection, which often involves social or emotional skills that have historically been associated with perceptions of untrustworthiness and unfairness (Lee, 2018). Therefore, organizations should carefully consider the task context when implementing ADM. The HITL design enhanced trust perceptions, and this intervention can be employed to mitigate trust concerns and maintain organizational attractiveness by organizations wishing to use ADM systems for tasks involving social or emotional skills. As ADM perceptions are context-dependent (Lee, 2018), we suggest that the results of the present study are particularly relevant in contexts where individuals are confronted with ADM (i.e., second parties; Langer & Landers, 2021) and in high-stakes decision-making processes.

Further, several other factors such as ADM modality or user expectations as well as external factors may influence the way in which organizations (can) use AI-based and ADM systems in the future. For example, proposals for AI regulation, such as mandatory human supervision (European Commission, 2021), could restrict the use of ADM to a support role. Another potential aspect influencing the perception and use of ADM is habituation. Specifically, as the usage of algorithms in selection procedures is quite new and trust typically develops over time (Cabiddu et al., 2022), trust might also increase as the use of

ADM in selection becomes more ubiquitous. Currently, however, relying on an HITL design and transparency may help overcome potential negative perceptions.

Although the present study demonstrates that HITL approaches are associated with higher levels of trust, HITL should not be presented as a panacea for all concerns about the use of algorithms in personnel selection. One reason for this caution is that the use of human oversight as a complementary approach to ADM assumes individuals can effectively supervise ADM (Green, 2022). However, Green (2022) found that humans are frequently not able to fulfill oversight functions, because they may expect algorithmic decisions to be less biased compared to HDM (Bonezzi & Ostinelli, 2021). Here, it is important that the involved humans have the capacity to identify false decisions (Schoeffer et al., 2024). Additionally, human oversight may even be used to exploit algorithmic decisions. For instance, Krügel et al. (2023) examined human supervision of ADM regarding moral decision-making and found that humans do not correct but instead exploit unethical decisions made by algorithms. Thus, while HITL might be a helpful approach to increase the trust perceptions of those affected by ADM, there is a need to integrate human oversight in a meaningful manner (Langer et al., 2025).

Additionally, since we simulated the decision-making process rather than incorporating an actual ADM system, we argue that trust perceptions are shaped by individuals' beliefs about whether they are being evaluated by algorithms or humans, rather than the actual decision-making process itself. This distinction is important, as organizations may choose not to disclose ADM usage to avoid resistance, which has already led to discussions about mandatory disclosure to increase transparency of decision-making processes (Renieris et al., 2024). Furthermore, humans may deflect responsibility onto algorithms to justify (unethical) decisions (Krügel et al., 2023). Both practices bear the risk of undermining trust and damaging the organization's image, but this mainly raises ethical concerns that should be critically examined by both organizations and policymakers (Renieris et al., 2024).

When examining trust in ADM, the question arises to what extent reliance on ADM systems is appropriate and in which cases stakeholders uncritically accept algorithmic decisions and thus employ blind faith in systems. This is particularly concerning as ADM systems are increasingly applied in high-stake contexts, such as personnel selection (König & Langer, 2022) or even medical diagnosis (Park & Han, 2018), and by the fact that several studies demonstrate the overreliance of individuals on algorithmic decisions (Klingbeil et al., 2024). As stated above, organizations should carefully examine how to meaningfully integrate ADM systems and HDM. This includes organizational measures to help individuals detect errors in algorithmic decisions (Jussupow et al., 2021; Langer et al., 2025) and to foster calibrated trust, where stakeholders appropriately adjust their level of trust to the reliability of the system (Muir, 1994). Here, organizations could encourage stakeholders to trust ADM systems while maintaining a critical awareness of their limitations.

Our results indicate that reactions to ADM are indeed impacted by the outcomes of decision-making processes, and the interventions to increase trust in ADM function differently depending on these outcomes. Specifically, negative trust perceptions were observed when ADM was used for decisions resulting in unfavorable outcomes. We hence advise practitioners to exercise caution when communicating negative outcomes made with or by algorithms. As transparency significantly contributed to trust perceptions, this might mitigate negative perceptions of ADM usage. From a managerial perspective, this is important for organizations even when rejecting applicants, because the negative perception of selection procedures can lead to adverse word-of-mouth, ultimately affecting the quality of the organization's applicant pool (Stockman et al., 2020). Practitioners should therefore tailor interventions based on decision outcomes. Furthermore, even though we used personnel selection as a sample context, there are several other organizational contexts, where customers, employees, or other stakeholders might be confronted with an algorithmic decision that might be unfavorable. Our findings could thus be beneficial for other areas where second parties are affected by ADM.

3.5.3 Limitations and future research avenues

Three primary limitations should be considered when interpreting our findings. First, the study was conducted in the context of personnel selection, which involves certain boundary conditions. Specifically, we focused on applicant reactions, excluding other key stakeholder groups such as managers and employees. Additionally, high-stakes decision-making contexts such as personnel selection may lead to stronger trust or fairness concerns compared to other domains. Research has demonstrated that reactions to ADM depend on the specific context or task (Araujo et al., 2020; Lee, 2018; Mahmud et al., 2022). Therefore, while our findings contribute valuable insights within the realm of personnel selection, they may not fully capture the broader range of decision-making tasks that ADM systems are used for in this context. Furthermore, within personnel selection, ADM can be implemented using different modalities, such as video-based systems (e.g., Su et al., 2021), which could yield very different trust-related outcomes compared to the approach used in our study. Thus, we cannot generalize our findings but instead encourage researchers to assess perceptions in different contexts and for different tasks. It would be valuable to gain insights into whether the effects we found in the present study are consistent when using another task or context, and whether transparency and an HITL design generally help to mitigate negative trust perceptions. Further, as our study included a German sample, cross-cultural differences could not be addressed, thereby limiting the generalizability of our results.

Second, there are methodological limitations to our study. Participants were asked to imagine they had applied for a trainee position, potentially undermining the external validity of the results. Older participants might not have identified with the scenario, as trainee positions are usually intended for younger applicants. Additionally, participants had no actual stakes in the task, limiting the generalizability to real-world situations. However, we implemented an interactive design with the intention to create a more realistic setting. In doing so, we kept the design as close as possible to a real application process by incorporating tasks and providing feedback. Nevertheless, the experiment still simulated an online assessment center, making participants believe that they were being evaluated by an algorithm or a human. In future research, field experiments might provide more insights regarding behavioral outcomes of applicants, increasing validity. Moreover, employing field experiments could potentially address the challenge faced in the current study, where feedback was provided randomly to control for participant performance in each group. Consequently, the feedback was not contingent on actual performance, potentially leading to confusion among participants regarding the evaluation rationale. However, in all scenarios, we informed participants before starting the assessment center that their feedback also depended on the performance of other participants, so that participants were not able to predict their actual performance. Additionally, in real-world application processes, applicants may also feel they performed well yet still receive negative feedback. In this sense, the experimental design may still reflect realistic applicant experiences.

Finally, as transparency impacted trust depending on the decision outcome, the design of interaction and communication between ADM systems and recipients is of great importance. Here, the present study did not explore all relevant research questions regarding this aspect, such as type of explanation or prior knowledge on ADM systems (Bitzer et al., 2023). Specifically for ADM systems, further research should assess additional antecedents of transparency to gain a deeper understanding of their perceptions. From a theoretical point of view, this study serves as a starting point for examining the mechanisms underlying the relationship between decision outcomes and trust-building interventions.

3.6 Conclusion

Although ADM systems are increasingly utilized within organizations and are associated with numerous benefits, stakeholders, particularly second parties (Langer & Landers, 2021), often perceive their use as untrustworthy (Lee, 2018). Our study demonstrates that these trust perceptions are not solely dependent on the type of task, as previous empirical research has shown (Lee, 2018; Mahmud et al., 2022), but might also be impacted by the decision outcome. Our general model corroborated prior findings, indicating that ADM is negatively associated with trust, while HITL systems are positively associated with trust, subsequently affecting organizational attractiveness. In terms of decision outcome, ADM usage is perceived with less trust when applicants are rejected, and transparency positively impacts trust. Conversely, accepted applicants did not exhibit negative reactions to ADM usage, and HITL even increased trust. Transparency had no effect on these stakeholders, indicating that both the source and explanation might be less critical when the outcomes are favorable. In conclusion, the results of our study demonstrate that the effect of trust-building interventions varies by decision outcome. We hope this research model inspires future investigations into additional antecedents and mechanisms to further explore the factors underlying trust perceptions of ADM systems.

4 Justice evaluations of algorithmic management: The role of prior discrimination experience (Essay III)[10]

Abstract

Algorithmic management (AM) is increasingly used and offers several benefits to organizations, such as increased efficiency of processes and greater accuracy. However, employees oftentimes associate AM use negatively with justice. To understand how justice evaluations of AM are shaped, this study investigated the impact of prior human and algorithmic discrimination experience on this relationship. Three online experiments (n_1 = 82, n_2 = 83, n_3 = 216) demonstrated that AM use is negatively associated with justice evaluations. Study 1 shows that the negative effect of AM use on justice evaluations was weakened if participants had prior human discrimination experience. Study 2 does not show a moderating effect of prior algorithmic discrimination experience on the relationship between AM use and justice evaluations. Study 3 replicates these findings by demonstrating that algorithmic discrimination is perceived less negatively compared to human discrimination. The qualitative responses show that individuals associated algorithmic discrimination with biased data and opacity, whereas human discrimination was associated with favoritism and increased human bias. These findings suggest that justice evaluations of AM systems are shaped by individuals' prior experiences. Understanding these patterns can help organizations design and implement AM systems that promote justice.

Keywords:

Algorithmic Management; Discrimination; Justice; Anticipatory Injustice

[10] **Authors:** Moritz, J. M. & Wehner, M. C.
Under review at *Business & Information Systems Engineering*
A similar version of this manuscript was presented at the following conferences:

- 32nd European Conference on Information Systems, Paphos, Cyprus, 2024
- 84th Annual Meeting of the Academy of Management, Chicago, US, 2024

Awarded best paper runner-up, European Conference on Information Systems 2024

4.1 Introduction

The recent growth in the use of algorithms and artificial intelligence (AI)-based systems in organizations is altering the future of work (Kadolkar et al., 2024). One specific form of AI-based systems, algorithmic management (AM), describes the delegation of coordination and control functions traditionally performed by managers to AM systems (Möhlmann et al., 2021). AM relies on large data models (Möhlmann et al., 2021) and is currently used primarily by digital labor platforms, with Uber, Airbnb, and Upwork as prominent examples (Benlian et al., 2022). However, with the increasing datafication of organizational processes, AM is now being adopted in traditional organizations to manage permanent employees (Hirsch et al., 2024). As such, understanding reactions to AM use in traditional organizations is becoming increasingly important.

Although AM systems offer several benefits to organizations, such as greater accuracy and increased efficiency (Daugherty & Wilson, 2018), interactions with humans is a major challenge in the development and implementation of AM (Glikson & Woolley, 2020). While some studies have found insignificant or positive effects regarding justice evaluations of AM in comparison to human management (HM; Lee, 2018; Logg et al., 2019), research has mostly revealed negative justice perceptions of AM (Langer & Landers, 2021). In the literature, negative reactions to AM, which often occur when individuals see algorithms err, are subsumed under the term algorithm aversion (Castelo et al., 2019; Dietvorst et al., 2015). AM has been negatively associated with justice perceptions (McGuire & De Cremer, 2023) and more specifically with interpersonal (Langer & Landers, 2021) and procedural justice when used for complex rather than simple tasks (Nagtegaal, 2021). Negative justice perceptions, in turn, have been related to decreased worker well-being (Kinowska & Sienkiewicz, 2022), decreased decision acceptance (Lind, 2001), and a more negative reputation of the organization using AM systems (Klöpper & Messer, 2025; McGuire & De Cremer, 2023).

While the relationship between justice evaluations and AM is well established, few studies have examined how prior experiences shape justice evaluations. However, in human–human contexts, theoretical assumptions, such as the model of justice expectations (Bell et al., 2004), suggest that individuals use their prior experiences to infer how justly they will be treated in the future. Thus, we examine the impact of prior discrimination experiences to understand how justice evaluations of AM are shaped.

In the context of recruiting, prior discrimination experiences by humans led to a decrease in algorithm aversion (Fleiß et al., 2024; Koch-Bayram et al., 2023; Pethig & Kroenung, 2022; Schulte Steinberg & Hohenberger, 2023). However, these studies have relied predominantly on personal characteristics such as gender as indirect indicators of discrimination experiences (e.g., Pethig & Kroenung, 2022). Besides this methodological consideration, it is noteworthy that, to date, these investigations have largely been limited to recruitment settings, raising the question of whether such findings can be generalized to broader applications of AM. Additionally, while prior research considers discrimination by human decision-makers, the effects of prior discrimination by AM systems remain largely unexplored. We propose that it is highly relevant to account for the source of discrimination and examine both human and algorithmic discrimination for three reasons.

First, from a theoretical standpoint, individuals may attribute human-like intentionality and moral agency to AM systems. This may be explained by the computers-are-social-actors (CASA) framework (Nass & Moon, 2000; Nass et al., 1994), which states that individuals treat computers, and by extension other technologies (Gambino et al., 2020), similarly to other humans. Consequently, they employ similar interactional rules and expectations when interacting with AM systems. If previous human experiences of discrimination impact AM reactions as shown in prior literature (e.g., Fleiß et al., 2024; Koch-Bayram et al., 2023), then algorithmic discrimination, which is processed similarly, can be expected to have comparable effects. Without theorizing this possibility, existing models of justice evaluations in algorithmic contexts remain conceptually incomplete.

Second, prior research demonstrates that the attribution of decisions to discrimination increases when a decision itself and its underlying processes are opaque or ambiguous. In human–human contexts, this is referred to as attributional ambiguity, whereby stigmatized or stereotyped individuals attribute decisions differently (i.e., to discrimination or prejudice) compared to individuals with no stigmatization or stereotyping experiences (Crocker et al., 1991). This may be transferable to human–algorithm contexts, wherein the opacity (Cobonpue et al., 2024) and lack of reasoning (Binns et al., 2018) often associated with AM systems may drive the association between algorithmic decisions and discrimination.

Third, examining discrimination experiences not only by humans but also by AM systems enables a better understanding of how justice evaluations of AM are shaped. This is particularly relevant as justice evaluations impact several organizational outcomes, such as organizational commitment, productivity (Viswesvaran & Ones, 2002), organizational attractiveness (Köchling et al., 2025), and turnover intention (Acikgoz et al., 2020), and thus determine organizational success. Besides organizational outcomes, workers and employees may face negative consequences, such as reduced well-being (Kinowska & Sienkiewicz, 2022). As understanding the source of perceived injustice can inform targeted interventions to mitigate negative effects, there is a need to understand how justice evaluations are shaped. To address these concerns, we ask the following:

Research question: *How do prior experiences of (1) human discrimination and (2) algorithmic discrimination impact the relationship between AM and justice evaluations?*

To offer a theoretical understanding, we draw on the model of justice expectations by Bell et al. (2004) and the idea of anticipatory injustice (Shapiro & Kirkman, 2001) to theorize why human and algorithmic discrimination impact justice evaluations. In doing so, first, we extend both theories from human–human contexts to human–algorithm contexts. Second, we provide empirical insights by having conducted three experimental studies that examined the effect of (1) prior human discrimination, (2) prior algorithmic discrimination, and (3) the experimental comparison of both types of discrimination experience. To further advance our understanding, we used open-ended qualitative questions to gain deeper insight into potential mechanisms underlying the proposed relationships. Third, based on the qualitative responses, we identify future research avenues and derive practical implications. In doing so, we contribute to a deeper understanding of the role prior discrimination experience plays in shaping the relationship between AM use and justice perceptions.

4.2 Theoretical background

4.2.1 Justice perceptions of algorithmic management

Justice can be defined as "the degree to which one's company or top management is perceived to act consistently, equitably, respectfully, and truthfully in decision contexts" (Colquitt & Rodell, 2015, p. 188) and is categorized into distributive, informational, interpersonal, and procedural justice (Colquitt, 2001). While distributive justice describes justice evaluations of an outcome, informational justice concerns the information and explanations provided for a decision (Colquitt, 2001). Interpersonal justice refers to the quality of interpersonal treatment, and, finally, procedural justice refers to whether a procedure followed to make a decision is perceived as just (Colquitt & Rodell, 2015). In the present paper, we focus on the perceived interaction between AM and individuals as well as the procedures followed to make a decision. Thus, we include informational, interpersonal, and procedural justice.

Workers' justice evaluations are essential to organizations. Here, prior research demonstrates that justice evaluations are negatively related to turnover intention and positively to performance (Moon, 2017) as well as job satisfaction and effort (Bankins et al., 2022; Lind, 2001). Additionally, decisions perceived as just are more likely to be accepted by workers (Lind, 2001). These are important organizational consequences, especially when considering the increasing use of AM. However, most research reveals that AM is negatively associated with justice evaluations, especially when used for tasks associated with social or emotional skills (Lee, 2018) and when tasks are considered complex rather than simple (Nagtegaal, 2021).

Theoretically, this may be explained by the CASA framework (Nass & Moon, 2000; Nass et al., 1994), which postulates that individuals use the same interactional rules known from human–human contexts when interacting with technology. Consequently, individuals apply stereotypes and assign personality traits and norms to computers and, by extension, other technologies, such as algorithmic systems (Gambino et al., 2020). However, human–human interactions differ from human–computer interactions; for example, algorithms do not naturally provide explanations for their decisions (Binns et al., 2018). Here, expectations regarding transparency (i.e., informational justice), personal and adequate treatment (i.e., interpersonal justice), and the procedures followed to make a decision (i.e., procedural justice) may not be met, resulting in weaker justice perceptions. Consequently, a mismatch between expectations and the outcome of human–algorithm interaction can negatively affect justice.

Further, we assume that the expectations and interactional rules proposed in the CASA framework are especially relevant when computers or algorithms perform tasks typically undertaken by humans as is the case for most tasks conducted by AM systems. Employees may perceive AM systems as less capable of understanding and responding to their social and emotional needs compared to human counterparts, leading to feelings of being treated unjustly or impersonally (Binns et al., 2018; Langer & Landers, 2021; Lee, 2018). Consequently, we propose the following:

Hypothesis 1: *Compared to HM, the use of AM is negatively associated with (a) informational justice, (b) interpersonal justice, and (c) procedural justice.*

4.2.2 Prior human discrimination experience

According to Allport et al. (1954), discrimination describes the denial of equal treatment to individuals or groups. Victims of discrimination at work often ruminate and seek explanations for what happened (Snape & Redman, 2003), making them more sensitive to injustice and vigilant to future violations (Dipboye & Colella, 2013). Several negative organizational consequences can follow when individuals feel discriminated against. For example, employees are more likely to engage in counterproductive work behavior (Follmer et al., 2023), whereas organizations often face increased turnover rates and damage to their reputation (Dipboye & Colella, 2013).

The role of prior discrimination regarding justice evaluations may be explained by the justice expectations model (Bell et al., 2004) and the concept of anticipatory injustice (Shapiro and Kirkman 2001). Bell et al. (2004) argue that individuals develop expectations about justice based on their previous experiences of organizational treatment. These expectations, in turn, influence how they evaluate future interactions. This aligns with Shapiro and Kirkman’s (2001) idea of anticipatory injustice, which posits that individuals who have experienced unjust treatment in the past are more likely to anticipate and interpret future encounters similarly. Consequently, in cases in which individuals have been discriminated against by human supervisors, we propose that they prefer AM systems over a human counterpart. This preference may stem from perceiving AM as being objective and consistent (Chamorro-Premuzic & Ahmetoglu, 2016) in contrast to a human manager, who may be seen as potentially biased and unjust, especially considering past treatment. Thus, we propose the following:

Hypothesis 2*: Prior human discrimination experience moderates the relationship between AM and (a) informational justice, (b) interpersonal justice, and (c) procedural justice, such that the negative impact of AM use on justice evaluations is reduced for those individuals who have experienced human discrimination.*

4.2.3 Prior algorithmic discrimination experience

Besides human discrimination, individuals can experience discrimination through AM systems (Kellogg et al., 2020; Kordzadeh & Ghasemaghaei, 2021). Here, algorithmic discrimination can emerge from various sources, such as in the software development phase, when, for example, organizations define goals that purposely exclude specific groups (Garcia et al., 2024). Besides this, algorithmic discrimination could occur during the data processing phase, specifically when biased data or sampling methods are used, or during the development phase, and discrimination may be apparent depending on how the systems' results are used in an organization (Garcia et al., 2024).

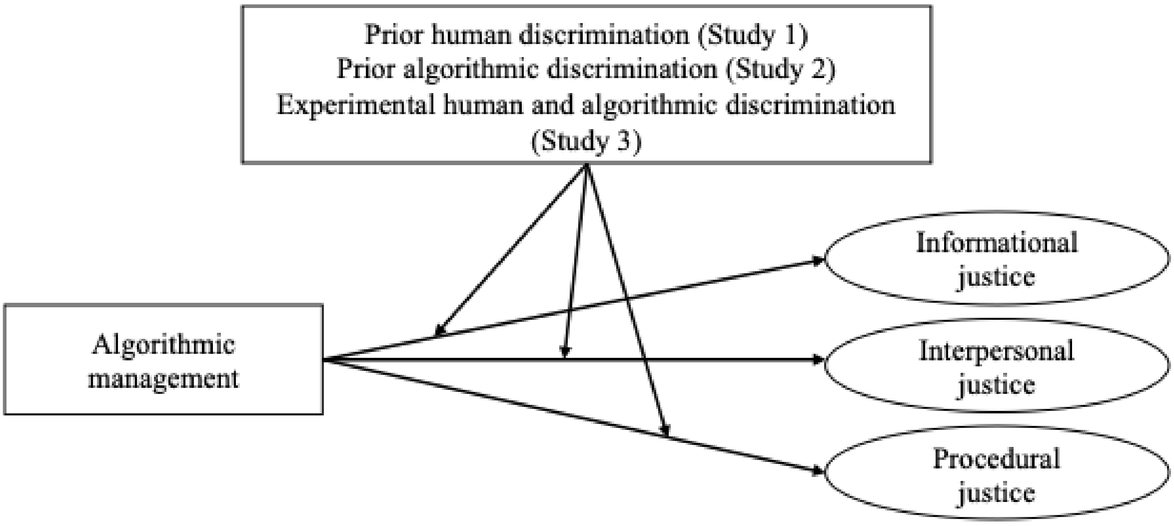
In the context of AM, algorithmic discrimination may be particularly important, as workers may have limited opportunities to address perceived discrimination from AM systems (Kellogg et al., 2020). According to the CASA framework, individuals apply interactional norms and expectations from human–human interactions to their interactions with technology, including AM systems. Therefore, theories developed in human–human contexts, such as the model of justice expectations (Bell et al., 2004) and the idea of anticipatory injustice (Shapiro & Kirkman, 2001), may also extend to human–algorithm interactions. As a result, individuals who have experienced discrimination by AM systems in the past may cognitively process these events similarly to human discrimination experiences, leading them to expect to be treated unjustly again, which leads to negative justice evaluations. We thus propose the following:

***Hypothesis** 3: Prior algorithmic discrimination experience moderates the relationship between AM and (a) informational justice, (b) interpersonal justice, and (c) procedural justice, such that the negative impact of AM use on justice is strengthened for those individuals who have experienced algorithmic discrimination.*

4.3 Overview of studies

We conducted three experimental studies to test our research model (Figure 4-1). In Study 1, we experimentally manipulated the management entity (HM versus AM) to examine differences in justice evaluations and to investigate the proposed moderating effect of prior human discrimination experience and thus examined hypotheses 1 and 2. In Study 2, we replicated this experiment, but we examined the moderating effect of prior algorithmic discrimination experience on the relationship between AM and justice evaluations, testing hypothesis 1 and 3. Given that we relied on the prior experiences of discrimination of our respondents in Study 1 and Study 2, we conducted a follow-up study (Study 3a) to manipulate discrimination experience and replicate our findings. Afterward, we analyzed the same respondents' answers to qualitative, open-ended questions to explore potential mechanisms underlying our findings (Study 3b).

Figure 4-1: Research model



Note. Reference group, algorithmic management: human management (= 0).

4.4 Study 1

We conducted a single-factor between-subjects experiment to test the theoretical research model (Figure 4-1) and applied covariance-based structural equation modeling (SEM) with the R-package lavaan (Rosseel, 2012).

4.4.1 Methods

4.4.1.1 Experimental procedure and sample

Within the experiment, each participant was randomly assigned to one of two written scenarios. The scenarios were fictitious situations from everyday working life in which participants were asked to imagine they worked for a company and were given feedback in the course of annual performance review meetings by either HM or an AM system (depending on the assigned scenario). We used the fictional company "Marzeo AG" and corresponding URL (www.marzeo.de) developed by Evertz et al. (2019). This allowed us to rule out that the participants would find contradictory information about the company online. Scenario descriptions for all three studies are presented in the Appendix C1-C3.

We recruited an online sample with $n = 86$ German participants. All participants had to have working experience in order to take part in the survey. None of the participants failed the manipulation checks. Because we assessed perceived discrimination by a (former) supervisor, we had to exclude four participants, as they indicated they did not have a supervisor before, leaving us with a final sample size of $n = 82$ ($M_{age} = 34.82$; $SD_{age} = 14.95$; male = 32.92%) participants.

4.4.1.2 Measures

We used established multi-item measures to measure all our constructs and adapted the items that measured justice dimensions according to the respective scenario. We measured informational, interpersonal, and procedural justice with an adapted version of the organizational justice scale by Colquitt (2001), which was translated to German and validated by Maier et al. (2007). A sample item for interpersonal justice measured in the HM scenario is “My manager treated me with respect.” A sample item for informational justice is “My manager tailored the feedback to my individual needs” and, for procedural justice, “I have been able to express my views and feelings during the performance review meeting.” We assessed prior human discrimination experience using an adapted version of the discrimination scale by Snape and Redman (2003). We modified the scale to assess all kinds of discrimination experience at work and to assess discrimination by a (former) supervisor. A sample item is “A supervisor has treated me less favorably than other workers.” We used experience with AI as a control variable. We assessed all items using a 7-point Likert scale ranging from “strongly disagree” to “strongly agree.” Cronbach’s alphas of all constructs exceed the threshold of 0.70 (Cortina, 1993), indicating the reliability of our scales (informational justice: $\alpha = 0.84$; interpersonal justice: $\alpha = 0.92$; procedural justice: $\alpha = 0.85$; prior discrimination experience: $\alpha = 0.88$; experience with AI: $\alpha = 0.85$). To test our treatments, we applied dummy coding to dichotomize the scenarios (AM = 1, HM = 0).

Although we randomly assigned participants to the scenarios, we calculated an analysis of variance (ANOVA) before testing our hypotheses. This was done to ensure that demographic variables and prior human discrimination experience did not significantly differ between groups, and no significant differences were found.

We used confirmatory factor analysis (CFA) to verify several validity criteria. The CFA demonstrated a satisfactory fit (χ^2 = 60.31; df = 32; χ^2/df = 1.88; p <.001; CFI = .93; RMSEA = .10; SRMR = 0.05; Hair et al., 2014). All factor loadings were significant and ranged between .61 and .97. Average variance extracted for all constructs is above the threshold of .50 (informational justice: .58; interpersonal justice: .84; procedural justice: .68), indicating convergent validity (Fornell & Larcker, 1981; Hair et al., 2014), and composite reliabilities are above the threshold of .60 (Bagozzi & Yi, 1988) for all our constructs (informational justice: .84; interpersonal justice: .94; procedural justice: .87). Additionally, discriminant validity is demonstrated for all constructs, as the average variance extracted exceeded squared correlations with other constructs (Fornell & Larcker 1981).

