Understanding Algorithmic Fairness: An Investigation of How Stakeholders Construct Algorithmic Fairness in the HR Context

Stefanie Knippschild

Business Information Systems

A thesis submitted to fulfil the requirements of the degree of Doctor of Philosophy

The University of Sydney

The University of Sydney Business School

2026

Statement of Originality

This is to certify that the content of this thesis is my own work. This thesis has not been submitted for any other degree or purpose.

I certify that the intellectual content of this thesis is the product of my own work, and that all assistance received in preparing this thesis and all sources have been acknowledged.

The research involving human data reported in the thesis was assessed and approved by the University of Sydney Human Research ethics committee. Approvals are as follows:

- 2023/HE000862

Stefanie Knippschild

Authorship Attribution Statement

This thesis contains material that is published, under review or prepared for submission.

Chapter 4 of this thesis is published as:

"**Knippschild, S., Boell, S., Riemer, K. and Peter, Sandra** (2025) '*Finding algorithmic fairness: An analysis of how the literature constructs algorithmic fairness from different stakeholder perspectives*' ACIS 2024 Proceedings"

- I led the study, the analysis of the literature, and co-wrote drafts of the manuscript. Authorship is by contribution. All authors contributed significantly to this work.

Chapter 5 of this thesis is published as:

"**Knippschild, S., Boell, S., Riemer, K. and Peter, S.** (2023) '*Fairness for hire: An investigation of how people analytics vendors construct fairness in the employment lifecycle*' ACIS 2023 Proceedings"

- I led the study, collected and analysed data and co-wrote drafts of the manuscript. Authorship is by contribution. All authors contributed significantly to this work.

Chapter 6 of this thesis is published as:

"**Knippschild, S., Riemer, K., Boell, S. and Peter, S.** (2025) '*Making fairness work: How stakeholders developing, implementing and using algorithmic decision-making systems construct algorithmic fairness in the HR context*'" ACIS 2025 Proceedings.

- I led the study, collected and analysed data and co-wrote drafts of the manuscript. Authorship is by contribution. All authors contributed significantly to this work.

Chapter 7 of this thesis is published as:

"**Knippschild, S., Riemer, K., Boell, S. and Peter, S.** '*Too artificial for HR? Understanding HR professionals' concerns about the adoption of algorithmic decision-making systems*'" Submitted to the Australasian Conference on Information Systems (ACIS) 2025.

- I led the study, collected and analysed data and co-wrote drafts of the manuscript. Authorship is by contribution. All authors contributed significantly to this work.

I am the corresponding author for the publications above. I can attest that permission to include the published and other material has been granted by all authors.

This work was supported by the WNS Global BIS Scholarship.

Student name: Stefanie Knippschild

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Supervisor name: Kai Riemer

Artificial Intelligence Acknowledgement

During the preparation of the thesis, I used ChatGPT for the purposes of language enhancement. The use of this generative AI tool includes paraphrasing, sentence structure, and spelling.

I confirm that where text was modified by generative AI, the content was reviewed for possible errors, inaccuracies, and bias. I take full responsibility for the submitted thesis and ensure the work is my own and have used generative AI within the parameters of use (refer to the University of Sydney generative AI guide for researchers).

Abstract

Algorithmic fairness has received growing scholarly and public attention in recent years, as numerous examples demonstrate that algorithmic decision-making (ADM) systems, which were originally intended to mitigate human biases, can instead introduce new forms of unfair treatment. In this thesis, I respond to calls to approach fairness from a sociotechnical perspective by examining how algorithmic fairness is constructed by stakeholders involved in developing, implementing, and using ADM systems in the Human Resources (HR) context. The research consists of four interconnected studies, each presented as a separate manuscript.

In the first manuscript, I conduct an organising review of the algorithmic fairness literature, analysing which stakeholders are considered in the algorithmic discourse and how algorithmic fairness is conceptualised from different stakeholder perspectives taken in the literature. The review finds that most scholarship focuses on perceptions of decision-subjects, primarily to understand how developers and organisational users construct algorithmic fairness. This finding reveals how notions of algorithmic fairness vary according to the stakeholder lens applied.

The second manuscript employs a thematic analysis of software supplier, particularly people analytics websites, to explore how algorithmic fairness is constructed publicly. The findings show that algorithmic fairness is acknowledged but often remains undefined or decontextualised, indicating limited transparency. The findings also highlight that algorithmic fairness concerns are concentrated around high-impact HR processes, with partial coverage potentially resulting in incomplete considerations across different HR operations.

In the third manuscript, I conduct interviews with HR professionals, software developers, people analysts, and Artificial Intelligence (AI) and HR consultants, to examine stakeholder understandings of algorithmic fairness and the factors shaping those

understandings. The analysis shows that algorithmic fairness is constructed through both social and technical dimensions, underscoring the need for a sociotechnical approach. While some algorithmic fairness concepts are consistently valued across stakeholder groups, divergent and sometimes conflicting interpretations also emerge.

The fourth manuscript is guided by the research constraints I encountered, namely that HR professionals were not yet using ADM in practice. To address this constraint, I conduct interviews to investigate these concerns and the underlying factors shaping them, thereby identifying critical barriers that must be addressed for effective ADM system implementation. These findings further illustrate the interplay between algorithmic fairness and real-world adoption, showing that algorithmic fairness influences HR decision-making regarding ADM, which provides an understanding of how concerns about algorithmic fairness affect technology integration.

Overall, this thesis argues that algorithmic fairness cannot be addressed through a one-size-fits-all approach. Instead, conflicting stakeholder perspectives must be recognised and managed. By incorporating stakeholders often overlooked in prior research, this thesis provides a comprehensive analysis of how algorithmic fairness is constructed in practice, highlighting where understandings align, where they conflict, and where concerns about algorithmic fairness, alongside broader concerns about the use of ADM systems, may shape reasons for reluctance towards adoption.

Acknowledgements

This thesis would not have been possible without the support, guidance, and encouragement of so many people, and I am deeply grateful to each of you.

First and foremost, I would like to thank my supervisors, Kai Riemer, Sebastian Boel, and Sandra Peter. Your guidance, constructive feedback, and support have shaped my academic journey and enabled me to complete this thesis. I am sincerely grateful for your time, patience, and commitment.

To my mum, without you none of this would have been possible. I remember our walk with Lilly, when I first mentioned the idea of pursuing a PhD in Sydney, without imagining it would actually become reality. Your words that you would always support me, no matter what, gave me the courage to apply. I am truly grateful for your love and encouragement, even though it meant accepting the risk that I might stay in Australia for longer. This gratitude also extends to the rest of my family for their continuous support.

I also owe a huge thank you to the wonderful PhD community at the University of Sydney, especially within the Business Information Systems discipline. I have been fortunate to find not only colleagues but also friends.

Ben Breeds, your tireless efforts to help me, often under difficult circumstances, has meant more than I can say. Dirk Horkova, our many conversations while I was sitting at the desk in front of your office have left a lasting positive impression. Your wisdom has profoundly shaped both my worldview and research ideas. I hope we will continue working together in the future. Rafaelle Ciriello, thank you for your valuable feedback and for always keeping your office door open I am grateful that our conversations have grown into a meaningful friendship.

Virginia Springer, your positivity and encouragement were a constant source of motivation, and I am so thankful that our paths crossed. You made life at, and beyond, the University so much better. I do not know what I would have done without you. To my "HDR Ladies Who Lunch", thank you for the laughter, the honesty, and the reassurance that I was not alone. Each of you made this journey more enjoyable. Lisa Asher, I am grateful that we went from Zoom classes to actually talking to each other in person. You are a true friend, and I am grateful to have shared this journey with you. Emily Boxall, meeting you at ACIS 2024 was one of the best things that happened. Thank you incredibly much for spending countless weekends together on campus towards the end of this journey. I cannot wait for more adventures outside uni. Caitlin Gardner, sharing similar struggles brought us closer together and even to spending some time together in Germany and at your wedding. I am grateful for the friendship that grew from that.

To the most important person at the University, Rick Sullivan (also known as the "unofficial supervisor of all PhD students"). I truly cannot imagine how I would have managed all of this without your wisdom, support, and humour. The countless conversations and shared laughter have profoundly shaped both my research and my personal growth. You were always there for me, not matter what. I am beyond grateful to consider you not only a friend but also my brother.

Outside the PhD bubble, I am thankful for everyone who reminded me there is life outside of research. Anouk, Michelle, Dan, Martina, and Juraj (together with lovely little Lara), thank you all for your kindness, encouragement, and for grounding me throughout this journey. To Chris, I will never forget that first coffee where we both talked about moving to Sydney. Look at us now! I am truly grateful for your support and always believing in me. To Paul, thank you for giving me a home when I did not have one. I do not know what I would have done without you helping me through this difficult situation, especially because it came

at the worst possible moment while I was trying to finish this thesis. To Vivien, meeting you and becoming close friends has been one of the best things that happened during this time. Even though it has not been that long, your constant encouragement, kindness, and support carried me through the final months. I know I can always count on you.

Finally, if I have missed anyone, please know that I am just as grateful for your encouragement and support.

Note: Copy Editing was provided for Chapters 1, 2, 3, and 8 as well as the links to thesis narrative by the professional online thesis editor Dr. David Harris at writewellaustralia.com.au. This service was used for copy editing only.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Algorithmic Fairness has emerged as a central concern in the use of algorithmic decision-making (ADM) within HR processes, such as hiring (Köchling & Wehner, 2020), performance assessment (Gal et al., 2019; Giermindl et al., 2022), employee monitoring, and work scheduling (Parent-Rocheleau & Parker, 2022). Advocates of ADM systems highlight their potential to mitigate human biases and enhance efficiency and scalability compared to human decision-making (Köchling & Wehner, 2020). This potential is particularly emphasised in contexts subject to discrimination based on gender, race, or age (Demuijnck, 2009; Triana et al., 2021; Yeung et al., 2021), where reliance on human emotions, intuitions, and prejudices increases the likelihood of biased, incorrect, and unfair outcomes (Lee, 2018; Rivera, 2012). Nevertheless, examples from practice have shown that ADM systems are not free from bias. A widely cited example is Amazon's hiring algorithm, which systematically disadvantaged female candidates by excluding resumes referencing women because the system had been trained on imbalanced data biased towards men (Dastin, 2018). This example shows that algorithmic systems are far from being neutral and can embed biases from their training data, design choices, or deployment contexts (Kordzadeh & Ghasemaghaei, 2022), thereby replicating and exacerbating historical inequalities (Barocas & Selbst, 2016).

In HR, such biases are particularly consequential because discriminatory outcomes affect individuals' livelihoods, hinder workforce diversity (Hunkenschroer & Luetge, 2022), and undermine trust in the decision quality (Lee, 2018). These concerns have further contributed to scepticism among potential users, such as recruiters, who are increasingly attentive to the ethical implications of ADM systems and, in some cases, exhibit an aversion to adopting them in practice (Mahmud et a., 2022).

In this context, the call across different disciplines, such as computer sciences and Information Systems (IS), for the advancement of research on algorithmic fairness, has prompted extensive debate on how algorithmic fairness should be conceptualised, defined, and operationalised (Dolata et al., 2022). Current research reflects three main perspectives. The *technical perspective* defines fairness through mathematical notions such as statistical parity, disparate impact, or equalised odds, often with limited attention to social context (Barocas & Selbst, 2016; Dolata et al., 2022). The *social perspective* draws on various social notions of fairness, including the organisational justice theory or ethical reasoning, and frames algorithmic fairness in terms of equality or equity (Green, 2022; Holm, 2023; Veale et al., 2018). However, both perspectives are limited. The technical perspective neglects social aspects, such as human perception or social contexts, while the social perspective overlooks the roles of formalised mathematical notions of fairness. Consequently, scholars increasingly advocate for a *sociotechnical perspective*, which acknowledges that decision-making systems operate at the intersection of technology and society and that fairness emerges through the joint optimisation of technical design and human interpretation (Dolata et al., 2022; Holstein et al., 2019; Selbst et al., 2019).

Research on algorithmic fairness in HR remains limited, with most studies focusing on decision-subjects' perceptions, such as those of job applicants (Lavanchy et al., 2023) or frontline workers (Zhou et al., 2023), with only a few studies focusing on other stakeholders, like software developers (Kasinidou et al., 2021; Kleanthous et al., 2022) or HR professionals (Feldkamp et al., 2023). While these stakeholders may hold different or conflicting views of what constitutes a fair decision, existing research rarely examines how these perspectives interact. Moreover, there are very few research explorations of how algorithmic fairness is constructed, enacted, and contested in organisational practice, or how technical and social dimensions intersect.

And while use of ADM systems in HR is widely discussed, their adoption in practice remains limited. Existing explanations often draw on the concept of algorithm aversion, which refers to conscious or unconscious rejection of algorithmic decisions in favour of human judgment (Mahmud et al., 2022) and encompasses a general predisposition against algorithms even prior to any direct interaction with them (Jussupow et al., 2024). While prior research has identified a range of individual and organisational factors associated with adoption concerns, such as concerns about lack of fairness or transparency (Ochmann et al., 2024), it remains unclear whether such reluctance reflects a fundamental rejection of ADM systems or a more provisional state of *not-yet-adoption*. Recent developments in the diffusion of generative AI tools, such as ChatGPT, provide a point of comparison. These systems have rapidly transitioned from limited use to widespread adoption in both personal and professional contexts, suggesting that increased accessibility, usability, and familiarity can significantly influence user acceptance (Agrawal, 2024; Bick et al., 2025). This raises the possibility that the limited uptake of ADM systems in HR may not be solely attributable to inherent resistance, but rather to insufficient exposure, perceived relevance, or demonstrated value. Consequently, understanding whether non-adoption reflects persistent concerns or a transitional phase remains a critical area for further investigation.

This thesis addresses these research gaps by investigating algorithmic fairness in the HR context through a sociotechnical lens (Dolata et al., 2022). In particular, it examines how different stakeholders, who design, implement, and are subject to ADM systems, construct algorithmic fairness. In doing so, this thesis addresses the following central research question:

> *"How do stakeholders designing, implementing, and using algorithmic decision-making systems construct algorithmic fairness in the organisational HR context?"*

To answer this question, I conducted a series of studies, each presented in the form of research manuscripts. In the course of my study, it became obvious that the construction of fairness is both materially influenced by a stakeholder's experience with algorithmic technologies and can in turn impact decisions to adopt algorithmic technologies (or not). As a result, across the course of this thesis I examine a set of additional research questions as shown in Table 1.1. As this thesis follows a publication-based format, some degree of overlap across chapters is inherent and should be considered when reading each manuscript.

Table 1.1: Overview of Research Questions

Chapter	Research Question	Research Question
Chapter 1: Introduction	*RQ 1*	*"How do stakeholders designing, implementing, and using algorithmic decision-making systems construct algorithmic fairness in the organisational HR context?"*
Chapter 4: Manuscript 1	*RQ 2*	*"What stakeholder groups are considered in existing research?"*
Chapter 4: Manuscript 1	*RQ 3*	*"How does existing research construct algorithmic fairness from different stakeholder perspectives in the organisational context?"*
Chapter 5: Manuscript 2	*RQ 4*	*How is fairness implicitly or explicitly constructed by People Analytics vendors in practice?"*
Chapter 5: Manuscript 2	*RQ 5*	*"How are the employment lifecycle phases impacted by fairness?"*
Chapter 6: Manuscript 3	*RQ 6*	*"What factors are considered in the construction of algorithmic fairness by stakeholders developing, implementing, and using algorithmic decision-making systems in the organisational HR context?"*
Chapter 6: Manuscript 3	*RQ 7*	*"How do these factors construct different understandings of algorithmic fairness?"*
Chapter 7:	*RQ 8*	*"What are concerns about the adoption of algorithmic*

Chapter	Research Question	Research Question
Manuscript 4		*decision-making systems by HR professionals?"*
Chapter 7: Manuscript 4	*RQ 9*	*"How are those concerns shaped?"*

To answer the overarching RQ 1, the first manuscript entails an organising review of the literature on algorithmic fairness (Leidner, 2016) to synthesise existing knowledge and establish a foundation for stakeholder selection for subsequent studies. Building on this, I conduct a thematic analysis (Braun & Clarke, 2006) of software suppliers' websites to examine how algorithmic fairness is publicly constructed and which HR processes are most frequently associated with algorithmic fairness. These insights inform the selection of interview partners across different HR operations, such as recruitment and selection. I then conduct the first interview study to investigate how stakeholders, who are involved in the development, implementation, and use of ADM systems, construct and interpret algorithmic fairness, and the factors shaping these constructions and interpretations. This study provides the basis for examining the not-yet-adoption of ADM systems in HR. As outlined in Chapter section 2.4, the recruitment process and early interviews revealed that such systems are rarely applied in practice. This revelation prompted a shift in research focus towards understanding the reasons for current non-adoption, despite the widespread enthusiasm for these systems in theory, and linking these reasons back to broader concerns around algorithmic fairness.

The findings show that algorithmic fairness is not a single, clearly defined concept, but rather constructed in fragmented and often inconsistent ways (Dolata et al., 2022; Kleanthous et al., 2022; Mulligan et al., 2019; Starke et al., 2022). Across all studies, stakeholders frequently encountered challenges to articulate what fairness means in the context of ADM systems, often relying on related concepts such as transparency, objectivity, explainability, non-discrimination, or bias mitigation as proxies (Draude et al., 2020; Starke et al., 2022).

These constructions of fairness were strongly shaped by stakeholders' roles and expertise. Technically skilled stakeholders, including software developers and consultants, tended to adopt a sociotechnical perspective, integrating both technical measures of fairness and contextual or social considerations (Dolata et al., 2022; Draude et al., 2020). In contrast, HR professionals, particularly those with limited technical backgrounds, primarily framed fairness in terms of non-discrimination, equality of opportunity, and alignment with existing HR practices and legal frameworks (Feldkamp et al., 2023). Similar patterns were observed in the literature, which is often divided between technical approaches from computer science (e.g., Friedler et al., 2019; Verma & Rubin, 2018) and socially grounded approaches such as organisational justice theory (Juijn et al., 2023; Narayanan et al., 2024), with few studies integrating both perspectives (Dolata & Schwabe, 2024).

The misalignment between stakeholder perspectives creates barriers to the adoption and implementation of ADM systems. While all stakeholders value fairness, they often differ in how it should be operationalised and who should be responsible for ensuring it, reflecting a broader gap between ethical intentions and practical understanding (Corrêa et al., 2025). Many HR professionals also expressed hesitation due to concerns about reliability, transparency, and ethical implications, often based on limited familiarity rather than direct experience (Bucher, 2017; Mahmud et al., 2022). These findings suggest that non-adoption may not solely reflect resistance but also a state of not-yet-adoption, influenced by uncertainty and limited exposure.

Overall, this thesis demonstrates that algorithmic fairness in HR is a multidimensional, stakeholder-contingent construct. The results underscore the need for interdisciplinary dialogue, clearer conceptual frameworks, and transparent communication to support both understanding and adoption of ADM systems in organisational practice (Dolata et al., 2022; Feldkamp et al., 2023; Starke et al., 2022).

This thesis contributes to theory by advancing the study of algorithmic fairness in HR by adopting a sociotechnical perspective that integrates both technical and social dimensions. It extends research beyond decision-subjects to include stakeholders involved in the design, implementation, and use of algorithmic systems. Through a literature review, website analysis, and interviews, this thesis reveals how algorithmic fairness is constructed and interpreted, and it highlights asymmetries in understanding and the factors shaping stakeholders' perspectives. This thesis also provides domain-specific insights into a not-yet-adoption, showing how HR professionals' concerns are influenced by system type, perceived fairness, and contextual factors.

On a practical level, the research findings offer actionable guidance for designing and implementing fair algorithmic systems that align with diverse stakeholder expectations. They inform strategies to enhance transparency, explainability, and adoption, while supporting policymakers in clarifying fairness standards and operationalising ethical and legal compliance. By emphasising cross-stakeholder communication and engagement, this thesis highlights how collaboration and explainability are essential for the successful integration of ADM systems in HR.

1.1 Research Motivation

Despite the growing recognition of, and increased research into, algorithmic fairness in HR, there are some notable matters that remain unaddressed. In particular, there is a lack of research that compares and integrates the perspectives of multiple stakeholders because much of the existing literature focuses on decision-subjects' perceptions of algorithmic fairness, including job applicants (Lavanchy et al., 2023) and frontline workers (Zhou et al., 2023). There are very few examples of research focusing on other stakeholders, such as software developers (e.g., Kasinidou et al., 2021; Kleanthous et al., 2022) or HR managers (Feldkamp et al., 2023). It is important to integrate the views of other stakeholders involved in the development, implementation, and use of ADM systems within HR – software developers, HR professionals, consultants, and people analysts – because they may hold different and potentially even conflicting views of what constitutes a "fair" decision (Raghavan, 2024). Nevertheless, existing research rarely examines these views in relation to one another. As a result, algorithmic fairness is often studied in silos, either through the lens of abstract technical metrics or through individual-level perceptions, with insufficient attention given to the interactions and tensions between stakeholders.

A siloed approach also overlooks how algorithmic fairness is constructed, enacted, and contested in organisational HR practice. While studies focus on examining what factors influence the concept of algorithm aversion (i.e., an individual's preference for relying on their own or other humans' decisions rather than on algorithmic decisions (Dietvorst et al., 2015), researchers have not yet taken a constructionist approach to understanding how HR professionals construct algorithmic fairness. Therefore, with regard to algorithmic fairness, we do not understand how different understandings of fairness are shaped by personal, institutional, and technological contexts. Moreover, we do not understand how these differing interpretations influence the adoption and implementation of ADM systems, particularly whether limited uptake reflects a state of not-yet-adoption or a more fundamental perception that such systems are inherently biased and unfair, thereby discouraging their future use. Consequently, there is limited knowledge of how the sociotechnical dimensions of fairness intersect in HR, or of how stakeholders' negotiations of algorithmic fairness shape the practical use of these systems.

Although the use of ADM systems in HR has been widely discussed, this discourse often creates misleading impressions that these systems are also widely adopted in practice. In reality, adoption remains limited, and there is only limited research about why HR professionals in particular are hesitant to implement them (e.g., Maasland & Weißmüller, 2022). Existing research interprets this gap through the lens of "algorithm aversion" which has been examined in psychology and general management research (Mahmud et al., 2022). Algorithm aversion research typically highlights individuals' concerns about aspects, such as lack of trust (Lee, 2018), lack of transparency or the potential of unfairness and biases (Ochmann et al., 2024). At the same time, AI adoption research extends beyond individual attitudes of algorithm aversion by incorporating organisational-level factors. This literature highlights additional drivers and barriers, such as competitive pressures, market demand, and the anticipated benefits of AI systems (e.g., Hangl et al., 2023; Horani et al., 2025; Khanfar et al., 2026).

However, prior research has not focused much on HR contexts. Therefore, it remains unclear what the reasons are for why ADM systems are not widely adopted in HR contexts. It further is unclear whether HR professionals perceive ADM systems as inherently problematic, reflecting a stable form of aversion, or whether their reluctance is better understood as a lack of conviction regarding the systems' maturity, relevance, and practical value (Marocco et al., 2024). This distinction points to the possibility that what is commonly described as resistance may instead represent a state of not-yet-adoption, which does not necessarily indicate immediate rejection, but rather a temporary condition shaped by evolving knowledge and familiarity (Marocco et al., 2024).

This distinction is particularly important because prior research suggests that many commonly cited barriers, such as distrust or concerns about transparency, are not fixed. Instead, they can evolve as individuals gain experience with AI systems and develop a better understanding of their functionality and value, especially when initial interactions are perceived as beneficial (Kar & Kushwaha, 2023; Marocco et al., 2024).

Recent developments of generative AI tools, such as ChatGPT, provide a good example. Within a relatively short period, these systems have moved from niche innovations to widespread use in both personal and professional contexts (Agrawal, 2024; Bick et al., 2025). This shift suggests that increased accessibility, usability, and user familiarity can significantly reduce initial resistance and foster adoption (Bick et al., 2025), which can also be applied to other types of AI systems, such as ADM systems. Moreover, these factors may be better understood a transitional condition that can be overcome through usability, exposure, and institutional support (Bick et al., 2025; Cubric, 2020; Hangl et al., 2023; Horani et al., 2025). This indicates that different types of AI systems may follow distinct adoption trajectories, depending on how they are experienced by users.

As a result, we know little about why HR practitioners themselves might resist or refrain from using ADM systems in their everyday practice. Understanding these dynamics is essential because insights from psychology or general management cannot be uncritically applied to HR. Different organisational contexts and norms may shape distinctive patterns of adoption, resistance, or avoidance. These distinctions necessitate a more situated investigation into the current adoption state of ADM systems in HR.