We addressed common method bias with the use of a marker variable that was theoretically uncorrelated with the focal constructs (Lindell & Whitney, 2001). After correction of the correlation matrix for correlations with the marker variable, all correlations between focal constructs remained significant. Hence, the threat of common method bias is minimized within our data. We used demographic variables (i.e., age, gender) and experience with AI as control variables. We dummy-coded gender and used males as a reference group. Table 4-1 shows the means, standard deviations, and correlations of our study variables.

Table 4-1: Means, standard deviations, and correlations for variables in Study 1

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. Algorithmic management	0.49	0.50							
2. Informational justice	4.92	1.36	–.49**						
3. Interpersonal justice	5.19	1.44	–.67**	.60**					
4. Procedural justice	4.27	1.55	–.71**	.68**	.70**				
5. Prior human discrimination	4.13	1.69	.20	–.18	–.17	–.10			
6. Age	34.82	14.95	.09	–.25*	–.27*	–.17	.09		
7. Gender	0.67	0.47	.01	.25*	.08	.01	–.18	–.37**	
8. Experience with AI	4.81	1.45	–.00	.02	.10	.00	.06	–.57**	.15

Note. M and *SD* are used to represent mean and standard deviation, respectively. Reference group, algorithmic management: human management (= 0). Reference group, gender: males (= 0). * indicates p <.05; ** indicates p <.01.

4.4.2 SEM results

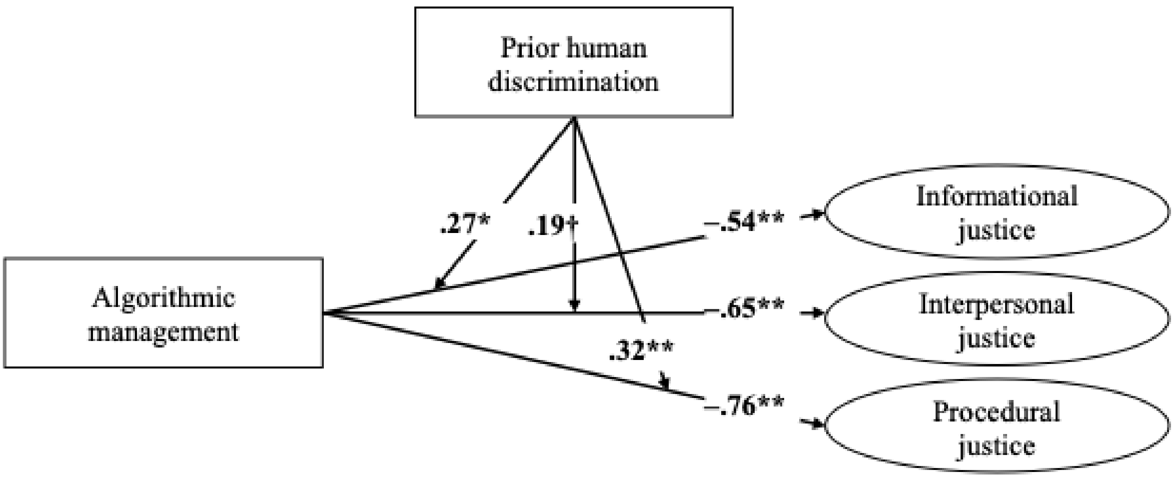
Due to our relatively small sample size, we applied the Swain correction (Boomsma & Herzog, 2013) to adjust for small-sample bias in the SEM fit indices, and our model showed a satisfactory fit (χ^2 = 115.15; *df* = 74; χ^2/*df* = 1.56; p = .002; CFI = .93; RMSEA = .08; SRMR = .06; Hair et al. 2014). The results demonstrate that, compared to HM, AM is negatively associated with informational (β = –.54; p <.001), interpersonal (β = –.65; p <.001), and procedural (β = –.76; p <.001) justice, supporting hypotheses 1a, 1b, and 1c. We found a moderating effect of prior human discrimination on the relationship between AM and both informational (β = .27; p = .030) and procedural (β = .32; p =.007) justice, whereas this moderating effect on the relationship between AM and interpersonal justice was significant only at a 10% p-level (β = .19; p = .066). This supports hypotheses 2a and 2c. Regarding our control variables, we found a significant effect of age on interpersonal justice (β = −.28; p =.012). The SEM results are depicted in Figure 4-2 and Table 4-2.

Table 4-2: SEM results, Study 1

Path	*B*	*SE*	β	*p*
Hypothesized effects				
AM → Informational justice	–1.45	(.28)**	–.54	<.001
AM → Interpersonal justice	–1.93	(.29)**	–.65	<.001
AM → Procedural justice	–2.55	(.51)**	–.76	<.001
Prior human discrimination → Informational justice	–.25	(.13)†	–.19	.055
Prior human discrimination → Interpersonal justice	–.22	(.12)†	–.15	.075
Prior human discrimination → Procedural justice	–.32	(.17)†	–.19	.055
AM x Prior human discrimination → Informational justice	.56	(.26)*	.27	.030
AM x Prior human discrimination → Interpersonal justice	.44	(.24)†	.19	.066
AM x Prior human discrimination → Procedural justice	.81	(.30)**	.32	.007
Control effects				
Age → Informational justice	–.02	(.01)	–.19	.104
Age → Interpersonal justice	–.03	(.01)**	–.28	.012
Age → Procedural justice	–.02	(.01)	–.14	.241
Gender → Informational justice	.55	(.32)†	.19	.090
Gender → Interpersonal justice	–.07	(.30)	–.02	.825
Gender → Procedural justice	–.21	(.36)	–.06	.557
Experience with AI → Informational justice	–.13	(.10)	–.14	.198
Experience with AI → Interpersonal justice	–.02	(.10)	–.02	.855
Experience with AI → Procedural justice	–.10	(.10)	–.09	.315

Note. *B* = unstandardized effect; *SE* = standard error; β = standardized effect. Reference group, algorithmic management: human management (= 0). Reference group, gender: males (= 0). † indicates p <.10; * indicates p <.05; ** indicates p <.01.

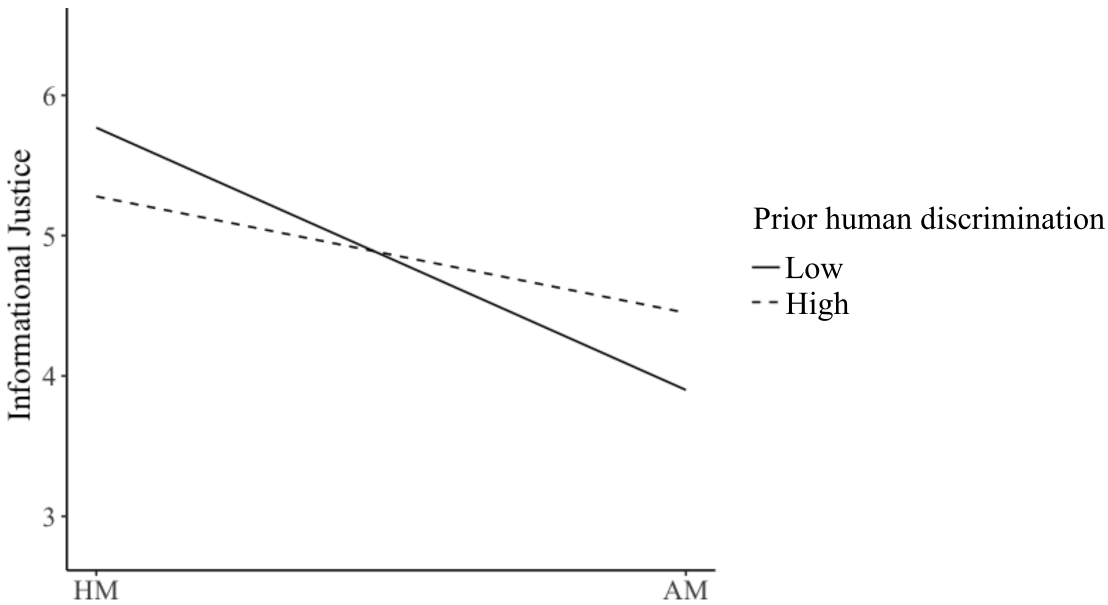
Figure 4-2: SEM results, Study 1



Note. Reference group, algorithmic management: human management (= 0). Control variables (not depicted): (1) age, (2) gender, (3) experience with AI. † indicates p <.10; * indicates p <.05; ** indicates p <.01.

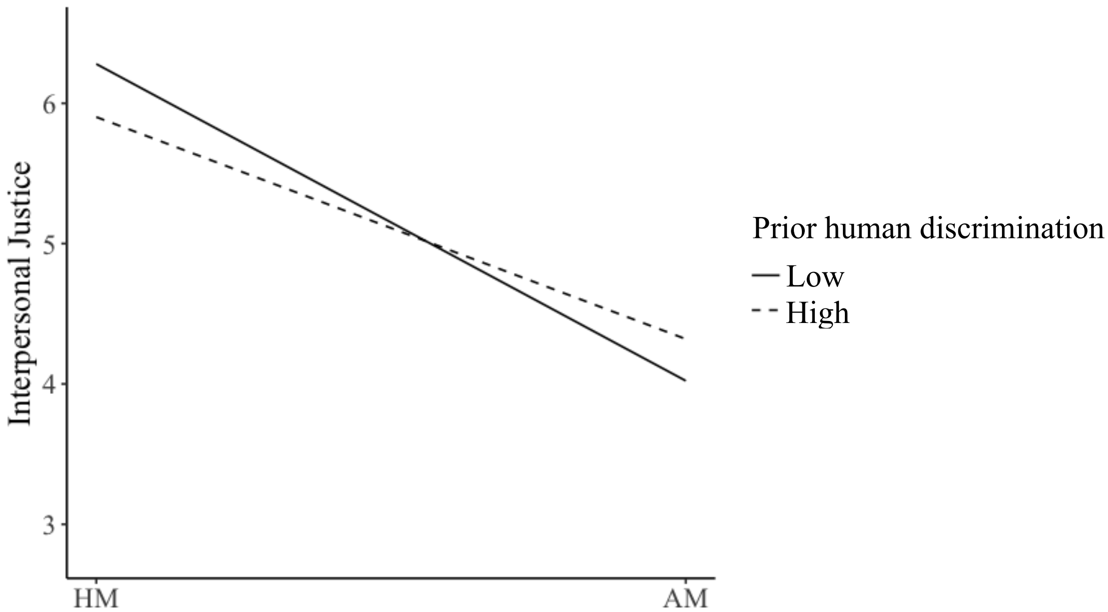
To provide a better understanding of the interactions, we illustrate the effects of human discrimination experience (operationalized using a median split) on the relationship between AM use and justice evaluations in Figures 4-3, 4-4, and 4-5.

Figure 4-3: Interaction effect of prior human discrimination on informational justice



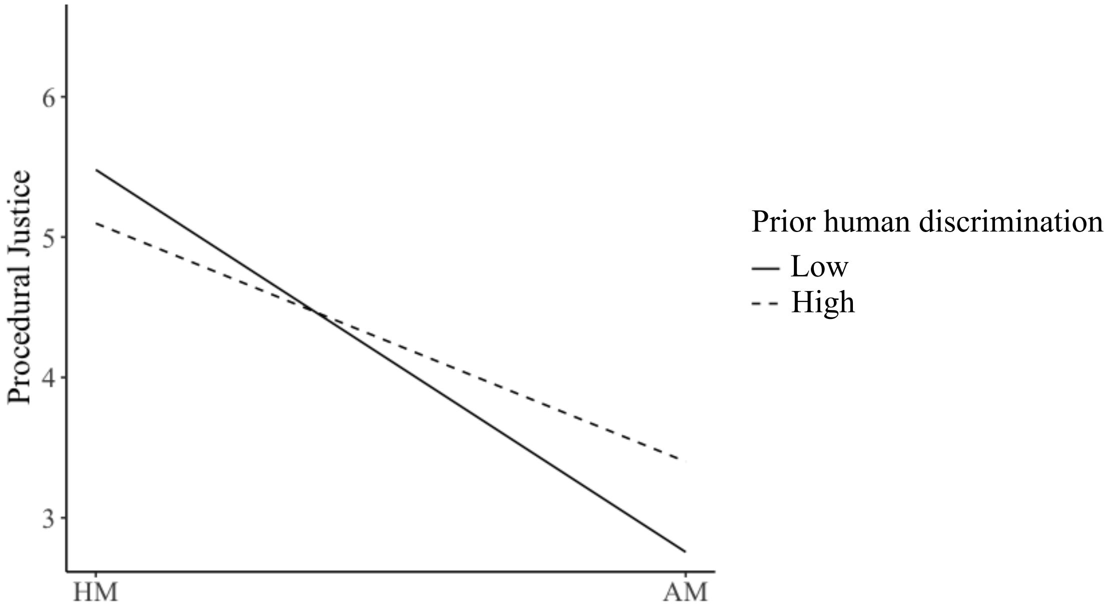
Note. HM = human management; AM = algorithmic management.

Figure 4-4: Interaction effect of prior human discrimination on interpersonal justice



Note. HM = human management; AM = algorithmic management.

Figure 4-5: Interaction effect of prior human discrimination on procedural justice



Note. HM = human management; AM = algorithmic management.

4.4.3 Brief discussion of Study 1

Study 1 demonstrates that individuals with higher levels of prior human discrimination experience perceive AM as less unjust compared to those with lower levels of such experience. This is aligned with findings in the recruiting context (e.g., Koch-Bayram et al., 2023; Schulte Steinberg & Hohenberger, 2023) and may be explained by the association of AM systems with being more consistent and objective (Koch-Bayram et al., 2023), two aspects that are particularly important regarding discrimination experiences. To compare the impact of prior human discrimination and prior algorithmic discrimination, we examined whether the moderating impact of prior discrimination is also apparent regarding algorithmic discrimination in Study 2.

4.5 Study 2

4.5.1 Methods

4.5.1.1 Experimental procedure and sample

In Study 2, we used the exact same experimental procedure and measurements. To examine the impact of prior algorithmic discrimination, we used Prolific Academic to recruit individuals who were frequently exposed to AM systems (i.e., gig workers). To ensure that only gig workers took part in the study, we employed a screening process whereby individuals were asked to describe their current job and the role of the AM system. All individuals were paid for taking part in the screening survey, but only those with AM exposure were invited to take part in the main study. We recruited an online sample with n = 89 German participants. Six participants were excluded for indicating that they faced a human manager although they were in the AM group and vice versa, leaving us with a final sample of n = 83 (M_{age} = 31.57; SD_{age} = 9.42; male = 56.62%) participants.

4.5.1.2 Measures

We used the same measures as in Study 1. To assess prior algorithmic discrimination, we adapted the scale used in Study 1 to refer only to AM systems. A sample item is "The algorithmic management system has treated me less favorably than other workers." Cronbach's alphas of all constructs exceed the threshold of .70 (Cortina, 1993), indicating the reliability of our scales (informational justice: α = .91; interpersonal justice: α = .95; procedural justice: α = .84; prior algorithmic discrimination: α = .94; experience with AI: α = .81). To test our treatments, we applied dummy coding to dichotomize the scenarios (AM = 1, HM = 0).

Similarly to the procedure of Study 1, we calculated an ANOVA and found no differences in the demographic variables or prior algorithmic discrimination experiences between our two groups.

To verify validity criteria, we calculated a CFA. The CFA fits the data well (χ^2 = 42.92; df = 32; χ^2/df = 1.34; p = .094; CFI = .98; RMSEA = .06; SRMR = .03; Hair et al., 2014). All factor loadings are significant and range between .74 and .98. The average variance extracted for all constructs is above the threshold of .50 (informational justice: .74; interpersonal justice: .88; procedural justice: .64), indicating that the convergent validity (Fornell & Larcker, 1981; Hair et al., 2014) and composite reliabilities are above the threshold of .60 (Bagozzi & Yi, 1988) for all our constructs (informational justice: .91; interpersonal justice: .96; procedural justice: .84). Additionally, discriminant validity is demonstrated for all constructs, as the average variance extracted exceeds squared correlations with other constructs (Fornell & Larcker, 1981). We followed the same procedure as in Study 1 to address common methods bias, used the same control variables, and dummy-coded gender (reference group = males and others). Table 4-3 shows the means, standard deviations, and correlations of our study variables.

Table 4-3: Means, standard deviations, and correlations of variables in Study 2

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. Algorithmic management	0.49	0.50							
2. Informational justice	5.12	1.33	$-.51^{**}$						
3. Interpersonal justice	5.13	1.44	$-.47^{**}$	$.84^{**}$					
4. Procedural justice	3.95	1.54	$-.46^{**}$	$.58^{**}$	$.53^{**}$				
5. Prior algo. discrimination	3.06	1.84	−.09	−.09	−.15	.13			
6. Age	31.57	9.42	.02	−.14	−.08	−.04	.20		
7. Gender	0.41	0.49	−.04	.16	.18	−.01	−.10	.16	
8. Experience with AI	5.81	1.15	−.11	.08	.04	.06	−.00	.01	−.09

Note. Algo. = algorithmic. *M* and *SD* are used to represent mean and standard deviation, respectively. Reference group, algorithmic management: human management (= 0). Reference group, gender: males and others (= 0). * indicates p <.05; ** indicates p <.01.

4.5.2 SEM results

Due to our relatively small sample size, we applied the Swain correction (Boomsma & Herzog, 2013) to adjust for small-sample bias in the SEM fit indices, and our model showed a good fit (χ^2 = 87.16; df = 74; χ^2/df = 1.18; p = .141; CFI = .98; RMSEA = .05;

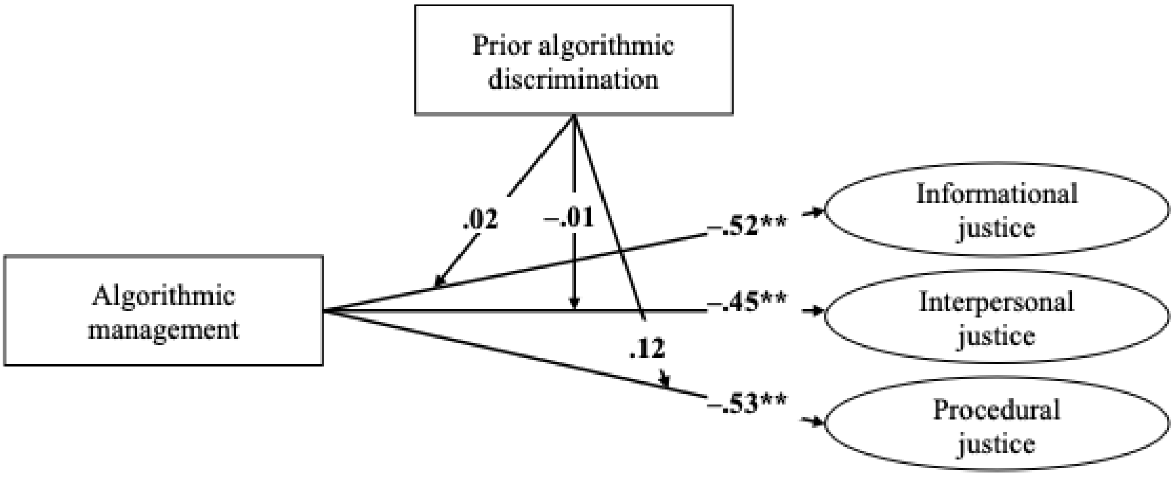
SRMR = .04; Hair et al. 2014). The results demonstrate that, compared to HM, AM is negatively associated with informational (β = −.52; p <.001), interpersonal (β = −.45; p <.001), and procedural (β = −.53; p <.001) justice, which replicates the findings of Study 1 and supports hypotheses 1a, 1b, and 1c. None of the hypothesized moderating effects of prior algorithmic discrimination affected the relationship between AM and justice. Thus, we have to reject hypothesis 3. Regarding our control variables, we found a significant effect of gender on informational (β = −.17; p =.024), and interpersonal (β = −.19; p =.028) justice. The SEM results are depicted in Figure 4-6 and Table 4-4.

Table 4-4: SEM results, Study 2

Path	B	*SE*	β	*p*
Hypothesized effects				
AM → Informational justice	−1.28	(.27)**	−.52	<.001
AM → Interpersonal justice	−1.04	(.27)**	−.45	<.001
AM → Procedural justice	−1.27	(.32)**	−.53	<.001
Prior algo. discrimination → Informational justice	−.14	(.15)	−.11	.356
Prior algo. discrimination → Interpersonal justice	−.18	(.13)	−.15	.179
Prior algo. discrimination → Procedural justice	.00	(.17)	.00	.984
AM x Prior algo. discrimination → Informational justice	.04	(.30)	.02	.890
AM x Prior algo. discrimination → Interpersonal justice	−.02	(.27)	−.01	.928
AM x Prior algo. discrimination → Procedural justice	.21	(.28)	.12	.447
Control effects				
Age → Informational justice	–.02	(.01)	−.14	.122
Age → Interpersonal justice	–.00	(.01)	−.03	.713
Age → Procedural justice	–.00	(.01)	−.04	.745
Gender → Informational justice	–.39	(.18)*	−.17	.024
Gender → Interpersonal justice	–.42	(.19)*	−.19	.028
Gender → Procedural justice	–.04	(.22)	−.02	.852
Experience with AI → Informational justice	–.06	(.10)	−.05	.585
Experience with AI → Interpersonal justice	–.02	(.10)	−.02	.856
Experience with AI → Procedural justice	.04	(.11)	.04	.722

Note. Algo. = algorithmic; B = unstandardized effect; *SE* = standard error; β = standardized effect. Reference group, algorithmic management: human management (= 0). Reference group, gender: males (= 0). † indicates p <.10; * indicates p <.05; ** indicates p <.01.

Figure 4-6: SEM results, Study 2



Note. Reference group, algorithmic management: human management (= 0). Control variables (not depicted): (1) age, (2) gender, (3) experience with AI. * indicates p <.05; ** indicates p <.01.

4.5.3 Brief discussion of Study 2

Study 2 demonstrates that individuals with high algorithmic discrimination experience do not react more aversely to AM systems compared to individuals with low algorithmic discrimination experience. The findings of Studies 1 and 2 thus indicate that discrimination by humans and algorithms is perceived differently. It follows that, although the CASA framework (Nass et al., 1994) postulates that there are several similarities between human–human and human–algorithm interactions, that framework may not be applicable here. Instead, different attributional processes may be apparent. Bonezzi and Ostinelli (2021) postulate that people believe algorithms neglect individual characteristics (which may lead to discriminatory decisions) and instead blindly apply rules to make decisions. This reasoning may explain why we found different moderating effects of discrimination in Study 1 and Study 2. However, we did not directly compare both discrimination experiences and were unable to control for the type and extent of discrimination. Thus, Study 3 experimentally manipulated discrimination experience and compared discrimination by humans and AM systems directly.

4.6 Study 3a

4.6.1 Methods

4.6.1.1 Experimental procedure and sample

Within the experiment, each participant was randomly assigned to one of four written scenarios. Similarly to Study 1 and Study 2, we manipulated type of management (HM versus AM). Additionally, we experimentally manipulated discrimination experience (no discrimination versus discrimination). In the discrimination scenario, participants received the general introduction that they worked for the Marzeo AG and were assigned tasks and got feedback from either a human manager or an AM system (depending on their assigned scenario). This general introduction was shown to all participants.

For the manipulation of discrimination, participants in the treatment groups read the following information (AM condition in parentheses): "Later this week, you read a report that reveals troubling patterns in the management practices (algorithmic management system) at Marzeo AG, including the manager (management system) that manages you and your team. The report shows that employees from certain demographic groups tend to receive more prestigious and challenging assignments. You notice that, despite your excellent track record, you are consistently assigned fewer challenging tasks compared to others, and this pattern aligns with the findings in the report. You begin to wonder whether the decisions of your manager (algorithmic management system) are biased."

In the follow-up, participants answered the study measures first and were then asked open-ended qualitative questions (see Study 3b). We debriefed the participants after the survey, indicating that the scenarios were entirely fictitious and that Marzeo AG was not a real company. The study protocol was reviewed and approved by the Ethics Committee of the German Association for Experimental Economic Research.

We recruited an online sample with $n = 216$ participants from the American working population. We used Prolific Academic to recruit our sample and directly screened out participants who failed our manipulation or attention checks and thus proceeded with a final sample size of $n = 216$ ($M_{age} = 36.13$; $SD_{age} = 13.17$; male = 51.85%; diverse = .01%) participants.

4.6.1.2 Measures

For informational, interpersonal, and procedural justice, we used the same measures that were used in Study 1. We controlled for attitude toward AI (Sindermann et al., 2021). Cronbach's alphas of all constructs exceed the threshold of .70 (Cortina, 1993), indicating the reliability of our scales (informational justice: $\alpha = .92$; interpersonal justice: $\alpha = .96$; procedural justice: $\alpha = .86$; attitude towards AI: $\alpha = .78$). To test our treatments, we applied dummy coding to dichotomize the scenarios (AM = 1, HM = 0; discrimination = 1, no discrimination = 0). Additionally, we dichotomized ethnic background to be able to include this variable in our analysis, following the approach of Koch-Bayram et al. (2023). We hence operationalized ethnic background as a dummy variable, with participants who selected "white" as their ethnic background as a reference group (= 0) compared to all other ethnic groups (e.g., Hispanic, Black, Asian = 1).

Although we randomly assigned participants to the scenarios, we calculated an ANOVA to examine differences regarding demographic variables and the four groups. We found significant differences regarding age ($F(3, 212) = 3.18$; $p = .025$), such that individuals in the AM × discrimination group were slightly younger ($n = 50$; $M = 32.20$; $SD = 11.20$) compared to those in the AM × no discrimination group ($n = 49$; $M = 39.98$; $SD = 13.57$). We also found significant differences regarding ethnic background ($F(3, 212) = 4.46$, $p <.01$), such that there were fewer individuals who identified as "white" in the AM × no discrimination scenario ($n = 49$; $M = .76$; $SD = .43$) compared to those in the AM ×

discrimination scenario (n = 50; M = .44; SD = .50). We did not find differences between the four groups regarding gender or attitude towards AI.

Again, we used CFA to verify several validity criteria. The CFA fits the data well (χ^2 = 44.52; df = 32; χ^2/df = 1.39; p = .070; CFI = .99; RMSEA = .04; SRMR = .03; Hair et al., 2014). All factor loadings are significant and range between .81 and .96. The average variance extracted for all constructs is above the threshold of .50 (informational justice: .76; interpersonal justice: .90; procedural justice: .68), indicating convergent validity (Fornell & Larcker, 1981; Hair et al., 2014), and composite reliabilities are above the threshold of .60 (Bagozzi & Yi, 1988) for all our constructs (informational justice: .93; interpersonal justice: .97; procedural justice: .87). Additionally, discriminant validity is demonstrated for all constructs, as the average variance extracted exceeds squared correlations with other constructs (Fornell & Larcker, 1981).

We followed the same procedure used in Study 1 and Study 2 to address common method bias. Additionally, we used demographic variables (age, gender, ethnic background, and attitude towards AI) as control variables to examine the impact of discrimination experience beyond these characteristics. We dummy-coded gender and used males and others as a reference group. Table 4-5 shows the means, standard deviations, and correlations of our study variables.

Table 4-5: Means, standard deviations, and correlations of variables in Study 3

Variable	M	SD	1	2	3	4	5	6	7	8
1. Algorithmic management	0.46	0.50								
2. Experimental discrimination	0.54	0.50	−.07							
3. Informational justice	4.57	1.58	−.03	$-.51^{**}$						
4. Interpersonal justice	4.87	1.72	.00	$-.50^{**}$	$.81^{**}$					
5. Procedural justice	4.01	1.64	−.02	$-.37^{**}$	$.78^{**}$	$.69^{**}$				
6. Age	36.13	13.17	−.01	$-.17^{*}$	.02	.03	−.05			
7. Gender	0.48	0.50	−.00	−.09	.07	.03	.01	$.15^{*}$		
8. Ethnic background	0.54	0.50	.11	$-.18^{**}$	.05	.04	−.05	$.32^{**}$	.12	
9. Attitude towards AI	3.33	0.81	−.06	.10	.07	.09	.07	−.07	−.13	$-.17^{*}$

Note. M and SD are used to represent mean and standard deviation, respectively. Reference group, algorithmic management: human management (= 0). Reference group, experimental discrimination: No discrimination (=0). Reference group, ethnic background: individuals who identify themselves as "white" (= 0). Reference group, gender: males and others (= 0). * indicates p <.05; ** indicates p <.01.

4.6.2 SEM results

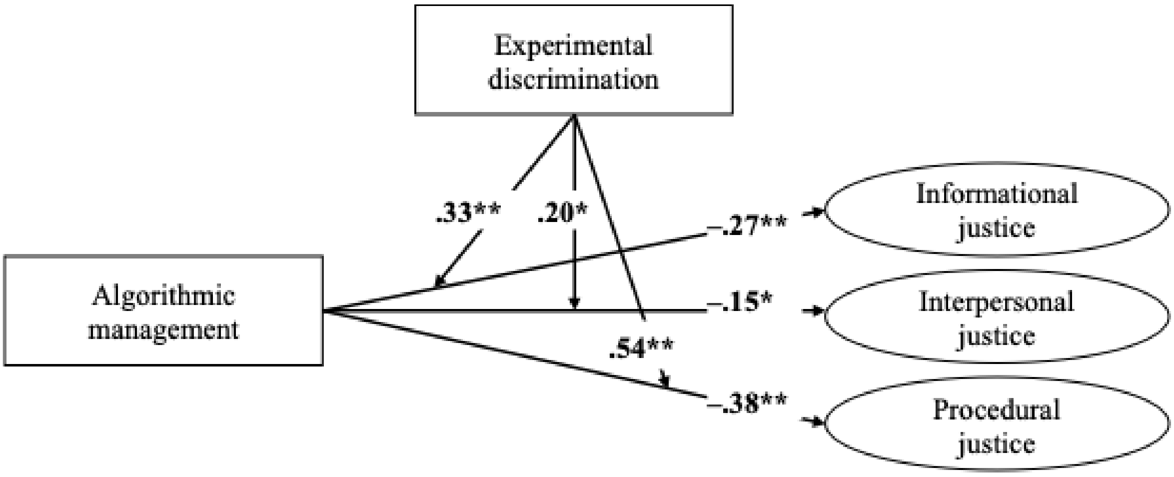
To test our research model, we calculated an SEM, and the overall model demonstrated a good fit (χ^2 = 100.58; df = 81; χ^2/df = 1.24; p = .069; CFI = .99; RMSEA = .03; SRMR = .02; Hair et al. 2014). The results demonstrate that, compared to HM, AM was negatively associated with informational (β = −.27; p <.001), interpersonal (β = −.15; p = .021), and procedural (β = −.38; p <.001) justice. We found a moderating effect of experimental discrimination on the relationship between AM and informational (β = .33; p = .001), interpersonal (β = .20; p =.032), and procedural (β = .54; p <.001) justice. The moderating effects reflect that although all individuals perceived discrimination as less just, this effect was weaker in the AM scenarios. We found no significant effects of our control variables. The SEM results are depicted in Table 4-6 and Figure 4-7. The moderating effects are depicted in Figures 4-8, 4-9, and 4-10.

Table 4-6: SEM results, Study 3

Path	*B*	*SE*	β	*p*
Hypothesized effects				
AM → Informational justice	−.67	(.17)**	−.27	<.001
AM → Interpersonal justice	−.37	(.16)*	−.15	.021
AM → Procedural justice	−.92	(.20)**	−.38	<.001
Experimental discrimination → Informational justice	−1.80	(.21)**	−.73	<.001
Experimental discrimination → Interpersonal justice	−1.53	(.19)**	−.64	<.001
Experimental discrimination → Procedural justice	−1.73	(.23)**	−.73	<.001
AM x experimental discrimination → Informational justice	.98	(.30)**	.33	.001
AM x experimental discrimination → Interpersonal justice	.58	(.27)*	.20	.032
AM x experimental discrimination → Procedural justice	1.52	(.33)**	.54	<.001
Control effects				
Age → Informational justice	−.01	(.01)	−.05	.365
Age → Interpersonal justice	−.00	(.01)	−.03	.569
Age → Procedural justice	−.01	(.01)	−.07	.325
Gender → Informational justice	.10	(.15)	.04	.504
Gender → Interpersonal justice	.00	(.14)	.00	1.000
Gender → Procedural justice	.00	(.16)	.00	.989
Ethnic background → Informational justice	.04	(.16)	.02	.800
Ethnic background → Interpersonal justice	−.01	(.14)	−.01	.935
Ethnic background → Procedural justice	−.14	(.17)	−.06	.385
Attitude towards AI → Informational justice	.17	(.10)	.11	.072
Attitude towards AI → Interpersonal justice	.18	(.10)	.13	.056
Attitude towards AI → Procedural justice	.10	(.10)	.07	.308

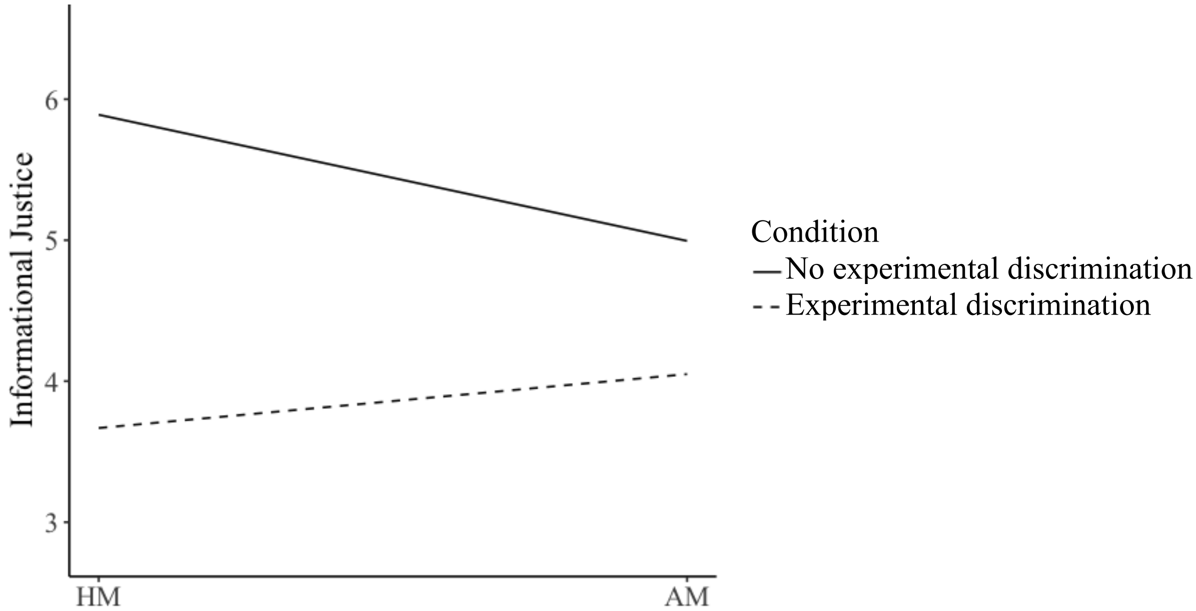
Note. B = unstandardized effect; *SE* = standard error; β = standardized effect. Reference group, algorithmic management: human management (= 0). Reference group, experimental discrimination: No discrimination (= 0). Reference group, ethnic background: individuals who identify themselves as “white” (= 0). Reference group, gender: males and others (= 0). * indicates p <.05; ** indicates p <.01.

Figure 4-7: SEM results, Study 3



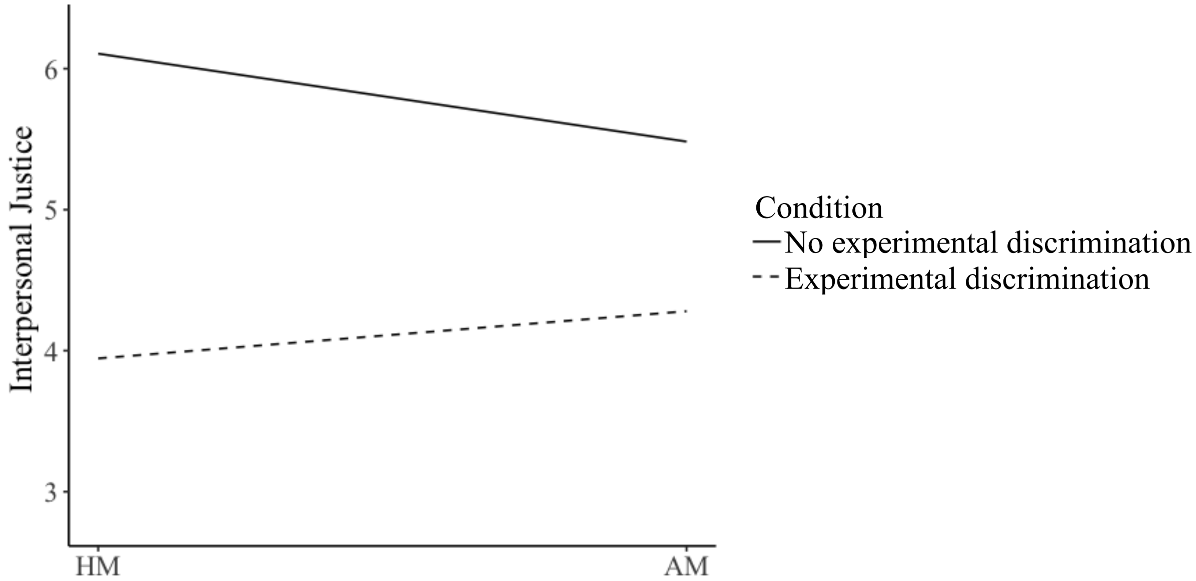
Note. Reference group, discrimination: no experimental discrimination experience (= 0). Reference group, algorithmic management: human management (= 0). Control variables (not depicted): (1) age, (2) gender, (3) ethnic background, (4) attitude toward AI. [*] indicates p <.05; [**] indicates p <.01.

Figure 4-8: Interaction effect of experimental discrimination on informational



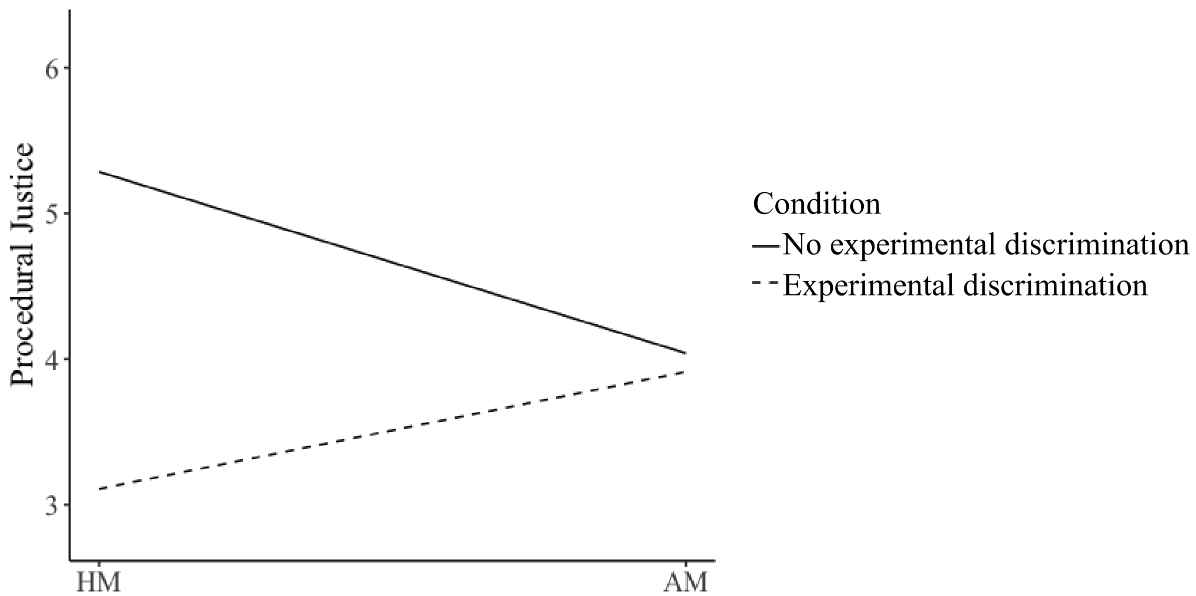
Note. HM = human management; AM = algorithmic management.

Figure 4-9: Interaction effect of experimental discrimination on interpersonal justice



Note. HM = human management; AM = algorithmic management.

Figure 4-10: Interaction effect of experimental discrimination on procedural justice



Note. HM = human management; AM = algorithmic management.

4.6.3 Brief discussion of Study 3a

Study 3a directly compared human and algorithmic discrimination experiences. The results demonstrate that human discrimination evokes more negative justice perceptions compared to discrimination by AM systems. Consistent with the findings of Studies 1 and

2, this shows that individuals seem to perceive human discrimination as worse than algorithmic discrimination. Bigman et al. (2023) suggest that algorithmic discrimination causes less moral outrage than human discrimination due to individuals being less likely to attribute prejudicial motivation to algorithms. Our findings may also be explained by the fact that algorithmic discrimination could be attributed to various sources, such as the developers or the organization using these systems (Garcia et al., 2024). To investigate the mechanisms underlying these results, we examined open-ended survey responses in Study 3b.

4.7 Study 3b

4.7.1 Qualitative analysis of open-ended survey responses

We followed the approach of Bankins et al. (2022) and asked participants several open-ended qualitative questions at the end of our online experiment. Specifically, we asked participants what they liked or disliked about the management type presented in their assigned scenario, and how their perceptions related to other types of management (AM or HM, depending on the scenario). Figure 4-11 depicts illustrative quotes from respondents (R) by experimental group.

Figure 4-11: Sample quotes per experimental group

Condition	No Discrimination	Discrimination
Human Management	"[…] For example, a human supervisor can understand sickness and childcare. They are able to reason and feel compassion and empathy for human situations." (R131), "Being managed by a human supervisor is more humanizing. I feel like it opens the floor for communication, which isn't possible with automated supervision." (R204), "I'd like that an algorithmic management system could provide fairer, more consistent decisions based on data, but I'd miss personal touch and understanding that a human supervisor offers, especially in handling emotions or unique situations." (R212)	"With a human manager I feel like they are more judgmental and biased and pick and choose who they like and dislike and can make the job harder." (R94). "Human managers can introduce issues related to bias, inconsistency, and emotional influence, while algorithmic systems may struggle with lack of emotional understanding, rigidity, and data bias. Both systems have strengths and weaknesses, and the ideal management approach might involve a balance between human judgment and algorithmic support." (R76)
Algorithmic Management	"I would like to be managed by an algorithmic management system because as a female, I've experienced some discriminations at work. […] my experiences with some supervisors haven't been favourable or fair at all which has shaped my decision here." (R21), "The system doesn't play favorites, and you get feedback that's based on your actual performance, not just what someone else thinks about you. It's also kind of nice to know exactly what tasks you have coming up without any ambiguity." (R96), "A human has the potential to be persuaded about a topic or situation. An algorithm doesn't have human aspects to appeal to." (R125)	"I would like it if an algorithmic management system acted based on logic and data when assigning tasks and giving feedback. I am, however, worried about the transparency of the algorithmic management system." (R2), "Despite being designed to be objective, an algorithm could still be biased if the data it's using reflects past biases." (R187), "It [the algorithmic management system] gives everyone tasks based on clear rules, so it feels like there's less chance of bias." (R192)

Note. Respondents were randomly assigned to one of the four scenarios and answered the following question: "What did you (dis-)like about the management type presented in the scenario?"

Regarding algorithmic discrimination, participants expressed two main concerns: (1) the lack of transparency in AM systems and (2) the potential inflexibility of algorithms. Here, one respondent stated, "I would like it if an algorithmic management system acted based on logic and data when assigning tasks and giving feedback. I am, however, worried about the transparency of the algorithmic management system. I think that employees must be able to see and access the parameters of the algorithm at any time so that potential biases can be identified and fixed to avoid something like the bias in the scenario from happening" (R2). Additionally, individuals expected that AM systems were inflexible and might not account for all circumstances when assigning work: "While they are efficient, algorithms might not account for individual circumstances or nuances, leading to mismanagement or unfair task assignments" (R4). Additionally, when discriminated against by AM systems, individuals often referred to AM systems as perpetuating inequalities, which may indicate that they do not necessarily attribute the discrimination to the AM system itself but rather to biased data stemming from humans. Here, R117 stated, "If the data that was used to train the algorithm is biased due to societal prejudices, the system would further strengthen these biases and amplify them." Interestingly, even in the group that faced discrimination by AM systems, respondents still described algorithms as more consistent and objective compared to humans. Issues associated with AM systems referred rather to inflexibility, opacity, and impersonal interactions.

Individuals in the HM discrimination groups described AM systems as more objective and less biased, particularly in contexts where discrimination has historically been more prevalent. As one respondent stated, "It [the AM system] would in a broad sense be far less likely to be biased or discriminatory based off multiple factors, like race, religion, gender, sexual orientation […] especially for minorities or marginalized members of society or the workforce" (R108). Regarding human discrimination, many respondents referred to subconscious bias, noting, "I associate human managers with explicit and implicit bias that could be conscious or subconscious decisions. It allows certain groups to benefit, but other groups to suffer" (R145), reflecting that human managers may not be aware of discriminatory practices. Additionally, respondents often reported fears of bias, favoritism, inconsistencies, and emotional factors impacting work processes and decision-making. One respondent stated, "With humans I'm always going to associate bias, bad decision making, [and] favoritism" (R142). Similarly, another responded stated, "Human managers may be influenced by personal preferences, unconscious biases, or emotions" (R156). Additionally, some respondents reasoned why human managers might discriminate: "Human managers can bring emotions into their decisions, which might lead to biases, such as favoring individuals from their own ethnic group" (R69).

In summary, the qualitative responses revealed that individuals interpret discriminatory behavior differently depending on whether it originates in human or algorithmic management. Discrimination by human managers was often seen as stemming from personal biases directed toward the individual, whereas discrimination by AM systems was generally perceived as impersonal and group based. One respondent stated, "A human can have a bias against you. An algorithm is just data. No more or less" (R45). This reflects the idea that, even if AM and HM discriminate similarly, people may believe that humans do so intentionally or because they have something against a specific person, whereas algorithms do not do so intentionally. Further, themes such as favoritism or (unconscious) "emotional" decisions are likely to be associated with issues deriving from HM, whereas issues deriving from AM were rather associated with biased data, opacity, and inflexibility.

4.8 Discussion

4.8.1 Main findings and theoretical contributions

The overarching goal of this research was to advance our understanding of how prior discrimination experiences shape justice perceptions of AM. Consistent with previous research, we generally found negative relationships between AM use and informational, interpersonal, and procedural justice (e.g., Höddinghaus et al., 2021; Lee, 2018). Furthermore, we found significant interaction effects for prior discrimination experiences, such that the negative effect of AM use on justice was mitigated if participants had previously experienced human discrimination. Thus, our results reveal that employees' prior discrimination experiences by human managers impacted their justice evaluations, whereas individuals who had not experienced discrimination before consistently preferred HM. This may be explained by the model of justice expectations (Bell et al., 2004) and the idea of anticipatory injustice (Shapiro & Kirkman, 2001), wherein individuals draw on past experiences to infer how justly they will be treated in the future, which in turn impacts their justice evaluations. Our results align with research findings in the context of recruiting, in which individuals with discrimination experience were more likely to perceive AM positively (Fleiß et al., 2024; Koch-Bayram et al., 2023; Schulte Steinberg & Hohenberger, 2023). This may be the case because individuals who have experienced discrimination by a supervisor in the past may prefer an algorithm's objectivity and precision (Wang & Benbasat, 2016; You et al., 2022), leading to the belief that algorithms are less likely to discriminate compared to a human counterpart.

Interestingly, we did not find the above-described moderating effect regarding prior algorithmic discrimination. Specifically, in Study 2, individuals who felt discriminated against by AM systems did not show stronger algorithm aversion. This was replicated, as Study 3 demonstrates that while discrimination always led to more negative justice evaluations, this effect was stronger when humans discriminated compared to AM systems. Consequently, our assumption that, based on the CASA framework (Nass et al., 1994), discrimination experiences shape justice perceptions independently of the management type, was not applicable, thus challenging the current understanding and wide applicability of CASA.

Instead, these results may be explained by different attribution processes regarding the source of discrimination as reflected in our qualitative data. First, individuals may not associate algorithmic discrimination solely with the AM system itself but may also consider other sources of discrimination, such as biased data stemming from humans. Second, most respondents reflected that they believed AM systems to be more consistent, whereas humans were believed to "play favorites" (R96) and "be persuaded about a topic" (R125). Additionally, individuals associated human discrimination with intentional processes and seemed to take the discrimination more personally, whereas algorithmic discrimination was generally believed to be unintentional. This may further be strengthened as individuals often ruminate after experiencing discrimination to understand why they were treated unjustly (Crocker et al., 1991). Here, Bigman et al. (2023) suggest that algorithms are often perceived to possess no motivation to discriminate, because algorithms may be perceived as lacking a mind compared to humans. Thus, individuals may perceive algorithmic discrimination as less driven by prejudice, whereas prejudice is more readily assumed when humans discriminate (Bigman et al., 2023), leading to differing perceptions of discrimination depending on the source. Overall, this suggests that attribution of intention plays a central role in how individuals with discrimination experience form justice perceptions.

The finding that algorithmic discrimination was perceived as less severe than human discrimination may also be explained by the decontextualization associated with algorithms. Although prior research shows that decontextualization is a driver of negative justice perceptions regarding algorithmic decision-making (Newman et al., 2020), the perception that algorithms make decisions without being able to integrate contextual factors, such as personal characteristics, may lead to the belief that they are not able to discriminate intentionally (Bonezzi & Ostinelli, 2021). Thus, decontextualization may act as a double-edged sword: it lowers overall justice evaluations, yet it may also lead individuals to believe that algorithms lack discriminatory intent.

4.8.2 Practical implications

The results of this study provide several implications for organizations. First, while the use of AM offers several advantages for organizations, Studies 1 and 2 demonstrate that algorithm aversion prevails. Organizations should thus thoroughly review which tasks can be managed by AM and which should remain under human oversight or be conducted by a human counterpart to minimize potential backlash from employees, such as algoactivism

(Kellogg et al., 2020). This is particularly relevant for organizations, as employees are less likely to follow decisions perceived as unjust (Bankins et al., 2022; Lind, 2001).

Second, we found that prior human discrimination experience mitigates negative justice evaluations of AM. This is a significant insight, as it implies that novel technologies such as AM may be instrumental in reducing fears of experiencing discrimination again. By providing a more consistent and objective decision-making process, AM has the potential to create a workplace environment perceived as more impartial and just. This could be particularly impactful in settings where human biases have previously led to unequal treatment.

Third, our results demonstrate that while discrimination experience always led to negative justice perceptions, individuals perceived human discrimination as worse compared to algorithmic discrimination. Interestingly, our qualitative responses revealed that even in the experimental group that faced discrimination by AM systems, respondents still described AM systems as more consistent and objective than humans. Nevertheless, it is important to approach this with caution. While AM systems can reduce the subjectivity associated with human decision-making, they are not immune to biases, especially those that may be inadvertently embedded in their algorithms (Bonezzi & Ostinelli, 2021). Therefore, the design and implementation of these technologies must be carefully monitored and regularly audited for fairness and impartiality. Additionally, the fact that algorithmic discrimination was perceived as less severe than human discrimination carries the risk that people may not perceive algorithmic discrimination at all and that the supposed neutrality and objectivity of algorithms may obscure underlying biases and reduce people’s motivation to question or challenge unfair outcomes.

Fourth, our findings suggest that discrimination experiences negatively impact justice perceptions. This highlights the need for organizations to address discrimination proactively through clear antidiscrimination policies, transparency in decision-making processes (including algorithmic systems), and accessible channels for employees to voice concerns. Such measures can mitigate negative perceptions and rebuild trust, which is particularly important as our findings show that individuals draw on past experiences to evaluate justice.

4.8.3 Limitations and future research avenues

Three primary limitations should be considered when interpreting our results. First, we used experimental approaches that may have led to methodological limitations. Specifically, we asked participants to imagine that they were facing the scenarios presented in the studies, potentially undermining external validity. Nevertheless, experimental designs allow the manipulation of individual variables and measure respective reactions to them (Aguinis & Bradley, 2014), which would have been difficult to carry out in a field experiment. Further, as the use of AM in traditional organizations is currently rather uncommon, it may have been difficult to assess field data at all. However, with an increasing use of AM systems, field experiments should be conducted in the future to increase external validity.

Second, in the present paper, we focus solely on prior discrimination experience by human managers or AM systems. Hence, we did not distinguish different types of discrimination. Future research could thus focus on different kinds of discrimination, such as gender discrimination (Schulte Steinberg & Hohenberger, 2023), age discrimination (Snape & Redman, 2003), or discrimination outside the workplace.

Finally, the three experiments did not empirically test the mechanisms underlying the observed moderating effects. However, based on our qualitative data, we propose potential mechanisms that future research could investigate. Specifically, future studies could examine whether individuals attribute discrimination by humans and algorithms differently. This question is particularly relevant, as algorithmic discrimination is often seen as a byproduct of biased training data or flawed system design (Garcia et al., 2024), factors perceived as external to the algorithm itself. In contrast, discrimination by human managers may be viewed as more intentional and less open to reinterpretation. This difference in attribution could influence how individuals evaluate justice and respond to the respective forms of discrimination. Additionally, some respondents reasoned that human managers may discriminate by favoring individuals from their own ethnic or social group, reflecting a form of in-group favoritism. This perception may help explain why discrimination by AM systems was viewed as less severe, as AM systems are not perceived as capable of forming such in-group biases. Future research could explore whether the absence of perceived in-group favoritism in AM contributes to greater tolerance or less negative reactions to algorithmic discrimination.

4.9 Conclusion

Our work provides valuable insights for both research and practice regarding justice perceptions of AM in light of prior discrimination experiences. First, we found that the implementation of AM is associated negatively with informational, interpersonal, and procedural justice among employees. This underscores the importance of carefully considering how such technologies are introduced and managed in the workplace. Organizations need to be aware of the potential issues that may arise and proactively address them through transparent communication and inclusive decision-making processes. Second, our findings reveal that prior human discrimination experiences moderate the relationship between AM use and justice perceptions. This finding points to the nuanced nature of how different groups of employees may perceive AM, depending on their past experiences. Third, as we did not find this moderating effect regarding algorithmic discrimination, the two types of discrimination seem to be perceived differently. Our qualitative results shed light on potential mechanisms by implying that individuals associated human discrimination with personal issues and favoritism, whereas algorithmic discrimination was perceived as discriminatory against whole groups, without targeting single individuals.

5 Understanding worker resistance to algorithmic management: The role of perceived algorithmic discrimination (Essay IV)[11]

Abstract

Algorithmic management (AM) is increasingly adopted in organizations. While AM can enhance efficiency and accuracy, prior research shows that workers often associate AM use with reduced autonomy, diminished well-being, and discriminatory decision-making. In response, some individuals may engage in resistance behaviors (i.e., algoactivism), such as actively manipulating AM systems. Drawing on the justice model of counterproductive work behavior and the model of justice expectations, we propose that individuals who have experienced injustice (i.e., perceived algorithmic discrimination) are more likely to engage in algoactivism as a way to counteract anticipated future injustices. To test our hypotheses, we conducted a four-wave survey (n_1 = 276, n_2 = 200, n_3 = 155, n_4 = 124). The results show that perceived algorithmic discrimination is positively associated with algoactivism, and that this relationship is fully mediated by justice expectations and justice evaluations. These findings underscore the importance of understanding the antecedents of resistance behaviors in AM contexts and have implications for both worker well-being and organizational practices.