1.2 Purpose & Contribution

In exploring ADM in HR, this thesis sheds light on how various stakeholders interpret and operationalise algorithmic fairness. This perspective moves beyond broad assumptions and responds to recent calls for the sociotechnical examination of algorithmic fairness. This research shows how algorithmic fairness understandings vary across stakeholders and often diverge from established technical or social definitions, and how these differences influence both system design and organisational adoption.

Theoretically, this thesis provides a structure to the multidisciplinary landscape of algorithmic fairness by synthesising and comparing different stakeholder perspectives on algorithmic fairness, as outlined in Chapter 4. This chapter provides a nuanced understanding of the factors shaping conceptions of algorithmic fairness depending on the stakeholder lens taken. By doing so, it shows that literature identifies various different stakeholders, including managers, developers, regulators and organisations, while decision-subjects are mainly investigated to inform the developer's and manager's perspective.

Chapter 5 and 6 together extend this understanding of different stakeholder perspectives by shifting attention from decision-subjects to those who actively develop, implement, and use ADM systems. A thematic analysis (Braun & Clarke, 2006) of software suppliers' websites reveals how design actors publicly frame fairness in ADM, while interviews highlight how algorithmic fairness is understood by those stakeholders. Together, these findings offer a rich and multi-layered account of how different stakeholders construct fairness, revealing asymmetries in understanding that can inform both theory and practice.

Chapter 7 adds further depth into my research by examining HR professionals' reasons for resisting ADM systems. While much of the literature treats algorithm aversion in general terms, without focusing on a specific domain (e.g., Mahmud et al., 2022), the study underpinning this chapter provides a domain-specific perspective. By concentrating on HR professionals, the study contributes to theory by offering insights into the particular concerns held by this stakeholder group. I identify several reasons uniquely pertinent to the HR context, including stakeholders' apprehensions about discrimination by ADM systems, which may result in talent loss and a consequent reduction in workforce diversity. My findings also indicate that the type of algorithmic system influences levels of reluctance towards adoption, which is a factor that has been insufficiently explored in prior research. These findings also show that perceptions of fairness and acceptance depend on the type of system in use, thereby identifying a dimension that has been underexplored in prior work.

On a practical level, the thesis provides guidance for designing and evaluating algorithmic systems that various stakeholders view as fair, thereby also addressing concerns about the use of ADM systems. The literature review in Chapter 4 highlights the range of fairness considerations across technical, regulatory, and social perspectives. I argue that this understanding is crucial because it informs the development and evaluation of ADM systems with respect to algorithmic fairness. Specifically, my findings provide valuable guidance on designing fair algorithms by highlighting the differing emphases that stakeholders place on aspects of algorithmic fairness. Moreover, the findings of the interview analysis in Chapter 5 contribute to the design and evaluation of ADM systems by enabling better alignment with varied stakeholder understandings of algorithmic fairness. My findings further offer actionable insights for developers and organisations, which can be drawn from my study of the not-yet-adoption of ADM systems in the HR context, detailed in Chapter 7. This provides a foundational basis for designing ADM systems and implementation strategies that directly address end-user concerns, particularly by enhancing transparency, explainability, and change management processes.

This thesis also offers value for policymakers by clarifying how fairness is constructed differently across contexts, and by highlighting the need for precise communication around fairness, transparency, and legal compliance. By recognising the influence of diverse stakeholder perspectives, as outlined in Chapter 4, Chapter 5, and Chapter 6, this thesis supports regulators in drafting clearer policies that avoid conflating related concepts, and in operationalising fairness more effectively. The findings in this thesis further underscore the criticality of effective change management practices that support the responsible adoption of

ADM systems in HR, helping ensure that future legislation and organisational practices are both nuanced and practicable.

Finally, this thesis highlights explainability as a critical condition for both algorithmic fairness understanding and ADM system adoption. As I show in Chapter 6 and Chapter 7, stakeholders often equated algorithmic fairness with transparency, and HR professionals' lack of trust was strongly linked to a lack of transparency and poor system explainability. These findings suggest that sustained engagement by software suppliers and organisations is essential to bridge these gaps and ensure that ADM systems are both perceived as fair and confidently adopted.

Overall, this thesis shows that the construction of algorithmic fairness is shaped by stakeholders' diverse beliefs, levels of familiarity, and interpretations, and that algorithmic fairness is not a uniform concept but one negotiated across different contexts and perspectives. By examining how these perspectives diverge, overlap, and influence both system design and adoption, this thesis highlights algorithmic fairness as a contested and context-dependent concept rather than a fixed standard. In doing so, this thesis further demonstrates that the success of ADM systems depends not only on technical design but also on how fairness is communicated and operationalised across organisational settings.

1.3 Thesis Structure

The thesis has a combination of traditional chapter and research manuscripts as outlined in Table 1.2. This introductory chapter establishes the context and foundational knowledge that supports the four research manuscripts in subsequent chapters. Given that this thesis follows a publication-based format, some repetition across chapters is inherent and unavoidable, particularly in manuscript-based chapters, as these contributions were developed and published as standalone works in IS conference proceedings. Moreover, the manuscripts use the first-person plural "we" rather than "I," reflecting both their collaborative development with my supervisors and the practices of co-authored academic publications.

Table 1.2: Thesis Structure

Chapter 1: Introduction *Purpose:* 1. Introduce the research context. 2. State research motivation, purpose, and contribution.
Chapter 2: Methods *Purpose:* 1. Describe my research philosophy. 2. Describe my research design, including data collection and data analysis methods, and the decision to select the organisational HR context as my research setting. 3. Describe research constraints encountered during data collection. 4. Highlight ethical considerations as part of research approach.
Chapter 3: Background Literature *Purpose:* Introduce foundational knowledge about concepts of algorithmic systems, algorithmic bias, algorithmic fairness, algorithm aversion, and the use algorithmic systems in the HR context.
Chapter 4: Construction of Algorithmic Fairness by Prior Literature (Manuscript 1) *Purpose:* 1. Identify stakeholder groups that are focused on as part of the discourse about algorithmic fairness. 2. Uncover what shapes the construction of algorithmic fairness depending on the stakeholder perspective that is taken in the literature. *Theoretical Contribution:* Understanding and comparing algorithmic fairness from various stakeholder perspectives that are taken in the literature. *Practical Contribution:* Understanding what aspects different stakeholders consider fair to inform the design and assessment of fair ADM systems.
Chapter 5: Construction of Algorithmic Fairness through Software Suppliers' Websites (Manuscript 2) *Purpose:* 1. Explore how software suppliers publicly articulate fairness on their websites. 2. Uncover which employment lifecycle phases consider algorithmic fairness. *Theoretical Contribution:* 1. Contribute to the discourse about algorithmic fairness from a software supplier perspective. 2. Identify the HR operations in which fairness plays a more important role. *Practical Contribution:* Contribute to a better understanding of algorithmic fairness in practice, thereby providing guidance to companies planning to implement software in HR.
Chapter 6: Construction of Algorithmic Fairness of Stakeholders Developing, Implementing, and Using Algorithmic Decision-making Systems in the Organisational HR Context (Manuscript 3) *Purpose:*

1. Explore how different stakeholders understand algorithmic fairness. 2. Explore what shapes these understandings of algorithmic fairness. *Theoretical Contribution:* 1. Understand construction of algorithmic fairness by stakeholders responsible for developing, implementing, and using ADM systems. 2. Identifying which stakeholders construct fairness from which perspectives, helping to reveal gaps in fairness considerations and to inform strategies to improve knowledge. *Practical Contribution:* 1. Clarify the diverse meanings of fairness, acknowledging that there is no single universally accepted definition. 2. Assist in designing and evaluating ADM systems to better align with varied understandings of algorithmic fairness. Therefore, articulate a more precise definition and explanation of fairness in future legislation.
Chapter 7: Not-Yet-Adoption of Algorithmic Decision-Making Systems in the Organisational HR Context (Manuscript 4) *Purpose:* 1. Explore HR professionals' concerns about the use of ADM systems in HR practices. 2. Uncover the elements that shape these concerns. *Theoretical Contribution:* 1. Identify interconnected concerns as central barriers to ADM adoption in HR. 2. Advance theory by identifying new concerns not widely discussed in the literature. *Practical Contribution:* 1. Provide a foundation for designing AI systems and implementation strategies that address end-user concerns. 2. Underscore the importance of clear communication around legal compliance and ethical considerations to support responsible AI adoption in HR.
Chapter 8: Discussion and Synthesis *Purpose:* 1. Summarise the thesis. 2. Provide insights based on how the chapters relate to each other. 3. Demonstrate theoretical and practical contributions. 4. Highlight limitations and provide future research avenues.

Chapter 2 discusses my research philosophy and the rationale behind my methodological choices. I outline my ontological and epistemological research view and provide information about data collection and data analysis for each manuscript chapter. I also address ethical considerations and constraints that I encountered during my research that has led me to further investigate the adoption concerns about ADM systems in HR as outlined in Chapter 6.

Chapter 3 outlines background literature to introduce knowledge about the nature of algorithms and what different types of algorithmic systems exist. The chapter also introduces the concepts of algorithmic bias and algorithmic fairness, which are closely related (Panarese et al., 2025). In outlining algorithmic fairness, I further focus on explaining the technical perspective, the social perspective, and the sociotechnical perspective on algorithmic fairness, which shapes my research approach towards investigating algorithmic fairness from a sociotechnical perspective. I then outline the concepts of algorithm aversion and the use of algorithmic systems in the HR context, along with associated fairness issues and concerns. I do so to highlight the understanding of algorithmic fairness in past studies. The background literature also provides the foundation for understanding the current status of research in the field and potential gaps in that research.

In Chapter 4, I report on my organising review of the literature to make the broad field of algorithmic fairness more comprehensible (Leidner, 2016). I categorise the literature using the stakeholder perspectives that are taken in the respective studies. By doing so, this chapter explores how algorithmic fairness is constructed and shaped depending on the stakeholder view taken in the research. This chapter provides an understanding of which stakeholders have been studied and how exactly they have been studied, which informs subsequent research manuscripts in the selection of stakeholders and in establishing foundational knowledge on the construction of algorithmic fairness in the literature (Webster & Watson, 2002).

Following this literature review, Chapter 5 outlines my first empirical study. I conduct a thematic analysis of software suppliers' websites (Braun & Clarke, 2006) to understand how software suppliers, in particular people analytics vendors, construct algorithmic fairness. This analysis shows that fairness is predominantly constructed through a variety of terms, such as "absence of bias" or "objectivity", and that if fairness is directly mentioned, there is no common definition of the term fairness in this context. This chapter also identifies that recruitment and performance assessments are the processes most prone to unfairness, which is why fairness has predominantly been discussed as part of those processes on software suppliers' websites.

In Chapter 6, I report on my second empirical study. I interviewed stakeholders who design, implement and use ADM systems – software developers, HR professionals, people analysts, AI consultants and HR consultants – to investigate their constructions of algorithmic fairness. I conducted 30 semi-structured interviews to explore the interview participants' understandings of algorithmic fairness and how these various understandings are shaped. I reveal several different factors that influence how algorithmic fairness is constructed, including laws and regulations, cultural influences, algorithm aversion, and considerations relating to technical fairness. I also explore how algorithmic fairness was spoken about during the interviews and analyse the factors shaping the varied understandings of algorithmic fairness. These varied understandings included fairness as anti-discrimination, transparency and explainability, inclusivity, objectivity and the absence of bias, and as dependent on social norms. I further reveal a tension between differing constructions of algorithmic fairness, specifically, between fairness understood as objectivity, as also identified in Chapter 5, and fairness viewed as context-dependent and shaped by social norms. The former construction frames algorithmic fairness in terms of the objective assessment of decision-subjects, emphasising neutrality and the elimination of bias. However, this perspective often overlooks the contextual nuances and lived realities of individual cases. For example, evaluating candidates for potential, such as those with non-linear career trajectories (Hunkenschroer & Luetge, 2022), cannot be fully captured within a purely objective framework. In contrast, stakeholders with technical expertise in ADM systems tended to adopt a more contextual and sociotechnical conceptualisation of fairness. This conceptualisation acknowledges that fairness must be responsive to diverse social contexts and dynamic normative expectations, rather than relying on static, one-size-fits-all definitions rooted solely in objectivity (Dolata et al., 2022).

In Chapter 7, I report on my exploration of the reasons for concerns about adopting ADM systems by HR professionals. This study took a different approach in response to the research constraints associated with the low use of algorithms by HR professionals (see Chapter 2.4). I show that underlying factors, like the lack of familiarity and understanding of algorithmic systems, significantly contribute to ethical apprehensions. Specifically, concerns that ADM may unfairly disadvantage certain individuals or groups, thereby potentially reducing diversity within the workforce, were particularly pronounced among participants with limited technical knowledge. This chapter complements the theoretical and empirical insights on algorithmic fairness by highlighting how limited familiarity and reliance on second-hand knowledge shape ethical concerns, risk perceptions, and resistance to technological change.

In Chapter 8, I summarise the research findings and provide a cohesive narrative that connects the constructions of fairness to real-world challenges in implementing ADM systems. This narrative offers a comprehensive understanding of both why algorithmic fairness is constructed in varied ways and how these constructions impact the adoption and integration of ADM systems. I also demonstrate the theoretical and practical contributions of these findings, discuss research limitations, and suggest areas for future research.

Chapter 2

Methodology

This chapter outlines the methodological foundations and research design underpinning my thesis, which investigates how different stakeholders in the organisational HR context construct algorithmic fairness. The chapter is structured as follows. First, in Chapter section 2.1, I present the research philosophy and design. I elaborate on the relational ontology approach I take in my research and on how social constructionism informs the investigation of algorithmic fairness within sociotechnical systems. In Chapter section 2.2.1, I justify the choice of the organisational HR context as the empirical setting for this research. In Chapter sections 2.2.2 and 2.2.3, I provide a detailed outline of the data collection and analysis procedures employed in both the theoretical and empirical components of the research. In Chapter section 2.3 I outline ethical considerations for my research, followed by a reflection on practical constraints in Chapter section 2.4. I encountered these constraints particularly during the recruitment of interview participants and they have informed my research design, especially the development of my research study that focuses on the not-yet-adoption of ADM systems in HR practices.

2.1 Research Philosophy

The methodology employed in this thesis is grounded in a relational ontology, which rejects the separation of the social and the material and instead emphasises their constitutive entanglement (Cecez-Kecmanovic et al., 2014; Orlikowski, 2007; Orlikowski & Scott, 2015; Scott & Orlikowski, 2014). This ontological stance is paired with a social constructionist epistemology, which posits that knowledge and reality are co-constructed through social interactions and practices (Berger & Luckmann, 1991). As such, algorithmic fairness is approached here not as a fixed or universal standard but as a contextually contingent construct shaped by individual and collective experiences, perspectives, and practices.

2.1.1 Ontology

My research is grounded in a relational ontology, which, in contrast to substantialist ontology, understands reality not as composed of discrete, pre-existing entities, such as individuals, technologies, or institutions, but as continuously constituted through dynamic and situated relations and practices (Cecez-Kecmanovic et al., 2014; Orlikowski, 2007; Orlikowski & Scott, 2015; Scott & Orlikowski, 2014). Within this ontological framework, technological artefacts, such as algorithmic systems in organisational contexts, are understood through the lens of sociomateriality; not as isolated technical objects, but as sociomaterial entities whose meaning and function emerge through their entanglement with human practices and organisational routines. Materiality, in this sense, refers to the ways in which technologies are embedded in, and co-constitutive of, the social contexts in which they are developed, deployed, and used. Rather than being viewed as passive or neutral tools, material artefacts, such as algorithmic systems, are seen as active participants in shaping organisational processes, decision-making practices, and normative understandings, such as algorithmic fairness, while simultaneously being influenced by social processes (Orlikowski, 2007; Orlikowski & Scott, 2015).

This relational ontology challenges the techno-centric and human-centred approaches to technology. A techno-centric perspective tends to prioritise the functions of technology, often portraying it as a neutral and deterministic force that shapes human action. This perspective risks overlooking the broader social, historical, and cultural contexts in which technology is embedded. On the other hand, a human-centred perspective focuses on how individuals interpret, adopt, and interact with technology, treating technological artefacts as meaningful only insofar as they are shaped by human intention and sensemaking. While this perspective recognises variation in how technology is used and understood, it tends to privilege human agency and underplays the performative role of material artefacts.

In contrast, a sociomaterial perspective emphasises the inseparable entanglement of the social and material aspects. Rather than viewing technology and human actors as separate entities that influence one another, this approach sees them as mutually constituted and inseparable in practice. Therefore, the social is always already material, and the material is inherently social, each shaping and being shaped by the other. This ontological stance challenges the dualistic thinking common in information systems and organisational research, where technology is either conceptualised as an external, objective force acting upon the social, or as a passive artefact interpreted solely through human agency. A sociomaterial lens, by contrast, highlights the co-constitution of people and technologies, enabling a more situated and nuanced understanding of algorithmic systems within organisational life (Orlikowski & Scott, 2015).

Applying a sociomaterial lens to my research to understand how algorithmic fairness is constructed by different stakeholders, enables a more nuanced understanding that moves beyond binary conceptions of algorithmic fairness as either a technical property inherent to algorithms or a purely normative, socially constructed concern. This perspective highlights the inseparable entanglement of the social and technological aspects of algorithmic fairness, and emphasises that algorithmic systems are not merely passive tools but active participants in shaping algorithmic fairness (Dolata et al., 2022). Instead, algorithmic fairness is understood as an emergent and context-dependent phenomenon, shaped through the dynamic interplay of technological affordances, organisational routines, institutional logics, and the values and practices of diverse stakeholders. This perspective is particularly relevant in HR contexts, where ADM systems not only mirror existing organisational values and biases but also actively participate in shaping decisions, such as those related to hiring, promotion, and performance assessments, and thereby reinforce or challenge broader societal norms around equity and justice (Dolata et al., 2022; Newman et al., 2020).

Although this relational ontology is grounded in the broader sociotechnical tradition, it places greater emphasis on the performative nature of fairness by highlighting how algorithmic fairness is enacted and brought into being through sociomaterial practices, rather than treated as a fixed attribute that can be embedded into systems in advance. Accordingly, my research investigates how stakeholders understand algorithmic fairness and how these understandings are embedded in, and enacted through, technological artefacts, institutional processes, and everyday organisational practices. My research also considers how algorithms themselves actively shape conceptions of fairness, for instance by revealing patterns that were previously unobserved, thereby enabling new possibilities for diversity and equity.

2.1.2 Epistemology

For the purposes of the research underpinning this thesis, I adopt a social constructionist epistemology, which argues that knowledge is not discovered as an objective truth that exists independently of human perception, but is instead socially produced and contextually defined. Therefore, knowledge and reality are not given but emerge through human practices, their interpretations, and exchanges between people and their environments. Meaning is created through ongoing processes of engagement, and what is understood as "real" or "true" depends on the social, cultural, and historical settings in which knowledge is produced, including the communities and institutions that influence how individuals perceive reality. Language plays a vital role in this process as the principal medium through which meaning is negotiated, institutionalised, and sustained. Language objectifies human experience and enables the preservation and transmission of knowledge across time and space. It allows for the construction of shared typifications and categories, making complex social realities appear stable and taken-for-granted. In this sense, language both conveys and structures social reality (Berger & Luckmann, 1991; Segre, 2016). Furthermore, knowledge and the realities it defines are embedded within power structures, such as political, legal, or religious institutions, that govern what is accepted as legitimate knowledge and influence how social realities are maintained and transformed. These institutions do not merely reflect reality. Instead. they actively participate in its construction by legitimising certain worldviews over others. Through processes of institutionalisation and legitimation, dominant groups can naturalise their interpretations of reality, embedding them into social structures and everyday routines, and thereby reinforcing existing hierarchies, and limiting alternative ways of knowing and being (Berger & Luckmann, 1991). As a result, knowledge is not universal or fixed, but fluid, negotiated, and collectively formed through shared human experience. Moreover, multiple meanings and realities can exist through negotiated social processes and interactions (Berger & Luckmann, 1991; Zhao, 2020). Therefore, human experience and the objects of that experience are inseparable, meaning that cannot be adequately understood in isolation from the other (Berger & Luckmann, 1991). Constructionist epistemology stands in clear opposition to objectivist paradigms, which assert that meaning exists independently of human consciousness and that the primary goal of research is to uncover these external, objective truths. Unlike objectivism, which privileges explanation through empirical verification and seeks to identify causal mechanisms, constructionism is primarily concerned with understanding how individuals and groups make sense of their experiences and social realities (Berger & Luckmann, 1991; Crotty, 1998; Hazelrigg, 1986). This distinction is particularly important for the study of complex social phenomena, where meaning is mediated by culture, context, and power relations, which can render reductive explanations insufficient (Berger & Luckmann, 1991) .

Given the wide range of interpretations of algorithmic fairness within existing literature (Dolata et al., 2022), a social constructionist approach is particularly well-suited to investigating how various stakeholders construct and make sense of algorithmic fairness within organisational HR contexts. A social constructionist approach acknowledges algorithmic fairness as a context-dependent and socially constructed phenomenon, shaped by interactions among individuals, technologies, and institutional environments. In adopting this epistemological stance, my research prioritises an exploration of how algorithmic fairness is constructed, negotiated, and interpreted in practice. This exploration involves examining the diverse and often contested ways in which stakeholders, such as HR professionals or software suppliers, understand and engage with fairness in the context of ADM. Consequently, fairness is approached not as a fixed or technical outcome but as an emergent property that is co-constituted through organisational routines, technological artefacts, and broader socio-cultural norms.

2.1.3 Sociotechnical Lens

Adopting these ontological and epistemological approaches in my research aligns with recent calls within Information Systems research to frame algorithmic fairness through a sociotechnical lens (Dolata et al., 2022). Sociotechnical systems theory emphasises the joint optimisation of technical and social elements. Technical elements encompass the tasks, processes and technologies necessary to convert inputs into outputs, while social elements focus on human attributes, including values, skills and attitudes as well as interpersonal relationships, incentive structures, and organisational hierarchies (Bostrom & Heinen, 1977). The outcomes of sociotechnical systems therefore emerge through the interplay between their technical and social components (Bostrom & Heinen, 1977; Cecez-Kecmanovic et al., 2014; Orlikowski, 2007), whereby technologies, such as algorithmic systems, are embedded in and shaped by complex social systems (Dolata et al., 2022; Draude et al., 2020). With regards to algorithmic systems, I therefore cannot only look at the algorithm from a technical perspective because the algorithm itself is situated in social contexts. Humans shape and affect the outcomes of algorithmic decisions (Dolata et al., 2022; Selbst et al., 2019) and they develop the algorithms, which means that they can implement their own prejudices and biases into those algorithm (O'Neil, 2016). Moreover, I cannot only view the algorithm from a social lens because of its underlying technical elements, which can include, for example, biases in data or mathematical notions of fairness (Dolata et al., 2022; Draude et al., 2020). I, therefore, have to understand how technical and social elements mutually influence each other and how they co-produce algorithmic fairness through the interactions between algorithms and humans, institutional policies, user practices, and broader social structures. Adopting this perspective enables a more comprehensive analysis that incorporates the technical logic and the social processes, power relations, and contextual conditions, that influence how fairness is defined, implemented, and experienced in practice (Dolata et al., 2022; Draude et al., 2020).

Accordingly, this research does not isolate technical elements from social elements but instead investigates how they are entangled in shaping stakeholders' perceptions and enactments of fairness in the organisational HR process. Social constructionism provides the conceptual grounding to examine these interactions, which allows for a more situated understanding of how algorithmic fairness is experienced and made meaningful. By focusing on these sociotechnical entanglements, my research offers a nuanced account of algorithmic fairness as a lived, negotiated, and socially embedded phenomenon that cannot be fully captured through purely technical or normative definitions alone.

2.2 Research Design

In the following sections, I outline the rationale for selecting the organisational HR context as the context of my research in Chapter section 2.2.1. Then, in Chapter section 2.2.3, I present the data collection processes undertaken for both the theoretical and empirical components of this study, specifically, the literature review and the three empirical papers. This section includes an account of the strategies used for sourcing and selecting theoretical materials, as well as the methodological approaches employed in gathering empirical data. In Chapter section 2.2.4, I explain the data analysis techniques I applied to both theoretical and empirical data. In Chapter section 2.3, I attend to ethical considerations, especially those arising from the interview-based components of the research. Finally, in Chapter section 2.4, I reflect on the constraints I encountered during the research process and discuss how these challenges influenced the development of my overall research approach.