Keywords:

Algorithmic Management; Discrimination; Algoactivism; Justice

---

[11] **Author:** Moritz, J. M.
Submitted to *Journal of the Association for Information Systems*
A similar version of this manuscript was presented at the following conference:

- 22nd European Congress of Work and Organizational Psychology, Prague, Czech Republic, 2025

5.1 Introduction

The use of algorithms within organizations is altering how workers are managed. Here, algorithmic management (AM) refers to the transfer of coordination and control functions, traditionally handled by managers, to algorithmic systems (Möhlmann et al., 2021). AM is a central feature of the platform economy, used primarily to control and match workers with clients (Benlian et al., 2022). However, the application of AM is increasingly being extended to traditional organizations (Hirsch et al., 2024), broadening its scope and impact. While AM offers potential advantages for organizations, such as increased efficiency and accuracy (Benlian et al., 2022; Kellogg et al., 2020), it is frequently associated with negative perceptions. These include physical and mental health risks (Deng & Galliers, 2024) and reduced autonomy (Kinowska & Sienkiewicz, 2022). Prior research also shows that AM systems are often perceived as unjust (e.g., Bankins et al., 2022; Deng & Galliers, 2024; Langer & Landers, 2021). One reason for this may be the presence of discriminatory practices or biases embedded in algorithmic systems (Kordzadeh & Ghasemaghaei, 2021), which contribute to the belief that AM systems may perpetuate existing inequalities (Kellogg et al., 2020).

In response to perceived injustice, workers may engage in resistance behaviors such as algoactivism, which refers to tactics aimed at influencing or resisting AM systems (Kellogg et al., 2020). Algoactivism can range from passive forms of resistance, such as voicing complaints about AM on online platforms (de Jong et al., 2025), to more active tactics, such as disabling GPS or phones to manipulate work assignments, for example, in ride-hailing contexts (Jarrahi & Sutherland, 2019). Although worker resistance is not a new phenomenon in the literature (e.g., Cohen & Diamant, 2019; Hodson, 1995; Kelloway et al., 2010), findings on counterproductive work behavior in human–human contexts may not fully apply to AM systems. Specifically, individuals working under AM, especially in the platform economy, may not perceive themselves as being directly monitored by a supervisor. Instead, they may experience a sense of autonomy or see themselves as self-employed (Deng & Galliers, 2024), which could foster the perception that engaging in algoactivism carries minimal risk or oversight. Furthermore, AM systems are often characterized by opacity (Jarrahi et al., 2021; Möhlmann et al., 2023), which may prompt workers to make sense of how algorithms function. This understanding, in turn, enables them to engage in algoactivism, as it requires at least some knowledge of the algorithm's functionalities (Cobonpue et al., 2024).

While prior research has focused on examining AM characteristics and workers' sensemaking of AM systems (e.g., Jarrahi et al., 2021; Jarrahi & Sutherland, 2019; Möhlmann et al., 2023), as well as describing algoactivism (e.g., Cobonpue et al., 2024), the mechanisms underlying algoactivism have yet to be empirically tested. Examining the motivations behind algoactivism is particularly important, as a lack of theoretical understanding risks fragmenting the research landscape and makes it difficult to compare empirical findings across studies. This gap is especially significant given the novel dynamics of worker resistance in AM contexts and the uncertainty over whether traditional theories of resistance, developed in human–human contexts, are applicable to human–algorithm interactions. To address this gap, we examine the relationship between prior experiences of injustice (i.e., feeling discriminated against), justice expectations, justice evaluations, and their impact on algoactivism.

Second, from a practical perspective, when organizations fail to understand the motivations behind resistance behaviors, they struggle to address the root causes. Moreover, resistance behaviors often spread within and across organizations, altering prevailing values, norms, and behaviors (Searle, 2022). In the case of algoactivism, this spread is already observable on several online platforms, where, for instance, ride-hailing drivers express complaints about AM systems and share tips on how to manipulate them (de Jong et al., 2025), potentially encouraging others to adopt similar behaviors. Given the serious consequences of algoactivism for organizations, such as reputational damage when customers are mistreated, or the risks posed by active manipulation of algorithms (Cobonpue et al., 2024), it is critical to understand the conditions under which algoactivism occurs.

Third, examining the mechanisms underlying algoactivism is essential because these practices are associated with negative outcomes for both workers and organizations. At the individual level, worker resistance is linked to decreased well-being and increased burnout symptoms (Searle, 2022), which are themselves associated with lower job performance (Corbeanu et al., 2023). At the organizational level, worker resistance is positively associated with turnover intention and negatively associated with productivity, profit, and customer satisfaction (Carpenter et al., 2021). By understanding the drivers of algoactivism, organizations can develop mitigation strategies to reduce its occurrence and associated harms. In sum, resistance behaviors have detrimental effects at both the individual and organizational levels, making it essential to understand what drives such behaviors in human–algorithm contexts. This need is particularly urgent given the expanding use of AM beyond platform-based companies into traditional organizations. Only by understanding why individuals engage in algoactivism can we develop both theoretical insights and practical recommendations for addressing it. To address these concerns, we ask:

Research question: *How do perceived algorithmic discrimination, justice expectations, and justice evaluations impact algoactivism within organizations?*

To offer a theoretical understanding of algoactivism, we draw on the justice model of counterproductive work behavior (Cohen & Diamant, 2019), which is rooted in equity theory (Adams, 1963). This model provides a framework for understanding how perceptions of workplace mistreatment, such as ostracism or discrimination, can lead to resistance behaviors like sabotage or theft, or in the context of AM, to algoactivism. The justice model of counterproductive work behavior is particularly relevant to our study, as it highlights the role of justice expectations in shaping worker behavior, suggesting that individuals are more likely to resist when they expect or perceive unjust treatment based on prior experiences of injustice (Kim et al., 2023). In our study, we conceptualize perceived algorithmic discrimination as a specific and salient form of injustice. Given that AM systems have previously been associated with discriminatory practices (Kellogg et al., 2020), we thus examine perceived algorithmic discrimination as a key experience that shapes individuals' justice expectations and evaluations, which in turn influence their resistance behaviors. To investigate these relationships, we conducted a four-wave survey among workers who regularly interact with AM systems in their daily work. In doing so, we offer three main contributions.

First, we offer a theoretical understanding of the mechanisms underlying algoactivism. Drawing on the justice model of counterproductive work behavior (Cohen & Diamant, 2019), we theorize and test perceived algorithmic discrimination, justice expectations, and subsequent justice evaluations as factors impacting algoactivism. In doing so, we extend traditional theories of worker resistance, developed in human–human contexts, to the emerging human–algorithm context. Second, we provide empirical insights by conducting a four-wave survey among workers who regularly interact with AM systems. By leveraging a four-wave survey design, we capture how workers' justice expectations and evaluations evolve over time and how these perceptions shape their resistance behaviors. Our study thus moves beyond prior research that primarily relies on cross-sectional or qualitative data, offering a more nuanced understanding of the mechanisms underlying algoactivism. Third, we highlight the practical implications of algoactivism for organizations. Our findings indicate that worker resistance in algorithmic contexts is driven not only by perceived injustices but also by evolving justice perceptions. By understanding the mechanisms underlying algoactivism, organizations can develop targeted mitigation strategies. Specifically, enhancing algorithmic transparency, ensuring fairness in algorithmic decision-making, and involving workers in the design of AM systems may help reduce resistance behaviors and their negative effects.

5.2 Theoretical background and hypotheses development

5.2.1 Algoactivism

Algoactivism refers to tactics aimed at exercising individual or collective resistance against AM systems (Kellogg et al., 2020). Although organizations are familiar with non-cooperation and other forms of resistance, workers may engage in resistance behaviors differently when interacting with AM systems rather than a human manager. In the context of AM, algoactivism can be divided into four categories: (1) individual practical action, (2) platform organization, (3) discursive framing, and (4) legal mobilization (Kellogg et al., 2020). Prior research demonstrates that workers may actively manipulate AM systems by observing the inputs and outputs of the algorithm. Based on this knowledge, individuals may adjust their behavior to improve outcomes (Jarrahi & Sutherland, 2019). In ride-hailing contexts, for instance, workers might avoid algorithmic tracking by logging off to prevent long journeys (Kellogg et al., 2020), influence customers to give them higher ratings (Cobonpue et al., 2024), or decline poorly paid jobs (Deng & Galliers, 2024).

Cobonpue et al. (2024) describe four "triggers" for algoactivism in the context of platform organizations. First, work environment tensions, such as conflicts between job autonomy and supervision, may increase the likelihood of engaging in algoactivism. Second, tensions between transparency and opacity may also prompt resistance, as workers may feel exploited when algorithmic rewards or punishments are opaque or incomprehensible. Third, customer conflicts and economic triggers, such as low piece rates, can drive workers toward resistance behaviors. Fourth, the transfer of risks traditionally borne by employers, for example, health risks during the COVID-19 pandemic, is increasingly shifted onto workers, which may further drive algoactivism. Although these triggers may be specific to the platform economy, the broader notion that certain management practices can provoke worker resistance has also been documented in traditional organizations (Carpenter et al., 2021).

Theoretically, these patterns can be explained by workers' perceptions of such triggers. The justice model of counterproductive work behavior (Cohen & Diamant, 2019) suggests that individuals are more likely to engage in resistance behaviors in response to perceived injustice. These responses are often attempts to restore equity (Adams, 1963; Cohen & Diamant, 2019), a concept supported both meta-analytically (Colquitt et al., 2013) and empirically, as studies show that perceived injustice predicts counterproductive work behavior (Priesemuth et al., 2013). Building on these findings, we propose that algoactivism may be a consequence of perceived injustice in the workplace, particularly in the form of algorithmic discrimination. This aligns with the idea that specific management practices can trigger algoactivism (Cobonpue et al., 2024). We suggest that these triggers may lead workers to feel unjustly treated and, as a result, to engage in algoactivism as a means of restoring equity and guarding against future violations.

5.2.2 Discrimination experiences

Discrimination refers to the unequal treatment of individuals or groups based on prejudice (Allport et al., 1954). This behavior often stems from both explicit and implicit biases that target characteristics such as race, gender, age, religion, sexual orientation, or disability. In the workplace, discrimination can have detrimental consequences for both employees and organizations (Sanchez & Brock, 1996). Perceived discrimination is positively associated with feelings of powerlessness, stress, health problems, and turnover intentions (Dipboye & Colella, 2013). Furthermore, victims of discrimination often seek to understand the reasons behind their experiences (Snape & Redman, 2003), which can further increase emotional and cognitive stress. At the organizational level, perceived discrimination may lead to greater resistance behaviors (Follmer et al., 2023; Sanchez & Brock, 1996) ultimately affecting organizational outcomes.

In the context of AM, workers may experience a form of discrimination that originates from the systems themselves (Kellogg et al., 2020). This type of discrimination can manifest, among other ways, through gender or racial stereotyping embedded in algorithmic processes (Kellogg et al., 2020). Such discrimination may stem from multiple sources, including biased training data, organizational specifications, and sampling methods used during the algorithm's development (Garcia et al., 2024). Importantly, workers often have limited avenues to challenge or correct these perceived injustices, which can further motivate forms of resistance like algoactivism (Kellogg et al., 2020).

To better understand this phenomenon, it is important to distinguish the sources of perceived discrimination. Workers may attribute discrimination to several factors: the system's developers, the organization implementing the system, human biases embedded in training data (Garcia et al., 2024), or the AM system itself. In this study, we define *perceived algorithmic discrimination* as the belief that an algorithmic system denies equal treatment to a person or group. Specifically, it refers to a form of discrimination in which individuals view the AM system itself—rather than human bias, developers, or the organization—as the primary source of discriminatory treatment.

According to the justice model of counterproductive work behavior (Cohen & Diamant, 2019), individuals compare their inputs to the outputs they receive compared to others. Here, perceived imbalance might lead to feeling treated unjustly. This perceived injustice can, in turn, lead to resistance behaviors aimed at restoring equity or retaliating against anticipated future mistreatment (Cohen & Diamant, 2019). In the context of AM, this restoration of justice may manifest as resistance behaviors like algoactivism. By engaging in algoactivism, individuals may attempt to correct perceived imbalances, thereby re-establishing a sense of justice. We thus hypothesize:

Hypothesis 1: *Perceived algorithmic discrimination is positively associated with algoactivism.*

5.2.3 Justice and anticipatory injustice

In organizational settings, justice is defined as "the degree to which one's company or top management is perceived to act consistently, equitably, respectfully, and truthfully in decision contexts" (Colquitt & Rodell, 2015, p. 188) and is commonly categorized into distributive, procedural, informational, and interpersonal justice (Colquitt, 2001). For organizations, justice is of high importance, as it is associated with severe consequences. Specifically, workers are more likely to accept and comply with decisions they perceive as fair (Lind, 2001), and perceptions of justice are associated with important organizational outcomes such as turnover and job performance (Moon, 2017).

The model of justice expectations (Bell et al., 2004) and the concept of anticipatory injustice (Shapiro & Kirkman, 2001) help explain how justice evaluations are formed in organizational contexts. According to Bell et al. (2004), individuals develop justice expectations based on past treatment, which in turn shape how they evaluate justice in future interactions. This aligns with Shapiro and Kirkman's (2001) idea of anticipatory injustice, which suggests that individuals who have experienced unjust treatment in the past are more likely to expect and perceive similar treatment in the future. This process is grounded in confirmatory bias (Snyder & Swann, 1978), which posits that people tend to interpret new information in ways that confirm their existing beliefs. Supporting this view, Sanchez and Brock (1996) empirically showed that individuals who have experienced workplace discrimination are more likely to anticipate and detect discriminatory treatment than those without such prior experiences.

When individuals have perceived discrimination in the past, they may develop expectations of future unjust or discriminatory treatment. These expectations can persist regardless of the actual actions or intentions of supervisors, organizations, or, by extension, algorithms. Drawing on theories originally developed in human–human contexts, we argue that these dynamics also apply to human–algorithm interactions. The computers-are-social-actors (CASA) framework (Nass & Moon, 2000; Nass et al., 1994) suggests that individuals tend to interact with technologies, such as AM systems, as if they were interacting with human agents (Gambino et al., 2020). According to this perspective, people unconsciously apply the same social norms and interactional expectations to technology that they do in human relationships. As a result, individuals may project expectations of injustice onto algorithms, treating them as social actors capable of discriminatory behavior. We therefore propose that individuals form justice expectations based on prior experiences of perceived algorithmic discrimination and use these expectations to evaluate how justly they believe they will be treated. If individuals have felt discriminated against in the past, we suggest they are more likely to expect unjust treatment again. Hence, we propose:

Hypothesis 2: *Perceived algorithmic discrimination is negatively associated with justice expectations.*

Furthermore, we propose that justice expectations impact justice evaluations. Consistent with confirmatory bias (Snyder & Swann, 1978) and the concept of anticipatory injustice (Shapiro & Kirkman, 2001), individuals are more likely to interpret experiences in ways that confirm their existing (justice) expectations. Accordingly, if individuals expect to be treated justly, they may be more likely to interpret outcomes as just. We thus propose:

Hypothesis 3: *Justice expectations are positively associated with justice evaluations.*

According to the justice model of counterproductive work behavior (Cohen & Diamant, 2019), individuals seek to restore balance when they perceive disparities between their inputs (e.g., effort, compliance) and outputs (e.g., rewards, opportunities). In the context of AM, such imbalances can trigger resistance behaviors, including algoactivism. Specifically, individuals may engage in tactics such as withholding data, manipulating algorithmic inputs, or voicing complaints on online platforms to regain a sense of justice (Cobonpue et al., 2024; de Jong et al., 2025). Negative justice evaluations thus function as a psychological mechanism that motivates workers to push back against AM systems. Conversely, when individuals feel they have been treated justly, there is little or no need to counteract future violations or restore equity, making them less likely to engage in algoactivism. We therefore propose:

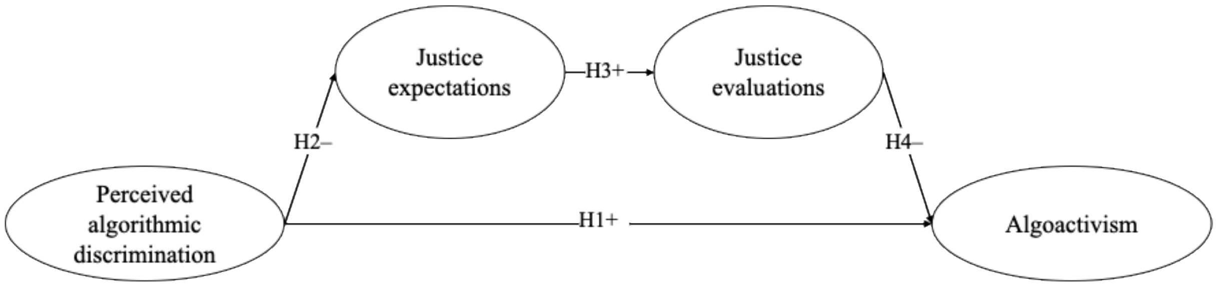
Hypothesis 4: *Justice evaluations are negatively associated with algoactivism.*

Based on the model of justice expectations (Bell et al., 2004) and the concept of anticipatory injustice (Shapiro & Kirkman, 2001), we propose that both justice expectations and justice evaluations mediate the relationship between perceived algorithmic discrimination and algoactivism. Individuals who have experienced injustice in the past, specifically through encounters with algorithmic discrimination, are likely to develop lower justice expectations based on these experiences, which in turn reduce their justice evaluations. Decreased justice expectations and evaluations may then drive resistance behaviors as individuals attempt to counteract potential future violations and restore a sense of justice (Cohen & Diamant, 2019). We thus propose the following hypothesis:

Hypothesis 5: *(a) Justice expectations and (b) justice evaluations mediate the relationship between perceived algorithmic discrimination and algoactivism.*

The research model and hypotheses are depicted in Figure 5-1.

Figure 5-1: Research model



Note. Control variables (not depicted): Age, gender, ethnic background, social desirability.

5.3 Method

To test our research model (Figure 5-1), we collected data across four waves and analyzed it using covariance-based structural equation modeling (SEM) with the lavaan package in R (Rosseel, 2012).

5.3.1 Data collection and sample

We collected data from the U.S. working population using Prolific Academic, an online platform for participant recruitment. To include only participants with experience using AM systems, we followed the approach of Becker (2025) and Alizadeh et al. (2023), using a screening item to assess whether participants were exposed to AM systems; only those who confirmed exposure were allowed to proceed with the survey. Before data collection began, we provided each participant with a definition of AM systems along with examples. This ensured a shared understanding, even when AM systems were referred to differently in participants' respective workplaces. We also informed participants that the survey focused on their personal experiences with AM systems.

To mitigate common method bias (Podsakoff et al., 2003), we collected data at four separate time points, allowing us to assess all variables over time. Participants were contacted for subsequent survey waves at five-day intervals. We used participants' unique

Prolific IDs to contact them and anonymously match their responses across all waves. Perceived algorithmic discrimination, control variables, and demographic information were assessed at t_1; justice expectations at t_2; justice evaluations at t_3; and algoactivism at t_4. We directly excluded all participants that failed any of the attention checks. In total, 276 participants completed the first questionnaire, 200 completed the second, 155 completed the third, and 124 completed the fourth. Following the approach of Edwards et al. (2024), we used full-information maximum likelihood (FIML) to address attrition across data collection points. FIML leverages all available data, enhancing statistical power, reducing bias in parameter estimates, and lowering Type I error rates compared to listwise deletion (Enders & Bandalos, 2001). Accordingly, our final analysis included all 276 participants (M_{age} = 37.28, SD_{age} = 11.37; female = 52.54%, other = 0.01%).

To assess potential effects related to attrition, we conducted *t*-tests to compare demographic variables between participants who completed all four waves and those who only participated in the first wave. We found no significant differences in gender, education, income, or ethnic background. However, we did find a significant difference in age, ($F(1, 274) = 5.41$, $p = .02$), with participants who only completed the first wave ($n = 77$; $M = 34.66$, $SD = 11.27$) being younger than those who completed at least two waves ($n = 199$; $M = 38.30$, $SD = 11.78$). Accordingly, we included age as a control variable in all subsequent analyses.

5.3.2 Measures

We used multi-item scales to assess our constructs. To measure perceived algorithmic discrimination, we adapted the perceived personal discrimination scale developed by Schulte Steinberg and Hohenberger (2023), modifying the items to specifically address discrimination by AM systems. A sample item is: "Prejudice against specific groups has affected me personally due to the actions or decisions of an automated management system." To measure justice expectations and justice evaluations, we used a scale developed by Wang et al. (2023), adapting the items to consistently refer to "automated management system." A sample item for justice expectations is: "I expect that decisions made by the automated management system are applied consistently." The same items were used to assess justice evaluations, with the phrase "I expect that" removed. We operationalized algoactivism using the active resistance subdimension of the algorithmic resistance scale developed by de Jong et al. (2024). A sample item is: "I deliberately change my behavior to make sure the automated management system does not punish me."

We included age, gender, and ethnic background as control variables to examine the impact of perceived algorithmic discrimination beyond these personal characteristics. Following Koch-Bayram et al. (2023), we operationalized ethnic background as a dummy variable, using participants who identified as "White" as the reference group in comparison to other ethnic groups (e.g., Hispanic, Black, Asian). We also controlled for social desirability, as individuals scoring high on this trait may be less likely to report engagement in algoactivism. To assess this, we used the short version of the Marlowe-Crowne Social Desirability Scale (Strahan & Gerbasi, 1972). All of our scales exceeded the Cronbach's alpha threshold of .70 (Cortina, 1993; perceived algorithmic discrimination: $\alpha = .95$; justice expectations: $\alpha = .90$; justice evaluations: $\alpha = .87$; algoactivism: $\alpha = .95$; social desirability: $\alpha = .72$).

To confirm validity criteria, we conducted a confirmatory factor analysis. The results indicated a satisfactory fit ($\chi^2 = 288.51$, $df = 224$, $\chi^2/df = 1.29$, $p < .001$, CFI = .97, RMSEA = .03, SRMR = .05; Hair et al., 2014). The average variance extracted (AVE) for all constructs exceeded the recommended threshold of .50 (perceived algorithmic discrimination = .84, justice expectations = .67, justice evaluations = .58, algoactivism = .69; Fornell & Larcker, 1981). Composite reliabilities are above .50

(perceived algorithmic discrimination = .95; justice expectations = .91; justice evaluations = .87, algoactivism = .95; Bagozzi & Yi, 1988). Discriminant validity is given for all constructs as the average variance extracted exceeds squared correlations with other constructs (Fornell & Larcker, 1981). Table 5-1 presents the means, standard deviations, and correlations among all study variables.

Table 5-1: Means, standard deviations, and correlations

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. Perceived Algorithmic Discrimination	2.22	1.59							
2. Justice Expectations	6.07	1.04	$-.27^{**}$						
3. Justice Evaluations	5.38	1.12	$-.26^{**}$	$.50^{**}$					
4. Algoactivism	2.68	1.51	$.23^{**}$	−.17	$-.28^{**}$				
5. Age	37.28	11.73	.00	.05	$-.18^{*}$	−.06			
6. Gender	0.53	0.50	−.10	$.15^{*}$	−.07	$-.18^{*}$	.09		
7. Ethnic Background	0.50	0.50	−.09	.03	−.04	−.03	.11	.04	
8. Social Desirability	4.40	0.96	$-.21^{**}$	.13	$.20^{*}$	$-.22^{*}$	.03	−.03	$-.15^{*}$

Note. M and *SD* represent mean and standard deviation, respectively. Reference group, Gender: male (= 0). Reference group, ethnic background: individuals who identify themselves as "white" (= 0). * Indicates $p < .05$. ** indicates $p < .01$.

5.4 Results

5.4.1 SEM results

In order to test hypothesis 1, we conducted a structural equation model (SEM) examining the relationship between perceived algorithmic discrimination and algoactivism, including all control variables. The model showed a good fit ($\chi^2 = 136.79$, $df = 112$, $\chi^2/df = 1.22$, $p < .001$, CFI = .98, RMSEA = .03, SRMR = .06; Hair et al., 2014). The results demonstrated a significant positive relationship between perceived algorithmic discrimination and algoactivism (β = .23, $p = .036$), supporting hypothesis 1. None of the control variables were significantly associated with algoactivism.

We then calculated a second SEM, including justice expectations and justice evaluations as mediators to test the full serial mediation depicted in Figure 5-2. This model also demonstrated satisfactory fit ($\chi^2 = 422.12$, $df = 306$, $\chi^2/df = 1.38$, $p < .001$, CFI = .95, RMSEA = .04, SRMR = .07; Hair et al., 2014). We found a significant negative effect of perceived algorithmic discrimination on justice expectations (β = −.28, $p = .002$), supporting hypothesis 2. Additionally, justice expectations were positively associated with justice evaluations (β = .56, $p < .001$), supporting Hypothesis 3. Justice evaluations, in turn, were negatively associated with algoactivism (β = −.27, $p = .014$), supporting hypothesis 4.

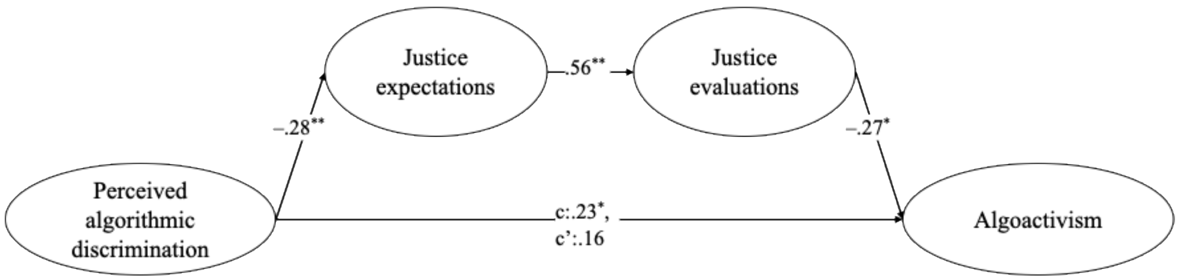
Regarding indirect effects, perceived algorithmic discrimination negatively affected justice evaluations via justice expectations (β = −.16, $p = .007$). Justice expectations also negatively impacted algoactivism via justice evaluations (β = −.15, $p = .014$). The serial mediation from perceived algorithmic discrimination to algoactivism through justice expectations and justice evaluations was also significant (β = .04, $p = .048$), supporting hypotheses 5a and 5b. After including the mediators, the direct relationship between perceived algorithmic discrimination and algoactivism was insignificant (β = .16, $p = .150$), indicating full mediation. In summary, all hypotheses were supported. Regarding control variables, we found that gender significantly predicted justice expectations (β = .14, $p = .049$) and justice evaluations (β = −.17, $p = .019$). Age was negatively associated with justice evaluations (β = −.21, $p = .008$). Ethnic background and social desirability did not show any significant effects. The results are depicted in Table 5-2 and Figure 5-2.