2.2.1 Selection of the HR Context

The choice to focus on the HR context is closely linked to my broader research motivation as outlined in Chapter section 1.1. I aimed to conduct research in a domain that is relevant and timely and therefore actively debated in academic literature and reflected in real-world practice. During my exploration of the topic and the applied contexts, I encountered various conceptualisations of algorithmic fairness, which are applied across multiple domains (Dolata et al., 2022; Starke et al., 2022), including finance, where algorithms assist in loan approval decisions (Fuster et al., 2022; Garcia et al., 2024; Kozodoi et al., 2022), and judicial systems, where risk assessment tools like COMPAS are employed to predict a risk score for recidivism rates (Angwin et al., 2016; Chouldechova, 2017; Larson et al., 2016). Although these contexts are critically important, due to the vulnerability of certain individuals or protected groups to unfair treatment and discrimination, I deliberately chose to examine algorithmic fairness within the HR domain. This decision was informed by the growing prominence of algorithmic systems in HR processes (see Chapter section 3.5.1), and the universal relevance of this context, since nearly everyone engages with job application procedures at some point in their lives. Therefore, HR operations provide a crucial and compelling setting for investigating algorithmic fairness because discriminatory practices in this domain can have significant negative consequences; consequences that not only affect individuals' career opportunities, such as through unfair hiring or promotion decisions, but also impact their overall health and well-being (Mara et al., 2021; Triana et al., 2021).

Given that HR decisions inherently involve judgments about diverse individuals, some form of intentional or unintentional discrimination is often unavoidable (Köchling & Wehner, 2020). Intentional discrimination may arise from explicit preferences or statistical assumptions, such as favouring candidates who share the decision-maker's background or making gender-based assumptions regarding parental leave. Conversely, unintentional discrimination includes unconscious biases and stereotypes that subtly influence decisions without overt intent (Demuijnck, 2009). There are various forms of discrimination, including biases related to race, gender, age, and socioeconomic or ethnic backgrounds (Demuijnck, 2009; Triana et al., 2021; Yeung et al., 2021).

The wide discussion in academic literature of the use of ADM systems in HR (for example, Jabagi et al., 2025; Köchling & Wehner, 2020; Meijerink et al., 2021; Newman et al., 2020) reflects the importance of understanding how algorithmic fairness is conceptualised, enacted, and experienced, given that algorithmic decisions are likely to affect a broad range of individuals engaging with the labour market (Jabagi et al., 2025; Parent-Rocheleau & Parker, 2022). Consequently, investigating algorithmic fairness in HR holds both scholarly significance and practical applicability for a broad population.

2.2.2 Data Collection

Manuscript 1: Literature Review (Chapter 4)

I decided to conduct an organising literature review (Leidner, 2016) as part of my research project, to gain a deeper understanding of algorithmic fairness in literature (Webster & Watson, 2002) prior to conducting interviews with different stakeholders to explore their current understanding of algorithmic fairness (Hart, 1998; Levy & Ellis, 2006). This literature review was undertaken to capture the multifaceted perspectives on algorithmic fairness that are presented in the interdisciplinary literature in order to gain an understanding of how algorithmic fairness is constructed across different literature depending on the stakeholder view that it is taken.

Because a comprehensive literature review encompasses relevant research on a topic and integrates perspective from various disciplines (Snyder, 2019; Webster & Watson, 2002), I undertook a broad literature search across a range of academic databases, targeting journals and conferences within information systems, computer science, organisational studies, and ethics. This wide disciplinary scope was essential to ensure that the review encompassed both the technical and the social dimensions of algorithmic fairness. The initial search strategy employed a combination of keywords related to algorithmic fairness, such as "algorithmic fairness," "algorithmic bias," "fairness in AI," and "fairness in automated decision-making systems" to identify literature relevant to my research question. To maintain rigour and relevance, inclusion criteria were established focusing on peer-reviewed publications that explicitly addressed fairness in the context of algorithmic or automated decision-making systems, particularly within organisational or employment settings. The temporal scope of the search was also defined to capture the most recent and relevant contributions, and to accommodate for the evolving nature of algorithmic fairness discourse (Levy & Ellis, 2006; Okoli, 2015; Snyder, 2019).

Following the compilation of an initial corpus, a multi-stage screening process was implemented. This process involved a preliminary review of titles and abstracts to exclude articles that were out of scope or lacked substantive discussion of fairness. Subsequently, a full-text review was conducted to confirm eligibility, with particular attention paid to how the studies conceptualised and operationalised algorithmic fairness and which stakeholder groups they addressed. The rigorous screening ensured that only studies offering explicit engagement with algorithmic fairness constructs and stakeholder perspectives were retained for deeper qualitative analysis (Snyder, 2019).

To further enrich the data set, snowball sampling techniques were employed, whereby references in key articles were examined for additional relevant studies. This iterative approach expanded the corpus beyond initial search hits and captured influential works that may not have surfaced through keyword queries alone (Webster & Watson, 2002). In total, the final sample included a carefully curated set of publications that represent a broad, yet coherent overview of how algorithmic fairness is constructed and discussed across different stakeholder viewpoints.

Manuscript 2: Analysis of Software Supplier Websites (Chapter 5)

In my first empirical study, I undertook a thematic analysis (Braun & Clarke, 2006) of how software suppliers, particularly those involved in people analytics software, construct and communicate notions of algorithmic fairness on their publicly accessible websites. I examined how software suppliers described algorithmic fairness and determined whether fairness is mentioned at all, in order to understand whether fairness plays a role in algorithmic systems. To identify relevant software suppliers, I conducted a targeted search using Capterra, an open-access digital marketplace that connects people analytics vendors with companies seeking HR technology solutions (Capterra, n.d.). My search focused on suppliers who operate in the Australian market, which meant that I searched Australian software suppliers, along with global software suppliers who offer their services or software in Australia. To ensure the inclusion of software capable of making algorithmically driven decisions, I filtered the search to include software suppliers offering products featuring trend analysis and/or predictive modelling capabilities. After applying these filters, my final dataset comprised 47 software supplier websites.

Manuscript 3 & Manuscript 4: Semi-Structured Interviews (Chapters 6 and 7)

To address the research question on how stakeholders develop, implement, and use ADM systems, the required target interview group included software developers, HR professionals, people analysts, and AI and HR consultants. Initially, my focus was on exploring how different stakeholders within the organisational HR context, particularly HR managers and software developers, construct their understanding of algorithmic fairness. A central aspect of this exploration was to compare stakeholders who actively use and implement algorithmic systems in their daily HR operations with those who do not, to assess whether hands-on experience and familiarity with these systems influence their perceptions of fairness. I also aimed to investigate whether considerations of fairness are integrated into organisational discussions surrounding the adoption of ADM systems in HR practices. However, I encountered practical constraints, particularly relating to HR professionals not using ADM systems in practice.

This included mainly the identification of HR professionals who actively used ADM systems in practice. Many potential participants declined to participate due to a lack of experience with such systems. Even among those interviewed, it became evident that ADM systems were rarely used or discussed within their organisations or professional contexts, indicating a limited uptake of ADM systems in the HR context. It is also important to consider the timing of the data collection. Recruitment for interview participants for Manuscript 3 and Manuscript 4 commenced in 2024, at a point when the AI landscape differed substantially from that at the later stages of this research in 2025. The rapid evolution and widespread adoption of AI tools, especially generative AI, such as ChatGPT, shows how quickly the technological landscape has shifted in the past years (Agrawal, 2024; Bick et al., 2025). This suggests that conducting the study at a later stage may have resulted in different levels of exposure to, and engagement with, ADM systems among HR professionals.

Consequently, I decided to focus data collection on HR professionals who are not currently using these systems, which I have further discussed in Chapter section 2.4. To recruit individuals for interviews, I used two different methods. The primary recruitment method to recruit interview participants was through LinkedIn. I searched for HR professionals, such as recruiters or talent and development professionals, people analysts, consultants, and software developers. The second recruitment method was to meet relevant people in person – via attending and networking at conferences and fairs, such as a HR conference, which also had software suppliers as attendees, and via conversations with recruiters at a careers fair – and invited them to participate.

I initially aimed to recruit participants based in Australia and Germany as my primary focus groups. However, difficulties in accessing suitable interview candidates, particularly as many appeared reluctant to discuss fairness, led me to broaden the interview candidate pool. Consequently, interviews were also conducted with participants located in the United States, Canada, the United Kingdom, and New Zealand, most of whom were employed by globally operating companies. Participants working in such organisations offered valuable insights, not only into the Australian and German contexts but also into international markets, thereby enriching the breadth of perspectives captured in this study. A detailed overview of all participants is provided in Appendix 2A.

As I wanted to gain an understanding of the of the participants' beliefs and thoughts while keeping the interview structure and questions flexible, I decided to conduct semi-structured interviews (Myers & Newman, 2007; Walsham, 1995). The interview questions were developed based on the relevant stakeholder group. This means that I prepared different questions for HR professionals, software developers, and HR and AI consultants, including people analysts working in consulting, respectively. However, I ensured that the interview remained flexible so that, depending on what my interviewee said, I was able to uncover interesting aspects (Myers & Newman, 2007) (see Appendix 2B). The interview protocol for all participants began with questions about their general professional experience, their experience with algorithmic systems if they had any, and their understanding of algorithmic fairness. However, when HR professionals or HR consultants did not have prior experience with algorithmic systems, I discussed potential scenarios in which algorithmic systems could be used, such as hiring decisions as part of recruitment, and whether/how their understanding of fairness would change if an algorithm came into play in those scenarios. As most of the participants had no, or little, experience with algorithmic systems, most interviews focused on the scenarios rather than discussing real life experience. I elaborate more on how the participants changed by research approach in Chapter section 2.4. Where participants had no experience with ADM systems, the interviews also focused on participants' beliefs about why they are not yet used in practice. Software developers, consultants, and people analysts, were also questioned about why they believed ADM systems are not extensively used in practice (see Appendix 2B). The empirical studies reported on in Chapter 6 and Chapter 7 are underpinned by the results of these interviews.

While acknowledging that in-person interviews facilitate better clarification and richer interaction between interviewer and interviewee (Irvine et al., 2012), and provide a more nuanced understanding of the interviewee due to their physical presence and the interviewer's access to the richness of non-verbal cues (Denham & Onwuegbuzie, 2013), the interviews underpinning this thesis were conducted online via Zoom or Microsoft Teams. All but one participant gave their permission for the interviews to be recorded. The recorded interviews were conducted in English or German and were transcribed either through the Zoom or Microsoft Teams recording function directly or through Microsoft Stream. The participant who did not give permission for recording, allowed me to take notes, and these notes were used as an interview protocol and further analysed.

2.2.3 Data Analysis

Manuscript 1: Literature Review (Chapter 4)

The literature review was conducted as an organising review. According to Leider (2016), an organising literature review aims to describe and synthesise existing literature around key themes, concepts, or dimensions, often from various disciplines. Because of the breadth of the relevant literature, organising literature reviews do not argue to be entirely comprehensive, but to synthesise and organise a broad body of literature to improve comprehensibility.

The literature analysis focused on extracting information about the stakeholders involved, their respective concerns and perceptions regarding algorithmic fairness, and the epistemological lenses through which fairness was framed. Through this approach, I synthesised a comprehensive and nuanced understanding of algorithmic fairness, grounded in evidence from existing studies. Specifically, I organised the literature based on the stakeholder perspectives emphasised across prior research, thereby providing a structured lens for understanding how the literature frames fairness depending on the stakeholder perspective taken.

To analyse and organise the selected articles into different topics, each paper was carefully read and examined to determine the stakeholder groups on which it focused. Stakeholders were identified in terms of those directly investigated as subjects within the studies, such as job applicants or algorithm users, and those implicitly addressed or referenced within the literature, including organisational actors, such as management or HR professionals. This dual approach allowed for a comprehensive mapping of stakeholder representation within the discourse on algorithmic fairness.

To categorise the literature, each paper was assigned to one or more stakeholder perspectives. For instance, a study was classified under the “managerial perspective” when it explored algorithmic fairness perceptions among employees with managerial implications (e.g., Newman et al., 2020). This classification process was iterative, involving multiple rounds of review to ensure completeness and accuracy, thereby minimising the risk of errors and omission (Myers & Newman, 2007). Notably, this iterative process also revealed that many studies addressed multiple stakeholder groups, highlighting the intersectionality and overlapping understandings of algorithmic fairness.

I then extended the analysis to investigate how algorithmic fairness was constructed in the reviewed studies, focusing on the epistemological perspective by each stakeholder group. Specifically, the review considered whether algorithmic fairness was framed predominantly through a technical lens (highlighting mathematical definitions and algorithmic features), through a social lens (highlighting normative, ethical, or organisational justice considerations), or through a sociotechnical lens (with technical and social elements interconnected and mutually influential (Dolata et al., 2022)).

Based on this analysis, the synthesis of the findings in Chapter 4 highlighted the diversity of algorithmic fairness perspectives and demonstrates how stakeholder positions influence these interpretations, offering a robust framework to guide further empirical research.

Manuscript 2: Analysis of Software Supplier Websites (Chapter 5)

As part of the first empirical study, I conducted a thematic analysis to understand how software suppliers of algorithmic systems, particularly people analytics vendors, publicly frame fairness on their websites. A thematic analysis enables researchers to identify themes and categories across various studies. A theme in this context reflects a meaningful element within the data that is relevant to the research question and indicates a recurring pattern or shared understanding across the studies (Braun & Clarke, 2006). I specifically analysed a dataset of 21 websites using an inductive thematic analysis. This approach was chosen to enable patterns and meanings to emerge directly from the data, without being restricted by pre-existing theoretical frameworks. Rather than starting with a set of predefined categories, I engaged with the content openly and reflexively, analysing whether fairness was either directly mentioned or indirectly implied through other terms and descriptions, and thereby allowing the themes to be shaped by how software providers presented fairness on their websites (Braun & Clarke, 2006). This work included analysing descriptions of product features, ethical claims, user-oriented benefits, and fairness-related terminology. This analysis enabled me to identify recurring themes and to reveal which phases, such as recruitment, performance evaluation, and offboarding, were most frequently associated with discussions of fairness.

This method aligned with the exploratory nature of my research and my broader aim of understanding how algorithmic fairness is constructed in the organisational HR context. Throughout the analysis, I remained aware of my role in the interpretive process, recognising that the generation of meaning is iterative and shaped by both the data and my position as a researcher (Braun & Clarke, 2006; Walsham, 1995).

Manuscript 3 & Manuscript 4: Semi-Structured Interviews (Chapters 6 and 7)

Both studies employed an inductive approach to qualitative data analysis, by drawing on a constructionist interpretation of grounded theory to develop a contextually situated understanding of algorithmic fairness and concerns about adopting ADM systems. The purpose of applying this approach was to generate theory that is grounded in the experiences and contextual interpretations of stakeholders who engage with algorithmic systems, rather than producing generalisable results (Charmaz, 2006). This approach builds on the epistemological foundations of social constructionism, which argues that reality is socially produced and maintained through ongoing human interaction. Within this framework, knowledge is understood as emerging from the interplay between individuals, institutional settings, and technological artefacts (Berger & Luckmann, 1991). Accordingly, algorithmic fairness, as well as the not-yet-adoption of ADM systems, are not treated as a fixed or universally accepted principles but as a fluid and evolving constructs shaped by organisational practices, sociotechnical dynamics, and normative discourses. This perspective enabled a rich and nuanced examination of how algorithmic fairness and the reasons for not-yet-adoption of ADM systems are interpreted, enacted, and contested across different contexts.

To analyse the data for both studies, I applied a constructionist approach to grounded theory, which views theory development as a process shaped by the interactions between researchers and participants. This approach focuses on the active role of researchers in interpreting and constructing meaning from the data, and eschews the idea that findings are discovered objectively, as discussed by Glaser and Strauss (1967) (Charmaz, 2006, 2008). Grounded in a social constructionist perspective, this approach recognises that knowledge is not fixed or universal but shaped by context, social processes, and the experiences of those involved. The approach allows me, as researcher, to be flexible and encourages ongoing reflection on my position and influence within the study. The approach also focuses the researcher on generating in-depth, context-sensitive insights that reflect the complexities of participants' realities, instead of seeking universal explanations (Charmaz, 2006, 2008, 2020). This quality of a constructionist approach makes it particularly useful for exploring nuanced and evolving issues, such as algorithmic fairness and the reasons behind the not-yet-adoption of ADM tools within sociotechnical systems.

The interview data were analysed using NVivo. The analysis began with the segmentation of the interview transcripts into smaller, meaningful parts, guided by the interview protocol. Initial coding focused on participants' work contexts, their interactions with ADM systems, their interpretations of algorithmic fairness, and the concerns they expressed about using ADM systems. Rather than applying predefined categories, codes were developed inductively through close engagement with the data. These early insights formed the basis for iterative rounds of coding, during which themes were refined and relationships between concepts explored. The process remained reflexive throughout, and I acknowledging my role in shaping interpretations as researcher. To confirm that theoretical saturation was reached, coding continued until further analysis no longer yielded new conceptual developments (Charmaz, 2006).

During the interview data analysis for both studies, I paid particular attention to the sociotechnical dimensions of participants' experiences. This includes analysing how organisational structures, technological affordances, and professional roles influenced participants' understandings of fairness in ADM systems. I also analysed participants' concerns about using ADM systems. By examining how algorithmic fairness and the reasons for not-yet-adoption are constructed within specific sociotechnical configurations, both studies contribute to a more nuanced and situated theorisation of ADM systems and their fairness considerations in practice, and they highlight the multiplicity of meanings and the conditions under which they emerge.

2.3 Ethical Considerations

Prior to conducting the interviews, human ethics approval was obtained in accordance with the University of Sydney's ethical guidelines through a formal application process. My ethics application [2023/HE000862] was approved on 5th February 2024. This process required a clear articulation of the research objectives and design, including the identification of potential stakeholder groups and the methods for approaching them. As part of the ethics application, interview questions were developed for each stakeholder group (see Appendix 2B). Additionally, considerations around data de-identification and secure storage were addressed to protect participant confidentiality. A research information statement outlining the research's purpose and design, alongside a participant consent form, were prepared and provided to interview participants as part of the recruitment process. Consent forms were completed before interviews commenced to confirm agreement to audio recording, with verbal consent also obtained at the start of each session. Interview participants were given the opportunity to ask questions and informed of their right to pause or discontinue the interview at any time. Where permission was provided, interviews were audio-recorded, transcribed, and anonymised. Following transcription, audio files were securely destroyed, and transcripts were stored in accordance with university data management policies. In cases where interview participants declined to be recorded, anonymised notes were taken after receiving participant permission to do so.

2.4 Research Constraints

During the course of this research, I encountered several constraints that significantly influenced both the development of the research design and the overall direction of the study. Firstly, it was challenging to identify HR professionals willing to discuss issues of fairness, which led me to broaden the interview pool with respect to the countries in which participants were located (see Chapter section 2.2.2). Additionally, I aimed to engage HR professionals who actively used ADM systems in practice. However, during the recruitment process, it became apparent that most HR professionals contacted did not use such systems. Consequently, many declined to participate, indicating that they were not suitable candidates for the study. Among those who agreed to be interviewed, it quickly became evident during the interviews that they also did not use ADM systems in practice. Several participants noted that these systems were rarely discussed within their organisations or at industry events. For example, one interviewee reported attending a major HR conference in Germany where no AI software vendors were present, which indicated that ADM systems have not yet become a mainstream topic in practical HR contexts.

After encountering this issue repeatedly, and while preserving the core focus of the research on the construction of algorithmic fairness, I decided to adjust my research approach and interview questions (Appendix 2B) and to include HR professionals who do not actively use ADM systems as participants. The semi-structured format of the interviews allowed me to redirect the discussion towards hypothetical or potential scenarios involving ADM systems. This refinement enabled me to explore participants' perceptions of the possible benefits and drawbacks of such systems, the reasons organisations might choose not to adopt them, and how fairness is conceptualised in relation to ADM, even in the absence of direct experience.

Despite the limitations in accessing participants with direct experience of ADM systems, I was still able to provide insights into how different stakeholders construct the concept of algorithmic fairness. This outcome was achieved through a focus on potential, rather than actual, use cases. This shift is not uncommon in this field because much of the existing research relies on experimental or scenario-based designs to investigate fairness perceptions. For example, prior studies have explored how employees, managers, and software developers, might respond to hypothetical decisions made by algorithmic systems (for example, see Feldkamp et al., 2023; Kleanthous et al., 2022). In this way, my modified approach remained consistent with established methodologies in the field and allowed me to examine how limited familiarity with ADM systems may itself shape stakeholders' constructions of fairness.

The refinements to my research strategy revealed an unexpected and compelling area of inquiry: the not-yet-adoption of ADM systems in HR processes. During data collection and in subsequent discussions after the data collection phase, I identified a gap between the academic discourse on ADM systems, which often focuses on their potential to enhance efficiency in HR operations, such as recruitment (Upadhyay & Khandelwal, 2018), and the apparent lack of extensive implementation in organisational practice. To ensure that my findings were not an isolated artefact of my interview sample, I consulted with other researchers working in the same field, who reported similar challenges in accessing HR professionals with practical experience of using ADM systems. These consultations confirmed that the limited adoption of these technologies is a wider phenomenon.

In response to these findings, I extended the scope of my research to include an examination of the not-yet-adoption of ADM systems in HR settings. This extension allowed me to analyse the factors contributing to resistance, or hesitation, among HR professionals, including their concerns about ethics, bias, discrimination, and workforce diversity. These considerations play a critical role in shaping organisational decisions not to implement ADM systems. In this way, the key constraint that initially posed a significant obstacle to the study ultimately led to valuable new insights and expanded the scope of the research in a meaningful way.

In summary, this chapter has outlined the methodological foundations and research design of my thesis, showing how a relational ontology and social constructionist epistemology inform the investigation of algorithmic fairness. By adopting a sociotechnical lens, the study foregrounds the entanglement of social and technological elements. The chapter has detailed the rationale for selecting HR as the empirical setting, and the processes of data collection and data analysis. Finally, I have reflected on the practical constraints encountered during interview participant recruitment, which led to an expanded focus on the not-yet-adoption of ADM systems.

Chapter 3

Background Literature

To establish an understanding of algorithmic fairness, I first introduce the main types of algorithms and algorithmic systems by distinguishing, first, between traditional algorithms and machine learning algorithms and, second, between different types of algorithmic systems, as each type of system raises distinct fairness concerns. This chapter also introduces the concepts of algorithmic bias and algorithmic fairness.

Given that the concept of algorithmic fairness has been considered in numerous disciplines, each applying different definitions and priorities, I compare perspectives from the literature (Barocas & Selbst, 2016; Dolata et al., 2022; Green, 2022; Holm, 2023; Holstein et al., 2019; Makarius et al., 2020; Selbst et al., 2019; Veale et al., 2018). The *technical perspective*, primarily researched by computer scientists, applies different technical notions of fairness on the group level and on the individual level. While, this perspective is rigorous, it neglects the influence of social context on fairness outcomes. To fill this void, the *social perspective*, typically used in the social sciences, addresses equality, equity, and fairness through frameworks such as organisational justice theory. However, this perspective does not address the technical constraints or challenges inherent in algorithmic systems. This disciplinary divide has led scholars to advocate for a *sociotechnical perspective*, which integrates technical and social considerations to jointly optimise human and algorithmic components.

I outline the concept of algorithm aversion and focus on why individuals often resist algorithmic systems. One key reason for algorithm aversion is the perception that such systems treat individuals less fairly than human decision-makers in comparable contexts.

I also review relevant literature on the use of algorithmic systems in HR and how fairness has been examined in these settings. This literature includes studies on applications across different stages of the HR lifecycle, such as recruitment and performance evaluation, and on the perceptions of various stakeholder groups. These overviews establish the empirical and conceptual background for the research context, as previously elaborated in Chapter section 2.2.1.

3.1 Algorithms

3.1.1 Traditional Algorithms

Traditional algorithms support organisations in processing and interpreting large datasets by aggregating and structuring information into a usable form (Chen et al., 2012; Namvar et al., 2023). These algorithms consist of specific rules and instructions programmed by human developers to perform calculations, analyse inputs, and solve predefined problems (Collins et al., 2021). Unlike modern machine learning models, which can identify patterns in unstructured and evolving datasets, traditional algorithms rely exclusively on structured and static data. Their functionality is strictly determined by the quality and precision of the instructions defined at the development stage. Consequently, their accuracy and effectiveness are largely contingent on the developer's understanding of the problem domain and the extent to which this understanding is accurately translated into the algorithm's logic (Namvar et al., 2023).

Traditional algorithms are particularly effective in environments where processes are stable, predictable, and governed by clearly defined rules, enabling efficient data aggregation, sorting, and analysis (Chen et al., 2012). They are especially valuable in contexts requiring transparency and reproducibility because each step in the decision-making process can be traced to explicitly encoded rules. However, this reliance on predefined logic limits their applicability in situations involving complex, ambiguous, or rapidly changing data. Despite these constraints, the interpretability and consistent performance of traditional algorithms ensure that they remain fundamental components of information systems and decision-support applications (Namvar et al., 2023).

3.1.2 Machine Learning Algorithms

In contrast with traditional algorithms, machine learning algorithms generate new knowledge and support informed decision-making by learning from data and detecting non-linear relationships and patterns within it (Namvar et al., 2023; van den Broek et al., 2021). They can process and analyse large-scale datasets to produce decisions and predictions that enhance organisational operations and productivity (Chen et al., 2012; Cheng et al., 2019; Kordzadeh & Ghasemaghaei, 2022; Tarafdar et al., 2022), without necessarily being coded by humans (Collins et al., 2021). Machine learning models can be categorised as shallow or deep. Shallow models, such as linear regression, support vector machines, decision trees, or neural networks with up to two layers, and they require the manual selection of relevant features prior to learning (Afiouni, 2019; Enholm et al., 2022; Herm et al., 2023; Wang et al., 2019). In contrast, deep learning models, employing complex multi-layer neural networks, can automatically derive relevant features and uncover latent data structures without manual intervention (Afiouni, 2019; Enholm et al., 2022).

Machine learning algorithms can also be classified according to their supervised or unsupervised learning approach. The primary distinction lies in whether the training data for the algorithm is labelled. In supervised learning, algorithms are trained on datasets where each input is paired with a known output, enabling them to learn the mapping between variables. These algorithms are typically used for classification or regression tasks. Classification assigns items to predefined categories either in binary form, where instances are classified into two groups, or in multi-class form, involving more than two categories. Decision trees and random forests are frequently used in this context. A practical example of supervised learning is credit scoring, whereby historical customer data is analysed to predict the likelihood of default, typically framed as a binary classification problem (Bao et al., 2019; Jo, 2021). Unsupervised learning, on the other hand, operates on unlabelled datasets and focuses on uncovering hidden structures or patterns in the data. The most common unsupervised learning technique is clustering, which groups items based on similarity metrics defined by the algorithm. Another widely used method is the k-means algorithm, which iteratively adjusts group assignments to minimise intra-cluster variance (Bao et al., 2019).

AI represents a broader domain within computer science that seeks to design and develop systems that are able to perform task which usually require human intelligence, while enhancing their performance through learning and experience (Meijerink et al., 2021), which is particularly the case for AI systems that are built upon machine learning algorithms (Sarker, 2022). In this way, AI extends the concept of an algorithm by incorporating data-driven learning, autonomy, and, in many cases, probabilistic reasoning, which constitute different types of AI systems. For example, generative AI is designed to create new content, such as images, text or audio, by learning patterns in underlying data distributions (Sarker, 2022) or advanced large language models, like ChatGPT, which generate human-like text by predicting word sequences (Budhwar et al., 2023; Stahl & Eke, 2024). Another key category is predictive AI, which leverages statistical and machine learning methods to forecast outcomes or support decision-making (Sarker, 2022). Predictive AI is widely used in credit scoring (Wu et al., 2021) and recruitment (Kelan, 2024). Lastly, natural language processing enables AI to understand, interpret, and generate human language, facilitating applications like automated interview evaluation (McHugh et al., 2020; Sarker, 2022).

3.2 Algorithmic Bias

Algorithmic bias is often referred to as an algorithm's outputs that advantage or disadvantage particular individuals or groups compared to other individuals or groups without a valid or justified reason (Akter et al., 2021; Kordzadeh & Ghasemaghaei, 2022). These biases mostly affect marginalised or underprivileged groups, including, for example, people with disabilities or people of colour (UK Equality Act, 2010). Algorithmic bias can emerge from various sources, and each affects the fairness and reliability of AI systems in different ways, as outlined in Table 3.1.

Table 3.1: Types of Algorithmic Bias

Category	**Bias Type**	**Definition**
Data	Sampling Bias	Misrepresentation of the population due to biased sampling practices.
	Representation Bias	Underrepresentation or exclusion of specific groups or attributes in the dataset.
Methods	Correlation vs. Causation	Algorithm mistakes correlation for causation, producing misleading inferences.
	Developer's Bias	Human assumptions and values embedded in model design, coding, or feature engineering.
User	Automation Bias	Overreliance on algorithmic outputs without adequate critical evaluation.
	Confirmation Bias	Acceptance of algorithmic results that align with prior beliefs, reinforcing existing biases.

One major source of algorithmic bias is bias in the data itself, which occurs when the information used to train algorithms reflects existing social inequalities or fails to adequately represent certain groups. This bias is often referred to as representation bias (Akter et al., 2021; Mehrabi et al., 2021). Consequently, training datasets may be biased, inaccurate, or incomplete (Dennehy et al., 2023). Sampling bias refers to the failure to select an adequate sample that reflects certain groups, which means that some populations are overrepresented while others are underrepresented. Sampling bias results in models that perform better for certain groups than for others (Akter et al., 2021).

Bias can also originate from the methods employed. While machine learning algorithms can detect correlations and patterns without human intervention, they are unable to infer causality, which increases the risk of confusing correlation with causation (Akter et al., 2021). Additionally, design bias occurs when developers' choices about features, objectives, or parameters inadvertently privilege certain outcomes or groups (Friedman & Nissenbaum, 1996). This bias is particularly relevant for traditional algorithms, where rules and instructions are explicitly programmed by human developers (Collins et al., 2021). In such cases, personal prejudices, such as the unfounded belief that men are more efficient workers than women, can be embedded into the decision logic, leading to systematic discrimination (Kordzadeh & Ghasemaghaei, 2022; Sun et al., 2020; Vanhée & Borit, 2022). Furthermore, the demographic composition of development teams can itself be a source of bias. The predominance of male developers in the field increases the risk that algorithms reflect and reinforce male-centric perspectives, while excluding insights that may arise from more diverse backgrounds. A lack of diversity can therefore perpetuate inadequate design choices and exacerbate bias. Addressing this issue requires the inclusion of development teams with varied gender, cultural, and disciplinary perspectives to improve fairness and mitigate bias (Piorkowski et al., 2021; Schulenberg et al., 2023).

In addition to biases embedded in data and design, human interaction with algorithmic systems can also perpetuate unfairness. Automation bias, which refers to the tendency to over-rely on algorithmic outputs without sufficient critical evaluation, can allow errors or discriminatory outcomes to go unchallenged (Parasuraman & Manzey, 2010). Moreover, algorithms can reinforce human confirmation bias as individuals are more inclined to accept algorithmic decisions that align with their pre-existing beliefs. This dynamic reinforces the importance of developing models with strong explanatory capacity, which enables users to understand and challenge outputs where necessary (Akter et al., 2021).

3.3 Algorithmic Fairness

In the following, I outline how algorithmic fairness has been discussed in the literature across various domains, with particular attention to three key perspectives: the technical perspective, the social perspective, and the sociotechnical perspective. This overview highlights how understandings of algorithmic fairness vary depending on the perspective adopted, while also outlining scholars' critiques of each approach.

Algorithmic fairness is generally understood as "the absence of any prejudice or favouritism towards an individual or a group based on their intrinsic or acquired traits in the context of decision-making" (Mehrabi et al., 2021, p. 11). Because fairness lacks a universally accepted definition, there is extensive scholarly debate about what constitutes algorithmic fairness and how it can be achieved. A notable point in this debate is the frequent and interchangeable use of the terms *fairness* and *justice*, particularly in organisational justice theory, and despite arguments that the two terms are not synonymous. Goldman and Cropanzano (2015) propose that *justice* should relate to adherence to rules and regulations, whereas *fairness* should refer to the perception of those rules and regulations. Nonetheless, given that much of the literature I draw upon uses the two terms interchangeably, I will predominantly refer to the concept as "algorithmic fairness."

3.3.1 Technical Perspective

Most research on algorithmic fairness originates from computer science and information technology, and it takes a technical perspective when defining algorithmic fairness. Within this perspective, researchers distinguish between *group-level* and *individual-level* fairness. Group-level fairness assesses outcomes for protected groups or evaluates disparities caused by prediction errors, while individual-level fairness concerns the consistent treatment of similar individuals. This section outlines definitions of group-level fairness based on protected groups and prediction errors, followed by individual-level fairness, then concludes with the main challenges associated with taking the technical perspective.

Group-Level Fairness based on Protected Groups

On a group-level, notions of fairness focus on attributes of protected groups that are vulnerable to discrimination. These groups include gender reassignment, age, disability, marriage and civil partnership, race, pregnancy and maternity, sex, sexual orientation, and religion or belief (UK Equality Act, 2010).

One example of a notion of fairness that relies on attributes of protected groups, is *statistical parity*, which requires that individuals receive a beneficial label or positive predicted class regardless of their group membership. In other words, members of protected and privileged groups should have an equal probability of being placed in the positive class (Corbett-Davies et al., 2017; Dolata et al., 2022; Haas, 2019.; Verma & Rubin, 2018). *Conditional statistical parity* is met when individuals of protected and privileged groups get allocated to the positive predicted class with the same probability given certain risk factors (Corbett-Davies et al., 2017; Verma & Rubin, 2018). Building on this idea *disparate impact* also examines fairness between groups. However, instead of comparing absolute probabilities, *disparate impact* evaluates the *relative rate* between these probabilities for protected and privileged groups, thereby capturing proportional differences in outcomes (Corbett-Davies et al., 2017; Dolata et al., 2022; Feldman et al., 2015; Haas, 2019).

Group-Level Fairness based on Prediction Errors

From the technical perspective, notions of fairness also extend to an algorithm's prediction errors, which can be assessed using a confusion matrix summarising true positives, false positives, false negatives, and true negatives, as outlined in Table 3.2 (Herm et al., 2023).

Table 3.2: Confusion Matrix

	Actual Positive	Actual Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Based on these metrics, the true positive rate (TPR), false positive rate (FPR), false negative rate (FNR) and the true negative rate (TNR) and the accuracy can be calculated as follows:

$$True\ Positive\ Rate = Recall = Sensitivity = \frac{TP}{TP + FN}$$

$$False\ Positive\ Rate = FPR = \frac{FP}{TN + FP}$$

$$True\ Negative\ Rate = Specificy = TNR = \frac{TN}{TN + FP}$$

$$False\ Negative\ Rate = FNR = \frac{FN}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

A notion of fairness derived from the confusion matrix and its associated prediction errors is the *equality of accuracy,* also referred to as *equalised odds.* This concept requires that the algorithm achieve equal decision accuracy across protected and privileged groups. This requirement means that the true positive rates and false positive rates should be the same for both groups, thereby indicating that the algorithm's predictive performance does not systematically advantage or disadvantage either group (Corbett-Davies et al., 2017; Dolata et al., 2022; Feldman et al., 2015; Glymour & Herington, 2019; Haas, 2019; Pleiss et al., 2017; Verma & Rubin, 2018).

Another commonly applied notion of fairness is *equality of opportunity* or *equal opportunity* which, as opposed to equalised odds, focuses only on the true positive rate. Under this definition, an algorithm is considered fair if the true positive rate for the protected group matches that of the privileged group (Agarwal et al., 2018; Haas, 2019; Hardt et al., 2016).

Individual-Level Fairness

To define fairness at the individual level, the technical perspective adopts a similar approach to group-level fairness but focuses on individuals regardless of group membership. Individual fairness is achieved when an algorithm treats two individuals similarly (Dolata et al., 2022; Haas, 2019). In other words, individuals who are equal with respect to a particular task or characteristic should receive similar outcomes (Dwork et al., 2012; Saxena et al., 2019). For example, a recidivism algorithm would treat two different individuals the same if they are equal in certain attributes (e.g., length of their sentence, or number of prior convictions) (Dolata et al., 2022). Several metrics can be used to assess individual-level fairness, including the Theil index (Haas, 2019) and the Lipschitz condition (Dwork et al., 2012; Mehrabi et al., 2021; Saxena et al., 2019).

Challenges of the Technical Fairness Perspective

Different notions of fairness can be understood as reflecting the diverse interests of stakeholders affected by an algorithm (Wong, 2020). However, some mathematical notions of fairness are mutually incompatible (Corbett-Davies et al., 2017; Dwork et al., 2012; Glymour

& Herington, 2019; Teodorescu et al., 2021; Wong, 2020) and they often overlook the specific contexts in which decisions are made (Fazelpour & Lipton, 2020; Jacobs & Wallach, 2021). For example, Selbst et al. (2019) point out that a high false negative rate might have different severities depending on the context and normative values. A high false negative rate in an automated hiring process ends the process for potential candidates immediately, resulting in them having fewer chances of finding a suitable occupation, whereas a high false negative rate in the context of criminal justice causes individuals to be imprisoned by mistake. Therefore, the debate about algorithmic fairness should not only include the technical perspective, but also consider the values of those affected by the algorithm, particularly the most vulnerable, to reduce social injustice (Kasirzadeh, 2022; Wong, 2020). This consideration is especially important because ADM increasingly permeates socially significant contexts, such as in organisational monitoring, hiring, and termination (Parent-Rocheleau & Parker, 2022; Pritchard et al., 2014), in healthcare (Esteva et al., 2017), and in the public sector (Veale et al., 2018). Therefore, relying solely on a technical perspective is insufficient, as it fails to capture social dynamics and contextual factors that shape how these systems are used and experienced (Dolata et al., 2022).

3.3.2 Social Perspective

The social perspective must be explored because the technical perspective on algorithmic fairness does not consider the social context. In this section, I outline the concepts of equality and equity as foundational approaches to understanding fairness when taking the social perspective. Each concept offers a distinct basis upon which to interpret algorithmic fairness. As part of this overview of the social fairness perspective, I also discuss organisational justice theory, which is a widely accepted framework for analysing different types of fairness in relation to the interplay between employees and technological change within organisations.

Concept of Equity and Equality

Numerous scholars have examined the definition of fairness both within and beyond the context of AI development. From a philosophical standpoint, a key debate concerns whether algorithmic fairness should be grounded in the concept of equity or the concept of equality. The concept of equity involves the recognition of differences between individuals and the contention that they should be treated differently according to their specific needs to ensure they have an equal opportunity for success. Aristotle referred to equity as “proportional equality”, an understanding of equality that focuses less on the final distribution of goods and more on whether that distribution accounts for and respects individual differences. Consequently, a central question in defining fairness is whether it is achieved by treating individuals differently in light of their differences, or by treating everyone the same regardless of those differences (Lee & Baykal, 2017; Pazner & Schmeidler, 1978; Robert et al., 2020).

Unlike equity, the concept of equality does not consider differences between individuals and therefore treats all individuals the same. According to those who espouse egalitarianism, all human individuals are the same in their worth and therefore should be treated equally (Binns, 2018; Green, 2022; Holm, 2023). Aristotle referred to equality as “numerical equality”, meaning that every individual must be treated identically (Lee & Baykal, 2017). Within this concept, it is important to distinguish between *formal equality* and *substantive equality.* Formal equality argues that two individuals, who are equal in specific aspects, should have the same opportunities despite their original position in their community (Dolata et al., 2022; Green, 2022; Rawls & Kelly, 2003). However, formal equality does not allow inequalities that may influence the outcome of a decision. As outlined by Green (2022), a university admissions process that assesses each university applicant only by their grades, does not consider the sources of inequalities, such as unequal opportunities or the quality of the education that the applicants received in the past. On the other hand, substantive equality considers these inequalities by determining and removing different social and institutional hierarchies (Green, 2022).

If we adopt a *psychological view*, fairness is described as the perception of an individual that compares their individual input or output to the input or output of other individuals who are considered to be similar to that individual, a so called relational partner (Adams, 1963, 1965; Dolata et al., 2022), which could, for example, refer to two employees that work at the same career level. Consequently, inequity derives from potential differences in the *actual* input or output and the *perceived* input or output from the individual itself and the other individuals. These differences in perceptions are difficult to predict because they are influenced by cultural aspects and historical aspects (Adams, 1963; Festinger, 1957).

Organisational Justice Theory

Building on the psychological perspective and Adams' theory of fairness (1963), the organisational justice theory was developed to encompass distributive justice, procedural justice, and interactional justice. This theory has been widely adopted in information systems research to analyse different types of fairness in the context of the interrelation between employees and technological change within organisations (Dolata et al., 2022; Joshi, 1989; Li et al., 2014).

Distributive fairness considers whether an individual perceives the outcome of a decision or the allocation of resources, such as their salary, as fair. In the context of ADM, research has primarily focused on distributive fairness, with companies selecting algorithms designed to achieve equity by ensuring fair outcomes for each individual (Robert et al., 2020).

Procedural fairness is associated with the perception of the decision-making process' fairness, and describing the rules and methodologies underpinning how a decision was made (Dolata et al., 2022; Helberger et al., 2020; Robert et al., 2020). In the context of ADM, a growing body of literature focuses on procedural fairness and how it can be promoted (e.g., Grgić-Hlača et al., 2018; Lee et al., 2019; Ötting & Maier, 2018). According to Rawls, "practice is just or fair, then, when it satisfies the principles which those who participate in it could propose to one another for mutual acceptance under the aforementioned circumstances" (1958, p. 178). In other words, ensuring procedural fairness in algorithmic systems involves clearly communicating the decision-making rules, methodologies, and criteria, incorporating the values of those affected by the decision into the process, and providing explanations for the decision outcomes. Furthermore, individuals need to have the opportunity to respond to the decision and the opportunity to provide feedback (Lee et al., 2019). Consequently, numerous researchers have emphasised the importance of *transparency* for improving perceived fairness among individuals who are affected by the decision-making process (Lee et al., 2019; Leewis & Smit, 2023; Ochmann et al., 2024; Parent-Rocheleau & Parker, 2022).

Interactional fairness – commonly divided into *interpersonal fairness* and *informational fairness* – is concerned with an employee's perception of how they are treated by their company. Interpersonal justice relates to whether individuals feel they are treated with respect, while informational justice refers to the adequacy and clarity of information provided to help individuals understand how fairness is ensured. In comparison to distributive justice and procedural justice, less research has been conducted on interactional justice in an ADM context (Robert et al., 2020; Starke et al., 2022). Bankins et al. (2022) have shown that individuals might be more concerned with the ADM process rather than a respectful and dignified treatment because of a lack of ability to apply fairness language to an algorithm. The authors argue that an individual's perception of fairness of a decision made by an algorithm, compared to a decision made by a human, also depends on the outcome of the decision. As such, a positive outcome of an algorithmic decision is perceived to be fairer than a negative outcome of a human decision.

3.3.3 Sociotechnical Perspective

While the social perspective expands the discussion of fairness beyond purely technical metrics by incorporating human values, perceptions, and contexts, it treats the algorithm and the social environment as separate entities. In contrast, a sociotechnical perspective recognises that the technical and social dimensions are interdependent and must be examined together. In this section, I outline the general concept of the sociotechnical perspective and the importance of a sociotechnical framing. Furthermore, I outline examples of the role humans can play as the operators in the algorithmic based decision-making process.

Numerous scholars have highlighted the need to combine the technical and social practices to get a more appropriate understanding of algorithmic fairness because it is highly dependent on the social context (Fazelpour & Lipton, 2020; Jacobs & Wallach, 2021; Madaio et al., 2022; Veale et al., 2018), and emphasised the need to understand algorithmic fairness from a sociotechnical perspective (Dolata et al., 2022; Green, 2022; Holstein et al., 2019; Makarius et al., 2020; Selbst et al., 2019). Unlike the human-in-the-loop concept, which integrates human knowledge, human supervisory control, and human oversight, into operations of computational systems (Dodge et al., 2019; Dolata et al., 2022; Grønsund & Aanestad, 2020; Teodorescu et al., 2021; Wu et al., 2022), the sociotechnical perspective includes the engagement of the technical and human components to achieve joint optimisation and generate an effective sociotechnical system within a given context (Dolata et al., 2022; Lee, 2004; Makarius et al., 2020; Sarker et al., 2019). Because the technical and social dimensions of ADM are mutually interdependent, the human-in-the-loop framing is insufficient because it primarily addresses control and oversight rather than the broader collaborative dynamics between human and algorithmic agents (Dolata et al., 2022; Lee, 2004; Makarius et al., 2020; Sarker et al., 2019). Dolata et al. (2022) call the engagement of the technical and the social components in ADM “human-algorithm-assembles as collective moral agents” (2022, p. 765).

Moreover, to include personal values in the debate about algorithmic fairness, Starke et al. (2022) and Rahwan (2018) propose the "society-in-the-loop" principle, which emphasises the need to integrate societal values and institutional contexts into ADM systems. The principle is defined as the "human-in-the-loop plus social contract" (Rahwan, 2018, p. 5), where the social contract encompasses conflicting values of various stakeholders, which has been left out in the original human-in-the-loop principle.

This sociotechnical framing supports the interaction between technology and humans because they both influence the outcome of ADM processes, for example, an algorithm-based risk assessment model and a judge (Dolata et al., 2022; Selbst et al., 2019). This framing facilitates the examination of fairness as an end-to-end process, incorporating both technical and social dimensions. For example, Selbst et al. (2019) argue that an output of an automated risk assessment in a single jurisdiction should not be interpreted as the outputs of the criminal justice system of the United States itself. This system is influenced by human decisions (judges in this example). Hence, it is important to also include a human perspective from all stakeholders affected by the decision. Nevertheless, some of the human decision-makers either ignore recommendations by the algorithm entirely, or they might not consider these recommendations consistently (Selbst et al., 2019; Starke et al., 2022).

Incorporation of Humans into the Algorithmic Decision-Making Process

Ensuring fairness in ADM necessitates incorporating a comprehensive representation of human decision-making into the system design (Selbst et al., 2019). With such systems, information, in particular data, constitutes a central element, as all algorithmic processes depend upon it. However, data cannot be regarded as independent or neutral. Hence, a critical view on data by humans is required to decrease the risk of relying on algorithms to make decisions alone (Chatterjee et al., 2021; Dolata et al., 2022). Teodorescu et al. (2021) suggest that the operators of the models, such as judges or managers, should be included in the development process of the algorithm alongside the actual developers (Pinch & Bijker, 1984). For instance, managers could define target fairness metrics aligned with organisational objectives, such as increasing the proportion of women hired. This approach could address issues like those observed in Amazon's hiring algorithm, which disadvantaged female candidates due to biased historical training data (Dastin, 2018). By setting an optimal fairness rate or potentially allowing a higher false-positive rate for women, decision-making could be better aligned with equity goals. Furthermore, human–algorithm collaboration can offset the limitations inherent in each part when considered in isolation, such as the human capacity to contextualise decisions and detect outliers, thereby reducing bias (Teodorescu et al., 2021). However, achieving an optimal fairness rate requires that managers possess a sufficient understanding of the algorithm's functioning (Grønsund & Aanestad, 2020; Teodorescu et al., 2021), which is known to be difficult because most algorithms are 'black boxes' (Sartori & Theodorou, 2022).

Teodorescu et al. (2021) define four different approaches to achieving a successful interaction between managers and machine learning applications. The first approach includes determining specific fairness objectives, selecting the definition of fairness the managers want to use, and deciding who they want to prioritise – for example, women – and communicating those things to the developers. The second approach stresses the need to support managers, that do not have a detailed technical knowledge, such as training and education programs, to ensure that they can provide proactive oversight. The third approach encompasses training and the determination of incentives for decision-makers to ensure fairness and informed reliance. The fourth approach highlights the importance of auditing decisions made by humans when they rely on machine learning systems.

3.4 Algorithm Aversion

In the following section, I outline the concept of algorithm aversion to establish an understanding of why individuals may be reluctant to integrate ADM systems into their work practices. Reluctance can arise from a range of factors, including perceptions that ADM systems are inherently unfair or biased, as well as limited understanding or familiarity with their operation. Developing an understanding of these factors is essential because it offers critical insight into how perceptions of ethical concerns, particularly those related to algorithmic fairness, can influence the adoption of algorithmic systems in HR practices, as discussed in Chapter 7.

3.4.1 Definitions of Algorithm Aversion

Prior literature has extensively discussed algorithm aversion, and this work has led to different definitions that reflect differences in research contexts and the selection of dependent variables (Jussupow et al., 2024). One of the earliest definitions of algorithm aversion describes an individual's preference for relying on their own or other humans' decisions rather than on algorithmic decisions (Dietvorst et al., 2015). Subsequent research has refined this conceptualisation by incorporating the role of perceived algorithmic errors and noting that individuals' attitudes may shift once they interact with algorithms and encounter their limitations. From this perspective, algorithm aversion can be understood as a reluctance to rely on algorithmic outputs or advice because the systems are perceived as imperfect (Burton et al., 2020). Another definition frames algorithm aversion as the conscious or unconscious rejection of algorithmic decisions in favour of human judgment (Mahmud et al., 2022), which encompasses a general predisposition against algorithms even prior to any direct interaction with them (Jussupow et al., 2024). For my research, I adopt this anticipatory form of algorithm aversion because it aligns with my focus on HR professionals, many of whom had not yet adopted ADM systems.