Table 5-2: SEM results

Path				
Hypothesized effects	*B*	*SE*	β	*p*
Algorithmic discrimination → Algoactivism	.17	(.12)	.16	.150
Algorithmic discrimination → Justice expectations	–.30	(.10)**	–.28	.002
Justice expectations → Justice evaluations	.67	(.12)**	.56	<.001
Justice evaluations → Algoactivism	–.23	(.10)*	–.27	.014
Algorithmic discrimination → Justice expectations → justice evaluations	–.20	(.07)**	–.16	.007
Justice expectations → Justice evaluations → Algoactivism	–.16	(.06)*	–.15	.014
Algorithmic discrimination → Justice expectations → Justice evaluations → Algoactivism	.05	(.02)*	.04	.048
Control effects				
Age → Justice expectations	.00	(.01)	.04	.578
Age → Justice evaluations	–.02	(.01)**	–.21	.008
Age → Algoactivism	–.01	(.01)	–.09	.362
Gender → Justice expectations	.30	(.15)*	.14	.049
Gender → Justice evaluations	–.44	(.19)*	–.17	.019
Gender → Algoactivism	–.37	(.19)	–.17	.051
Ethnic Background → Justice expectations	.04	(.15)	.02	.796
Ethnic Background → Justice evaluations	.01	(.20)	.01	.953
Ethnic Background → Algoactivism	–.03	(.20)	–.02	.875
Social Desirability → Justice expectations	.11	(.08)	.10	.181
Social Desirability → Justice evaluations	.20	(.11)	.15	.072
Social Desirability → Algoactivism	–.17	(.10)	–.15	.090

Note. *B* = unstandardized effect, *SE* = standard error, β = standardized effect. Reference group, Gender: male (= 0). Reference group, ethnic background: individuals who identify themselves as "white" (= 0).* Indicates $p < .05$. ** indicates $p < .01$.

Figure 5-2: SEM results



Note. Control variables (not depicted): Age, gender, ethnic background, social desirability. ** indicates $p < .01$, * indicates $p < .05$.

5.5 Discussion

5.5.1 Main findings and theoretical contributions

Our research advances the understanding of the antecedents of algoactivism. The findings demonstrate that individuals who have experienced algorithmic discrimination are more likely to engage in algoactivism, and that this relationship is fully mediated by the proposed mediators justice expectations and justice evaluations. These findings align with the model of justice expectations (Bell et al., 2004) and the concept of anticipatory injustice (Shapiro & Kirkman, 2001), which suggest that individuals who have encountered unjust treatment in the past—such as discrimination—use these experiences to infer how justly they will be treated in the future. In line with Shapiro and Kirkman's (2001) notion of anticipatory injustice, individuals who expect to be treated unjustly again are more likely to perceive and detect future violations. This is evidenced in the present study by the significant negative relationship between prior experiences of algorithmic discrimination and expectations of justice. These findings challenge previous research, which often overlooks the role of individuals' past experiences and expectations in shaping justice evaluations of AM systems or related behavioral responses such as algoactivism. By demonstrating that previous experiences and justice-related expectations shape justice evaluations, our research contributes to the literature by emphasizing the importance of temporally embedded perceptions in understanding how individuals respond to AM.

Interestingly, the model of justice expectations (Bell et al., 2004) and the concept of anticipatory injustice (Shapiro & Kirkman, 2001), originally developed in the context of human–human interactions, also appear to apply to interactions with AM systems. This finding is consistent with the CASA framework, which suggests that humans apply the same interactional rules used in interpersonal relationships to their interactions with computers and algorithms (Gambino et al., 2020; Nass & Moon, 2000; Reeves & Nass, 1996). As such, the CASA framework provides a theoretical rationale for why expectations and perceptions shaped in human–human contexts may similarly influence engagement with AM systems. Accordingly, individuals may not only anticipate discriminatory behavior from human supervisors but may also project similar expectations onto algorithmic systems, perceiving them as capable of unjust or biased decision-making.

Regarding algoactivism, we found both a direct positive relationship between perceived algorithmic discrimination and algoactivism, as well as an indirect relationship mediated by justice expectations and justice evaluations. The direct relationship can be explained by the justice model of counterproductive work behavior (Cohen & Diamant, 2019), which suggests that individuals who feel treated unjustly, such as through discrimination, engage in behaviors aimed at restoring justice or taking revenge. In terms of the indirect effects, experiences of discrimination may reduce individuals' expectations of being treated justly in the future, which in turn influences their justice evaluations. This is consistent with the concept of anticipatory injustice (Shapiro & Kirkman, 2001), whereby individuals who expect future mistreatment are more likely to interpret subsequent interactions as unjust. As a result, they may engage in algoactivism to preemptively counteract expected future violations.

Taken together, our findings demonstrate that individuals who perceive algorithmic discrimination are more likely to engage in algoactivism as a means of restoring justice. This pattern aligns with the justice model of counterproductive work behavior (Cohen & Diamant, 2019), the model of justice expectations (Bell et al., 2004), and the concept of anticipatory injustice (Shapiro & Kirkman, 2001). Notably, although these theories were originally developed in human–human contexts, our results support their applicability in human–algorithm interactions.

5.5.2 Practical implications

Our results reveal several important practical implications for organizations. First, when discrimination is apparent, it is essential to identify its source. Workers may experience discrimination because the AM systems in use are in fact discriminatory, for example, due to historical data that reflects biased patterns (Kellogg et al., 2020). In such cases, organizations must take proactive steps to identify and address perceived algorithmic discrimination. This can be achieved through regular audits of AM systems to ensure fairness and eliminate biases, particularly those involving sensitive data.

Second, workers may perceive discrimination because they attribute algorithmic decisions to biased treatment, even when the systems themselves are not explicitly discriminatory. This may occur when individuals have experienced discrimination in the past and, consistent with anticipatory injustice (Shapiro & Kirkman, 2001), expect to be treated unjustly again. Such attribution is more likely when algorithmic decisions are ambiguous or difficult to interpret. This aligns with the triggers of algoactivism proposed by Cobonpue et al. (2024), who argue that tensions between transparency and opacity in AM systems can drive resistance. When decisions lack transparency, individuals may be more likely to infer discriminatory intent. Increasing the transparency of algorithmic evaluation processes can therefore help mitigate these misattributions. Transparent systems can reduce the likelihood that employees feel unjustly treated, which in turn may reduce the engagement in resistance behaviors such as algoactivism.

Third, given that perceived injustice leads to resistance, organizations should implement strategies to restore justice. This includes creating formal channels for employees to provide input and feedback on algorithmic decision-making processes and ensuring that concerns about discriminatory practices are addressed promptly and transparently. In practice, workers have already begun forming platforms to share information about AM systems, their functions, and resistance behaviors (de Jong et al., 2024). Organizations and system developers can benefit by engaging with these discussions and incorporating employee feedback into system design. Doing so may help reduce algoactivism, benefiting both organizational performance and employee well-being.

5.5.3 Limitations and future research avenues

Four primary limitations should be considered when interpreting our results. First, one limitation of our study lies in the potential impact of social desirability bias on self-reported data. As shown in Table 5-1, several study constructs were significantly correlated with social desirability. Specifically, we found a negative association between social desirability and prior algorithmic discrimination experience, suggesting that individuals high in social desirability may be less inclined to report experiences of discrimination. This may be because disclosing such experiences could reflect negatively on their employer or the AM system in place. Similarly, social desirability was negatively correlated with algoactivism and positively correlated with justice evaluations, implying that those higher in social desirability may report higher justice perceptions and lower engagement in resistance behaviors. These patterns indicate that social desirability may distort the accuracy of self-reported data on sensitive topics such as justice, perceived algorithmic discrimination, and resistance behaviors. To address this limitation, future research should consider alternative data collection methods that mitigate the effects of social desirability bias. For example, behavioral or observational data on algoactivism may yield more objective insights into resistance practices.

Second, we focused on perceived algorithmic discrimination in general and did not distinguish between specific types, such as gender-based discrimination (Schulte Steinberg & Hohenberger, 2023) or age-based discrimination (Snape & Redman, 2003). Future research should explore how different forms of discrimination, as well as demographic variables, impact justice expectations and evaluations, and consequently, resistance behaviors like algoactivism. Similarly, we assessed justice expectations and evaluations as single, global constructs and did not differentiate between the four established subdimensions of organizational justice (i.e., interpersonal, informational, procedural, and distributive justice; Colquitt, 2001). Future research could examine whether specific justice subdimensions act as mediators in the relationship between perceived algorithmic discrimination and resistance. With knowledge on specific subdimensions, it might be easier for organizations to target injustice specifically, for example by increasing transparency when negative evaluations of informational justice are apparent. Additionally, while our analysis focused on individual-level algoactivism, collective resistance practices, such as coordinated strikes or organized algorithmic manipulation via worker forums, are well-documented (Cobonpue et al., 2024; Kellogg et al., 2020). Given that algorithmic discrimination often disproportionately affects marginalized groups (e.g., through biased rating systems), future research should examine how collective sensemaking and group-based resistance interact with systemic algorithmic biases.

Third, although we recruited participants from the U.S. working population who reported experience with AM systems, we relied on self-reports collected via an online survey. While we did include several screening and open-ended questions to confirm AM exposure (e.g., job type, management structure), we cannot fully guarantee the comparability of participants' experiences with AM systems. Future studies could strengthen internal validity by conducting field research to verify AM exposure and ensure consistent experience among participants. Nonetheless, the diversity of our sample and the consistent support for all hypotheses may suggest a degree of generalizability compared to using a sample from workers of a single company.

Fourth, the study did not explore whether any variables moderated the proposed relationships. As a result, we were unable to assess variables that may strengthen or weaken the associations between perceived algorithmic discrimination, justice mechanisms, and algoactivism. However, research from human–human contexts has identified several relevant moderators of resistance behaviors that may be applicable in human–algorithm contexts as well. For instance, personality traits may play a role; prior research has shown that conscientiousness is negatively associated with resistance behaviors (Jensen & Patel, 2011). This presents an avenue for future studies to explore. Notably, our results showed a negative relationship between female gender and reported algoactivism. Since prior research has found that women tend to score higher in conscientiousness than men (e.g., Keiser et al., 2016; Vianello et al., 2013), it is possible that personality traits such as conscientiousness may already be reflected in our findings. Further research is needed to explore such moderating effects directly.

5.6 Conclusion

Our findings highlight the detrimental effects of perceived discrimination in the workplace, particularly its potential to trigger resistance behaviors such as algoactivism. Specifically, individuals who have experienced algorithmic discrimination are more likely to engage in algoactivism, a relationship driven by diminished justice expectations and justice evaluations. These results demonstrate that justice expectations and norms, originally developed in the context of human–human interactions, are also applicable in human–algorithm contexts. This study contributes to the growing literature on algorithmic management by identifying the psychological mechanisms that underlie the relationship between perceived algorithmic discrimination and resistance behavior. In doing so, it extends established organizational justice theories into the domain of human–algorithm interaction. From a practical perspective, our findings underscore the importance of increasing algorithmic transparency, ensuring fairness in algorithmic decision-making processes, and establishing effective communication channels for employees to voice concerns. Addressing these areas may help organizations reduce perceived injustice and mitigate resistance behaviors such as algoactivism. While this study provides a robust foundation, future research should address its limitations, particularly by examining different dimensions of justice, to deepen our understanding of the drivers and dynamics of algoactivism.

6 Conclusion of the dissertation

This dissertation presented four essays aiming for a deeper understanding of the perceptions of ADM and their boundary conditions, including (1) interaction characteristics, (2) interventions to address algorithm aversion, and (3) prior discrimination experiences, as well as (4) subsequent responses, including algoactivism. Based on the SOR model (Mehrabian & Russell, 1974), the presented essays explored the perceptions of ADM (i.e., justice, fairness, trust, trustworthiness, affective reactions) and the subsequent responses (i.e., algoactivism, pursuit intention, organizational attractiveness). Through a meta-analysis (Essay I), four vignette-based experiments (Essays II and III), and a four-wave survey (Essay IV), this dissertation contributes to a nuanced understanding of these relationships.

6.1 Summary of the research findings

Essay I presented a meta-analysis on reactions to ADM use in the HRM domain and proposed an overarching theoretical model to explain these reactions and their boundary conditions. Comprising 53 studies with 73 samples (n = 24,578), the results indicated that ADM use was negatively associated with system-related reactions (i.e., justice, fairness, trust, trustworthiness) and organization-related reactions (i.e., organizational attractiveness, pursuit intention), demonstrating that algorithm aversion is more prevalent in the HRM context. Moreover, findings on boundary conditions suggested that negative reactions were more likely to occur when decisions are automated, than when decisions are augmented. The results indicated that individuals also reacted more negatively when they were confronted with ADM decisions than when they collaborated with ADM systems. Based on the CASA framework (Nass & Moon, 2000; Reeves & Nass, 1996), Essay I reasons that theories originally used in human–human contexts can also be applied in human–algorithm contexts, as humans interact with algorithms as if they engage with other humans. From the meta-analytical synthesis and the overarching theoretical model, Essay I concludes several implications for future research, such as examining the impact of cultural differences and interventions that focus on successful human–algorithm collaboration.

Based on the future research avenues identified in Essay I, Essay II used a vignette-based experiment (n = 483) to examine the effectiveness of trust-building interventions (i.e., transparency and a human-in-the-loop design) depending on outcome favorability. The results revealed that, when applicants were rejected, ADM use was negatively associated with trust. This relationship was positively impacted by transparency, whereas this effect was not evident in the participant group that was accepted. Conversely, when applicants were accepted, they did not perceive a significant difference between a human manager and ADM, and even experienced algorithm appreciation when a human-in-the-loop design was used. This effect was not evident in the group that was rejected. Consistent with attribution theory (Weiner, 1985) and self-serving bias (Zuckerman, 1979), these results indicate that individuals are more likely to question the application process, and thus the decision entity, when they receive a rejection, as failure is more often attributed to external causes.

Essay III investigated prior discrimination experience as a boundary condition impacting justice perceptions of ADM. Two vignette experiments (n_1 = 82, n_2 = 83) demonstrated that prior human discrimination lowered algorithm aversion, whereas prior algorithmic discrimination did not impact justice perceptions. The third follow-up study (n_3 = 216) compared both types of discrimination and demonstrated that human discrimination had a greater impact on justice perceptions than algorithmic discrimination. Open-ended qualitative responses indicated that individuals interpret discriminatory behavior differently depending on whether it stems from humans or algorithms. Discrimination by humans was often perceived as driven by personal biases directed at the individual, whereas discrimination by algorithmic systems was generally seen as impersonal and as targeting broader groups, which might explain the differences in reactions.

Essay IV used a four-wave survey ($n_{t1} = 276$, $n_{t2} = 200$, $n_{t3} = 155$, $n_{t4} = 124$) to examine the consequences of algorithm aversion. The results demonstrated that prior algorithmic discrimination was negatively associated with justice expectations and justice evaluations, which aligns with the model of justice expectations by Bell et al. (2004) and the justice model of counterproductive work behavior (Cohen & Diamant, 2019). Individuals with prior algorithmic discrimination experience were more likely to expect unjust treatment again, leading to lower justice evaluations, which aligns with confirmation bias (Snyder & Swann, 1978). These reactions, in turn, increased individuals' engagement in algoactivism, indicating that individuals might use algoactivism to counteract (perceived) injustice.

6.2 Theoretical implications

By synthesizing initial findings and extending research on perceptions of ADM, along with their boundary conditions and subsequent responses, this dissertation offers three main theoretical implications. Based on the findings of the four essays, these implications include the role of boundary conditions in ADM reactions, their conceptualization, and the impact of and differentiation between human and algorithmic discrimination as a key boundary condition.

6.2.1 Understanding stakeholder reactions through boundary conditions

As evident throughout all four essays, algorithm aversion was more prevalent than algorithm appreciation. However, Essays I–III showed that reactions to ADM depend on boundary conditions. Essay I demonstrated that algorithm aversion was strengthened when algorithms were used to automate rather than augment decisions, and when stakeholders were confronted with ADM decisions instead of collaborating with ADM systems. Additionally, Essay I provided initial evidence that reactions to ADM differ across stakeholder groups with negative findings for employees and applicants compared to non-significant findings for human resource managers.

This might be explained by self-determination theory (Deci & Ryan, 1985), which proposes autonomy as a basic need. In cases when algorithms take over decision-making, individuals might experience a loss of autonomy, which can lead to algorithm aversion (Burton et al., 2020). This perspective could also help explain differing reactions among stakeholder groups. For instance, human resource managers are more likely to use ADM systems rather than to be subject to their decisions (Langer & Landers, 2021), allowing them to benefit from the efficiency gains of ADM without experiencing a loss of autonomy. However, as this dissertation focuses on decision-making in the organizational context, the tasks involved might also be associated with human skills such as social or emotional skills. According to Lee (2018), ADM use for the aforementioned tasks is more likely to be perceived as unjust or untrustworthy. Consequently, besides self-determination theory, these findings could also be attributed to the tasks involved and whether they are perceived as suitable to be conducted by algorithms instead of humans.

This dissertation extends prior research by exploring interventions to target algorithm aversion in the context of outcome favorability. Essay II demonstrated that participants who were rejected reported higher levels of trust when the decision-making process was highly transparent compared to when there was no transparency. This was not apparent when participants were accepted. Furthermore, the human-in-the-loop design increased trust in the accepted group more than in the rejected group, also revealing that the effectiveness of the intervention depends on outcome favorability. This advances prior research that tested interventions independently from the outcome of decisions made (e.g., Schlicker et al., 2021; Schoeffer et al., 2022) and stresses the importance of controlling for outcome favorability.

6.2.2 Conceptualization of stakeholder reactions to ADM

Drawing on the SOR model (Mehrabian & Russell, 1974), ADM use and associated characteristics represent external stimuli that impact perceptions, which result in (behavioral) responses with severe consequences for organizations. For instance, Essay II showed that trust perceptions relate positively to organizational attractiveness. Essay IV further revealed that negative justice perceptions of ADM can motivate employees to engage in forms of active resistance (i.e., algoactivism). As such, algoactivism represents a behavioral response to internal perceptions of justice and thus reflects a growing manifestation of algorithm aversion, where stakeholders do not only perceive aversion but actively resist ADM systems. On a broader level, this dissertation demonstrates the importance of examining not only the perceptions of ADM use, but also the (behavioral) responses in order to understand organizational consequences. Here, the SOR model can be applied across contexts and guides the understanding of initial perceptions of stimuli and subsequent responses.

Furthermore, the dissertation addresses a central issue in the research field of ADM, namely, that theories used to explain human–algorithm interactions are often adapted from those developed to explain human–human interactions (e.g., organizational justice theory; Colquitt, 2001). To bridge this gap, Essay I introduced an overarching theoretical model grounded in the CASA framework (Nass et al., 1994; Reeves & Nass, 1996). According to CASA, individuals interact with technology and, by extension, with algorithms (Gambino et al., 2020), as if they were engaging with other humans. Thus, interacting with ADM, humans apply stereotypes, follow social norms, and interactional rules, and attribute human-like characteristics to ADM systems. Consequently, as human–algorithm interactions mimic human–human interactions, theoretical explanations can also be transferred. This line of reasoning contributes to the theoretical understanding of how individuals react to ADM and enables scholars to examine human–algorithm interaction with a stronger theoretical foundation.

Drawing on this reasoning, this dissertation extends the justice model of counterproductive work behavior (Cohen & Diamant, 2019) from human–human contexts to human–algorithm contexts. Here, Essay IV demonstrated a negative relationship between justice evaluations and engagement in algoactivism. This suggests that individuals might use algoactivism to counteract perceived or expected injustice, even if injustice might originate from an ADM system. This extends the applicability of the justice model of counterproductive work behavior and deepens previous research by underlining the risks that negative justice perceptions pose to organizations.

On a conceptual level, this dissertation further contributes to the understanding of ADM perceptions by differentiating between fairness and justice, two constructs that are frequently used interchangeably in organizational research (Beugré, 2009). From a theoretical perspective, justice refers to rule-based evaluations that build a basis for fairness perceptions, while fairness itself refers to an individual's subjective perception, outlining two separate constructs (Goldman & Cropanzano, 2015). Empirically, results from Essays I and II demonstrated that ADM use is negatively associated with perceptions of both justice and fairness. The meta-analysis conducted in Essay I, however, revealed that correlation coefficients differed, and the negative relationship was more pronounced regarding justice evaluations. Interestingly, the findings revealed that ADM use had the strongest negative relationship with interactional justice, indicating that ADM use might particularly trigger evaluations of informational and interpersonal justice, emphasizing the importance of measuring these constructs separately.

6.2.3 Differences in reactions to human and algorithmic discrimination

This dissertation, specifically Essays III and IV, extends prior research by offering a more nuanced perspective on the role of discrimination experience in justice evaluations of ADM use in the management context. Specifically, whereas prior research in this area used personal characteristics, such as gender (Pethig & Kroenung, 2022) or belonging to a minority group (Lima et al., 2025), as a proxy for having experienced discrimination, Essay III directly assessed experiences of both human and algorithmic discrimination. The findings demonstrated that prior human discrimination experience led to a decrease in algorithm aversion, which is aligned with prior findings in the context of personnel selection (Koch-Bayram et al., 2023; Schulte Steinberg & Hohenberger, 2023).

Interestingly, Essay III demonstrated that human discrimination evokes more negative justice perceptions than algorithmic discrimination. Qualitative responses suggested that this difference may stem from attribution processes, whereby participants associate algorithmic discrimination with indirect sources, such as biased data, developers working on the systems, or organizational practices, but link human discrimination directly to the individual decision-maker. Additionally, these findings may also be explained by the decontextualization associated with algorithms. Although prior research shows that decontextualization is a driver of negative justice perceptions regarding ADM (Newman et al., 2020), the perception that algorithms make decisions without being able to integrate contextual factors, such as personal characteristics, may lead to the belief that they are not able to discriminate intentionally (Bonezzi & Ostinelli, 2021). Thus, decontextualization may act as a double-edged sword, lowering overall justice evaluations but also leading individuals to believe that algorithms lack discriminatory intent.

These findings highlight that perceptions of injustice are shaped not only by the outcome but also by the (perceived) decision source. By showing that prior discrimination experience impacts reactions to ADM, particularly in relation to perceived justice, Essay III identifies prior experiences as an important boundary condition. This challenges the prevailing focus in ADM research, which emphasizes situational factors (e.g., decision transparency or stakes of a decision; Langer, König, & Papathanasiou, 2019; Schoeffer et al., 2022), while largely neglecting individuals' prior experiences. Consequently, when examining justice evaluations of ADM, relevant prior experiences should be included or controlled for.

6.3 Practical implications

In light of the increasing ADM use in organizations, this dissertation offers important practical implications. While ADM is associated with several potential benefits for organizations, all four essays demonstrated that algorithm aversion is currently more prevalent than algorithm appreciation. Interestingly, robustness checks in Essay I found that algorithm aversion is consistent across Germany and the U.S., with non-significant findings in India. This highlights the importance of understanding cultural differences when implementing ADM across organizations operating globally. In Germany and the U.S., organizations might need to be more cautious and might use targeted interventions, for example, increased transparency, to mitigate algorithm aversion.

Furthermore, algorithm aversion is demonstrated when algorithms automate rather than augment decisions, and when affected stakeholders are not able to collaborate with the system but instead are confronted with decisions. Given that the expected benefits of ADM can only be fully leveraged when individuals perceive decisions made as fair and accept those decisions (Bankins et al., 2022; Lind, 2001), organizations need to carefully consider when and how they make use of ADM systems. When using ADM for tasks that are known to evoke algorithm aversion, they should implement respective interventions. Here, Essay II demonstrated that both transparency and a human-in-the-loop design impact perceptions of

ADM use. However, the findings suggest that the effectiveness of trust-building interventions depends on outcome favorability. Consequently, trust-building interventions, such as transparency, might not work equally well across all contexts. Organizations should carefully evaluate whether interventions are effective, especially when used for tasks associated with more or less favorable outcomes (e.g., using ADM for resumé screening, where applicants are either rejected or accepted).

Moreover, practitioners should be aware that incorporating a human-in-the-loop design does not offer a universal remedy for the challenges associated with ADM in organizational settings. Research increasingly highlights several difficulties when integrating human decision-making with ADM. For example, studies show that people often exploit unethical algorithmic decisions (Krügel et al., 2023) and may struggle to recognize algorithmic bias (Bonezzi & Ostinelli, 2021; Green, 2022). For instance, Langer et al. (2025) argue that effective intervention requires humans to first develop the ability to detect errors in algorithmic outputs. Consequently, organizations face the task of teaching employees the necessary knowledge and competencies to interact with ADM systems in an informed and productive manner. Furthermore, organizations could foster calibrated trust, meaning that stakeholders are able to tie their level of trust to the reliability of the respective system (Muir, 1994) and thus remain cautious when using ADM.

Additionally, results of Essay III demonstrated that prior human discrimination reduces negative justice perceptions of ADM. This indicates that algorithms could play a role in reducing fears of facing discrimination again. By offering a more consistent and seemingly objective decision-making process, ADM has the potential to foster a work environment that is perceived as fairer and more impartial (Choung et al., 2023). This may be especially relevant in contexts where human bias has previously contributed to unequal treatment. In line with signaling theory (Spence, 1973), organizations using ADM can signal a commitment to fairness and objectivity in decision-making. This may shape stakeholders' reactions to organizational values, which might be particularly helpful in contexts where discrimination has historically been prevalent. At the same time, this carries the risk that people may not perceive algorithmic discrimination at all and that the supposed neutrality and objectivity of algorithms may obscure underlying biases, reducing the motivation to question or challenge unfair outcomes (Bonezzi & Ostinelli, 2021). This underlines the need for careful design and ongoing monitoring of ADM systems to ensure they operate fairly and transparently, which ultimately represents the basis for the effective use of ADM in organizations.

Furthermore, Essay III showed that while discrimination experiences generally reduce justice perceptions, individuals evaluate human discrimination more negatively than discrimination by algorithms. Interestingly, in qualitative responses, even participants who experienced discrimination from algorithms described these systems as more consistent and objective than humans. However, these findings need to be treated with caution. While ADM may help reduce the subjectivity of human decision-making, it is not free from bias, especially when bias is unintentionally built into the algorithm itself (Bonezzi & Ostinelli, 2021). Again, this demonstrates that organizations employing ADM systems need to carefully and critically monitor their use.

Essay IV showed that algorithmic discrimination is negatively associated with expectations and evaluations of justice, and positively associated with algoactivism. While counterproductive work behavior is well-established in management research (Cohen & Diamant, 2019), ADM introduces new forms and triggers of resistance (Cobonpue et al., 2024), making it important for organizations to understand when individuals engage in algoactivism. Certain characteristics of ADM, such as its opacity (Jarrahi et al., 2021), might contribute to increased engagement in algoactivism (Cobonpue et al., 2024). To mitigate such resistance, organizations should consider the information employees need to feel informed and fairly treated. This is relevant as the results of Essays I and III illustrate that ADM evoked negative informational and interpersonal justice perceptions, thereby highlighting the importance of these two justice dimensions in this particular context.

Additionally, as Essay IV revealed that prior algorithmic discrimination impacts expectations and evaluations of justice, organizations should implement strategies aimed at restoring perceived justice. This could include creating opportunities for employees to give input and feedback on ADM processes, and ensuring that concerns, especially those related to discrimination, are addressed promptly and transparently. In fact, several platforms, organized by workers, already exist in which they exchange views on ADM systems, their functions, and resistance behaviors (de Jong et al., 2024). Feeding these discussions and concerns back into the organization and to system developers could offer valuable insights into emerging forms of algoactivism and how to respond to them. Ultimately, this might help create systems that are both more effective and more acceptable for workers.

While this dissertation has primarily focused on worker and employee reactions to the use of ADM, results from the meta-analysis provide initial evidence that reactions to ADM vary across stakeholder groups. Specifically, none of the negative relationships between ADM use and system-related or organization-related reactions were statistically significant for managers using ADM systems. A possible explanation is that managers typically collaborate with ADM systems rather than being directly affected by algorithmic decisions. Consequently, although this stakeholder group may not require interventions aimed at reducing aversion, organizations should still ensure that managers are trained to use ADM systems effectively. In this context, organizations should consider ways to train their workforce to enhance AI literacy (Carolus et al., 2023) and to foster awareness of ethical risks associated with ADM. Helping managers develop a sense of competence when using

ADM systems and signaling to employees that their managers use ADM in an informed and responsible way could also reduce algorithm aversion. With the increasing use of ADM systems in more traditional organizations, upskilling managers to effectively utilize ADM systems will be a crucial resource in the future.