3.4.2 Reasons for Algorithm Aversion

As algorithmic bias and concerns about algorithmic fairness are key drivers of algorithm aversion, I will examine these alongside other factors that shape human reluctance to use ADM systems. These drivers include pre-existing beliefs about algorithms (Dietvorst et al., 2015; Jussupow et al., 2024), a lack of familiarity (Mahmud et al., 2022), a lack of trust due to their opacity (Omrani et al., 2022), fears of diminished control over decision-making (Mahmud et al., 2022; Schaap et al., 2024), resistance to change (Mahmud et al., 2022; Sutherland et al., 2016), and ethical concerns about potential bias or unfairness in algorithmic processes (Jauernig et al., 2022; Jones-Jang & Park, 2022).

Underlying Beliefs

Underlying beliefs about the differences between human and ADM strongly influence algorithm aversion (Dietvorst et al., 2015; Jussupow et al., 2024). Human decision-making is often associated with intuition informed by prior experiences, which individuals believe leads to decisions that are more accurate, innovative, and context specific. By contrast, ADM is frequently perceived as counterintuitive, impersonal, and standardised through predefined statistical parameters (Jussupow et al., 2024), despite often producing more consistent outcomes than human decision-making (Dietvorst & Bharti, 2020). A common criticism is that algorithms struggle to respond effectively to unforeseen situations that require flexibility and innovation (Dietvorst et al., 2015; Jussupow et al., 2024), which is further reinforced by individuals expecting decisions and predictions to be perfect (Dawes, 1979; Einhorn, 1986; Highhouse, 2008). However, even when algorithms outperform humans by committing fewer errors, people often maintain the belief that human decision-making is inherently more adaptable and capable of correcting mistakes more rapidly than algorithms, which are constrained by systematic and standardised processes (Dietvorst et al., 2015). Especially in ethical decision-making, human decision-making is preferred as humans are able to include morality and empathy (Jauernig et al., 2022), whereas algorithms are perceived to be less empathetic (Luo et al., 2019; Mahmud et al., 2022).

Lack of Familiarity

These general beliefs about human and ADM are shaped by multiple factors, including familiarity with algorithmic systems and demographic characteristics such as age. Research shows that individuals with prior experience of using algorithmic systems tend to exhibit lower levels of aversion and are more willing to rely on algorithmic outputs compared to those with no, or limited, exposure (Mahmud et al., 2022). In cases where individuals lack direct familiarity with ADM, their beliefs are often informed by an "algorithmic imaginary" (Bucher, 2017, p. 30) rather than by substantive knowledge (Jussupow et al., 2024). The concept of an algorithmic imaginary refers to how individuals envision algorithms in terms of their ideal form and intended functionality (Jussupow et al., 2024). Negative experiences with algorithmic systems have been shown to further intensify aversion (Mahmud et al., 2022). Age-related differences have also been observed, with older individuals more likely to perceive algorithms as ineffective or irrelevant (Araujo et al., 2020; Mahmud et al., 2022), which, in turn, is associated with lower trust in such systems (Lourenço et al., 2020).

Lack of Trust

Trust forms the foundation for a successful relationship between humans and non-humans, such as algorithmic systems, similar to relationships between humans (Madhavan & Wiegmann, 2007; Omrani et al., 2022). In this context, trust can be conceptualised as cognitive trust and emotional trust. Cognitive trust develops through factors such as reliability, tangibility, transparency, and task characteristics, whereas emotional trust arises from anthropomorphism, whereby human-like qualities are attributed to algorithmic systems, often leading individuals to overlook cognitive trust factors, such as transparency and reliability (Glikson & Woolley, 2020; Omrani et al., 2022). When new technologies, such as AI systems, are implemented, humans need to trust the unknown – something that they are unable to fully comprehend cognitively (Hoff & Bashir, 2015). This challenge is particularly pronounced for machine learning algorithms, such as neural networks, which are commonly described as "black boxes" due to their opacity. A lack of transparency is widely cited as a barrier to adoption, especially among users with limited understanding of ADM processes (Omrani et al., 2022). Compared to algorithms, human decision-makers are often trusted more because they can provide direct explanations for their reasoning, thereby enabling access to their thought process. In contrast, accessing the rationale behind algorithmic decisions is more difficult, particularly for users without technical expertise, thereby reducing trust. Increasing explainability and transparency is therefore critical to fostering cognitive trust in algorithmic systems (Mahmud et al., 2022; Omrani et al., 2022).

Emotional trust, in contrast, often reflects a general aversion to algorithmic systems, driven by broader distrust or negative preconceptions, such as the belief that machines should not be entrusted with decision-making (Mahmud et al., 2022; Prahl & Van Swol, 2021). Moreover, trust varies according to task type. Individuals tend to have higher trust in algorithms that perform objective tasks, based on quantifiable and measurable facts, compared to those that perform subjective tasks, which require intuition and interpretation. This difference is due to a perception that algorithms are less effective at handling subjectivity (Castelo et al., 2019). Consequently, a lack of trust, arising from opacity, limited understanding, and pre-existing beliefs, can significantly inhibit the adoption of algorithmic systems, with distrust often manifesting as bias against their use and resulting in non-implementation (Omrani et al., 2022; Turel & Kalhan, 2023).

Fear of Diminished Control

Controllability of algorithmic systems is another factor influencing individuals' willingness to acceptance and use algorithmic technologies (Schaap et al., 2024). Individuals generally seek to maintain control over their environment and the decisions being made (Dzindolet et al., 2002; Mahmud et al., 2022), especially in domains where they are typically held responsible for the outcome (Schaap et al., 2024). Algorithm aversion is therefore closely tied to the perceived agency of an algorithm. The means that the less influence users feel they have over a system, the lower their acceptance becomes, as their perceived lack of control undermines their trust and willingness to rely on the algorithm (Schaap et al., 2024). Since ADM systems are often perceived as offering less control than human decision-makers (Stowers et al., 2017), individuals may believe the locus of control lies outside the ADM system, and attribute responsibility instead to external stakeholders, such as the organisation implementing the technology, the end users, or the developers (Jones-Jang & Park, 2022; Shank & DeSanti, 2018). The perception of control can be enhanced through a better understanding of how the algorithm functions and performs, as well as through opportunities for limited intervention in the decision-making process without altering its core functioning. Consequently, both researchers and practitioners frequently advocate for a "human-in-the-loop" approach, allowing human oversight to be retained as a means of maintaining user agency and enhancing acceptance (Burton et al., 2020).

Resistance to Change

Another reason for algorithm aversion identified in the literature is a general aversion towards technologies, which can result from a bias against algorithms as well as the human tendency for egocentric trimming, meaning a general preference for prioritising one's own decisions over those of others (Mahmud et al., 2022; Sutherland et al., 2016). This tendency, combined with a reluctance to change established behaviours when new technologies are introduced, and a broader resistance to innovation, particularly when individuals are satisfied with existing work practices, can lead to algorithm aversion and resistance (Mahmud et al., 2022, 2023). This resistance, which occurs when individuals doubt the effectiveness of the innovation and feel insecure about its outcomes, results in delayed adoption or even active resistance. Active resistance refers to consciously opposing an innovation, whereas passive resistance occurs when individuals acknowledge the innovation but show no interest in adopting it (Mahmud et al., 2023; Ram & Sheth, 1989). Moreover, general aversion towards the implementation of ADM systems can be understood through functional and psychological barriers. Functional barriers relate to perceived changes associated with adoption, such as alterations in values or increased risks. For example, individuals may consider the innovation to be of low value or high risk. In contrast, psychological barriers are rooted in individual beliefs, often linked to tradition or perceived threats to one's image, that may conflict with adoption (Arif et al., 2020; Mahmud et al., 2023).

Ethical Concerns

Individuals also express concerns regarding the algorithm's efficacy in detecting potential discrimination and its adherence to fairness constraints (Jauernig et al., 2022; Jones-Jang & Park, 2022). Research also indicates that individuals tend to favour human decision-making over ADM because they perceive human decisions as more authentic and ethical (Jago, 2019; Jauernig et al., 2022). A common concern is that algorithms may perpetuate or exacerbate existing human biases and unfairness, given that they are developed by humans (Jauernig et al., 2022). Ethical concerns are particularly salient when algorithms are applied to subjective tasks that require moral judgement and ethical reasoning (Jauernig, et al., 2019; Mahmud et al., 2022; Lee 2018). These concerns arise because machines do not have empathy and make decisions on statistic parameters, which could lead to unfair decisions (Jauernig et al., 2022). Such perceptions are closely tied to trust, especially as past incidents have demonstrated that algorithms can produce discriminatory or unjust outcomes (Dastin, 2019), which undermines confidence in these systems (Omrani et al., 2022; Starke et al., 2022). Ethical concerns are often rooted in the presence of bias within systems, which stem from developers' own biases (Xivuri & Twinomurinzi, 2023) or from biased training data (Mahmud et al., 2023; Verma et al., 2021) and result in incorrect or unfair decisions for those affected. Moreover, tensions can arise when statistical definitions of fairness embedded by developers differ from the fairness concepts held by end-users or decision-subjects (Lee & Baykal, 2017).

3.5 Algorithmic Systems and Algorithmic Fairness in the HR Context

3.5.1 Algorithmic Systems in the HR Context

ADM in human resource management involves the application of algorithms to select one option from a set of possible options. Such systems provide informational outputs that assist or augment human decision-makers. Assisting includes providing information, such as descriptive statistics about the workforce, to inform decisions, which is often conceptualised as people analytics (Giermindl et al., 2022). Augmentation of decision-making includes generating predictive outputs that enable HR professionals to anticipate the potential consequences of current decisions on future outcomes (Meijerink et al., 2021) . These processes can be summarised as algorithmic management, or "the large-scale collection and use of data on a platform to develop and improve learning algorithms that carry out coordination and control functions traditionally performed by managers" (Möhlmann et al., 2021, p. 2001).

The concept of algorithmic management has been primarily explored in the context of the gig economy, which connects workers and customers for short-term contracts (Jabagi et al., 2019). However, the use of algorithmic systems in HR now occurs in a wide range of organisational contexts, influencing job design, autonomy, and employee well-being (Parent-Rocheleau & Parker, 2022; Sullivan et al, 2024). Algorithmic systems can either fully automate HR processes, replacing human involvement entirely (Leicht-Deobald et al., 2019), or operate collaboratively by automating specific tasks and supporting HR professionals in daily operations (Parent-Rocheleau & Parker, 2022). In the latter case, decision-making authority is shared between humans and algorithms (Tinguely et al., 2023).

The literature identifies the core applications of algorithmic systems in a variety of fields. Algorithms are widely used for monitoring, such as continuously tracking employees' behaviours and outputs to generate detailed, real-time insights (Parent-Rocheleau & Parker, 2022). Algorithms set goals by dynamically assigning tasks or performance targets (Lee et al., 2015; Parent-Rocheleau & Parker, 2022). Algorithms assess and rank employee performance and highlight areas for improvement (Budhwar et al., 2022; Manoharan et al., 2011), often in real time, while also forecasting potential future performance by employing diverse metrics, ranging from emotion analysis (Pinheiro et al., 2017), to internet browsing activity (Angrave et al., 2016), to social network engagement (Leicht-Deobald et al., 2019), to levels of work engagement (Burnett & Lisk, 2019). Scheduling systems optimise the allocation of shifts and tasks by factoring in workload, worker availability, and predicted demand (Hoshino et al., 2018; Parent-Rocheleau & Parker, 2022). In terms of compensation, algorithms calculate pay or rewards based on performance metrics, productivity measures, and customer feedback (Parent-Rocheleau & Parker, 2022; Rani & Furrer, 2021; Wood et al., 2019). Certain systems are capable of making automated termination decisions by deactivating workers whose performance falls below algorithmically determined thresholds, which is a practice long associated with the gig economy but now increasingly observed in other sectors (Parent-Rocheleau & Parker, 2022). Additionally, such systems can predict the likelihood of employee turnover, thereby enabling the development of targeted retention strategies (Meijerink et al., 2021).

In the traditional organisational HR context, the use of ADM systems has also been studies as part of core HR functions, such as recruitment processes, which include detecting talent and the attraction of potential candidates as part of the outreach stage (Hunkenschroer & Luetge, 2022), automatic screening or ranking of resumes to identify the most suitable job candidate (Budhwar et al., 2022; Hunkenschroer & Luetge, 2022; Rieskamp et al., 2023), and making future job performance predictions of job candidates (Bogen, 2019; Hunkenschroer & Luetge, 2022). Moreover, algorithms can replace the human interviewer during video interviews, and they can analyse interviews by providing insights into applicants' personality traits based on their movements and tone of voice (Hunkenschroer & Luetge, 2022; Köchling et al., 2021; Van Esch & Black, 2019). AI systems can also be used for recruitment administrative tasks, such as communicating with job applicants (Hunkenschroer & Luetge, 2022; Van Esch & Black, 2019), and parsing CVs to extract relevant information (Hunkenschroer & Luetge, 2022). In other HR operations, such as learning and development, algorithmic systems can deliver personalised training recommendations tailored to employees' existing skills and career goals (Tinguely et al., 2023).

Although ADM offers benefits, such as improved efficiency (Leicht-Deobald et al., 2019) and reduced HR operational costs (Vrontis et al., 2022), it simultaneously raises ethical concerns (Douglas et al., 2024) and may undermine the perceived integrity of decision-making processes (Leicht-Deobald et al., 2019). These risks are linked to the phenomenon of algorithm aversion, whereby individuals perceive algorithmic decisions to be less fair than those made by humans (Lee, 2018). Such perceptions – discussed in the Chapter 4 – can erode trust in, and increase resistance to the implementation of, ADM systems (Omrani et al., 2022).

3.5.2 Algorithmic Fairness in the HR Context

As outlined in Chapter section 3.3, algorithmic fairness has been studied from multiple different perspectives. Emerging from the Information Systems field, there has been a focus on investigating algorithmic fairness in the HR context. In this regard, most of the existing literature has concentrated on specific applications, particularly in recruitment, and has frequently addressed ethical concerns (Köchling & Wehner, 2020; Narayanan et al., 2024; Rieskamp et al., 2023; Robert et al., 2020; Starke et al., 2022; Fabris et al., 2024). The use of ADM systems has been associated with the expectation of reducing or eliminating human biases from HR processes (Martin, 2019). However, several examples have demonstrated that these technologies can reproduce, or even amplify, existing inequities. For instance, job advertisements have in some cases been disproportionately displayed to male users, thereby systematically excluding female candidates from equal opportunities to apply (Köchling & Wehner, 2020). Similarly, the Amazon hiring algorithm became a prominent example of gender bias when historical training data, reflecting past discriminatory hiring practices, led to the systematic downgrading of female applicants' resumes (Dastin, 2018).

While recruitment has been the most studied domain, fairness issues have also been observed in other HR functions. In performance evaluation, algorithmic rating can reproduce discrimination, particularly through racial stereotyping and gender bias, leaving workers disadvantaged and with limited opportunities to challenge such assessments (Kellogg et al., 2020). Moreover, algorithmic systems have been criticised for being opaque and not being able to recognise context. For example, if an algorithm cannot recognise certain obstacles that might make work more difficult, such as road blocks in the food delivery context, certain workers are disadvantaged and receive worse performance assessments (Meijerink et al., 2021). In scheduling, the use of algorithmic management has been criticised because it may result in social isolation and irregular working hours (Wood et al., 2019).

Research into algorithmic fairness in HR has predominantly investigated the perspectives of specific stakeholder groups and their perceptions of fairness. These stakeholders include employees (Zhou et al., 2023) and job applicants (Lavanchy et al., 2023) as decision-subjects, HR managers as end-users of algorithmic systems (Feldkamp et al., 2023), and software developers responsible for designing future algorithmic HR tools (Kasinidou et al., 2021; Kleanthous et al., 2022). Many of these studies have applied organisational justice theory to investigate perceived fairness, and they have focused on distributive, procedural, informational, and interactional justice (e.g., Acikgoz et al., 2020; Narayanan et al., 2024;

Newman et al., 2020; Ochmann et al., 2024). Within this framework, distributive and procedural justice have received the most scholarly attention (Robert et al., 2020).

Most of these studies find that, despite the expectation that ADM would produce more neutral and bias-free outcomes than human judgment, there is persistent scepticism among stakeholders (Martin, 2019; Newman et al., 2020). Individuals frequently express a preference for human decision-making over ADM, with purely algorithmic outcomes perceived as less fair than identical outcomes involving some form of human participation (Acikgoz et al., 2020; Lavanchy et al., 2023; Lee, 2018; Newman et al., 2020). Only very few studies show that individuals do not perceive differences in fairness perceptions between a human decision-maker or an ADM system (e.g., Suen et al., 2019).

The preference for human decision-making over ADM appears to be linked to several factors, including the belief that algorithms fail to capture qualitative attributes (Newman et al., 2020), the algorithm's inability to adapt to nuanced contextual circumstances (Grange, 2022; Longoni et al., 2019), and the absence of human intuition (Lee, 2018). Moreover, individuals' perceptions of fairness can be influenced by their level of algorithmic literacy. Specifically, those with limited familiarity may distrust algorithmic outputs more readily (Lavanchy et al., 2023). While involving a human decision-maker can mitigate negative perceptions, studies consistently show that purely human decisions are still regarded as more legitimate and fair, even when participants acknowledge human susceptibility to bias and error (Newman et al., 2020).

These perceptions connect closely to the phenomenon of algorithm aversion, where individuals tend to favour human judgment because they perceive it as more authentic, empathetic, and ethically grounded, regardless of its actual accuracy (Jago, 2019; Jauernig et al., 2022). In the HR context, this aversion may be amplified by the high-stakes nature of employment-related decisions, the personal implications for those affected, and the historical expectation that such decisions involve human deliberation and moral reasoning. Recent work also suggests that perceptions of fairness in algorithmic HR systems are not static but may evolve with increased exposure, transparency in system design, and opportunities for affected individuals to contest or appeal decisions (Binns et al., 2018; Lee, 2018). However, empirical studies investigating such longitudinal shifts remain scarce.

Existing research has predominantly examined algorithmic fairness either from a technical perspective, by focusing on formalised fairness metrics, or from a social perspective, by applying theories such as organisational justice (Dolata et al., 2022). Far less attention has been devoted to a sociotechnical perspective that acknowledges how fairness is shaped through the interaction of technical systems and social contexts, particularly by incorporating the views of multiple stakeholders (Dolata et al., 2022). To address this gap, my research seeks to explore how different stakeholders construct algorithmic fairness by examining the perspectives they adopt, and the factors that influence these constructions, through a sociotechnical lens. In the following sections, I outline this approach in greater detail, while noting that some concepts and literature, particularly regarding algorithmic fairness, are revisited in subsequent chapters to provide context for the empirical studies. Any repetition is intended to ensure clarity and continuity.

Chapter 4

Finding Algorithmic Fairness: An Analysis of how the Literature Constructs Algorithmic Fairness from Different Stakeholder Perspectives

Link to Thesis Narrative

In Chapter 1, I positioned my research in reference to algorithmic fairness and the not-yet-adoption of ADM systems in the organisational HR context. In Chapter 2, I outlined my relational ontological and epistemological position and research design. In doing so, I acknowledged that researchers can influence the interpretation of research outcomes (Charmaz, 2006, 2008). In Chapter 3, I provided an overview of background information on ADM systems, algorithmic fairness, and their use in HR practice, which this Chapter draws upon. The literature reviewed demonstrates that past research has predominantly focused on either the technical or the social dimensions of algorithmic fairness in isolation. However, scholars have increasingly called for an integrated approach that considers these dimensions jointly. This includes in particular the adoption of a sociotechnical perspective to better understand and optimise algorithmic fairness (Dolata et al., 2022). Moreover, there has, to date, been very limited use of social constructionist theory to study algorithmic fairness. This lack of social constructionist focus limits our capacity to understand what shapes different understandings of algorithmic fairness and the associated concerns about the use of ADM systems. This gap provoked me to explore how past studies have investigated algorithmic fairness and which stakeholders they have identified. By conducting a thematic analysis of the literature (Leidner, 2016), I highlight the different perspectives from which algorithmic fairness has been explored in research as well as how algorithmic fairness is considered and understood depending on the stakeholder perspective taken. Consequently, this chapter advances the understanding of algorithmic fairness in literature and the diverging scholarly narratives around the phenomenon. The work presented in this chapter positions my empirical research on algorithmic fairness, reported in Chapter 5, Chapter 6 and Chapter 7, in reference to my relational onto-epistemic worldview. Furthermore, this chapter highlights how different stakeholders prioritise aspects of algorithmic fairness, providing guidance for the design of ADM systems, informing policymakers and regulators on how to communicate algorithmic fairness.

State: Published in Australasian Conference on Information Systems (ACIS) Proceedings 2024.

Abstract

Algorithms are increasingly used by organisations to make decisions or recommendations about their operations. This can include operations, such as credit lending or recruitment. To date, algorithmic fairness has been discussed from multiple different lenses in the literature. However, Information Systems understand algorithmic fairness within a sociotechnical system involving different stakeholders. We thus need to understand how existing research socially constructs algorithmic fairness from different stakeholder perspectives in organisational contexts. Our literature review outlines what stakeholders are discussed and how they construct algorithmic fairness. We show that different aspects of algorithmic fairness are considered by different stakeholders. We also show that employees, as decision-subjects in an ADM system, are predominantly used as study objects to derive implications about algorithmic fairness for management and developers.

4.1 Introduction

Algorithms are increasingly used by organisations to make decisions or recommendations about their operations (Kordzadeh & Ghasemaghaei, 2022). This can include processes like credit lending (Saxena et al., 2019) or human resources (HR) operations, such as recruitment and selection processes or performance assessments (Gal et al., 2019; Giermindl et al., 2022). Therefore, managers increasingly rely on algorithmic decision-making (ADM), which have the risk to be biased and treat humans unfairly (Lee, 2018). With multiple cases arising, in which an algorithm discriminated a particular group of people or individuals (Giermindl et al., 2022), scholars have engaged in ongoing discussions about what algorithmic fairness (AF) is and how it can be ensured. This includes understanding AF from a technical standpoint but also the social notions of fairness of AF, such as distributive or procedural fairness perceptions (Lee, 2018) or philosophical reasoning (Feldkamp et al., 2023).

Information Systems (IS) understands AF within a sociotechnical system. A sociotechnical system implies that technology, such as the algorithm itself, as well as social aspects, which include different stakeholders, interact with each other (Dolata et al., 2022). We argue that the understanding of AF depends not only on technical or social notions of fairness but is also shaped by different stakeholder perspectives. These stakeholders include groups that have a direct impact on the algorithm or are somehow affected by it, such as the algorithm's developer, the algorithm's users, e.g., management, an algorithm's decision-subjects, e.g., employees, or regulators (Dolata et al., 2022). These various stakeholders can have conflicting interests, which can cause differences in the construction of AF. For example, developers might design ADM systems with a broad application in mind and therefore prioritising technical fairness notions over specific contexts (Dolata et al., 2022). Conversely, individuals affected by the decision, might have a different understanding of fairness depending on their own context (Dolata et al., 2022).

Moreover, fairness can have a different meaning for different individuals. Thus, what is fair for one individual does not have to be fair for another individual as their experiences, context as well as social and technical interactions shape their meaning of fairness (Feldkamp et al., 2023) For instance, for a developer an algorithmic decision might be understood as fair as long as it applies technical fairness notions correctly. However, the decision itself does not have to be perceived as fair by individuals that are directly affected by the decision, because the decision might not take context into account (Morse et al., 2022). Therefore, depending on the stakeholder group, we might have different fairness outcomes. Thus, if we understand AF as socially constructed, we cannot solely focus on technical or social notions of fairness. Instead, we should include the impact stakeholders themselves might have on the understanding of AF. This is because the meaning of fairness can vary across stakeholders which is shaped by the context they are in, their interests, the experiences they have made, as well as existing social and technological interactions. Therefore, different stakeholders will construct their own conception of AF (Berger & Luckmann, 1966; Charmaz, 2008).

While several literature reviews discuss what AF is and how it can be defined by either considering technical, social, or sociotechnical aspects or by only focusing on a specific stakeholder group (e.g., Köchling & Wehner, 2020; Narayanan et al., 2024; Robert et al., 2020), none of the literature reviews have investigated how AF is constructed in the existing literature depending on the stakeholder's perspective that is taken. As we want to understand how AF is socially constructed, we should include importance of the stakeholder's perspective that is taken in the literature. Therefore, we ask the following research questions: What stakeholder groups are considered in existing research?; How does existing research construct AF from different stakeholder perspectives in the organisational context?

To do so, we analysed existing research that discuss the organisational context or could be applied to an organisational context. Based on our analysis, we identified different stakeholder groups to categorise how AF is constructed by the literature and discuss how AF is constructed by these.

The remainder of the papers is organised as follows. By analysing literature reviews about AF published to date, we provide the underlying research background, which emphasises the motivation for our research. We then explain our methodology to conduct the literature review and analyse the selected papers. We discuss our results per identified stakeholder group. Our paper ends with a discussion of our results and a conclusion.

4.2 Research Background

To provide an overview of the underlying research background, we outline in the following, how AF is currently discussed in the literature. We further show what existing literature reviews about AF have investigated, providing a motivation for our research.

Existing research mostly focuses on defining AF from different lenses, such as a technical lens, a social lens and a sociotechnical lens. AF from a technical lens has emerged from disciplines, like Computer Sciences, where different statistical notions of fairness are considered to investigate under which condition algorithms are technically fair. Group fairness treats individuals, that are part of a protected group based on sensitive attributes, such as gender or race, the same as people that belong to any other group (Calvi & Kotzinos, 2023; Dolata et al., 2022; Juijn et al., 2023). It includes technical notions of fairness based on protected attributes, such as statistical parity or disparate impact, and technical notions based on prediction errors, such as equalised odds or equality of opportunity (Dolata et al., 2022; Morse et al., 2022). Individual fairness considers whether similar individuals are treated similarly regardless of their membership to a particular group (Dolata et al., 2022; Hertweck & Heitz, 2021). Research about AF from a social lens has emerged as a solely technical view on AF does not consider influences of social contexts, even though algorithms are predominantly applied in social contexts (Robert et al., 2020). Therefore, AF is discussed based on philosophical perspectives, such as the concepts of equity and equality or from psychological aspects, such as perceptions of fairness by individuals (Dolata et al., 2022). Moreover, AF can also be analysed from a sociotechnical lens, as analysing technical aspects and social aspects in separation might not be appropriate enough as ADM is neither a solely technical task, nor a solely social task. Therefore, AF is understood as the combination of technical aspects and social aspects while considering the interactions between technical and social aspects (Dolata et al., 2022).

By analysing existing literature reviews about AF, we find that they either focus on the perception of decision-subjects, such as employees, or that they do not focus on a selected stakeholder group at all. Moreover, some literature reviews focus on specific contexts, such as the recruitment context, whereas others look at a broad organisational context. Therefore, we found that none of the literature reviews compare different stakeholder groups and discuss how AF is constructed by existing literature based on the stakeholder perspective (see Table 4.1).

Table 4.1: Overview of Current Literature Reviews on AF

Literature Review	Theme	Context	Focused Stakeholder Group
Giermindl et al. (2022)	Perils of People Analytics Use	Organisational Context	Organisations & Employees
Heyder et al. (2023)	Ethical Management of Human-AI Interaction	Organisational Context	No focus group
Hunkenschroer & Luetge (2022)	Ethicality of AI Applications	HR Recruitment & Selection	No focus group
Köchling & Wehner (2020)	Unfairness & Discrimination of Algorithmic Decision-Making Systems	HR Recruitment & Development	No focus group
Kordzadeh & Ghasemaghaei (2022)	Algorithmic Bias, Fairness Perceptions	Organisational Context	No focus group
Lämmermann et al. (2022)	AI Fairness at a Subgroup-Level	No clear context	No focus group
Narayanan et al. (2024)	Fairness Perceptions	Organisational Context	Decision-Subjects (e.g., employees)
Rieskamp et al. (2023)	Strategies to address Unfairness	HR Hiring	No focus group
Robert et al. (2020)	AI & Organisational Justice Types	Organisations	No focus group
Starke et al. (2022)	Fairness Perceptions	Organisational Criminal Justice, University Admissions, Loan Decisions, Targeted Advertisement, Allocations of Donations Context	No focus group
Zhou et al. (2023)	Negative Effects of AI-enabled Human Resource Management	Organisational Context	Employees

Existing literature reviews on fairness perceptions structure their review according to the organisational justice theory, which is concerned with explaining individuals' reactions to unfairness or inequities (Robert et al., 2020) and which "reflects the degree to which one's company or top management is perceived to act consistently, equitably, respectfully, and truthfully in decision contexts" (Kordzadeh & Ghasemaghaei, 2022, p.396). Applying the organisational justice theory helps to categorise and analyse fairness perceptions of different individuals, which is important as they can impact the effective implementation of algorithms in workplaces (Narayanan et al., 2024). In addition to analysing different types of fairness perceptions, ways to redress unfairness are also discussed in the literature. Applying the concepts of retributive fairness and restorative fairness helps to understand what needs to be done to restore fairness once unfairness is perceived by individuals (Robert et al., 2020). To restore fairness or to enhance the perception of fairness, transparency, voice and explainability are discussed, as they can increase trust in an algorithm and the ability to interact with ADM systems (Narayanan et al., 2024; Robert et al., 2020).

Moreover, HR operations discussed in the literature mainly include the review of ethics of AI-enabled recruiting, which shows that ethical principles, such as privacy and confidentiality, transparency and respect, should be applied in those operations, which is also supported by laws and regulations that consider privacy and anti-discrimination (Hunkenschroer & Luetge, 2022; Köchling & Wehner, 2020). Reviewing the negative effects of AI-enabled HR operations reveals that the use of AI, such as People Analytics (PA) tools, can illude the feeling of having control as the users believe that data represents the reality accurately. Moreover, PA tools promise to exclude human prejudices and thus increase fairness. However, as the underlying assumptions of those tools become untraceable and hard to understand for the users, transparency is reduced, which can lead to bigger ethical and moral challenges (Giermindl et al., 2022; Zhou et al., 2023).

Considering AF from a subgroup-level in the technical perspective has emerged in the literature, as scholars criticise that a two-level technical perspective, that only considers group-level fairness and individual-level fairness, is not sufficient (Lämmermann et al., 2022). Applying a subgroup-level perspective helps to consider multiple dimensions simultaneously, which can increase the understanding of AF as a whole. Moreover, the technical lens on AF in the literature also discusses measures to mitigate algorithmic bias and therefore enhance AF, such as pre-processing, in-processing or post-processing measures (Lämmermann et al., 2022).

4.3 Research Approach

To answer the research question about how existing research constructs AF from different stakeholder perspectives in the organisational context, we conducted an organising review, which aims to make a vast body of literature from various disciplines comprehensible in order to synthesise the findings in the literature (Leidner, 2018). Our search databases included Scopus, Web of Science and Academic Search Ultimate. They cover the major academic journals from the domains of IS, Management & Organisational Science, Ethics and Computer Science. Furthermore, we manually checked the main IS conferences – ICIS, ECIS, AMCIS, PACIS, ACIS – as well as the leading conferences on Artificial Intelligence (AI) – KDD, FAccT, ICML, and HCI.

We conducted a search based on the keywords “algorithmic fairness”, “AI fairness”, “algorithmic bias”, “AI bias”, “algorithmic justice”, “AI justice”, “algorithmic injustice” and “AI injustice”. Because fairness and justice are often used interchangeably in the literature (Ochmann et al., 2024) and because AF seeks to mitigate bias (Dolata et al., 2022), we included those terms in our keyword search. We searched for papers published between 2018 and 2024, as the topic of AF only became much more prominent in the recent years (van Berkel et al., 2023). The search was carried out in January 2024. After applying category filters, which included Computer Sciences, Business & Management, Law, Social Sciences, Decision Sciences and Political Sciences, the searches overall yielded 2,806 hits. After scanning the titles of each paper for relevance and removing duplicates, the number of papers narrowed down to

420. Papers that clearly addressed our keywords were included as well as papers that implicitly applied to our keyword search, for instance, when discrimination was discussed. Papers that clearly addressed contexts, such as healthcare and court decisions, were excluded, as our research aims to address how AF is constructed in business corporations. We then analysed the abstracts of the remaining papers, which lead to a result of 198 papers. After reading through the papers, we excluded papers that focused only on the technical aspects of AF and papers that could not be applied to a business corporation context. This led to a final sample of 27 papers.

We read the final sample of papers to identify and analyse which stakeholders are discussed and how AF is considered by the literature depending on the identified stakeholder group. The identification of stakeholders included stakeholders, that were directly studied, such as job applicants as study objects, and stakeholders that were addressed by the papers, such as management. We then analysed the papers with regard to their construction of AF, particularly, whether they construct AF from a technical lens, a social lens or a sociotechnical lens, based on the respective stakeholder group. Based on this analysis, we organised the papers with regard to the respective stakeholder group and their construction of AF (Leidner, 2018).

4.4 Results

We identified five different stakeholder groups that are addressed by different papers. These include management, as the algorithm's users, employees, as the algorithm's decision-subjects, developers, regulators, and an organisation understood as a whole.

Managerial Perspective

Existing research, that has considered AF from a managerial perspective, wants to address management's concerns about performance and reputation. As higher perceived AF positively correlates with organisational commitment and performance (Newman et al., 2020; Yu et al., 2023) as well as how satisfied individuals are with the decision-maker (Ochmann et al., 2024), the managerial perspective constructs AF through the perception of fairness by individuals that are affected by a decision, i.e., the decision-subject. In the literature, this includes employees and job applicants (e.g., Juijn et al., 2023; Lavanchy et al., 2023). Therefore, to increase organisational performance, management needs to understand how they can treat their individuals fair to, for instance, attract employees and to retain talent (Newman et al., 2020).

Moreover, AF from the managerial perspective is also influenced through AF perceptions of the decision-makers, for examples HR management who relies on decisions or recommendations made by algorithms (Feldkamp et al., 2023). To understand differences in decision-subjects' perceptions and decision-makers' perceptions, past research (Feldkamp et al., 2023; Lavanchy et al., 2023; Newman et al., 2020) have investigated different types of decision-making systems, which include human-only decisions, algorithm-only decisions and human-algorithm decisions. The results overall indicate that individuals, both decision-subjects and decision-makers, perceive human-only decisions as fairer than algorithm-only decisions. They further indicate that, even though ADM processes are perceived to derive more consistent decisions compared to humans and thus are perceived to be procedurally fairer, consistency can lead to a failure to consider and contextualise a decision-subject's qualitative attributes, such as potential or leadership skills and therefore are perceived as reductionistic by the decision-subject and the decision-maker (Feldkamp et al., 2023; Newman et al., 2020). Moreover, the findings have implications for management as algorithm aversion by decision-subjects can lead to lower organisational commitment and a lower performance of individuals, which overall impact an organisation's performance (Newman et al., 2020; Yu et al., 2023). Therefore, management needs to ensure that decision-subjects, such as employees or job applicants perceive the ADM processes and outcomes as fair.

To avoid the risk of excluding context and qualitative characteristics, research has also investigated how individuals perceive algorithmic decisions that are augmented by humans.

Augmented decisions are perceived fairer compared to algorithm-only decisions because individuals perceive that humans can still intervene a decision solely made by an algorithm and therefore can take qualitative features and contexts into account. However, the perceptions of procedural fairness of such a hybrid process depends on the degree of which humans can control the algorithmic decision. If a human decision-maker can only adjust the algorithmic decision but overall has to rely on it, the decision is still perceived as unfair, whereas when a human has the option to consider the algorithmic decision, the decision-making process is perceived as fairer (Lavanchy et al., 2023; Newman et al., 2020).

Developer Perspective

AF from a developer perspective in the literature is mainly constructed through technical notions of fairness, followed by AF perceptions of an algorithm's decision-subjects and users as they can influence an algorithm's development process and therefore have implications for developers (Dolata et al., 2022; Morse et al., 2022). Technical notions of fairness are categorised into group fairness, individual fairness and subgroup fairness (Fazelpour & Lipton, 2020; Fleisher, 2021; Lämmermann et al., 2022). Group fairness treats individuals that are part of a protected group based on sensitive attributes, such as gender or race, the same as people that belong to any other group (Calvi & Kotzinos, 2023; Dolata et al., 2022; Juijn et al., 2023; Morse et al., 2022). Individual fairness considers whether similar individuals are treated similarly regardless of their membership to a particular group (Dolata et al., 2022; Hertweck & Heitz, 2021). Subgroup fairness considers multiple dimensions simultaneously, which can increase the understanding of AF as a whole (Lämmermann et al., 2022).

As different technical notions of fairness cannot be enforced all together at the same time, a developer chooses which fairness metric will be applied, which entails a moral justification and is highly context dependent (Hertweck & Heitz, 2021; Scantamburlo, 2021). Therefore, philosophical reasoning is additionally considered in the literature. For example, the application of statistical parity might be morally justifiable if variations in predictive features stem from unjust circumstances, such as historical injustices, potentially resulting in unjust disparities if left unaddressed, and if an intervention would not adversely affect the already disadvantaged group (Hertweck & Heitz, 2021).

Moreover, when considering AF from a developer perspective, existing research also takes the influence of laws and regulations into account, as technical notions of fairness are derived from legal aspects. This includes anti-discrimination laws, equality laws and data protection laws (Morse et al., 2022; Weerts et al., 2023). Therefore, most of the technical notions of fairness have been designed based on legal principles and ethical values, which impacts the construction of fairness in the developer perspective as well (Calvi & Kotzinos, 2023; Morse et al., 2022; Weerts et al., 2023).

However, a technical-only construction of AF might not be sufficient as ADM systems do not only consist of technical elements. Therefore, AF in the developer perspective is also considered through the sociotechnical theory, which considers technical and social aspects and their interdependencies within a system (Dolata et al., 2022). Stakeholders of an ADM system are social elements of a sociotechnical system. Therefore, perceptions of AF of different stakeholders are considered in the from a developer perspective, as their perceptions can influence the development and therefore how AF is constructed in the developer's perspective (Dolata et al., 2022; Heyder et al., 2023; Morse et al., 2022). Investigating decision-subjects' perceptions of technical AF notions reveal that treating similar individuals similarly is not necessarily perceived as fair because it excludes an individual's merit and contextual factors (Saxena et al., 2019), which is also supported by developers' perception of fairness (Kasinidou et al., 2021).

Employee Perspective

The employee perspective includes present employees and potential future employees, such as job applicants, in their role of decision-subjects, that are directly affected by an algorithmic decision or recommendation. The literature considers AF from an employee perspective through the employee's perception of AF (Acikgoz et al., 2020; Juijn et al., 2023; Lavanchy et al., 2023; Newman et al., 2020; Ochmann et al., 2024; Wang et al., 2020; Yu et al., 2023). The results indicate that even though humans tend to be prejudiced, decision-subjects perceive human decision-making systems as fairer compared to ADM systems, especially in recruitment and termination contexts, particularly because of the lack of transparency and the exclusion of personal factors. Even though present and potential future employees play a crucial role in the context of AF as they are directly affected by the algorithmic decision, the literature does not consider implications for decision-subjects. The implications of these studies only focus either on the actions that management has to implement to ensure fair treatment of decision-subjects or how decision-subjects' AF perception can influence the development process.

Regulator Perspective

Existing research on AF, that focuses on a regulator perspective, constructs AF through laws and regulations (Binns & Kirkham, 2021; Colmenarejo et al., 2022; Hoffmann et al., 2022; van Bekkum & Zuiderveen Borgesius, 2023; Weerts et al., 2023). As laws and regulations are the foundation for fairness metrics (Weerts et al., 2023), it is crucial to discuss how AF is constructed in the literature that takes a regulator's point of view. Laws and regulations that are focused on in the AF discussion are anti-discrimination laws and data protection laws.

Anti-discrimination laws prohibit direct and indirect discrimination. Direct discrimination by an ADM system includes treating an individual less favourably than another individual based on protected attributes, such as race and ethnicity. Indirect discrimination by an ADM system arises when an ADM system seems to be theoretically fair based on technical notions of fairness but eventually discriminates against protected groups or individuals. For example, fairness through unawareness, which is in theory a fair notion of fairness, can in practice identify correlations between attributes, which results in discrimination (Binns & Kirkham, 2021; Colmenarejo et al., 2022; Hoffmann et al., 2022; van Bekkum & Zuiderveen Borgesius, 2023; Weerts et al., 2023). To address discrimination, organisations are allowed to apply an exemption to the law that allows positive action but should not lead to positive discrimination against the protected group (Binns & Kirkham, 2021).

Data protection laws and regulations aim to protect personal and sensitive information, like political opinions, race and ethnicity or religious beliefs. Therefore, by protecting these types of information, other rights, such as anti-discrimination (van Bekkum & Zuiderveen Borgesius, 2023) or disability human rights law (Binns & Kirkham, 2021), are also protected. As an organisation must adhere to both, anti-discrimination laws and data protection laws, it is not allowed to collect protected attributes, unless specific consent is given by the data subject. When an organisation wants to assess whether its ADM accidentally discriminates based on protected attributes, such as ethnicity, collecting those data might be useful to audit the ADM, which is why potential exceptions to the law are discussed (van Bekkum & Zuiderveen Borgesius, 2023).

Organisational Perspective

Some literature focuses on the organisation as a whole without distinguishing between particular stakeholders (Dolata et al., 2022; Draude et al., 2020; Gkeredakis, 2022; Marjanovic et al., 2022). Therefore, we identified the organisation as a stakeholder itself. In this perspective, AF is constructed from a sociotechnical view where the interdependencies of technical aspects, social aspects as well as the organisation’s context, including an organisation's culture, should be considered and analysed as a whole system (Dolata et al., 2022; Draude et al., 2020).

When analysing AF from an organisational perspective by applying a performative sensemaking lens, AF can either displace, complement, or overrule existing practices and meanings of fairness. Displacing existing practices could include, for example, relying on ADM systems to overrule human decisions. Using ADM systems' recommendations for consulting purposes to generate new insights and combining them with human judgement is an example of complementing existing practices. Overruling existing practices includes the enactment of AF in a counter-performative way. This means using an ADM system but rejecting its decisions or recommendations, which makes AF irrelevant (Gkeredakis, 2022). Therefore, AF in the organisational perspective is constructed through its enactment, which is dependent on the interactions between technological aspects and social aspects. This indicates that the construction of AF and its meaning in practice is highly dependent on the context and its different actors. Moreover, it is subject to change, as actors and their perception of AF and their actions do not remain the same over time and as an organisation's context constantly changes (Dolata et al., 2022; Gkeredakis, 2022; Marjanovic et al., 2022).

4.5 Discussion

By reviewing literature about AF, we were able to show that existing research constructs AF fairness differently depending on the stakeholder group. Based on our literature analysis, we identified five stakeholder groups covered in the literature. Our results show that depending on the stakeholder perspective taken, different aspects of AF are emphasised in the construction of AF in the literature. These findings are important when we want to understand AF as sociotechnical construct identifying how notions of fairness are contingent on the socially perceived affordances of technology.

We were able to show that some aspects, in which AF is constructed, are similar across some stakeholder groups. For example, AF from both, a developer perspective and a managerial perspective, is constructed through the fairness perceptions of decision-subjects. Moreover, we showed that there is no one-size-fits-all approach to construct AF, but that multiple aspects have implications for the construction of AF in a particular group. For example, the developer's perspective is not only constructed through technical notions of fairness, but also legal implications and fairness perceptions.

In general, papers that construct AF from a managerial perspective and from a developer perspective dominate the literature. Particularly the construction of AF from a developer's point of view is highly discussed in the literature, as developers are the origin of the algorithm's creation and therefore it is important to understand how AF is constructed from their perspective in the first place before it is implemented in algorithms used by organisations.

In general, we could see that AF literature highly focuses on fairness perceptions by decision-subjects, particularly job applicants. Nevertheless, these decision-subjects are only used in papers to study different aspects of AF in order to derive implications for management or developers. The reason for this could be that decision-subjects are just affected by a decision and not able to make a decision themselves, which means that they still do not have a say on ADM systems. Therefore, implications of AF are either addressed to management that can make decisions and change organisational structures to, for instance, provide more transparency to decision-subjects (Lavanchy et al., 2023) or addressed to developers, that can address employees' perceptions of fairness in the development process (Morse et al., 2022).

We noticed that, while focusing on papers that specifically consider an organisational context or that could be applied to an organisational context, the reviewed papers mainly focus on HR operations, especially hiring processes. This domination of papers that consider HR operations might be because HR operations are naturally prone to discrimination because every

HR decision considers individual differences between people. Therefore, there is an inevitable discrimination part included (Cascio & Aguinis, 2013; Köchling & Wehner, 2020). We are aware that including papers from other domains, such as healthcare or judiciary, might have revealed different results in terms of the construction of AF. However, we decided to exclude papers from other domains, because we wanted to focus on the organisational context and its relevant stakeholders. Including other domains might have revealed different results in terms of the construction of fairness. However, as the context shapes the social construction of AF (Berger & Luckmann, 1966), the results in different contexts are not applicable for our research.

4.6 Conclusion

Our literature review provides new insights into how existing literature constructs AF from different stakeholder perspectives in the organisational context. By conducting a literature review on AF, we identified five stakeholder perspectives: the managerial perspective, the developer perspective, the employee perspective, the regulator perspective, and the organisational perspective. Our results indicate that AF is constructed differently in the literature depending on the stakeholder's perspective, even though there are some aspects of AF that are similar across different stakeholders. Our findings further indicate that developers and management are mainly discussed in the literature. Moreover, the findings reveal that employees as decision-subjects are stakeholders of an ADM system. However, they are mostly used to study perceptions of AF to derive implications for management or developers and not for employees themselves. Having this knowledge is important especially when we understand AF as socially constructed that includes various stakeholders Therefore, our literature review contributes to theory by filling this knowledge gap, especially by drawing awareness to the importance of including various understandings of AF depending on the stakeholder's perspective when researching AF. We contribute to practice as our results show which aspects of AF are considered by different stakeholders. This helps to understand how stakeholders might emphasise different aspects of AF. This can be especially important when developing and assessing ADM systems for AF in practice. Moreover, it helps policy makers and regulators to draw attention to the impact of stakeholders with regard to their understanding of AF.

4.7 References

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

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality*.

Binns, R., & Kirkham, R. (2021). How could equality and data protection law shape AI fairness for people with disabilities? *ACM Transactions on Accessible Computing, 14*(3), 1–32. https://doi.org/10.1145/3473673

Calvi, A., & Kotzinos, D. (2023). Enhancing AI fairness through impact assessment in the European Union: A legal and computer science perspective. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1229–1245. https://doi.org/10.1145/3593013.3594076

Cascio, W. F., & Aguinis, H. (2013). *Applied psychology in human resource management*. Pearson Education.

Charmaz, K. (2008). Constructionism and the grounded theory method. In J. A. Holstein & J. F. Gubrium (Eds.), *Handbook of constructionist research* (pp. 397–412). Guilford Press.

Colmenarejo, A. B., Nannini, L., Rieger, A., Scott, K. M., Zhao, X., Patro, G. K., Kasneci, G., & Kinder-Kurlanda, K. (2022). Fairness in agreement with European values: An interdisciplinary perspective on AI regulation. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 107–118. https://doi.org/10.1145/3514094.3534158

Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal, 32*(4), 754–818. https://doi.org/10.1111/isj.12370

Draude, C., Klumbyte, G., Lücking, P., & Treusch, P. (2020). Situated algorithms: A sociotechnical systemic approach to bias. *Online Information Review, 44*(2), 325–342. https://doi.org/10.1108/OIR-10-2018-0332

Fazelpour, S., & Lipton, Z. C. (2020). Algorithmic fairness from a non-ideal perspective. *AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 57–63. https://doi.org/10.1145/3375627.3375828

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*, 1–16. https://doi.org/10.1080/1359432X.2023.2169140

Fleisher, W. (2021). What's fair about individual fairness? *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 480–490. https://doi.org/10.1145/3461702.3462621

Gal, U., Riemer, K., Harth-Kitzerow, C., Aboud, C., & Briggs, S. (2019). *People analytics: Using data and algorithms to shape the employee experience*. https://ses.library.usyd.edu.au/bitstream/handle/2123/20082/People%20Analytics%20Report%202019.pdf?sequence=4&isAllowed=y

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

Gkeredakis, M. (2022). Fair algorithms in organizations: A performative-sensemaking model. *ICIS 2022 Proceedings*.

Hertweck, C., & Heitz, C. (2021). A systematic approach to group fairness in automated decision making. *2021 8th Swiss Conference on Data Science (SDS)*, 1–6. https://doi.org/10.1109/SDS51136.2021.00008

Heyder, T., Passlack, N., & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. *The Journal of Strategic Information Systems, 32*(3), 101772. https://doi.org/10.1016/j.jsis.2023.101772

Hoffmann, H., Vogt, V., Hauer, M. P., & Zweig, K. (2022). Fairness by awareness? On the inclusion of protected features in algorithmic decisions. *Computer Law & Security Review, 44*, 105658. https://doi.org/10.1016/j.clsr.2022.105658

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

Juijn, G., Stoimenova, N., Reis, J., & Nguyen, D. (2023). Perceived algorithmic fairness using organizational justice theory: An empirical case study on algorithmic hiring. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 775–785. https://doi.org/10.1145/3600211.3604677

Kasinidou, M., Kleanthous, S., Barlas, P., & Otterbacher, J. (2021). I agree with the decision, but they didn't deserve this: Future developers' perception of fairness in algorithmic decisions. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 690–700. https://doi.org/10.1145/3442188.3445931

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

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems, 31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Lämmermann, L., Zwickel, A., & Markgraf, M. (2022). AI fairness at subgroup level – A structured literature review. *ECIS 2022 Proceedings*.