6.4 Limitations

Four primary limitations should be considered when interpreting the results of this dissertation. First, all the data collected relied on a single source (i.e., the participants themselves). This could lead to common method bias (Podsakoff et al., 2003). However, all studies addressed common method bias by employing Harman's one-factor test and by using theoretically uncorrelated marker variables (Lindell & Whitney, 2001). Additionally, Essay IV employed a four-wave study design to mitigate common method bias (Podsakoff et al., 2012). Nevertheless, future research should aim to complement self-reported data with objective data sources, such as digital trace data (Ohme et al., 2024).

A combination of data sources increases validity and allows researchers to examine whether self-reported intentions (e.g., job pursuit and turnover intentions) translate into actual behavior, leading to greater insight into behavioral consequences of ADM use. In addition, this would meaningfully extend the findings based on the SOR model, as this dissertation solely assessed behavioral intentions or reports on behavior, whereas the response within the SOR model also relates to observable behavior (Mehrabian & Russell, 1974). Furthermore, researchers could combine data at the team- or organizational level to examine how collective factors, such as team dynamics or organizational culture, shape ADM reactions.

Second, as the presented essays relied on self-reported data, social desirability bias may be apparent. This could be particularly relevant since this dissertation includes measures that individuals high in social desirability might be reluctant to admit. For example, when assessing algoactivism, participants were directly asked whether they had manipulated the ADM system. This required them to openly acknowledge behavior that may not be seen as socially acceptable. As shown in Essay IV, social desirability was negatively associated with both prior discrimination experiences and algoactivism, and positively associated with justice evaluations. This suggests that individuals with high social desirability tendencies may be less likely to admit engaging in algoactivism or to experiencing discrimination, and might be more likely to report higher justice perceptions. These findings point to the risk that social desirability could skew self-reported data on sensitive topics such as justice perceptions, experiences of algorithmic discrimination, and resistance behaviors. Future research should thus consider alternative data collection methods that reduce the influence of social desirability. For example, drawing on behavioral or observational data related to algoactivism could offer more objective insights into how these dynamics unfold in practice.

Third, Essays II and III, and most of the primary studies included in the meta-analysis of Essay I were based on experimental designs. Although experimental designs increase internal validity, they tend to have low external validity, which threatens generalizability (Aguinis & Bradley, 2014). Additionally, vignette experiments require individuals to realistically imagine themselves in the described scenario. For example, in the scenario used in Essay II, participants imagined they applied for a trainee position. This specific scenario might be easier for younger participants to realistically imagine compared to older participants, as trainee positions are more commonly filled by younger individuals, which could ultimately influence the results. Nevertheless, the meta-analysis tested differences between vignette experiments and field experiments as a robustness check and did not find significant effects, indicating that methodological differences might not be as pronounced in this research field. Still, with the increasing use of ADM in organizations, future research should examine field data to increase external validity and generalizability. This would also enable the examination of the differences between organizations and how organizational culture impacts ADM adoption, which are often neglected in vignette experiments.

Fourth, this dissertation focuses solely on organizational contexts. This is an important limitation as research shows ADM perceptions are task- and context-dependent (Lee, 2018). Here, tasks associated with emotional and social skills are more likely to be perceived as untrustworthy or unjust than tasks associated with mechanical skills, such as work scheduling (Lee, 2018). This may account for the consistent finding across all essays that algorithm aversion is more prevalent than algorithm appreciation. Therefore, it should be noted that effects may vary depending on the task examined. Nevertheless, tasks and contexts that are perceived negatively might be particularly interesting to examine as they represent greater risks for organizations, making it important to identify the boundary conditions that impact these relationships and interventions to decrease algorithm aversion.

6.5 Future research avenues

Based on the findings of this dissertation, there are several avenues for future research. First, Essay I revealed that reactions to ADM depend on the type of interaction and gave initial evidence of differences between stakeholders, with significant negative reactions for applicants and employees, but non-significant findings for human resource managers. Overall, this indicates that reactions to ADM are contingent upon the interaction characteristics and the specific stakeholder group involved. However, it remains unclear why these two dimensions shape algorithm aversion and algorithm appreciation. Future research should therefore investigate the underlying mechanisms driving these effects. For example, self-determination theory (Deci & Ryan, 1985) could be used to test whether perceptions of autonomy or competence drive differences between stakeholders. Understanding not only when differences in ADM reactions occur but also why they occur, would enable a more comprehensive understanding of stakeholder perceptions and responses, facilitating more targeted approaches to addressing algorithm aversion.

Second, ADM use has shifted from platform-based organizations (e.g., Uber) to traditional organizations, where employees are now confronted with two management entities, namely their human manager and an algorithmic management system (Hirsch et al., 2024). This development offers several new research avenues, including investigations into managerial identity when management responsibility is shared with algorithms or when employees are managed by two different entities. Additionally, workers in the platform economy are typically aware that algorithms will manage them and often actively choose to engage in this form of work. This might not be the case in traditional organizations, where managers and employees may have had less exposure to ADM systems and as a result may be more hesitant or even resistant to algorithmic decisions. Future research could explore change management processes and examine how reactions to ADM differ between platform organizations and traditional organizations. Moreover, since prior research has shown that trust in ADM systems develops over time (Cabiddu et al., 2022), it would be particularly valuable to investigate the longitudinal dynamics of this shift.

Third, since this dissertation primarily focused on the organizational context and prior research has indicated that reactions to ADM are context-dependent (Lee, 2018; Mahmud et al., 2022), the research findings could be extended to other domains. For instance, as Essay II demonstrated that outcome favorability impacts interventions aimed at enhancing trust in ADM systems, future research could explore other settings where individuals face varying outcome favorability, such as the education sector. This would allow greater generalizability of findings on interventions to address algorithm aversion.

Additionally, this dissertation was unable to examine all boundary conditions that affect reactions to ADM. In particular, the perceived importance of a decision may play a crucial role, and future studies could investigate whether decisions that are subjectively more important (e.g., hiring) result in different responses compared to less critical decisions (e.g., task assignments). However, since decision importance is inherently subjective, future research should first assess perceptions of decision importance prior to examining their effects.

Fourth, as this dissertation included regional differences only as part of robustness checks, cultural and regional differences could be further addressed in future research. Following the distinction by North (1990), these differences could be explained by formal and informal institutions. While informal institutions refer to values and norms in a society, formal institutions comprise laws, regulations, and policies that shape organizational practices. Informal institutions, such as cultural values, could be assessed using the six dimensions of national culture by Hofstede (2011). In the context of ADM, both power-distance and individualism might impact perceptions (Mahmud et al., 2022). In addition, although the robustness checks in the meta-analysis in Essay I revealed no differences between the U.S. and Germany, future research could examine the impact of regulations on ADM reactions, particularly in the context of emerging regional differences in AI regulations. For example, in a comparison between the U.S. Algorithmic Accountability Act and the European AI Act, Mökander et al. (2022) state that the European regulation illustrates a "Europe-wide assurance ecosystem in detail" (p. 755), whereas the U.S. regulation only applies to large companies and leaves out certain regulations. With the European AI Act becoming compulsory for organizations in 2027 (Regulation (EU) 2024/1689, 2024), future research could examine pre- and post-conditions in comparison to countries outside of Europe.

6.6 Summary

This dissertation explored stakeholder reactions to the use of ADM in organizations and boundary conditions of these relationships. The results demonstrate that reactions to ADM use, although prominently associated with algorithm aversion, are not static phenomena but instead depend on boundary conditions (i.e., type of interaction, type of decision, interventions, and prior discrimination experiences). Interestingly, prior discrimination by humans decreased algorithm aversion, and discrimination by algorithms and humans was perceived differently. Furthermore, algorithm aversion led to engagement in algoactivism and lower organizational attractiveness and pursuit intention, suggesting that organizations using ADM systems should focus on an integration that is perceived as trustworthy and just. The results of this dissertation suggest that interventions to increase trust in ADM depend on outcome favorability. Summarizing, while ADM systems offer multiple benefits, organizations are well-advised to use these systems carefully and responsibly because they only profit from the expected benefits if ADM decisions are perceived to be trustworthy, just, and fair.

7 References

Acikgoz, Y. (2019). Employee recruitment and job search: Towards a multi-level integration. *Human Resource Management Review*, *29*(1), 1–13. https://doi.org/10.1016/j.hrmr.2018.02.009

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

Acikgoz, Y., Pollard, C., & Culver, M. (2024). *Applicant reactions to automated assessments: Moderation by job type and remote status.* Unpublished manuscript.

Adams, J. S. (1963). Towards an understanding of inequity. *The Journal of Abnormal and Social Psychology*, *67*(5), 422–436. https://doi.org/10.1037/h0040968

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

Aguinis, H., Gottfredson, R. K., & Wright, T. A. (2011). Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior*, *32*(8), 1033–1043. https://doi.org/10.1002/job.719

Ahn, D., Almaatouq, A., Gulabani, M., & Hosanagar, K. (2024, May 11–16, 2024). Impact of model interpretability and outcome feedback on trust in AI. *CHI '24: Proceedings of the CHI Conference on Human Factors in Computing Systems,* Honolulu, HI, USA. https://doi.org/10.1145/3613904.3642780

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alizadeh, A., Hirsch, F., Benlian, A., Wiener, M., & Cram, W. A. (2023). Measuring workers' perceptions of algorithmic control: Item development and content validity assessment. *Proceedings of the Hawaii International Conference on System Sciences (HICSS),* Maui, HI, USA.

Allport, G. W., Clark, K., & Pettigrew, T. (1954). *The nature of prejudice*. Addison-Wesley.

Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*, 611–623. https://doi.org/10.1007/s00146-019-00931-w

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Ausbildungspark Verlag. (2023). Logisches Denken: Zahlenreihen. [Logical thinking: Number sequences.]. Retrieved February 02, 2023, from https://www.ausbildungspark.com/einstellungstest/zahlenreihen/

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*, 74–94. https://doi.org/10.1007/BF02723327

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*, 857–875. https://doi.org/10.1007/s10796-021-10223-8

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior*, *45*(2), 159–182. https://doi.org/10.1002/job.2735

Basu, S., Majumdar, B., Mukherjee, K., Munjal, S., & Palaksha, C. (2023). Artificial intelligence–HRM interactions and outcomes: A systematic review and causal configurational explanation. *Human Resource Management Review*, *33*(1), Article 100893. https://doi.org/10.1016/j.hrmr.2022.100893

Becker, L. (2025). How transparency and value affect platform workers' workaround use and continuance intention. *Proceedings of the Hawaii International Conference on System Sciences (HICSS),* Kona, HI, USA.

Bedemariam, R., & Wessel, J. L. (2023). The roles of outcome and race on applicant reactions to AI systems. *Computers in Human Behavior*, *148*, Article 107869. https://doi.org/10.1016/j.chb.2023.107869

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. https://doi.org/10.2307/2533446

Bell, B. S., Ryan, A. M., & Wiechmann, D. (2004). Justice expectations and applicant perceptions. *International Journal of Selection and Assessment*, *12*(1-2), 24–38. https://doi.org/10.1111/j.0965-075X.2004.00261.x

Benlian, A., Wiener, M., Cram, W. A., Krasnova, H., Maedche, A., Möhlmann, M., Recker, J., & Remus, U. (2022). Algorithmic management: bright and dark sides, practical implications, and research opportunities. *Business & Information Systems Engineering*, *64*(6), 825–839. https://doi.org/10.1007/s12599-022-00764-w

Beugré, C. D. (2009). Exploring the neural basis of fairness: A model of neuro-organizational justice. *Organizational Behavior and Human Decision Processes*, *110*(2), 129–139. https://doi.org/10.1016/j.obhdp.2009.06.005

Bigman, Y. E., Wilson, D., Arnestad, M. N., Waytz, A., & Gray, K. (2023). Algorithmic discrimination causes less moral outrage than human discrimination. *Journal of Experimental Psychology: General*, *152*(1), 4–27. https://doi.org/10.1037/xge0001250

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. *CHI '18: Proceedings of the 2018 CHI conference on human factors in computing systems,* Montreal, QC. https://doi.org/10.1145/3173574.3173951

Bitzer, T., Wiener, M., & Cram, W. A. (2023). Algorithmic transparency: Concepts, antecedents, and consequences–A review and research framework. *Communications of the Association for Information Systems*, *52*(1), 293–331. https://doi.org/10.17705/1CAIS.05214

Black, J. S., & van Esch, P. (2021). AI-enabled recruiting in the war for talent. *Business Horizons*, *64*(4), 513–524. https://doi.org/10.1016/j.bushor.2021.02.015

Bonezzi, A., & Ostinelli, M. (2021). Can algorithms legitimize discrimination? *Journal of Experimental Psychology: Applied*, *27*(2), 447–459. https://doi.org/10.1037/xap0000294

Boomsma, A., & Herzog, W. (2013). *Correcting structural equation model fit statistics and indexes under small-sample and/or large model conditions*. Retrieved February 02, 2025, from https://aboomsma.webhosting.rug.nl/swain.pdf

Boston Consulting Group. (2024). *AI Adoption in 2024: 74% of Companies Struggle to Achieve and Scale Value*. Retrieved May 20, 2025, from https://www.bcg.com/press/24october2024-ai-adoption-in-2024-74-of-companies-struggle-to-achieve-and-scale-value

Breidbach, C. F. (2024). Responsible algorithmic decision-making. *Organizational Dynamics*, *53*(2), Article 101031. https://doi.org/10.1016/j.orgdyn.2024.101031

Brennen, A. (2020). What do people really want when they say they want "explainable AI?" We asked 60 stakeholders. *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA. https://doi.org/10.1145/3334480.3383047

Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., Boselie, P., Lee Cooke, F., Decker, S., & DeNisi, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, *33*(3), 606–659. https://doi.org/10.1111/1748-8583.12524

Budhwar, P., Malik, A., De Silva, M. T. T., & Thevisuthan, P. (2022). Artificial intelligence – Challenges and opportunities for international HRM: A review and research agenda. *The International Journal of Human Resource Management*, *33*(6), 1065–1097. https://doi.org/10.1080/09585192.2022.2035161

Bujold, A., Roberge-Maltais, I., Parent-Rocheleau, X., Boasen, J., Sénécal, S., & Léger, P.-M. (2024). Responsible artificial intelligence in human resources management: A review of the empirical literature. *AI and Ethics*, *4*(4), 1185–1200. https://doi.org/10.1007/s43681-023-00325-1

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Cabiddu, F., Moi, L., Patriotta, G., & Allen, D. G. (2022). Why do users trust algorithms? A review and conceptualization of initial trust and trust over time. *European Management Journal*, *40*(5), 685–706. https://doi.org/10.1016/j.emj.2022.06.001

Campion, E. D., Campion, M. A., Johnson, J., Carretta, T. R., Romay, S., Dirr, B., Deregla, A., & Mouton, A. (2024). Using natural language processing to increase prediction and reduce subgroup differences in personnel selection decisions. *Journal of Applied Psychology*, *109*(3), 307–338. https://doi.org/10.1037/apl0001144

Campion, M. A., & Campion, E. D. (2023). Machine learning applications to personnel selection: Current illustrations, lessons learned, and future research. *Personnel Psychology*, *76*(4), 993–1009. https://doi.org/10.1111/peps.12621

Canagasuriam, D., & Lukacik, E.-R. (2024). ChatGPT, can you take my job interview? Examining artificial intelligence cheating in the asynchronous video interview. *International Journal of Selection and Assessment*, *33*(1), Article e12491. https://doi.org/10.1111/ijsa.12491

Candrian, C., & Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior*, *134*, Article 107308. https://doi.org/10.1016/j.chb.2022.107308

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS-Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change-and meta-competencies. *Computers in Human Behavior: Artificial Humans*, *1*(2), Article 100014. https://doi.org/10.1016/j.chbah.2023.100014

Carpenter, N. C., Whitman, D. S., & Amrhein, R. (2021). Unit-level counterproductive work behavior (CWB): A conceptual review and quantitative summary. *Journal of Management*, *47*(6), 1498–1527. https://doi.org/10.1177/0149206320978812

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

Chacon, A., Larrain, M., & Moritz, J. M. (2024). *Algorithm aversion in career decision-making: Examining the effects of algorithmic terminology.* Unpublished manuscript.

Chamorro-Premuzic, T., & Ahmetoglu, G. (2016). The pros and cons of robot managers. *Harvard Business Review*. Retrieved May 10, 2024, from https://hbr.org/2016/12/the-pros-and-cons-of-robot-managers

Chandra, S., Shirish, A., & Srivastava, S. C. (2022). To be or not to be …Human? Theorizing the role of human-like competencies in conversational artificial intelligence agents. *Journal of Management Information Systems*, *39*(4), 969–1005. https://doi.org/10.1080/07421222.2022.2127441

Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology*, *90*(5), 928–944. https://doi.org/10.1037/0021-9010.90.5.928

Cheng, M. M., & Hackett, R. D. (2021). A critical review of algorithms in HRM: Definition, theory, and practice. *Human Resource Management Review*, *31*(1), Article 100698. https://doi.org/10.1016/j.hrmr.2019.100698

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, *29*(4), 387–396. https://doi.org/10.1007/s11065-019-09415-6

Choi, J., & Chao, M. M. (2024). For me or against me? Reactions to AI (vs. human) decisions that are favorable or unfavorable to the self and the role of fairness perception. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/01461672241288338

Choung, H., Seberger, J. S., & David, P. (2023). When AI is perceived to be fairer than a human: Understanding perceptions of algorithmic decisions in a job application context. *International Journal of Human–Computer Interaction*, *44*(2), 7451–7468. https://doi.org/10.1080/10447318.2023.2266244

Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., & Truong, L. (2023). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review*, *33*(1), Article 100899. https://doi.org/10.1016/j.hrmr.2022.100899

Cobonpue, J. S., Wurm, B., & Hess, T. (2024). Fighting back the algorithm: A Systematic Literature Review on Algoactivism. *Proceedings of the International Conference on Information Systems (ICIS)*, Bangkok, Thailand. https://aisel.aisnet.org/icis2024/lit_review/lit_review/16

Cohen, A., & Diamant, A. (2019). The role of justice perceptions in determining counterproductive work behaviors. *The International Journal of Human Resource Management*, *30*(20), 2901–2924. https://doi.org/10.1080/09585192.2017.1340321

Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *86*(2), 278–321. https://doi.org/10.1006/obhd.2001.2958

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., Hill, E. T., & De Cremer, D. (2023). Forever focused on fairness: 75 years of organizational justice in Personnel Psychology. *Personnel Psychology*, *76*(2), 413–435. https://doi.org/10.1111/peps.12556

Colquitt, J. A., & Rodell, J. B. (2015). Measuring justice and fairness. In R. S. Cropanzano & M. L. Ambrose (Eds.), *The Oxford handbook of justice in the workplace* (pp. 187–202). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199981410.013.8

Corbeanu, A., Iliescu, D., Ion, A., & Spînu, R. (2023). The link between burnout and job performance: A meta-analysis. *European Journal of Work and Organizational Psychology*, *32*(4), 599–616. https://doi.org/10.1080/1359432X.2023.2209320

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Crocker, J., Voelkl, K., Testa, M., & Major, B. (1991). Social stigma: The affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology*, *60*(2), 218–228. https://doi.org/10.1037/0022-3514.60.2.218

Cropanzano, R., Fortin, M., & Kirk, J. F. (2015). How do we know when we are treated fairly? Justice rules and fairness judgments. In M. R. Buckley, A. R. Wheeler, & J. R. B. Halbesleben (Eds.), *Research in personnel and human resources management*

(Vol. 33, pp. 279–350). Emerald Group Publishing Limited. https://doi.org/10.1108/S0742-730120150000033010

Czernietzki, C., Märtins, J., Westmattelmann, D., Grotenhermen, J.-G., Oldeweme, A., Baumeister, V., & Schewe, G. (2023). Navigating AI in personnel selection: A scenario-based study on applicants' perceptions. *International Conference on Information Systems,* Hyderabad, India.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554

Dalvi-Esfahani, M., Mosharaf-Dehkordi, M., Leong, L. W., Ramayah, T., & Kanaan-Jebna, A. M. J. (2023). Exploring the drivers of XAI-enhanced clinical decision support systems adoption: Insights from a stimulus-organism-response perspective. *Technological Forecasting and Social Change*, *195*, Article 122768. https://doi.org/10.1016/j.techfore.2023.122768

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

Davis, F. D., Bagozzi, R., & Warshaw, P. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

De Cremer, D., & McGuire, J. (2022). Human–algorithm collaboration works best if humans lead (because it is fair!). *Social Justice Research*, *35*(1), 33–55. https://doi.org/10.1007/s11211-021-00382-z

de Jong, A. W., Remus, U., & Weber, M. (2025). Measuring Algoactivism in Gig Work: Scale Development. *Proceedings of the European Conference on Information*

Systems (ECIS), Amman, Jordan. https://aisel.aisnet.org/ecis2025/algo_mgmt/algo_mgmt/1

de Jong, A. W., Weber, M., & Remus, U. (2024). Towards A Measurement Instrument for Assessing Resistance to Algorithmic Control: Conceptualization and Content Validity. *Proceedings of the European Conference on Information Systems (ECIS)*, Paphos, Cyprus. https://aisel.aisnet.org/ecis2024/track05_fow/track05_fow/13

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, *19*(2), 109–134. https://doi.org/10.1016/0092-6566(85)90023-6

Deepa, R., Sekar, S., Malik, A., Kumar, J., & Attri, R. (2024). Impact of AI-focussed technologies on social and technical competencies for HR managers – A systematic review and research agenda. *Technological Forecasting and Social Change*, *202*, Article 123301. https://doi.org/10.1016/j.techfore.2024.123301

Del Giudice, M., Scuotto, V., Orlando, B., & Mustilli, M. (2023). Toward the human – centered approach. A revised model of individual acceptance of AI. *Human Resource Management Review*, *33*(1), Article 100856. https://doi.org/10.1016/j.hrmr.2021.100856

Deng, X. N., & Galliers, R. D. (2024). Toward an understanding of gig work risks and worker agency on different digital labor platforms. *Journal of the Association for Information Systems*, *25*(5), 1163–1193. https://doi.org/10.17705/1jais.00894

Deriu, V., Pozharliev, R., & De Angelis, M. (2024). How trust and attachment styles jointly shape job candidates' AI receptivity. *Journal of Business Research*, *179*, Article 114717. https://doi.org/10.1016/j.jbusres.2024.114717

Di Battista, A., Grayling, S., Hasselaar, E., Leopold, T., Li, R., Rayner, M., & Zahidi, S. (2023). Future of jobs report 2023. *World Economic Forum*. Retrieved May 10, 2024, from https://www.weforum.org/reports/the-future-of-jobs-report-2023/

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, *5*(7), 809–828. https://doi.org/10.1080/21670811.2016.1208053

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

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

Dipboye, R. L., & Colella, A. (2013). *Discrimination at work: The psychological and organizational bases*. Psychology Press.

Dong, M., Bonnefon, J.-F., & Rahwan, I. (2024). Toward human-centered AI management: Methodological challenges and future directions. *Technovation*, *131*, Article 102953. https://doi.org/10.1016/j.technovation.2024.102953

Edwards, M. R., Zubielevitch, E., Okimoto, T., Parker, S., & Anseel, F. (2024). Managerial control or feedback provision: How perceptions of algorithmic HR systems shape employee motivation, behavior, and well-being. *Human Resource Management*, *63*(4), 691–710. https://doi.org/10.1002/hrm.22218

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

European Commission. (2021). Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. (COM/2021/206 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

Evertz, L., Kollitz, R., & Süß, S. (2019). Electronic word-of-mouth via employer review sites–the effects on organizational attraction. *The International Journal of Human Resource Management*, *32*(16), 3428–3457. https://doi.org/10.1080/09585192.2019.1640268

Feldkamp, T., Langer, M., Wies, L., & König, C. J. (2023). Justice, trust, and moral judgements when personnel selection is supported by algorithms. *European Journal of Work and Organizational Psychology*, *33*(2), 130–145. https://doi.org/10.1080/1359432X.2023.2169140

Feng, B., & MacGeorge, E. L. (2010). The influences of message and source factors on advice outcomes. *Communication Research*, *37*(4), 553–575. https://doi.org/10.1177/0093650210368258

Fessenko, D. S., & Jasperse, A. (2024). Ethics at the heart of AI regulation. *AI and Ethics*, *5*, 3387–3398. https://doi.org/10.1007/s43681-024-00562-y

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fleiß, J., Kubicek, B., & Thalmann, S. (2024). Better the devil you don't know. perceived discrimination increases positive evaluation of ai recruiting decisions. *Proceedings of the International Conference on Information Systems (ICIS)*, Bangkok, Thailand. https://aisel.aisnet.org/icis2024/soc_impactIS/soc_impactIS/17

Follmer, K. B., Geiger, M., Beatty, J. E., & Follmer, D. J. (2023). The consequences of not being me: Longitudinal examination of the relations among anticipated discrimination, authenticity, and counterproductive work behaviors. *Group & Organization Management*, *48*(4), 1058–1095. https://doi.org/10.1177/10596011221107720

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *Management Information Systems Quarterly*, *45*(3), 1527–1556. https://doi.org/10.25300/MISQ/2021/16553

Fumagalli, E., Rezaei, S., & Salomons, A. (2022). OK computer: Worker perceptions of algorithmic recruitment. *Research Policy*, *51*(2), Article 104420. https://doi.org/10.1016/j.respol.2021.104420

Furneaux, G. B. (2012). Task-technology fit theory: A survey and synopsis of the literature. In Y. K. Dwivedi & M. R. Wade (Eds.), Information systems theory:

Explaining and predicting our digital society (Vol. 1, pp. 87–102). Springer. https://doi.org/10.1007/978-1-4419-6108-2_5

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 71–85. https://doi.org/10.30658/hmc.1.5

Garcia, A. C. B., Garcia, M. G. P., & Rigobon, R. (2024). Algorithmic discrimination in the credit domain: what do we know about it? *AI & Society*, *39*(4), 2059–2098. https://doi.org/10.1007/s00146-023-01676-3

Gazi, M. A. I., Ibrahim, M., Masud, A. A., & Reza, S. M. A. (2025). Brand experiences and loyalty among young smartphone users: a serial mediation analysis. *Management Decision*. https://doi.org/10.1108/MD-12-2023-2321

Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, *31*(3), 410–435. https://doi.org/10.1080/0960085x.2021.1927213

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goldman, B., & Cropanzano, R. (2015). “Justice” and “fairness” are not the same thing. *Journal of Organizational Behavior*, *36*(2), 313–318. https://doi.org/10.1002/job.1956

Gomes, D., & Neves, J. (2011). Organizational attractiveness and prospective applicants' intentions to apply. *Personnel Review*, *40*(6), 684–699. https://doi.org/10.1108/00483481111169634

Gonzalez, M. F., Capman, J. F., Oswald, F. L., Theys, E. R., & Tomczak, D. L. (2019). “Where’s the IO?” Artificial intelligence and machine learning in talent management systems. *Personnel Assessment and Decisions*, *5*(3), Article 5. https://doi.org/10.25035/pad.2019.03.005

Gonzalez, M. F., Liu, W. W., Shirase, L., Tomczak, D. L., Lobbe, C. E., Justenhoven, R., & Martin, N. R. (2022). Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. *Computers in Human Behavior*, *130*, Article 107179. https://doi.org/10.1016/j.chb.2022.107179

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, *45*, Article 105681. https://doi.org/10.1016/j.clsr.2022.105681

Greenberg, J. (1987). A taxonomy of organizational justice theories. *Academy of Management Review*, *12*(1), 9–22. https://doi.org/10.5465/amr.1987.4306437

Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, *83*(2), 241–262. https://doi.org/10.1111/puar.13483

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, *29*(2), Article 101614. https://doi.org/10.1016/j.jsis.2020.101614

Gursoy, D., Chi, O. H., Lu, L., & Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management*, *49*, 157–169. https://doi.org/10.1016/j.ijinfomgt.2019.03.008

Haesevoets, T., Verschuere, B., Van Severen, R., & Roets, A. (2024). How do citizens perceive the use of Artificial Intelligence in public sector decisions? *Government Information Quarterly*, *41*(1), 101906. https://doi.org/10.1016/j.giq.2023.101906

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis (7th ed.)*. Pearson.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*(3), 639–683. https://doi.org/10.1111/j.1744-6570.2004.00003.x

Hidalgo, C. A., Orghian, D., Canals, J. A., De Almeida, F., & Martin, N. (2021). *How humans judge machines*. MIT Press.