Lavanchy, M., Reichert, P., Narayanan, J., & Savani, K. (2023). Applicants' fairness perceptions of algorithm-driven hiring procedures. *Journal of Business Ethics, 188*(1), 125–150. https://doi.org/10.1007/s10551-022-05320-w

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society, 5*(1), 205395171875668. https://doi.org/10.1177/2053951718756684

Leidner, D. (2018). Review and theory symbiosis: An introspective retrospective. *Journal of the Association for Information Systems, 19*(6), 552–567. https://doi.org/10.17705/1jais.00501

Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2022). Theorising algorithmic justice. *European Journal of Information Systems, 31*(3), 269–287. https://doi.org/10.1080/0960085X.2021.1934130

Morse, L., Teodorescu, M. H. M., Awwad, Y., & Kane, G. C. (2022). Do the ends justify the means? Variation in the distributive and procedural fairness of machine learning algorithms. *Journal of Business Ethics, 181*(4), 1083–1095. https://doi.org/10.1007/s10551-021-04939-5

Narayanan, D., Nagpal, M., McGuire, J., Schweitzer, S., & De Cremer, D. (2024). Fairness perceptions of artificial intelligence: A review and path forward. *International Journal of Human–Computer Interaction, 40*(1), 4–23. https://doi.org/10.1080/10447318.2023.2210890

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

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/10.1111/isj.12482

Rieskamp, J., Hofeditz, L., Mirbabaie, M., & Stieglitz, S. (2023). Approaches to improve fairness when deploying AI-based algorithms in hiring – Using a systematic literature review to guide future research. *Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2023.026

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human–Computer Interaction, 35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Saxena, N., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D., & Liu, Y. (2019). *How do fairness definitions fare? Examining public attitudes towards algorithmic definitions of fairness* (arXiv:1811.03654). arXiv. http://arxiv.org/abs/1811.03654

Scantamburlo, T. (2021). Non-empirical problems in fair machine learning. *Ethics and Information Technology, 23*(4), 703–712. https://doi.org/10.1007/s10676-021-09608-9

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society, 9*(2), 205395172211151. https://doi.org/10.1177/20539517221115189

van Bekkum, M., & Zuiderveen Borgesius, F. (2023). Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception? *Computer Law & Security Review, 48*, 105770. https://doi.org/10.1016/j.clsr.2022.105770

van Berkel, N., Sarsenbayeva, Z., & Goncalves, J. (2023). The methodology of studying fairness perceptions in artificial intelligence: Contrasting CHI and FAccT. International *Journal of Human-Computer Studies, 170*, 102954. https://doi.org/10.1016/j.ijhcs.2022.102954

Wang, R., Harper, F. M., & Zhu, H. (2020). *Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences* (arXiv:2001.09604). arXiv. http://arxiv.org/abs/2001.09604

Weerts, H., Xenidis, R., Tarissan, F., Olsen, H. P., & Pechenizkiy, M. (2023). Algorithmic unfairness through the lens of EU non-discrimination law: Or why the law is not a decision tree. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 805–816. https://doi.org/10.1145/3593013.3594044

Yu, J., Ma, Z., & Zhu, L. (2023). The configurational effects of artificial intelligence-based hiring decisions on applicants' justice perception and organisational commitment. *Information Technology & People, 38*(2). https://doi.org/10.1108/ITP-04-2022-0271

Zhou, Y., Wang, L., & Chen, W. (2023). The dark side of AI-enabled HRM on employees based on AI algorithmic features. *Journal of Organizational Change Management, 36*(7), 1222–1241. https://doi.org/10.1108/JOCM-10-2022-0308

Chapter 5

Fairness for Hire: An Investigation of how People Analytics Vendors Construct Fairness in the Employment Lifecycle

Link to Thesis Narrative

In Chapter 4, I investigated how algorithmic fairness is constructed in literature depending on the stakeholder view that is taken, as well as what influences the respective constructions. Building on this foundation, this chapter frames this research in line with my research philosophy as outlined in Chapter 2. I identify software developers as an important stakeholder group in designing fair algorithmic decision-making systems and note that it is important to understand how they construct algorithmic fairness. I demonstrate how software suppliers, particularly people analytics vendors, publicly articulate algorithmic fairness on their websites, and equate it with different terms. I further demonstrate that algorithmic fairness plays a more or a less important role depending on the employment lifecycle phase in which an algorithmic decision-making system can be used. In this regard, I show that algorithmic fairness on the websites is mainly discussed in relation to the recruitment and selection of job candidates. This finding foregrounds the rationale for the selection of interview candidates mainly working in recruitment and selection practices for the research reported in Chapters 6 and 7.

State: Published in Australasian Conference on Information Systems (ACIS) Proceedings 2023.

Abstract

People Analytics (PA) software is increasingly used by companies to gain more insights into their data and to enhance their decision-making processes. Many companies rely on external software provided by PA vendors to support processes, such as screening of job applications or assessing employee's performance. Using this software can impose multiple ethical risks. As there is only limited research about how fairness is constructed by PA vendors in practice, the aim of this research is to investigate the construction of fairness by PA vendors and to understand which phases of the employment lifecycle are impacted by fairness. To do so, we analysed 21 websites of PA vendors. We identified seven different concepts of fairness. Furthermore, we were able to identify the phases that are most and least impacted by fairness in practice. Thus, our findings reveal that fairness is rather understood as a "thing" than a process or a particular outcome of a decision. This is important to understand, as it makes a difference in how fairness is theoretically constructed and how it is actually constructed in practice, which provides us a more differentiated understanding of fairness in the employment lifecycle.

5.1 Introduction

Organisations are increasingly using algorithms to process and analyse big amounts of data to support their decision-making process and to make predictions about the future (Kordzadeh & Ghasemaghaei, 2022; Tarafdar et al., 2022). Within human resources (HR) operations, organisations often rely on external software provided by People Analytics (PA) vendors to support processes, such as screening of job applications or assessing employee's performance (Beulen et al., 2022; Gal et al., 2019; Giermindl et al., 2022; Parent-Rocheleau & Parker, 2022). As HR operations have a risk to be affected by discrimination and human bias, the belief is that algorithms are less biased and fairer than humans (Lee, 2018). However, it has been shown that this is not always the case and algorithms can also be biased and not fair (Giermindl et al., 2022).

This has not only resulted in ongoing discussions amongst scholars about how algorithmic fairness can be defined (Dolata et al., 2022) but also made regulators aware of the need to develop laws and regulations for safe, responsible and fair algorithms (Australian Government Department of Industry Science and Resource, 2023). However, as most research focuses on the construction of fairness on a theoretical level (Dolata et al., 2022), there is only limited research about how fairness is constructed in practice with a focus on the perception of fairness of different groups (Lee, 2018; Woodruff et al., 2018). Moreover, there is no research about how fairness is constructed by PA vendors. Hence, we set out to investigate research question (RQ) 1 *How is fairness implicitly or explicitly constructed by PA vendors in practice?*

Furthermore, as PA software is used across the employment lifecycle, there is currently only limited understanding about how fairness impacts the different phases (Parent-Rocheleau & Parker, 2022), that is, if fairness impacts the phases equally or whether there are differences across the phases. This is important to understand as fairness might have a bigger impact on specific phases compared to others. Furthermore, this can also have an impact on the construction of fairness. Therefore, we additionally ask RQ2 *How are the employment lifecycle phases impacted by fairness?*

We investigate PA software vendors whose software is available in Australia and whose software offers trend analysis and/or predictive modelling. To gain an understanding of how fairness is constructed, we conducted a thematic analysis to identify different concepts of how fairness is constructed implicitly or explicitly (Braun & Clarke, 2006; Nowell et al., 2017). To do so, we analysed the websites by conducting a search for key words, such as "fairness", "bias", "equality", "equity" and analysed each company's offerings, solutions, and blogposts to ensure that everything that was stated on the website was analysed. We then matched the concepts to the phases of the employment lifecycle to gain an understanding which phases are predominately impacted by fairness. While we were able to identify seven different concepts of fairness, we show that most PA vendors constructed fairness mainly as the absence or reduction of bias. Furthermore, our results indicate that fairness plays a significant role in the recruitment and selection phase and in the performance management phase of the employment lifecycle. This implies that those phases are more prone to fairness issues.

These findings are important as they contribute to a discussion about how fairness is constructed and how the phases in the employment lifecycle are impacted by fairness in practice by providing a more differentiated understanding of fairness through the identified concepts. Moreover, on a practical level it shows in what phases fairness plays a more significant role than others. Hence, emphasising the need to consider fairness in those phases. This provides guidance not only for PA vendors, but also for companies that want to integrate software into their HR operations.

In the following, we provide a literature review about algorithmic bias and algorithmic fairness as well as about how algorithms can be used across the employment lifecycle phases.

We then continue to describe our research approach and the results of our research. The paper ends with a discussion of the results and conclusion.

5.2 Literature Review

To understand how fairness is constructed in practice, we first outline the theoretical construction of fairness in the literature. Furthermore, as we want to find out how different phases of the employment lifecycle are impacted by fairness, we provide an explanation for each phase, including how algorithms can be used in those phases.

5.2.1 Algorithmic Bias & Algorithmic Fairness

Algorithmic bias is often referred to as outputs produced by an algorithm, that advantage or disadvantage particular individuals compared to other individuals without an underlying valid reason (Akter et al., 2021; Kordzadeh & Ghasemaghaei, 2022). Before we discuss the sources of algorithmic bias, we first need to distinguish between traditional algorithms and machine learning algorithms. Traditional algorithms are specific rules and codes that are used for calculations and other operations to solve specific problems. They are able to aggregate data to provide comprehensive information for rule-based decision-making processes, where those rules must be programmed by humans (Collins et al., 2021; Namvar et al., 2023). Machine learning algorithms on the other hand generate new data insights and enable informed decision-making through learning while identifying non-linear relationships and patterns in the underlying data (Namvar et al., 2023; van den Broek et al., 2021). They are used to process and analyse big amounts of data to make decisions and predictions to improve a company's operations and to increase productivity (Chen et al., 2012; Cheng et al., 2019; Kordzadeh & Ghasemaghaei, 2022; Tarafdar et al., 2022), without necessarily being coded by humans (Collins et al., 2021).

Based on the algorithm type, there are different sources of bias. As traditional algorithms are programmed by humans, they are the main source of bias. This can include biases due to human prejudice or racism (Kordzadeh & Ghasemaghaei, 2022; Vanhée & Borit, 2022). Moreover, as many of the algorithm developers are male, there is a lack of gender diversity among the developers. Traditional algorithms therefore entail the risk of favouring the contemplations of men that developed the algorithms. Having only men as a predominant development group can increase the risk of possible biases and inadequate algorithms because diverse perspectives are missing in the development process. Therefore, it is crucial to have diverse algorithm development teams with different capabilities and backgrounds (Piorkowski et al., 2021; Schulenberg et al., 2023).

As the human component does not exist in machine learning algorithms, the main sources of bias can be found in the data set that is used to train the algorithm. Societal or historical bias play a crucial role in this context as it can be entailed in the data (Akter et al., 2021). Hence, the training data itself can be biased, inaccurate or inadequate (Akter et al., 2021; Dennehy et al., 2023). Another source of bias in machine learning algorithms is method bias. As machine learning algorithms are able to identify correlations and patterns between variables without any necessary human interaction, but are not able to identify causality between variables, there is a risk that the underlying methods confuse correlation with causation (Akter et al., 2021).

Algorithmic fairness can be understood as "the absence of any prejudice or favouritism towards an individual or a group based on their intrinsic or acquired traits in the context of decision-making" (Mehrabi et al., 2021, p. 11). It aims to mitigate biases through technical metrics and human interaction with the technology (Dolata et al., 2022; Kasirzadeh, 2022). However, there is still a far-ranging discussion about what algorithmic fairness is and how it can be ensured. Dolata et al. (2022) identified three different perspectives of algorithmic fairness. Firstly, the *technical perspective* defines different notions of fairness based on a group-level or an individual-level. On a group-level, these notions of fairness identify attributes of protected groups, which are exposed to discrimination based on their protected groups or prediction errors. On an individual level, a similar approach is used by looking at individuals regardless of their membership to a protected group (Corbett-Davies et al., 2017; Dolata et al., 2022).

Secondly, the *social perspective* understands algorithmic fairness in social contexts through the concept of equality, which treats every individual the same (Binns, 2018; Green, 2022; Holm, 2023), and the concept of equity, which treats individuals differently to ensure that they receive what they need to become equal (Dolata et al., 2022; Green, 2022; Rawls & Kelly, 2003). Therefore, it is crucial to determine whether fairness is enacted when differences between individuals are considered or not (Lee & Baykal, 2017; Robert et al., 2020). Furthermore, the social perspective is also driven by the organisational justice theory, which has been accepted in Information Systems (IS) to analyse different types of fairness regarding the interrelation between employees and technological change within an organisation (Dolata et al., 2022; Joshi, 1989; Li et al., 2014). Those different types of fairness include distributive justice, procedural justice, interpersonal justice and informational justice (Li et al., 2014).

Thirdly, the *sociotechnical perspective* argues that decision-making is neither a purely technical task, because humans are either taking the decision or are affected by the decision, nor a purely social task, because of the technical components. Thus, the sociotechnical perspective includes the engagement of the technical and human components in joint optimisation that aims to generate an effective sociotechnical system within a given context (Dolata et al., 2022; Lee, 2004; Makarius et al., 2020; Sarker et al., 2019).

5.2.2 Algorithms in the Employment Lifecycle

To address the research questions how fairness is constructed by different PA vendors and how the employment lifecycle phases are impacted by fairness, it is necessary to explain the underlying assumptions of the employment lifecycle. As there are different conceptions of the employment lifecycle in the literature, we used the Talent Management Wheel (TMW) as our underlying framework to define the phases of the employment lifecycle (Stahl et al., 2012).

"Recruitment and Selection" (R&S) entails elements, such as the application processes and interview processes. Algorithms can be applied to select job applicants by reviewing their applications for specific required skills and experience (Gal et al., 2019; Marks, 2022). Moreover, they can exclude the job candidates who do not have the appropriate skillset and experience for a certain job description. By doing so, this procedure saves more time compared to a recruitment and selection process that is solely conducted by a human (Gal et al., 2019; Marks, 2022).

In the "Development and Training" (D&T) phase of the newly hired workforce, algorithms are used to identify whether there is a need for training and select the appropriate kind of trainings that are necessary for certain employees. Moreover, the algorithm can compare the experience of one employee to the experiences of their co-workers to identify training measures (Horesh et al., 2016; Parent-Rocheleau & Parker, 2022; Tambe et al., 2019). Data of similar employees is also used to make recommendations about the next possible and useful career steps for an employee (Kirimi & Moturi, 2016; Tambe et al., 2019).

"Performance Management" (PM) is crucial to identify, communicate and control the achievement of individual performance goals or group performance goals (Stahl et al., 2012). Algorithms are used to identify the performance goals of an employee and their performance targets by collecting data through algorithmic monitoring. The data is collected from the employee's previous performance (Angrave et al., 2016). Moreover, algorithmic monitoring in performance management can predict the future performance of an employee by relying on past data of that particular employee and their co-workers. An algorithm can also assess an employee’s performance to check whether an employee achieved their performance goals or not (Parent-Rocheleau & Parker, 2022).

“Talent Retention” (TR) identifies employees that are at risk of leaving the company. Algorithms are used to predict the risk of an employee wanting to leave the company (Lee, 2018). The data, that is used for the predictions, can include different sources, such as employee surveys, job tenure, patterns of communication as well as performance reviews (Silverman & Waller, 2015). Based on these predictions, the company can take preventative action (Lee, 2018).

“Compensation and Rewards” (C&R) can be helpful to attract new employees and retain the current workforce. Based on the performance ratings, an employee’s skillset and their experience, algorithms can also help to determine compensation and rewards, such as determining how much salary an employee will receive, salary raises as well as determining who will get promoted. By doing so, algorithms can also integrate the overall strategy of a company and external influences, such as customer satisfaction (Parent-Rocheleau & Parker, 2022; Wood et al., 2019). As algorithms can also predict future performance, they can also influence the decision whether and in what amount an employee will receive a bonus or a salary raise (Parent-Rocheleau & Parker, 2022).

To plan for the future, “Talent Review” (TRev) is important to identify the skills of the current workforce and plan the need for future employees as well as future leaders (Gal et al., 2019). Algorithms use the data of employees to assess the skillset of the current workforce and their need for specific training. This review will show whether specific skills are available in the current workforce or whether a company needs to hire new employees (Gal et al., 2019; Parent-Rocheleau & Parker, 2022).

5.3 Research Approach

To answer our research questions, we needed a dataset that includes PA vendors that offer PA software. To identify suitable vendors, we conducted an online search for PA software, that is available in Australia, on Capterra. Capterra is a free marketplace that connects PA vendors with software buyers. After filtering for “Available in Australia”, 86 PA vendors were displayed. As we only wanted to include software that is able to provide a trend analysis and/or predictive modelling, the results decreased to 62 vendors. Out of the 62 suppliers, 15 were excluded as they were not related to the employment lifecycle phases. This resulted in a final data set of 47 PA vendors.

To gain an understanding how fairness is constructed implicitly or explicitly, we conducted a thematic analysis by identifying different concepts in our dataset and analyse as well as organise the data according to those concepts (Braun & Clarke, 2006; Nowell et al., 2017). To identify the concepts, we analysed each website by starting with a simple word search for words like “fair”, “fairness”, “bias”, “equal”, “equality”, “equitable”, “equity”, “ethical” and “ethics” as well as “just” and “justice”. Furthermore, we read through each company’s offerings, solutions, and blogposts to ensure that everything, that was stated on the website, was analysed. We decided against using a software that analyses websites because we wanted to be open for other constructions of fairness that we might not have been aware of. Based on the analysis of the websites, we then developed different concepts of fairness according to the PA vendors’ understanding. By doing so, we considered all three algorithmic fairness perspectives as we wanted to understand which perspectives are applied in practice. We then matched those concepts to the TMW, that was used as our underlying framework to define the phases of the employment lifecycle, to gain an understanding which phases are predominately impacted by fairness.

5.4 Results

5.4.1 Construction of fairness

To analyse how fairness is constructed by different PA vendors, we selected the relevant paragraphs on the websites, in which fairness or anything related to fairness, was implicitly or explicitly mentioned. The analysis shows that 26 of the 47 software companies neither mention fairness implicitly nor do they mention fairness explicitly in HR processes. Even though this does not indicate that they do not consider the impact of fairness at all, we excluded those PA vendors from our dataset. For the remaining 21 PA vendors, we grouped the identified paragraphs into different concepts. An overview can be found in Table 5.1 and will be described in more detail in the following.

Table 5.1: PA Vendor's Construction of Fairness in the Employment Lifecycle

PA vendor	Offered Services in the Employment Lifecycle	Construction of Fairness	Employment Lifecycle Phase
Rippling	R&S D&T C&R	Fairness Equity	R&S C&R
Worknice	R&S PM	Limiting subjectivity & speculation	PM
OneModel (OneAI)	R&S PM	Ethics	No specific phase
Sage HR	R&S PM	Elimination of bias	PM
Engagedly	D&T PM	Objectivity Elimination of bias	PM
Value Beat	R&S TR	Reduction of bias Objectivity	R&S
intelliHR	R&S PM TR	Hidden bias	No specific phase
Tableau HR Analytics	R&S D&T	Equity	No specific phase
Visier	R&S D&T	Reduction of bias Fairness	TR TRev

PA vendor	Offered Services in the Employment Lifecycle	Construction of Fairness	Employment Lifecycle Phase
	TR TRev		
rexxsystems	R&S D&T PM C&R	Fairness	PM D&T
Teamspective	D&T PM	Elimination bias Reduction of bias	PM
Lanteria HR	R&S D&T PM TR	Fairness	PM TR
Compport	C&R	Equity Fairness	C&R
LutherOne	D&T PM	Elimination of bias	PM
Wisnio	R&S	Elimination of bias Reduction of bias Fairness Objectivity	R&S
iTrent	PM TRev	Equality	TRev
Fuel50	R&S	Elimination of bias	R&S
Workera	R&S D&T TRev	Elimination of bias	TRev
TalentIdentify	R&S D&T PM TR	Elimination of bias	R&S D&T
Orgvue	R&S TRev	Equality	R&S
ADP NextGen HCM	C&R TRev	Fairness	C&R

Fairness and the Absence of Bias

One third of the PA vendors use the word "fairness" to describe how the use of their software can increase fairness across the different phases of the employment lifecycle. This includes statements like *"evaluating candidates fairly", "setting fair performance and development goals"* and a software being *"simple and fair"*. However, as there is no single definition about fairness in the literature (Dolata et al., 2022; Feuerriegel et al., 2020), there was also no clear definition by one of the PA vendors about what fairness means for either the software itself or for humans that either use the software or are affected by the use.

Most PA companies consider *"reducing or eliminating biases"* as well as *"uncover hidden biases"* through PA in the processes across the employment lifecycle. While two of them specifically focus on excluding unconscious bias from different HR processes, eight other PA vendors state that their software is either able to eliminate bias completely or that processes can be made *"unbiased"* with the use of the software. Another PA vendor argues that their software is able to identify hidden biases in different HR processes, such as hiring and performance management. Moreover, the findings also show that "fairness" and "unbiased" are used jointly in a sentence, e.g., "*reduce bias, reward fairly,"*. This indicates that being fair and being unbiased do not have the same meaning. This understanding is in alignment with the literature that predominantly distinguishes between fairness and bias by acknowledging their close interrelation (Kordzadeh & Ghasemaghaei, 2022; Newman et al., 2020; Pfeiffer et al., 2023).

Objectivity and Limiting Subjectivity

Another concept that has been identified is objectivity as well limiting subjectivity. Whereas only one PA vendor considers limiting subjectivity and speculation to *"encourage action"*, three others construct fairness implicitly as providing objective feedback for employees as well as *"objective alignment"* in the recruitment process. The findings also show that bias, fairness and objectivity are used in combination, emphasising the statement that the use of the particular software will *"reduce bias through objective alignment"* or that an "*objective agreement"* can be achieved through setting "fair performance and development goals". Hence, on the one hand they argue that fairness and the reduction of bias can be achieved through objectivity. On the other hand, they argue that objectivity can be a result of fairness. However, despite multiple vendors of PA software argue that algorithms increase objectivity (Leicht-Deobald et al., 2019), there is still a far-ranging discussion in the literature that this is not the case in practice (Leicht-Deobald et al., 2019; Mehrabi et al., 2021; Pfeiffer et al., 2023).

Concept of Equality and Equity

There is a vast discussion in the literature about whether an algorithm should be developed in alignment with the concept of equality or the concept of equity. Hence, there is no common ground on whether a software should treat humans, that are affected by the decision-making processes, equal or equitable (Dolata et al., 2022; Lee & Baykal, 2017), as fairness mostly depends on the context (Jacobs & Wallach, 2021; Madaio et al., 2022; Veale et al., 2018). The findings show that fairness was constructed through providing equity mainly in the context of compensation management phase and talent retention phase by *"paying equity"*, making *"equitable and fair decisions"* as well as by providing *"competitive, equitable and fair pay practices to attract and retain talent"*. The concept of equality was mostly considered in the context of talent review, esp. a company's workforce planning, to *"get an overview of your employee demographics [...] to track your performance against equality and diversity"* as well as in the talent selection process by *"providing equal opportunity"*.

Ethical Decision-Making

Another concept that is connected to fairness, is the ethical AI (Glymour & Herington, 2019; Martin, 2019). One PA vendor argues having designed a platform that *"allows teams to make the best-informed, most-ethical people decisions because it is the only fully explainable machine learning platform, designed specifically for transparency and ethics"*, saying that the *"AI is trained ethically and transparently"*. However, as all of the other PA vendors, also this one has not explained what is meant by being "trained ethically".

5.4.2 Fairness in the Employment Lifecycle Phases

After gaining an understanding of how different PA vendors construct fairness implicitly or explicitly, we wanted to understand how different employment lifecycle phases are impacted by fairness. To do so, we matched the above identified concepts of fairness to each of the six employment lifecycle phases. Overall, the findings show that the PA vendors mostly consider the importance of fairness in the recruitment and selection process and in performance management.