Highhouse, S., Brooks, M. E., & Gregarus, G. (2009). An organizational impression management perspective on the formation of corporate reputations. *Journal of Management*, *35*(6), 1481–1493. https://doi.org/10.1177/0149206309348788

Highhouse, S., Lievens, F., & Sinar, E. F. (2003). Measuring attraction to organizations. *Educational and Psychological Measurement*, *63*(6), 986–1001. https://doi.org/10.1177/0013164403258403

Highhouse, S., Thornbury, E. E., & Little, I. S. (2007). Social-identity functions of attraction to organizations. *Organizational Behavior and Human Decision Processes*, *103*(1), 134–146. https://doi.org/10.1016/j.obhdp.2006.01.001

Hirsch, F., Alizadeh, A., Wiener, M., & Benlian, A. (2024). The uberization of work: non-platform workers' perceptions and legitimacy judgments of algorithmic control in hybrid control regimes. *Proceedings of the International Conference on Information Systems (ICIS)*, Bangkok, Thailand. https://aisel.aisnet.org/icis2024/digtech_fow/digtech_fow/2

Höddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, *116*, Article 106635. https://doi.org/10.1016/j.chb.2020.106635

Hodson, R. (1995). Worker resistance: An underdeveloped concept in the sociology of work. *Economic and Industrial Democracy*, *16*(1), 79–110. https://doi.org/10.1177/0143831X9501600104

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, *2*(1), Article 8. https://doi.org/10.9707/2307-0919.1014

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics*, *178*(4), 977–1007. https://doi.org/10.1007/s10551-022-05049-6

Jackson, S. E., & Schuler, R. S. (Eds.). (2003). *Managing human resources through strategic partnerships* (8th ed.). Southwestern.

Jacoby, J. (2002). Stimulus-Organism-Response Reconsidered: An Evolutionary Step in Modeling (Consumer) Behavior. *Journal of Consumer Psychology*, *12*(1), 51–57. https://doi.org/10.1207/S15327663JCP1201_05

Jain, R., Garg, N., & Khera, S. N. (2023). Comparing differences of trust, collaboration and communication between human-human vs human-bot teams: an experimental study. *CERN IdeaSquare Journal of Experimental Innovation*, *7*(2), 8–16. https://doi.org/10.23726/cij.2022.1387

Jarrahi, M. H., Newlands, G., Lee, M. K., Wolf, C. T., Kinder, E., & Sutherland, W. (2021). Algorithmic management in a work context. *Big Data & Society*, *8*(2). https://doi.org/10.1177/20539517211020332

Jarrahi, M. H., & Sutherland, W. (2019). Algorithmic management and algorithmic competencies: Understanding and appropriating algorithms in gig work.

Proceedings of the Information in Contemporary Society: 14th International Conference, Washington, DC, USA. https://doi.org/10.1007/978-3-030-15742-5_55

Jensen, J. M., & Patel, P. C. (2011). Predicting counterproductive work behavior from the interaction of personality traits. *Personality and Individual Differences*, *51*(4), 466-471. https://doi.org/10.1016/j.paid.2011.04.016

Johnson, R. D., Lukaszewski, K. M., & Stone, D. L. (2016). The evolution of the field of human resource information systems: Co-evolution of technology and HR processes. *Communications of the Association for Information Systems*, *38*(1), 533–553. https://doi.org/10.17705/1CAIS.03828

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). *semTools: Useful tools for structural equation modeling. R package version 0.5-1*. https://cran.r-project.org/web/packages/semTools/index.html

Jussupow, E., Benbasat, I., & Heinzl, A. (2024). An integrative perspective on algorithm aversion and appreciation in decision-making. *Management Information Systems Quarterly*, *48*(4), 1575–1590. https://doi.org/10.25300/MISQ/2024/18512

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research*, *32*(3), 713–735. https://doi.org/10.1287/isre.2020.0980

Kadolkar, I., Kepes, S., & Subramony, M. (2024). Algorithmic management in the gig economy: A systematic review and research integration. *Journal of Organizational Behavior*, 1–24. https://doi.org/10.1002/job.2831

Kares, F., König, C. J., Bergs, R., Protzel, C., & Langer, M. (2023). Trust in hybrid human-automated decision-support. *International Journal of Selection and Assessment*, *31*(3), 388–402. https://doi.org/10.1111/ijsa.12423

Kaushal, N., Kaurav, R. P. S., Sivathanu, B., & Kaushik, N. (2023). Artificial intelligence and HRM: Identifying future research agenda using systematic literature review and bibliometric analysis. *Management Review Quarterly*, *73*(2), 455–493. https://doi.org/10.1007/s11301-021-00249-2

Keiser, H. N., Sackett, P. R., Kuncel, N. R., & Brothen, T. (2016). Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns. *Journal of Applied Psychology*, *101*(4), 569–581. https://doi.org/10.1037/apl0000069

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kelloway, E. K., Francis, L., Prosser, M., & Cameron, J. E. (2010). Counterproductive work behavior as protest. *Human Resource Management Review*, *20*(1), 18–25. https://doi.org/10.1016/j.hrmr.2009.03.014

Keppeler, F. (2023). No thanks, dear AI! Understanding the effects of disclosure and deployment of artificial intelligence in public sector recruitment. *Journal of Public Administration Research and Theory*, *34*(1), 39–52. https://doi.org/10.1093/jopart/muad009

Keppeler, F., Borchert, J., Pedersen, M. J., & Nielsen, V. L. (2024). *How ensembling AI and public managers improves decision-making.* Unpublished manuscript.

Kim, Y., Cohen, T. R., & Panter, A. T. (2023). Workplace mistreatment and employee deviance: an investigation of the reciprocal relationship between hostile work environments and harmful work behaviors. *Group & Organization Management*, *48*(4), 1173–1202. https://doi.org/10.1177/10596011231151747

Kinowska, H., & Sienkiewicz, Ł. J. (2022). Influence of algorithmic management practices on workplace well-being–evidence from European organisations. *Information Technology & People*, *36*(8), 21–42. https://doi.org/10.1108/ITP-02-2022-0079

Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. *CHI '16: Proceedings of the 2016 CHI conference on human factors in computing systems*, San Jose, CA, USA. https://doi.org/10.1145/2858036.2858402

Kleinlogel, E. P., Schmid Mast, M., Jayagopi, D. B., Shubham, K., & Butera, A. (2023). "The interviewer is a machine!" Investigating the effects of conventional and technology-mediated interview methods on interviewee reactions and behavior. *International Journal of Selection and Assessment*, *31*(3), 403–419. https://doi.org/10.1111/ijsa.12433

Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, *160*, Article 108352. https://doi.org/10.1016/j.chb.2024.108352

Klöpper, M., & Messer, U. (2025). Empirical insights on perceived fairness violation in algorithm-supported managerial decisions. *Journal of Business Analytics*, 1–18. https://doi.org/10.1080/2573234X.2025.2482645

Koch-Bayram, I. F., Kaibel, C., Biemann, T., & Triana, M. d. C. (2023). </Click to begin your digital interview>: Applicants' experiences with discrimination explain their reactions to algorithms in personnel selection. *International Journal of Selection and Assessment*, *31*(2), 252–266. https://doi.org/10.1111/ijsa.12417

Koch-Bayram, I. F., & Kaibel, C. (2023). Algorithms in personnel selection, applicants' attributions about organizations' intents and organizational attractiveness: An experimental study. *Human Resource Management Journal*, *34*(3), 733–752. https://doi.org/10.1111/1748-8583.12528

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

Köchling, A., & Wehner, M. C. (2023). Better explaining the benefits why AI? Analyzing the impact of explaining the benefits of AI-supported selection on applicant responses. *International Journal of Selection and Assessment*, *31*(1), 45–62. https://doi.org/10.1111/ijsa.12412

Köchling, A., Wehner, M. C., & Ruhle, S. A. (2025). This (AI)n't fair? Employee reactions to artificial intelligence (AI) in career development systems. *Review of Managerial Science*, *19*, 1195–1228. https://doi.org/10.1007/s11846-024-00789-3

Köchling, A., Wehner, M. C., & Warkocz, J. (2023). Can I show my skills? Affective responses to artificial intelligence in the recruitment process. *Review of Managerial Science*, *17*, 2109–2138. https://doi.org/10.1007/s11846-021-00514-4

Koenig, N., Tonidandel, S., Thompson, I., Albritton, B., Koohifar, F., Yankov, G., Speer, A., Hardy III, J. H., Gibson, C., Frost, C., Liu, M., McNeney, D., Capman, J., Lowery, S., Kitching, M., Nimbkar, A., Boyce, A., Sun, T., Guo, F., . . . Newton, C. (2023). Improving measurement and prediction in personnel selection through the application of machine learning. *Personnel Psychology*, *76*(4), 1061–1123. https://doi.org/10.1111/peps.12608

König, C. J., & Langer, M. (2022). Machine learning in personnel selection. In S. Strohmeier (Ed.), *Handbook of research on artificial intelligence in human resource management* (pp. 149–167). Edward Elgar Publishing.

Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 1–22. https://doi.org/10.1080/0960085x.2021.1927212

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, *50*(1), 569–598. https://doi.org/10.1146/annurev.psych.50.1.569

Krügel, S., Ostermaier, A., & Uhl, M. (2023). Algorithms as partners in crime: A lesson in ethics by design. *Computers in Human Behavior*, *138*, 107483. https://doi.org/10.1016/j.chb.2022.107483

Kushwaha, A. K., Pharswan, R., Kumar, P., & Kar, A. K. (2022). How do users feel when they use artificial intelligence for decision making? A framework for assessing users' perception. *Information Systems Frontiers*, *25*, 1241–1260. https://doi.org/10.1007/s10796-022-10293-2

Lacroux, A., & Martin-Lacroux, C. (2022). Should I trust the artificial intelligence to recruit? Recruiters' perceptions and behavior when faced with algorithm-based recommendation systems during resume screening. *Frontiers in Psychology*, *13*, Article 895997. https://doi.org/10.3389/fpsyg.2022.895997

Langer, M., Baum, K., & Schlicker, N. (2025). Effective human oversight of AI-based systems: A signal detection perspective on the detection of inaccurate and unfair outputs. *Minds and Machines*, *35*(1), Article 1. https://doi.org/10.1007/s11023-024-09701-0

Langer, M., & König, C. J. (2018). Introducing and testing the Creepiness of Situation Scale (CRoSS). *Frontiers in Psychology*, *9*, Article 2220. https://doi.org/10.3389/fpsyg.2018.02220

Langer, M., König, C. J., Back, C., & Hemsing, V. (2023). Trust in artificial intelligence: Comparing trust processes between human and automated trustees in light of unfair bias. *Journal of Business and Psychology*, *38*, 493–508. https://doi.org/10.1007/s10869-022-09829-9

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

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

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

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*, Article 106878. https://doi.org/10.1016/j.chb.2021.106878

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence*, *296*, 1–58. https://doi.org/10.1016/j.artint.2021.103473

Lavanchy, M., Reichert, P., Narayanan, J., & Savani, K. (2023). Applicants' fairness perceptions of algorithm-driven hiring procedures. *Journal of Business Ethics*, *188*, 125–150. https://doi.org/10.1007/s10551-022-05320-w

Lee, J. D., & See, K. A. (2004). 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 & Society*, *5*(1), 1–16. https://doi.org/10.1177/2053951718756684

Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27–55). Springer. https://doi.org/10.1007/978-1-4613-3087-5_2

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review*, *23*(3), 438–458. https://doi.org/10.5465/amr.1998.926620

Liao, M., & Sundar, S. S. (2021). How should AI systems talk to users when collecting their personal information? Effects of role framing and self-referencing on human-AI interaction. *CHI '21: Proceedings of the 2021 CHI conference on human factors in computing systems*, Yokohama, Japan.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology*, *62*(10), e1–e34. https://doi.org/10.1016/j.jclinepi.2009.06.006

Lima, G., Grgić-Hlača, N., Langer, M., & Zou, Y. (2025). Lay Perceptions of Algorithmic Discrimination in the Context of Systemic Injustice. *CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems,* Yokohama, Japan. https://doi.org/10.1145/3706598.3713536

Lind, E. A. (2001). Fairness heuristic theory: Justice judgments as pivotal cognitions in organizational relations. In J. Greenberg & R. Cropanzano (Eds.), *Advances in organizational justice* (Vol. 56, pp. 56–88). Stanford University Press.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, *86*(1), 114–121. https://doi.org/10.1037//0021-9010.86.1.114

Liu, B., Wei, L., Wu, M., & Luo, T. (2023). Speech production under uncertainty: How do job applicants experience and communicate with an AI interviewer? *Journal of Computer-Mediated Communication*, *28*(4), Article zmad028. https://doi.org/10.1093/jcmc/zmad028

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

Long, N. (2003). *Development sociology: Actor perspectives*. Routledge.

Luo, W., & Zhang, Y. (2023). Why AI-enabled interviews might reduce candidates' job application intention? *Academy of Management Proceedings*, *2023*(1). https://doi.org/10.5465/AMPROC.2023.17884abstract

Maasland, C., & Weissmueller, K. S. (2022). Blame the machine? Insights from an experiment on algorithm aversion and blame avoidance in computer-aided human resource management. *Frontiers in Psychology*, *13*, Article 779028. https://doi.org/10.3389/fpsyg.2022.779028

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, Article 121390. https://doi.org/10.1016/j.techfore.2021.121390

Mahmud, H., Islam, A. N., Luo, X. R., & Mikalef, P. (2024). Decoding algorithm appreciation: Unveiling the impact of familiarity with algorithms, tasks, and algorithm performance. *Decision Support Systems*, *179*, Article 114168. https://doi.org/10.1016/j.dss.2024.114168

Maier, G. W., Streicher, B., Jonas, E., & Woschée, R. (2007). Gerechtigkeitseinschätzungen in Organisationen. *Diagnostica*, *53*(2), 97–108. https://doi.org/10.1026/0012-1924.53.2.97

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McGuire, J., & De Cremer, D. (2023). Algorithms, leadership, and morality: Why a mere human effect drives the preference for human over algorithmic leadership. *AI and Ethics*, *3*(2), 601–618. https://doi.org/10.1007/s43681-022-00192-2

McKinsey & Company. (2025). *The state of AI: How organizations are rewiring to capture value*. Retrieved May 10, 2025, from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Mehrabian, A., & Russell, J. (1974). *An approach to environmental psychology*. MIT Press.

Meijerink, J., & Bondarouk, T. (2023). The duality of algorithmic management: Toward a research agenda on HRM algorithms, autonomy and value creation. *Human*

Resource Management Review, *33*(1), Article 100876. https://doi.org/10.1016/j.hrmr.2021.100876

Meijerink, J., Boons, M., Keegan, A., & Marler, J. (2021). Algorithmic human resource management: Synthesizing developments and cross-disciplinary insights on digital HRM. *The International Journal of Human Resource Management*, *32*(12), 2545–2562. https://doi.org/10.1080/09585192.2021.1925326

Middleton, S. E., Letouzé, E., Hossaini, A., & Chapman, A. (2022). Trust, regulation, and human-in-the-loop AI: Within the European region. *Communications of the ACM*, *65*(4), 64–68. https://doi.org/10.1145/3511597

Mikulincer, M., & Shaver, P. R. (2005). Attachment security, compassion, and altruism. *Current Directions in Psychological Science*, *14*(1), 34–38. https://doi.org/10.1111/j.0963-7214.2005.00330.x

Mirowska, A. (2020). AI evaluation in selection effects on application and pursuit intentions. *Journal of Personnel Psychology*, *19*(3), 142–149. https://doi.org/10.1027/1866-5888/a000258

Mirowska, A., & Mesnet, L. (2022). Preferring the devil you know: Potential applicant reactions to artificial intelligence evaluation of interviews. *Human Resource Management Journal*, *32*(2), 364–383. https://doi.org/10.1111/1748-8583.12393

Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *Management Information Systems Quarterly*, *45*(4), 1999–2022. https://doi.org/10.25300/MISQ/2021/15333

Möhlmann, M., Alves de Lima Salge, C., & Marabelli, M. (2023). Algorithm sensemaking: How platform workers make sense of algorithmic management.

Journal of the Association for Information Systems, *24*(1), 35–64. https://doi.org/10.17705/1jais.00774

Mökander, J., Juneja, P., Watson, D. S., & Floridi, L. (2022). The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What can they learn from each other? *Minds and Machines*, *32*, 751–758. https://doi.org/10.1007/s11023-022-09612-y

Moon, K.-K. (2017). Fairness at the organizational level: Examining the effect of organizational justice climate on collective turnover rates and organizational performance. *Public Personnel Management*, *46*(2), 118–143. https://doi.org/10.1177/0091026017702610

Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy*, *7*(4), 33–35.

Moritz, J. M., & Schmidt, C. (2024). Trust in algorithmic management: The role of justice and prior discrimination experience. *Proceedings of the European Conference on Information Systems (ECIS)*, Paphos, Cyprus. https://aisel.aisnet.org/ecis2024/track05_fow/track05_fow/8

Moritz, J. M., Witte, J., & Gier-Reinartz, N. R. (2024). In algorithms we trust? Exploring the role of human-in-the-loop and social presence. *Academy of Management Proceedings*, *2024*(1). https://doi.org/10.5465/AMPROC.2024.10385abstract

Moritz, J. M., Zahs, D., Schmodde, L., Wehner, M., Kulkarni, S., & Chacon, A. (2025). AI usage in HRM: A cross-cultural investigation. *Academy of Management Proceedings*, *2025*(1).

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, *37*(11), 1905–1922. https://doi.org/10.1080/00140139408964957

Munnukka, J., Talvitie-Lamberg, K., & Maity, D. (2022). Anthropomorphism and social presence in Human–Virtual service assistant interactions: The role of dialog length and attitudes. *Computers in Human Behavior*, *135*, Article 107343. https://doi.org/10.1016/j.chb.2022.107343

Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly*, *38*(1), Article 101536. https://doi.org/10.1016/j.giq.2020.101536

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI conference on Human factors in computing systems,* Boston, MA, USA. https://doi.org/10.1145/191666.191703

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 & Human Decision Processes*, *160*, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

Nishii, L. H., Lepak, D. P., & Schneider, B. (2008). Employee attributions of the "why" of HR practices: Their effects on employee attitudes and behaviors, and customer satisfaction. *Personnel Psychology*, *61*(3), 503–545. https://doi.org/10.1111/j.1744-6570.2008.00121.x

Noble, S. M., Foster, L. L., & Craig, S. B. (2021). The procedural and interpersonal justice of automated application and resume screening. *International Journal of Selection and Assessment*, *29*(2), 139–153. https://doi.org/10.1111/ijsa.12320

North, D. C. (1990). *Institutions, institutional change, and economic performance*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808678

Oakes, P., & Turner, J. C. (1986). Distinctiveness and the salience of social category memberships: Is there an automatic perceptual bias towards novelty? *European Journal of Social Psychology*, *16*(4), 325–344. https://doi.org/10.1002/ejsp.2420160403

Ochmann, J., Michels, L., Tiefenbeck, V., Maier, C., & Laumer, S. (2024). Perceived algorithmic fairness: An empirical study of transparency and anthropomorphism in algorithmic recruiting. *Information Systems Journal*, *34*(2), 384–414. https://doi.org/https://doi.org/10.1111/isj.12482

Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2024). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Communication Methods and Measures*, *18*(2), 124–141. https://doi.org/10.1080/19312458.2023.2181319

Oostrom, J. K., Holtrop, D., Koutsoumpis, A., van Breda, W., Ghassemi, S., & de Vries, R. E. (2023). Applicant reactions to algorithm- versus recruiter-based evaluations of an asynchronous video interview and a personality inventory. *Journal of Occupational and Organizational Psychology*, *97*(1), 160–189. https://doi.org/10.1111/joop.12465

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, *5*, Article 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., & Mulrow, C. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLoS Med*, *18*(3), Article e1003583. https://doi.org/10.1371/journal.pmed.1003583

Pan, Y., & Froese, F. J. (2023). An interdisciplinary review of AI and HRM: Challenges and future directions. *Human Resource Management Review*, *33*(1), Article 100924. https://doi.org/10.1016/j.hrmr.2022.100924

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Park, H., Ahn, D., Hosanagar, K., & Lee, J. (2022). Designing Fair AI in Human Resource Management: Understanding Tensions Surrounding Algorithmic Evaluation and Envisioning Stakeholder-Centered Solutions. *CHI '22: Proceedings of the 22 CHI Conference on Human Factors in Computing Systems*, New Orleans, LA. https://doi.org/10.1145/3491102.3517672

Park, S. H., & Han, K. (2018). Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*, *286*(3), 800–809. https://doi.org/10.1148/radiol.2017171920

Pethig, F., & Kroenung, J. (2022). Biased Humans, (Un)Biased Algorithms? *Journal of Business Ethics*, *183*, 1–16. https://doi.org/10.1007/s10551-022-05071-8

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

Pieters, W. (2011). Explanation and trust: What to tell the user in security and AI? *Ethics and Information Technology*, *13*(1), 53–64. https://doi.org/10.1007/s10676-010-9253-3

Ployhart, R. E., & Harold, C. M. (2004). The applicant attribution-reaction theory (AART): An integrative theory of applicant attributional processing. *International Journal of Selection and Assessment*, *12*(1-2), 84–98. https://doi.org/10.1111/j.0965-075X.2004.00266.x

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Podsakoff, P. M., Podsakoff, N. P., Williams, L. J., Huang, C., & Yang, J. (2024). Common method bias: It's bad, it's complex, it's widespread, and it's not easy to fix. *Annual Review of Organizational Psychology and Organizational Behavior*, *11*, 17–61. https://doi.org/10.1146/annurev-orgpsych-110721-040030

Pomrehn, L., Göttel, V., & Wehner, M. C. (2024). *The bright and dark sides of algorithmic decision-making: The role of stereotypes and psychological distance of ADM in organizational attraction.* Unpublished manuscript.

Pomrehn, L., Schmodde, L., & Wehner, M. C. (2025). *An experimental study on reactions of different stakeholders to artificial intelligence.* Unpublished manuscript.

Pomrehn, L., & Wehner, M. C. (2025). Employee affective reactions to algorithmic management: How does context and algorithm transparency matter? *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*, Kona, HI, USA.

Priesemuth, M., Arnaud, A., & Schminke, M. (2013). Bad behavior in groups: The impact of overall justice climate and functional dependence on counterproductive work behavior in work units. *Group & Organization Management*, *38*(2), 230–257. https://doi.org/10.1177/1059601113479399

Prikshat, V., Malik, A., & Budhwar, P. (2023). AI-augmented HRM: Antecedents, assimilation and multilevel consequences. *Human Resource Management Review*, *33*(1), Article 100860. https://doi.org/10.1016/j.hrmr.2021.100860

Qu, X., Yao, X., & Zhang, W. (2023). How to define a good recruiter: A dual-process model of recruiter effect. *Current Psychology*, *42*(34), 30111–30124. https://doi.org/10.1007/s12144-022-04030-w

R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org

Ravid, D. M., White, J. C., Tomczak, D. L., Miles, A. F., & Behrend, T. S. (2023). A meta-analysis of the effects of electronic performance monitoring on work outcomes. *Personnel Psychology*, *76*(1), 5–40. https://doi.org/10.1111/peps.12514

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. https://doi.org/10.1037/0096-3445.118.3.219

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

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules for artificial intelligence (Artificial Intelligence Act), (2024). https://data.europa.eu/eli/reg/2024/1689/oj

Reicherts, L., Rogers, Y., Capra, L., Wood, E., Duong, T. D., & Sebire, N. (2022). It's good to talk: A comparison of using voice versus screen-based interactions for agent-assisted tasks. *ACM Transactions on Computer-Human Interaction*, *29*(3), 1–41. https://doi.org/10.1145/3484221

Renier, L. A., Mast, M. S., & Bekbergenova, A. (2021). To err is human, not algorithmic – Robust reactions to erring algorithms. *Computers in Human Behavior*, *124*, Article 106879. https://doi.org/10.1016/j.chb.2021.106879

Renieris, E. M., Kiron, D., & Mills, S. (2024). Artificial intelligence disclosures are key to customer trust. *MIT Sloan Management Review*. Retrieved December 02, 2024, from https://sloanreview.mit.edu/article/artificial-intelligence-disclosures-are-key-to-customer-trust/

Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. *Computer Law & Security Review*, *53*, Article 105966. https://doi.org/10.1016/j.clsr.2024.105966

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Roulin, N., Wong, O., Langer, M., & Bourdage, J. S. (2023). Is more always better? How preparation time and re-recording opportunities impact fairness, anxiety, impression management, and performance in asynchronous video interviews. *European Journal of Work and Organizational Psychology*, *32*(3), 333–345. https://doi.org/10.1080/1359432X.2022.2156862

Sanchez, J. I., & Brock, P. (1996). Outcomes of perceived discrimination among Hispanic employees: is diversity management a luxury or a necessity? *Academy of Management Journal*, *39*(3), 704–719. https://doi.org/10.2307/256660

Schlicker, N., Langer, M., Ötting, S. K., Baum, K., König, C. J., & Wallach, D. (2021). What to expect from opening up 'black boxes'? Comparing perceptions of justice between human and automated agents. *Computers in Human Behavior*, *122*, Article 106837. https://doi.org/10.1016/j.chb.2021.106837

Schmidt, F. L., Hunter, J. E., Pearlman, K., Hirsch, H. R., Sackett, P. R., Schmitt, N., Tenopyr, M. L., Kehoe, J., & Zedeck, S. (1985). Forty questions about validity generalization and meta-analysis: Commentary on forty questions about validity generalization and meta-analysis. *Personnel Psychology*, *38*(4), 697–798. https://doi.org/10.1111/j.1744-6570.1985.tb00565.x

Schoeffer, J., De-Arteaga, M., & Kuehl, N. (2024). Explanations, fairness, and appropriate reliance in human-AI decision-making. *CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA. https://doi.org/10.1145/3613904.3642621

Schoeffer, J., Kuehl, N., & Machowski, Y. (2022). "There is not enough information": On the effects of explanations on perceptions of informational fairness and trustworthiness in automated decision-making. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, Seoul, Republic of Korea. https://doi.org/10.1145/3531146.3533218

Schuler, H., & Stehle, W. (1983). Neuere Entwicklungen des Assessment-Center-Ansatzes—beurteilt unter dem Aspekt der sozialen Validität. [New uses for the assessment center approach: Opinions on the aspect of social validity.]. *Psychologie und Praxis*, *27*(1), 33–44.

Schulte Steinberg, A. L. S., & Hohenberger, C. (2023). Can AI close the gender gap in the job market? Individuals' preferences for AI evaluations. *Computers in Human Behavior Reports*, *10*, Article 100287. https://doi.org/10.1016/j.chbr.2023.100287

Searle, R. H. (2022). *Counterproductive Work Behaviors*. Oxford University Press.

Shapiro, D. L., & Kirkman, B. L. (Eds.). (2001). *Anticipatory injustice: The consequences of expecting injustice in the workplace*. Stanford University Press.

Sheridan, T. B. (1995). Human centered automation: oxymoron or common sense? *1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent systems for the 21st century*, Vancouver, BC, Canada.

Shulner-Tal, A., Kuflik, T., & Kliger, D. (2022). Fairness, explainability and in-between: understanding the impact of different explanation methods on non-expert users' perceptions of fairness toward an algorithmic system. *Ethics and Information Technology*, *24*(1). https://doi.org/10.1007/s10676-022-09623-4

Shulner-Tal, A., Kuflik, T., Kliger, D., & Mancini, A. (2024). Who made that decision and why? Users' perceptions of human versus AI decision-making and the power of explainable-AI. *International Journal of Human–Computer Interaction*, *41*(7), 4230–4247. https://doi.org/10.1080/10447318.2024.2348843

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI-Künstliche intelligenz*, *35*, 109–118. https://doi.org/10.1007/s13218-020-00689-0

Snape, E., & Redman, T. (2003). Too old or too young? The impact of perceived age discrimination. *Human Resource Management Journal*, *13*(1), 78–89. https://doi.org/10.1111/j.1748-8583.2003.tb00085.x

Snyder, M., & Swann, W. B. (1978). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of experimental social psychology*, *14*(2), 148–162. https://doi.org/10.1016/0022-1031(78)90021-5

Sondern, D., Arnholz, N., & Hertel, G. (2025). Employment negotiations with an algorithm? How AI as negotiation counterpart would affect negotiators' trust and subjective value expectations. *Conflict Resolution Quarterly.* https://doi.org/10.1002/crq.21472

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, *87*(3), 354–374. https://doi.org/10.2307/1882010

Steel, P., Beugelsdijk, S., & Aguinis, H. (2021). The anatomy of an award-winning meta-analysis: Recommendations for authors, reviewers, and readers of meta-analytic reviews. *Journal of International Business Studies*, *52*(1), 23–44. https://doi.org/10.1057/s41267-020-00385-z

Stockman, S., Van Hoye, G., & da Motta Veiga, S. (2020). Negative word-of-mouth and applicant attraction: The role of employer brand equity. *Journal of Vocational Behavior*, *118*, Article 103368. https://doi.org/10.1016/j.jvb.2019.103368

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of clinical psychology*, *28*(2), 191–193.

Su, Y.-S., Suen, H.-Y., & Hung, K.-E. (2021). Predicting behavioral competencies automatically from facial expressions in real-time video-recorded interviews. *Journal of Real-Time Image Processing*, *18*(4), 1011–1021. https://doi.org/10.1007/s11554-021-01071-5

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

Suen, H.-Y., & Hung, K.-E. (2023). Building trust in automatic video interviews using various AI interfaces: Tangibility, immediacy, and transparency. *Computers in Human Behavior*, *143*, Article 107713. https://doi.org/10.1016/j.chb.2023.107713

Sundar, S. S., & Nass, C. (2000). Source orientation in human–computer interaction: Programmer, networker or independent social actor? *Communication Research*, *27*(6), 683–703. https://doi.org/10.1177/009365000027006001

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, *61*(4), 15–42. https://doi.org/10.1177/0008125619867910

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, *8*, 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

van Esch, P., & Black, J. S. (2019). Factors that influence new generation candidates to engage with and complete digital, AI-enabled recruiting. *Business Horizons*, *62*(6), 729–739. https://doi.org/10.1016/j.bushor.2019.07.004

Vianello, M., Schnabel, K., Sriram, N., & Nosek, B. (2013). Gender differences in implicit and explicit personality traits. *Personality and Individual Differences*, *55*(8), 994–999. https://doi.org/10.1016/j.paid.2013.08.008

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Viswesvaran, C., & Ones, D. S. (2002). Examining the construct of organizational justice: A meta-analytic evaluation of relations with work attitudes and behaviors. *Journal of Business Ethics*, *38*, 193–203. https://doi.org/10.1023/A:1015820708345 von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, *34*(4), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

Votto, A. M., Valecha, R., Najafirad, P., & Rao, H. R. (2021). Artificial intelligence in tactical human resource management: A systematic literature review. *International Journal of Information Management Data Insights*, *1*(2), Article 100047. https://doi.org/10.1016/j.jjimei.2021.100047

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review. *International Journal of Human Resource Management*, *33*(6), 1237–1266. https://doi.org/10.1080/09585192.2020.1871398

Walther, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. In M. L. Knapp & J. A. Daly (Eds.), *The handbook of interpersonal communication* (4th ed., pp. 443–479). Sage.

Wang, G., Guo, Y., Zhang, W., Xie, S., & Chen, Q. (2023). What type of algorithm is perceived as fairer and more acceptable? A comparative analysis of rule-driven versus data-driven algorithmic decision-making in public affairs. *Government Information Quarterly*, *40*(2), Article 101803. https://doi.org/10.1016/j.giq.2023.101803

Wang, R., Harper, F. M., & Zhu, H. (2020). Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA.

Wang, W., & Benbasat, I. (2016). Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents.

Journal of Management Information Systems, *33*(3), 744–775. https://doi.org/10.1080/07421222.2016.1243949

Weber, P., Carl, K. V., & Hinz, O. (2023). Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature. *Management Review Quarterly*, *74*, 867–907. https://doi.org/10.1007/s11301-023-00320-0

Weiner, B. (1985). Attribution Theory. In *Human motivation* (pp. 275–326). Springer New York. https://doi.org/10.1007/978-1-4612-5092-0_7

Wesche, J. S., Hennig, F., Kollhed, C. S., Quade, J., Kluge, S., & Sonderegger, A. (2024). People's reactions to decisions by human vs. algorithmic decision-makers: the role of explanations and type of selection tests. *European Journal of Work and Organizational Psychology*, *33*(2), 146–157. https://doi.org/10.1080/1359432X.2022.2132940

Wiblen, S., & Marler, J. H. (2021). Digitalised talent management and automated talent decisions: The implications for HR professionals. *The International Journal of Human Resource Management*, *32*(12), 2592–2621. https://doi.org/10.1080/09585192.2021.1886149

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., & Hester, J. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Xu, X., Yu, S. J., Pang, N., Dou, C. X., & Li, D. (2022). Review on A big data-based innovative knowledge teaching evaluation system in universities. *Journal of Innovation & Knowledge*, *7*(3), Article 100197. https://doi.org/10.1016/j.jik.2022.100197

Yan, C., Chen, Q., Zhou, X., Dai, X., & Yang, Z. (2024). When the automated fire backfires: The adoption of algorithm-based HR decision-making could induce consumer's unfavorable ethicality inferences of the company. *Journal of Business Ethics*, *190*(4), 841–859. https://doi.org/10.1007/s10551-023-05351-x

You, S., Yang, C. L., & Li, X. (2022). Algorithmic versus human advice: Does presenting prediction performance matter for algorithm appreciation? *Journal of Management Information Systems*, *39*(2), 336–365. https://doi.org/10.1080/07421222.2022.2063553

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology*, *32*, 661–683. https://doi.org/10.1007/s13347-018-0330-6

Zhang, L., & Amos, C. (2023). Dignity and use of algorithm in performance evaluation. *Behaviour & Information Technology*, *43*(2), 401–418. https://doi.org/10.1080/0144929X.2022.2164214

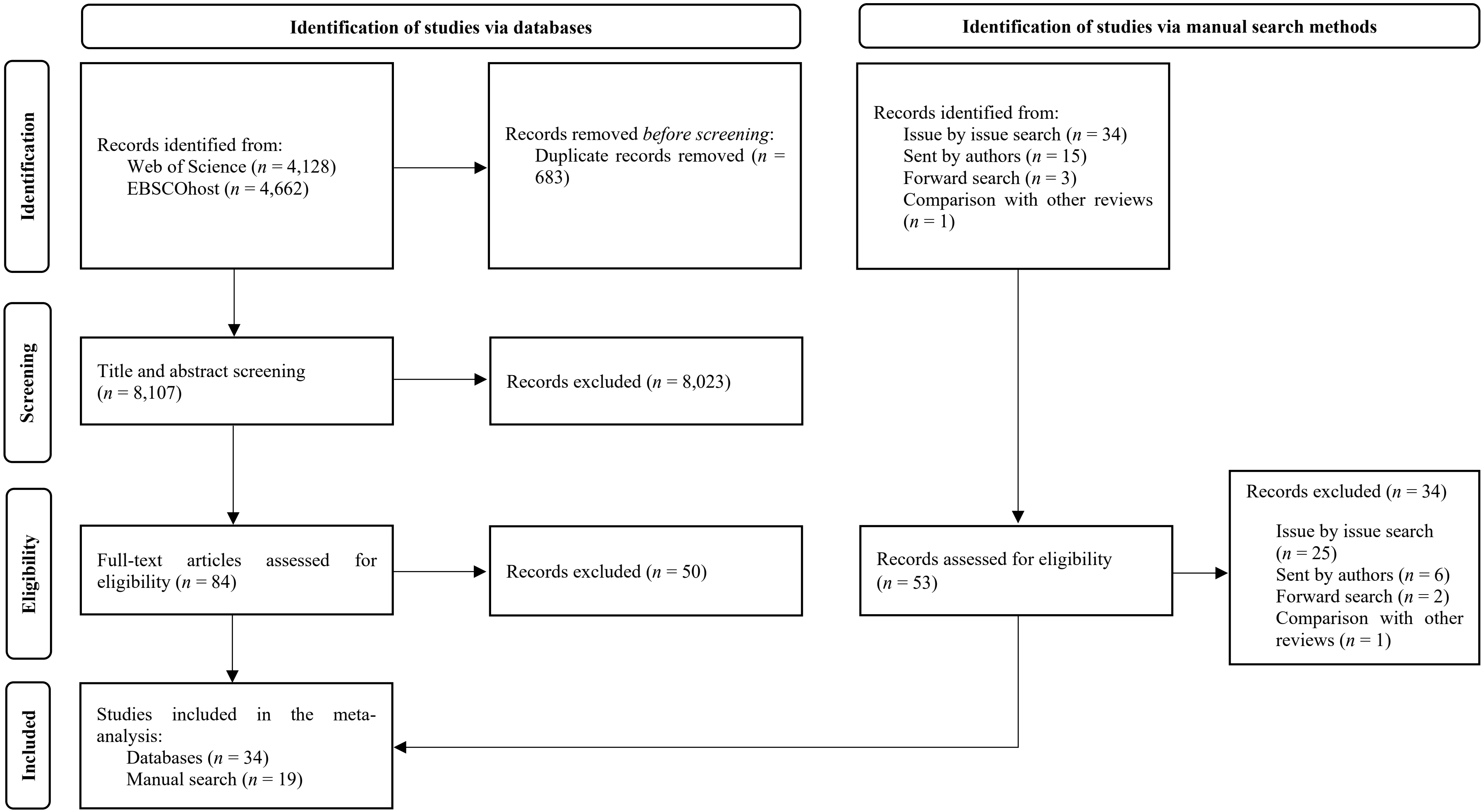
Zhang, P. (2013). The affective response model: A theoretical framework of affective concepts and their relationships in the ICT context. *Management Information Systems Quarterly*, *37*(1), 247–274. https://www.jstor.org/stable/43825945

Zheng, Y., & He, Y. (2024). The audience perception to AI voicebot news: An experimental analysis. *ICEMT '24 Proceedings of the 2024 8th International Conference on Education and Multimedia Technology,* Tokyo, Japan. https://doi.org/10.1145/3678726.3678764

Zuckerman, M. (1979). Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory. *Journal of personality*, *47*(2), 245–287. https://doi.org/10.1111/j.1467-6494.1979.tb00202.x

8 Appendix

8.1 Appendix A1 (Essay I) - Flow diagram



Note. The 53 studies included a total of 73 samples.

8.2 Appendix A2 (Essay I) - Overview of included studies

Author(s), year	Journal	Sample size	Type of stakeholder	Type of interaction	Type of task	Extent of decision	Personal impact	Random-ization
Acikgoz et al. (2020)	International Journal of Selection and Assessment	Study 1: 298, study 2: 225	Applicants	Confrontation	Interviews	Decision support	High	Yes
Acikgoz et al. (2024)	Unpublished	Interview sample: 276, Screening sample: 274	Applicants	Confrontation	Interview sample: interviews, screening sample: screening	Sole decision entity	High	Yes
Bankins et al. (2022)	Information Systems Frontiers	446	Employees	Confrontation	Career development	Sole decision entity	High	Yes
Bedemariam and Wessel (2023)	Computers in Human Behavior	309	Applicants	Confrontation	Recruiting	Sole decision entity	High	Yes
Canagasuriam and Lukacik (2024)	International Journal of Selection and Assessment	245	Applicants	Collaboration	Interviews	Decision support	High	Yes
Chacon et al. (2024)	Unpublished	336	Applicants	Confrontation	Recruiting	Decision support	High	Yes
Choi and Chao (2024)	Personality and Social Psychology Bulletin	Study 1: 177, study 2: 421, study 3a: 248, study 3b: 638, study 5: 1,092	Employees	Confrontation	Conflict resolution	Sole decision entity	High	Yes
Choung et al. (2023)	International Journal of Human-Computer Interaction	235	Applicants	Collaboration	Screening	Decision support	Low	Yes
Deriu et al. (2024)	Journal of Business Research	108	Applicants	Both	Interviews	Sole decision entity	High	Yes
Dong et al. (2024)	Technovation	504	Employees	Confrontation	Career development	Sole decision entity	High	Yes
Dutta and Mishra (2024)	Personnel Review	1,179	Employees	Collaboration	Career development	Decision support	High	No
Feldkamp et al. (2023)	European Journal of Work and Organizational Psychology	215	HR professionals	Collaboration	Screening	Decision support	Low	Yes

Author(s), year	Journal	Sample size	Type of stakeholder	Type of interaction	Type of task	Extent of decision	Personal impact	Random-ization
Fumagalli et al. (2022)	Research Policy	1,725	Applicants	Confrontation	Recruiting	Sole decision entity	High	Yes
Gonzalez et al. (2019)	Personnel Assessment and Decisions	192	Applicants	Confrontation	Recruiting	Sole decision entity	High	No informa-tion
Gonzalez et al. (2022)	Computers in Human Behavior	Study 1: 183, study 2: 163	Applicants	Confrontation	Recruiting	-	High	Yes
Höddinghaus et al. (2021)	Computers in Human Behavior	333	Employees	Confrontation	Career development	Sole decision entity	High	Yes
Kares et al. (2023)	International Journal of Selection and Assessment	Study 1: 170, study 2: 154	HR professionals	Collaboration	Study 1: screening, study 2: Career development	Decision support	Low	Yes
Keppeler (2023)	Journal of Public Administration Research and Theory	415 (reduced *N* for our included correlations)	Applicants	Confrontation	Sourcing	-	High	Yes
Keppeler et al. (2024)	Unpublished	Study 2: 538	HR professionals	Collaboration	Screening	Decision support	Low	Yes
Kleinlogel et al. (2023)	International Journal of Selection and Assessment	Study 1: 151, study 2: 148[12]	Applicants	Confrontation	Interviews	Decision support	High	Yes
Koch-Bayram et al. (2023)	International Journal of Selection and Assessment	Study 1: 209, study 2: 302	Applicants	Confrontation	Interviews	Sole decision entity	High	Yes
Koch-Bayram and Kaibel (2023)	Human Resource Management Journal	Study 1: 259, study 2: 342	Applicants	Confrontation	Screening	Sole decision entity	Study 1: low, study 2: high	Yes
Köchling and Wehner (2023)	International Journal of Selection and Assessment	200	Applicants	Confrontation	Recruiting	Decision support	High	Yes
Köchling et al. (2023)	Review of Managerial Science	160	Applicants	Confrontation	Recruiting	Decision support	High	Yes
Köchling et al. (2024)	Review of Managerial Science	280	Employees	Confrontation	Career development	Decision support	High	Yes

[12] We operationalized the study by Kleinlogel et al. (2023) as two studies due to the given data basis.

Author(s), year	Journal	Sample size	Type of stakeholder	Type of interaction	Type of task	Extent of decision	Personal impact	Random-ization
Lacroux and Martin-Lacroux (2022)	Journal of Business and Psychology	694	HR professionals	Collaboration	Screening	Decision support	Low	Yes
Langer , König and Hemsing (2019)	Journal of Managerial Psychology	124	Applicants	Confrontation	Interviews	Sole decision entity	High	Yes
Langer, König and Papathanasiou (2019)	International Journal of Selection and Assessment	123	Applicants	Confrontation	Interviews	-	Low	Yes
Langer, König, Sanchez, et al. (2019)	Journal of Managerial Psychology	148	Applicants	Confrontation	Interviews	-	High	Yes
Langer et al. (2023)	Journal of Business and Psychology	121	HR professionals	Collaboration	Screening	Decision support	Low	Yes
Lavanchy et al. (2023)	Journal of Business Ethics	Study 1: 249, study 2: 272, study 3: 282, study 4: 270	Applicants	Confrontation	Screening	Sole decision entity	High	Yes
Liu et al. (2023)	Journal of Computer-Mediated Communication	134	Applicants	Collaboration	Interviews	Sole decision entity	High	Yes
Luo and Zhang (2023)	Academy of Management Proceedings	302	Employees	Confrontation	Interviews	Sole decision entity	High	Yes
Maasland and Weissmueller (2022)	Frontiers in Psychology	156	HR professionals	Collaboration	Administrative	Decision support	Low	Yes
Mirowska (2020)	Journal of Personnel Psychology	184	Applicants	Confrontation	Interviews	Sole decision entity	High	Yes
Moritz and Schmidt (2024)	European Conference on Information Systems Proceedings	173	Employees	Confrontation	Management tasks	Sole decision entity	High	Yes
Moritz et al. (2024)	Academy of Management Proceedings	416	Applicants	Confrontation	Recruiting	Sole decision entity	High	Yes
Moritz et al. (2025)	Academy of Management Proceedings	252	Applicants	Confrontation	Recruiting	Decision support	High	Yes

Author(s), year	Journal	Sample size	Type of stakeholder	Type of interaction	Type of task	Extent of decision	Personal impact	Random-ization
Nagtegaal (2021)	Government Information Quarterly	Study 1: 109, study 2: 126	Employees	Confrontation	Study 1: administrative & career development, study 2: administrative & recruiting	Sole decision entity	Low	Yes
Noble et al. (2021)	International Journal of Selection and Assessment	360	Applicants	Confrontation	Screening	Sole decision entity	High	Yes
Oostrom et al. (2023)	Journal of Occupational and Organizational Psychology	Study 1: 172, study 2: 276	Applicants	Confrontation	Recruiting	Sole decision entity	High	No information
Pomrehn, et al. (2024)	Unpublished	405	Applicants	Confrontation	Recruiting	Decision support	High	Yes
Pomrehn et al. (2025)	Unpublished	HR professionals sample: 205, employees sample: 202	HR professionals sample: HR professionals, Employees sample: employees	HR professionals sample: collaboration, employees sample: confrontation	Recruiting, Layoff	Decision support	HR professsionals sample: low, employees sample: high	Yes
Pomrehn and Wehner (2025)	Hawaii International Conference on System Sciences (HICSS)	190	Employees	Confrontation	Career development	Sole decision entity	High	Yes
Renier et al. (2021)	Computers in Human Behavior	439	Applicants	Confrontation	Recruiting	Sole decision entity	Low	Yes
Schlicker et al. (2021)	Computers in Human Behavior	209	Employees	Confrontation	Administrative	Sole decision entity	High	Yes
Shulner-Tal et al. (2024)	International Journal of Human-Computer Interaction	3,068	HR professionals	Collaboration	Screening	Sole decision entity	Low	Yes
Sondern et al. (2025)	Unpublished	291	Applicants	Collaboration	Negotiation	Sole decision entity	High	No information
Suen and Hung (2023)	Computers in Human Behavior	152	Applicants	Confrontation	Interviews	Sole decision entity	High	Yes

Author(s), year	Journal	Sample size	Type of stakeholder	Type of interaction	Type of task	Extent of decision	Personal impact	Random-ization
Suen et al. (2019)	Computers in Human Behavior	180	Employees	Confrontation	Interviews	Sole decision entity	High	Yes
Wesche et al. (2024)	European Journal of Work and Organizational Psychology	270	Employees	Confrontation	Career development	Sole decision entity	High	Yes
Yan et al. (2024)	Journal of Business Ethics	Study 1: 192, study 2: 206	Employees	Confrontation	Layoff	Sole decision entity	Low	Yes
Zhang and Amos (2023)	Behavior & Information Technology	95	Employees	Confrontation	Career development	Sole decision entity	Low	Yes

Note. Recruiting means that the study included screening as well as interviews in the hypothetical scenario. A high personal impact means that participants were directly affected by the decision, a low personal impact means that others were affected by the decision.

8.3 Appendix B1 (Essay II) - Exemplary survey flow for the HITL condition

Introduction (for all conditions):

In this survey, you will participate in a digital assessment center and complete a series of assessment center tasks. Below you will find a brief description of the situation. Read through the description and try to imagine yourself in the situation as best you can. Following the assessment center, you will be asked to evaluate your experience with the digital assessment center, so please pay attention to the details.

Start of the digital assessment center:

You are currently in the application phase for a trainee position. You have applied to Marzeo AG for this position. You are already aware that the application process consists of two parts and that today the first part will take place in form of a digital assessment center. With the digital assessment center, Marzeo AG is testing a new online approach that enables the assessment process to be conducted digitally. For this purpose, a connection to the digital assessment center will be established and you will be connected online with the artificial intelligence Vesa that will accompany you during the tasks. In the assessment center, you are asked to solve various tasks, and you will receive feedback on your performance from the artificial intelligence Vesa after each task. Your performance will be evaluated in comparison to other participants.

At the end of the process, the artificial intelligence Vesa evaluates your performance, after which a representative from Marzeo AG's HR department reviews the results and determines whether you advance past the first round of selection.

"Connecting" to the digital assessment center:

V

Welcome to the assessment center and thank you for your interest in a trainee position at Marzeo AG.

V

I am the artificial intelligence Vesa and today I will guide you through the digital assessment center. In the following, you will be given various tasks that you should answer as quickly and correctly as possible. I will give you feedback after each task and evaluate your performance at the end of the assessment center. Afterwards, a person from Marzeo AG's HR department will check and decide whether you pass the first round of this selection procedure.

Task completion:

Please complete the following series of numbers: 150 100 50 0 _

100

50

-50

-100

Feedback:

V

Vesa is answering right now...

V

You were able to solve this quickly.

Participants completed 24 similar tasks and received feedback after each task.

Final decision:

V

Thank you for participating in the Assessment Center.
I have evaluated your tasks and a person from Marzeo AG's HR department has decided, based on your performance, that you passed the first round of the selection procedure. Congratulations!

8.4 Appendix B2 (Essay II) - Exemplary survey flow for the ADM condition

Introduction (for all conditions):

In this survey, you will participate in a digital assessment center and complete a series of assessment center tasks. Below you will find a brief description of the situation. Read through the description and try to imagine yourself in the situation as best you can. Following the assessment center, you will be asked to evaluate your experience with the digital assessment center, so please pay attention to the details.

Start of the digital assessment center:

You are currently in the application phase for a trainee position. You have applied to Marzeo AG for this position. You are already aware that the application process consists of two parts and that today the first part will take place in form of a digital assessment center. With the digital assessment center, Marzeo AG is testing a new online approach that enables the assessment process to be conducted digitally. For this purpose, a connection to the digital assessment center will be established and you will be connected online with the artificial intelligence Vesa that will accompany you during the tasks. In the assessment center, you are asked to solve various tasks, and you will receive feedback on your performance from the artificial intelligence Vesa after each task. Your performance will be evaluated in comparison to other participants.

At the end, the artificial intelligence Vesa evaluates your performance and determines whether you advance past the first round of selection.

"Connecting" to the digital assessment center:

V

Welcome to the assessment center and thank you for your interest in a trainee position at Marzeo AG.

V

I am the artificial intelligence Vesa and today I will guide you through the digital assessment center. In the following, you will be given various tasks that you should answer as quickly and correctly as possible. I will give you feedback after each task and evaluate your performance at the end of the assessment center.

Task completion:

Please complete the following series of numbers: 150 100 50 0 _

100

50

-50

-100

Feedback:

V

Vesa is answering right now...

V

You were able to solve this quickly.

Participants completed 24 similar tasks and received feedback after each task.

Final decision:

V

Thank you for participating in the Assessment Center.
I have evaluated your tasks.

For the evaluation, I considered the following aspects:
– speed of response
– accuracy of response
– number of positions available in the company
– your performance in relation to all other applicants

Based on your performance, I have decided, that you did not pass the first round of the selection procedure.

8.5 Appendix C1 (Essay III) - Scenario descriptions

All participants were introduced to the scenarios with the following text:

The following is a brief description of a situation. Please read the following description of the situation carefully and answer the questions as intuitively as possible.

This situation is independent of your current workplace, and we ask you to read the description carefully and try to imagine yourself in the situation as best you can. Questions regarding this situation will follow, so please pay attention to the details. Until further notice, all of the following questions should be answered based on the described situation.

8.6 Appendix C2 (Essay III) - Scenario descriptions for Study 1 and Study 2

Human management condition

You have been working at Marzeo AG for several years in the marketing department, handling a variety of tasks. You are very happy with your job and enjoy going to work every day.

At Marzeo AG, it is standard practice for every employee to have a **feedback meeting** with their manager at least twice per year. These meetings are intended to evaluate past performance, set goals, and identify potential areas for development. Your feedback meeting takes place in person with your manager. During the meeting, your **manager** shares their assessment of your performance with you.

You receive both praise and suggestions for development.

Algorithmic management condition

You have been working at Marzeo AG for several years in the marketing department, handling a variety of tasks. You are very happy with your job and enjoy going to work every day.

At Marzeo AG, it is standard practice for every employee to have a feedback meeting with their supervisor at least twice per year. These meetings are intended to evaluate past performance, set goals, and identify potential areas for development. In your case, the assessment of your performance and the feedback conversation are conducted by an **algorithm**.

You receive both praise and suggestions for development.

8.7 Appendix C2 (Essay III) - Scenario descriptions for Study 3

Human management condition

You have been working at Marzeo AG for several years in the marketing department, handling a variety of tasks. You are very happy with your job and enjoy going to work every day.

Each week begins with a team meeting on Monday in which you and your colleagues discuss upcoming tasks. During the meeting, your **manager** decides who will take on which responsibilities. Next week, you are assigned to support the development of a new marketing campaign.

Additionally, Marzeo AG ensures that every employee has at least two feedback meetings per year with their **manager**. These meetings are conducted in person and focus on evaluating performance, setting goals, and identifying potential areas for growth. In your recent feedback meeting, your **manager** shared an assessment of your work, providing both praise and constructive suggestions for improvement.

Algorithmic management condition

You have been working at Marzeo AG for several years in the marketing department, handling a variety of tasks. You are very happy with your job and enjoy going to work every day.

Each week begins with a team meeting on Monday in which you and your colleagues discuss upcoming tasks. **An algorithmic management system** determines who will take on which responsibilities. **The system** assigns you to support the development of a new marketing campaign next week.

Additionally, Marzeo AG ensures that every employee receives regular feedback on their work. Therefore, the **algorithmic management system** provides you with automated performance evaluations twice a year. The feedback is focused on evaluating performance, setting goals, and identifying potential areas for growth. In your recent performance evaluation, the **algorithmic management system** shared its assessment of your work, providing both praise and constructive suggestions for improvement.

Experimental discrimination condition (algorithmic management condition in parentheses)

Individuals in the discrimination condition received the following additional information:

Later this week, you read a report that reveals troubling patterns in the **management practices (the algorithmic management system)** at Marzeo AG, including the **manager (management system)** that manages you and your team. The report shows that employees from certain demographic groups tend to receive more prestigious and challenging assignments.

You notice that, despite your excellent track record, you are consistently assigned less challenging tasks compared to others, and this pattern aligns with the findings in the report. You begin to wonder whether the decisions of your **manager (the algorithmic management system)** are biased.