Recruitment and Selection

The findings show that most PA vendors acknowledge the importance of fairness in the recruitment and selection phase, such as evaluation of candidates as well as the selection of candidates. However, these PA vendors construct fairness differently in this phase. Two PA vendors explicitly use the term "fairness", e.g., "*you can evaluate candidates fairly and get measurable insights"*. Four construct fairness through the reduction or elimination of bias in the recruitment and selection phase, for example, *by "recruiting with intention and efficiency"* while *"reducing unconscious bias"*. The analysis also revealed that fairness is also constructed through equality in the talent selection process by providing *"equal opportunity"* and by ensuring that *"all candidates go through the same structured candidate selection process"*.

The results show that in the recruitment and selection phase compared to other phases, fairness overall is highly considered.

Development and Training

The findings indicate that fairness in the development and training phase is considered least important, as only two PA vendors mention fairness in relation to development or training. They state to *"set fair performance and development goals quickly and easily"* as well as *"take the guesswork and unconscious bias out of [...] development"*. However, as the development and training phase is related to the performance management phase (Gal et al., 2019), the results have to be understood in conjunction with the importance of fairness in performance management.

Performance Management

Performance management is highly impacted by fairness according to the findings. The results show that within performance management, fairness is primarily constructed as the reduction of bias and ensuring objectivity, with a specific focus on reducing human bias. Hence, most of the software providers argue that using an algorithm for performance management reduces human biases and therefore ensure an objective and fair feedback process. This can be seen in statements, such as *"360 feedback eliminates suspicions of bias by allowing employees to paint a holistic view of their own performance"*, *"AI ensures the feedback is accurate, objective and unbiased"*, *"say goodbye to irregular & unbiased feedback scattered across performance review forms"* as well as *"collect unbiased data for complete, easy and reliable performance evaluations"*.

However, even though there is a focus on excluding human biases from performance management and performance reviews, the findings reveal that most of the PA vendors still rely on human feedback. To reduce bias and ensure objectivity, they argue that conducting continuous performance reviews instead of only one annual performance review makes the process less biased. In this case, AI is used to send out automated reminders to request feedback or provide feedback frequently as well as to collect feedback from different sources, such as surveys and past feedback. In other cases, AI is used to assist managers or other colleagues in providing *"fair feedback"* by *"becoming better communicators"*. Hence, the software proposes general statements and phrases that can be used to provide feedback.

Talent Retention

The findings indicate that fairness does not have a big impact on talent retention, as only two PA vendors consider fairness. However, these suppliers only consider fairness in combination with compensation and rewards, as part of providing employees *"competitive, equitable and fair pay practices to attract and retain talent"* and *"reduce bias, reward fairly, and motivate the right employees with the right opportunities"*. Even though talent retention is highly related to other phases of the employment lifecycle, such as compensation and rewards (Gal et al., 2019), having only two PA vendors that consider this phase in relation with fairness, suggests that fairness is not an important element for retaining talent in a company.

Compensation and Rewards

Compensation and Rewards was after performance management and recruitment the third phase that is mostly impacted by fairness. Fairness is mainly constructed through the concept of equality by ensuring an equitable process. For example, by *"managing the entire compensation management process, including [...] equity allocation"*, and by *"making rewards fair and equitable"*. Moreover, one software can help to *"identify and address pay gaps and reward top performers fairly"*. However, as compensation and rewards are highly related to performance management (Gal et al., 2019), because performance can determine an employee's compensation and rewards, it is not clear how a fair pay can be ensured if human biases can still be entailed in performance reviews (Kordzadeh & Ghasemaghaei, 2022; Murphy, 2020; Newman et al., 2020).

Talent Review

As Talent Review entails other processes, such as workforce planning and gaining workforce insights (Gal et al., 2019), one PA vendor considers the importance of fairness as part of tracking a company's performance regarding equality and diversity among the workforce. Another PA vendor has created an *"unbiased way to measure your workforce"*. However, in comparison to the other phases of the employment lifecycle, only three suppliers consider implicit or explicit fairness in talent review, indicating that fairness also does not play a big role in talent review.

5.5 Discussion

By analysing 21 PA vendors that are available in Australia, we were able to identify seven concepts of fairness that were either constructed implicitly or explicitly. Moreover, we were able to match these concepts of fairness to the employment lifecycle phases to gain an understand what phases are impacted by fairness.

However, when matching the concepts of fairness to the employment lifecycle phases, we excluded three PA vendors that only considered fairness on a general level without providing insights to which of the employment lifecycle phase it belongs to. For example, having an "AI [that] is trained ethically and transparently" or ensuring a "*safe and equitable workplace*". Hence, it could be argued that those PA vendors generally consider the impact of fairness in every phase.

Moreover, the results indicate that most of the PA vendors construct fairness either implicitly by clearly describing that an outcome of a decision or a process is fair, or explicitly through the reduction of bias. Hence, understanding a fair outcome or a fair process without advantaging or disadvantaging specific individuals (Akter et al., 2021; Kordzadeh & Ghasemaghaei, 2022). However, it is not clear how this is conducted in practice, as it is unclear in most cases how the reduction or elimination of bias is ensured. For example, some PA vendors argue that a continuous performance review using a software is less biased than an annual review. However, since the continuous performance feedback for an employee is still provided by a human, there is still a risk of human bias (Newman et al., 2020). In those cases, the PA vendors do not provide information about how human bias can be eliminated or reduced in a continuous review.

Moreover, we can argue that none of the PA vendors view fairness from a technical or a sociotechnical perspective (Dolata et al., 2022), as neither of them provide information about how the software itself is fair, that is, what statistical notions of fairness are used (Barocas & Selbst, 2016; Dolata et al., 2022), nor do they provide information about the engagement of technical and human components in joint optimisation to generate an effective sociotechnical system (Dolata et al., 2022; Makarius et al., 2020; Sarker et al., 2019). Therefore, we can argue that fairness is primarily constructed through the social fairness perspective, which considers the influence of social contexts (Jacobs & Wallach, 2021; Madaio et al., 2022). Nevertheless, when we consider the employment lifecycle phases as different contexts, there is no consistent understanding of distributive fairness and procedural fairness across the phases. In most cases, fairness is rather understood as a "thing" than a process, e.g., *"say goodbye to irregular and biased feedback"*. Hence, there does not seem to be a clear understanding across the PA vendors about whether distributive fairness, procedural fairness (Helberger et al., 2020; Robert et al., 2020) or both should be ensured.

Our results further indicate fairness is more important in some phases of the employment lifecycle than in others. Hence, the importance can differ depending on the context. For example, when comparing the recruitment phase to the development and training phase, we could see that fairness was less considered in the latter. A reason for this can be the impact of decisions in an employment lifecycle phase on an individual. Whereas an unfair hiring decision has a big impact on an individual's life as it decides whether they are able to get a job or not, a decision about which employee will get particular development measures might not have such a big impact (Kordzadeh & Ghasemaghaei, 2022). Furthermore, we must also consider the interdependencies between different phases. Performance management, for example, is heavily impacted by fairness, whereas in comparison compensation and rewards is less impacted by fairness. However, since an individual's performance has an impact on its compensation and rewards (Parent-Rocheleau & Parker, 2022), we need to acknowledge that when a PA vendor considers fairness in the performance management phase, it must indirectly consider fairness in the compensation and rewards phase.

Nevertheless, this study helps us to better understand how fairness is constructed in practice by different PA vendors. This is important as we did not have this understanding in the past and as it has different implications for practice as different contexts have to be affected by fairness differently.

5.6 Conclusion

As most research focuses on the construction of fairness on a theoretical level (Dolata et al., 2022), there is only limited research about how fairness is constructed in practice with a focus on the perception of fairness of different groups (Lee, 2018; Woodruff et al., 2018). Moreover, there is little research about how fairness is constructed by PA vendors in practice. However, gaining this understanding is important as many companies do not develop their own software but rather rely on external PA vendors. Hence, it is important to understand how they construct fairness and how the employment lifecycle phases are impacted by fairness, as this has an impact on employees, their lives and wellbeing (Mara et al., 2021; Triana et al., 2021).

We were able to identify seven different concepts of fairness that mostly align with a theoretical construction of fairness from a social fairness perspective. Furthermore, we were able to identify the phases that are most and least impacted by fairness in practice. Thus, our results show that fairness and bias are important topics not only in theory but also in practice. Our findings also reveal that fairness is rather understood as a "thing" than a process or a particular outcome of a decision. This is an important finding, as it makes a difference in how fairness is theoretically constructed and how it is constructed in practice from an organisational justice theory point of view.

Our research contributes to the current discourse about algorithmic fairness on a theoretical level as it identifies how PA vendors construct fairness and how the employment lifecycle phases are impacted by fairness. Hence, the theoretical construction of fairness can be enhanced by our practical findings, especially by providing a more differentiated understanding of fairness in the employment lifecycle. On a practical level, it contributes to a better understanding of different concepts of fairness within the employment context. More importantly it shows in what phases fairness plays a more significant role than in others. Hence, emphasising the need to consider fairness in those phases. This provides guidance not only for PA vendors, but also for companies that want to integrate software into their HR operations.

5.7 References

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management, 60*, 102387. https://doi.org/10.1016/j.ijinfomgt.2021.102387

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal, 26*(1), 1–11. https://doi.org/10.1111/1748-8583.12090

Australian Government Department of Industry Science and Resource. (2023). *Safe and responsible AI in Australia*. https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *104 California Law Review, 671*(2016). https://doi.org/10.2139/ssrn.2477899

Beulen, E., Plugge, A., & van Hillegersberg, J. (2022). Formal and relational governance of artificial intelligence outsourcing. *Information Systems and E-Business Management, 20*(4), 719–748. https://doi.org/10.1007/s10257-022-00562-7

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research, 81*, 1–11. http://arxiv.org/abs/1712.03586

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chen, Chiang, & Storey. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165. https://doi.org/10.2307/41703503

Cheng, H.-F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019). Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300789

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management, 60*, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797–806. https://doi.org/10.1145/3097983.3098095

Dastin. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. In K. Martin (Eds.), *Ethics of data and analytics* (pp. 296–299). Auerbach Publications. https://doi.org/10.1201/9781003278290-44

Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y. K., Mäntymäki, M., & Pappas, I. O. (2023). Artificial intelligence (AI) and information systems: Perspectives to responsible AI. *Information Systems Frontiers, 25*(1), 1–7. https://doi.org/10.1007/s10796-022-10365-3

Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal, 32*(4), 754–818. https://doi.org/10.1111/isj.12370

Feuerriegel, S., Dolata, M., & Schwabe, G. (2020). Fair AI: Challenges and opportunities. *Business & Information Systems Engineering, 62*(4), 379–384. https://doi.org/10.1007/s12599-020-00650-3

Gal, U., Riemer, K., Harth-Kitzerow, C., Aboud, C., & Briggs, S. (2019). *People analytics: Using data and algorithms to shape the employee experience*. https://ses.library.usyd.edu.au/bitstream/handle/2123/20082/People%20Analytics%20Report%202019.pdf?sequence=4&isAllowed=y

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

Glymour, B., & Herington, J. (2019). Measuring the biases that matter: The ethical and casual foundations for measures of fairness in algorithms. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 269–278. https://doi.org/10.1145/3287560.3287573

Green, B. (2022). Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology, 35*(4), 90. https://doi.org/10.1007/s13347-022-00584-6

Helberger, N., Araujo, T., & de Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Computer Law & Security Review, 39*, 105456. https://doi.org/10.1016/j.clsr.2020.105456

Holm, S. (2023). Egalitarianism and algorithmic fairness. *Philosophy & Technology, 36*(1), 6. https://doi.org/10.1007/s13347-023-00607-w

Horesh, R., Varshney, K. R., & Yi, J. (2016). Information retrieval, fusion, completion, and clustering for employee expertise estimation. *2016 IEEE International Conference on Big Data (Big Data)*, 1385–1393. https://doi.org/10.1109/BigData.2016.7840746

Jacobs, A. Z., & Wallach, H. (2021). Measurement and fairness. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 375–385. https://doi.org/10.1145/3442188.3445901

Joshi, K. (1989). The measurement of fairness or equity perceptions of management information systems users. *MIS Quarterly, 13*(3), 343. https://doi.org/10.2307/249010

Kasirzadeh, A. (2022). Algorithmic fairness and structural injustice: Insights from feminist political philosophy. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 349–356. https://doi.org/10.1145/3514094.3534188

Kirimi, J. M., & Moturi, C. A. (2016). Application of data mining classification in employee performance prediction. *International Journal of Computer Applications, 146*(7), 28–35. https://doi.org/10.5120/ijca2016910883

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems, 31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Lee, A. S. (2004). Thinking about social theory and philosophy for information systems. In J. Mingers & L. Willcocks (Eds.), *Social theory and philosophy for information systems* (pp. 1–26). John Wiley & Sons.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society, 5*(1), 205395171875668. https://doi.org/10.1177/2053951718756684

Lee, M. K., & Baykal, S. (2017). Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1035–1048. https://doi.org/10.1145/2998181.2998230

Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics, 160*(2), 377–392. https://doi.org/10.1007/s10551-019-04204-w

Li, H., Sarathy, R., Zhang, J., & Luo, X. (2014). Exploring the effects of organizational justice, personal ethics and sanction on internet use policy compliance: Determinants of IUP compliance. *Information Systems Journal, 24*(6), 479–502. https://doi.org/10.1111/isj.12037

Madaio, M., Egede, L., Subramonyam, H., Vaughan, J. W., & Wallach, H. (2022). *Assessing the fairness of AI systems: AI practitioners' processes, challenges, and needs for support* (arXiv:2112.05675). arXiv. http://arxiv.org/abs/2112.05675

Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research, 120*, 262–273. https://doi.org/10.1016/j.jbusres.2020.07.045

Mara, L.-C., Ginieis, M., & Brunet-Icart, I. (2021). Strategies for coping with LGBT discrimination at work: A systematic literature review. *Sexuality Research and Social Policy, 18*(2), 339–354. https://doi.org/10.1007/s13178-020-00462-w

Marks, P. (2022). Algorithmic hiring needs a human face. *Communications of the ACM, 65*(3), 17–19. https://doi.org/10.1145/3510552

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607

Murphy, K. R. (2020). Performance evaluation will not die, but it should. *Human Resource Management Journal, 30*(1), 13–31. https://doi.org/10.1111/1748-8583.12259

Namvar, M., Intezari, A., Akhlaghpour, S., & Brienza, J. P. (2023). Beyond effective use: Integrating wise reasoning in machine learning development. *International Journal of Information Management, 69*, 102566. https://doi.org/10.1016/j.ijinfomgt.2022.102566

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 160940691773384. https://doi.org/10.1177/1609406917733847

Parent-Rocheleau, X., & Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resource Management Review, 32*(3), 100838. https://doi.org/10.1016/j.hrmr.2021.100838

Pfeiffer, J., Gutschow, J., Haas, C., Möslein, F., Maspfuhl, O., Borgers, F., & Alpsancar, S. (2023). Algorithmic fairness in AI: An interdisciplinary view. *Business & Information Systems Engineering, 65*(2), 209–222. https://doi.org/10.1007/s12599-023-00787-x

Rawls, J., & Kelly, E. (2003). *Justice as fairness: A restatement* (3rd ed.). Havard University Press.

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human–Computer Interaction, 35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS Quarterly, 43*(3), 695–719. https://doi.org/10.25300/MISQ/2019/13747

Silverman, R. E., & Waller, N. (2015). *The algorithm that tells the boss who might quit*. https://www.wsj.com/articles/the-algorithm-that-tells-the-boss-who-might-quit-1426287935

Stahl, G. K., Bjorkman, I., Farndale, E., Morris, S. S., Paawe, J., Stiles, P., Trevor, J., & Wright, P. (2012). Six principles of effective global talent management. *MIT Sloan Management Review*.

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

Tarafdar, M., Page, X., & Marabelli, M. (2022). Algorithms as co-workers: Human algorithm role interaction in algorithmic work. *Information Systems Journal, 33*(2), 232–267. https://doi.org/10.1111/isj.12389

Triana, M. del C., Gu, P., Chapa, O., Richard, O., & Colella, A. (2021). Sixty years of discrimination and diversity research in human resource management: A review with suggestions for future research directions. *Human Resource Management, 60*(1), 145–204. https://doi.org/10.1002/hrm.22052

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly, 45*(3).

Vanhée, L., & Borit, M. (2022). Viewpoint: Ethical by designer - How to grow ethical designers of artificial intelligence. *Journal of Artificial Intelligence Research, 73*, 619–631. https://doi.org/10.1613/jair.1.13135

Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3174014

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment and Society, 33*(1), 56–75. https://doi.org/10.1177/0950017018785616

Woodruff, A., Fox, S. E., Rousso-Schindler, S., & Warshaw, J. (2018). A qualitative exploration of perceptions of algorithmic fairness. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3174230

Chapter 6

Making Fairness Work: How Stakeholders Developing, Implementing, and Using Algorithmic Decision-Making Systems Construct Algorithmic Fairness in the HR Context

Link to Thesis Narrative

In Chapter 6, I reported on the results of the semi-structured interviews I used to investigate how algorithmic fairness is constructed by stakeholders who are involved in developing, implementing, and using ADM systems. In my study in Chapter 4, I found that these stakeholder groups are primarily informed by decision-subjects' fairness perceptions. This motivated the decision to gather first-hand information directly from those stakeholders. In Chapter 5, I investigated how software suppliers publicly construct algorithmic fairness on their websites, and this work provided a foundational understanding of how algorithmic fairness can be equated with different terms in practice. In this chapter, I report on an empirical study that aims to understand how understandings of algorithmic fairness are driven by various different factors, such as individuals' beliefs about, and their perceptions of, ADM. I find that, depending on the stakeholder group, their level of understanding of how algorithmic systems function as well as their familiarity with algorithmic systems influence how algorithmic fairness is understood.

State: Published in Australasian Conference on Information Systems (ACIS) Proceedings 2025. A previous version of this paper has been published as a Research-in-Progress Paper in Australasian Conference on Information Systems (ACIS) Proceedings 2024.

Abstract

Algorithmic decision-making (ADM) systems are increasingly integrated into Human Resource (HR) processes, such as recruitment and performance evaluations, aiming to enhance efficiency and reduce human biases. Yet these systems may introduce new forms of bias, raising concerns about algorithmic fairness. While debates address fairness from technical, social, and sociotechnical perspectives, little is known about how organisational stakeholders developing, implementing, and using ADM systems construct fairness in HR. This study examines perspectives of HR professionals, software developers, and AI/HR consultants through 30 semi-structured interviews. Using constructionist grounded theory and a sociotechnical lens, we analyse how algorithmic fairness is constructed, shaped by organisational contexts, roles, and interactions. Findings reveal diverse, contextually grounded understandings of algorithmic fairness, underscoring the need to consider multiple perspectives in ADM design, evaluation, and regulation in high-risk HR settings. The study advances theoretical insights into fairness as a sociotechnical construct and offers practical implications for equitable ADM implementation.

6.1 Introduction

Algorithmic decision-making (ADM) systems are increasingly being integrated into Human Resource (HR) to improve operational efficiency, for instance, by expediting the screening of incoming CVs and reducing related costs (Giermindl et al., 2022; Rieskamp et al., 2023). Additionally, ADM systems are often introduced with the objective of mitigating human biases, which are inherently present in HR decisions due to the human-to-human nature of these interactions (Ochmann et al., 2024). Nevertheless, real-world examples, such as Amazon's hiring algorithm, have demonstrated that these systems may themselves introduce new forms of bias (Dastin, 2018).

Academic scholars and practitioners have actively engaged in debates about the conceptualisation and assessment of fairness within ADM systems, commonly referred to as algorithmic fairness (Dolata et al., 2022; Mehrabi et al., 2021). The literature defines algorithmic fairness from a variety of perspectives, including technical, social, and sociotechnical perspectives. In the field of Information Systems (IS), algorithmic fairness is often framed through a sociotechnical lens, recognising the interaction between technological artefacts, such as hiring algorithms, and various social actors, including HR professionals and job applicants (Dolata et al., 2022). Regulatory initiatives, such as the EU Artificial Intelligence Act, further stress non-discrimination and fairness, requiring AI systems to include diverse actors, ensure equal access, and avoid prohibited biases: "AI systems are developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law" (European Parliament and European Council, 2024, p. 8).

Despite the growing body of literature on algorithmic fairness, relatively little is known about how fairness is constructed by stakeholders responsible for developing, implementing and using ADM systems in HR settings. Existing research predominantly concentrates on the perspectives of frontline workers and job applicants (Köchling & Wehner, 2020; Narayanan et al., 2024). While algorithmic fairness is of relevance across a range of application domains, this research focuses specifically on HR processes, which are widely recognised as being susceptible to discriminatory practices and are consequently categorised as high-risk under the EU AI Act (European Parliament and European Council, 2024). Key stakeholders in these settings include developers, end-users such as recruiters, and consultants who advise on or facilitate the implementation of ADM systems (Dolata et al., 2022).

The current literature has extensively explored how individuals affected by algorithmic decisions, such as job applicants or employees, referred to in our research as "decision-subjects", perceive fairness of these systems (Lavanchy et al., 2023; Narayanan et al., 2024; Ochmann et al., 2024). Less attention has been given to how stakeholders involved in the development, implementation, and use of these systems, such as end-users, software developers, and HR as well as AI consultants, construct their understanding of algorithmic fairness. By focusing exclusively on the perspectives of decision-subjects, existing studies risk overlooking the nuanced ways in which other critical stakeholders construct the concept of algorithmic fairness. Understanding this diversity in conceptualisations is essential for the development, evaluation, and assessment of ADM systems intended to be fair across stakeholder groups.

Algorithmic fairness should therefore not be understood as a uniform or purely technical or social construct. It is rather sociotechnically constructed, shaped by the experiences, organisational contexts, and sociotechnical interactions of diverse stakeholders. Therefore, to explore what is deemed "fair" by various stakeholders, we must investigate the broader sociotechnical dynamics that influence such understandings, beyond solely technical definitions or solely social definitions of fairness (Dolata et al., 2022).

Our research therefore focuses on the perspectives of HR professionals, software developer and HR and AI consultants, whose views significantly contribute to the construction of algorithmic fairness. Therefore, we ask the following research questions: *(1) What factors are considered in the construction of algorithmic fairness by stakeholders developing, implementing and using ADM systems in the organisational HR context? (2) How do these factors construct different understandings of algorithmic fairness?*

To address these questions, we conducted 30 semi-structured interviews with various stakeholders, who are responsible for developing, implementing and using ADM systems in the HR context. To analyse our data, we used constructivist grounded theory which accommodates the influence of context and the position of the researchers (Charmaz, 2006, 2008). We identified five aspects that influence the construction of algorithmic fairness by HR professionals, including the regulatory landscape, cultural influences, familiarity and experience with ADM systems, understanding of ADM systems as well as technical fairness considerations.

We then synthesised these aspects into 5 different constructions of algorithmic fairness by HR professionals, which involve algorithmic fairness 1) as anti-discrimination, 2) as equated with transparency and explainability, 3) as inclusivity, 4) as objectivity and the absence of bias, and 5) as dependent on context and social norms. Given that stakeholders' understandings of algorithmic fairness are shaped by their respective contexts, our study expands current theoretical insights by adopting a sociotechnical perspective on the construction of algorithmic fairness. We contribute to practice by outlining how different stakeholders construct algorithmic fairness and what factors inform the constructions. Recognising these differing perspectives is essential for the design and evaluation of ADM systems, as notions of fairness vary across stakeholder groups.

The remainder of the paper is structured as follows – we provide an overview of background literature on different perspectives on algorithmic fairness, as well as its application in the HR context. We then outline our methods for data collection and analysis. This is followed by our findings, the discussion of implications and contributions, and a conclusion.

6.2 Background Literature

We conduct a review of the existing literature to explore the competing perspectives of algorithmic fairness, highlighting distinctions among technical, social, and sociotechnical lenses. Furthermore, we situate ADM systems within the context of HR, drawing on prior research to examine how notions of fairness have been previously explored in this domain.

6.2.1 Perspectives on Algorithmic Fairness

Technical Perspective

The technical perspective emerged in Computer Science and focuses solely on statistical notions of fairness, typically on either group-level or individual-level fairness. Group-level fairness means that individuals, who are part of a protected group defined by sensitive attributes, such as age, gender or race, are treated equally by an ADM system, when compared with individuals belonging to another group (Calvi & Kotzinos, 2023; Dolata et al., 2022; Juijn et al., 2023). Individual fairness assesses whether an individual is treated the same as another individual regardless of their group membership (Dolata et al., 2022; Hertweck & Heitz, 2021). Some researchers advocate for subgroup-level fairness, which considers multiple dimensions simultaneously to help increase the understanding of algorithmic fairness as a whole (Lämmermann et al., 2022; Pessach & Shmueli, 2023).

However, there has been ongoing critique of viewing algorithmic fairness solely from a technical perspective, as the technical perspective may ignore the context, in which the ADM system is applied in. Therefore, scholars have called for integrating social aspects into the definition of algorithmic fairness, especially when ADM systems are used in social contexts (Dolata et al., 2022).

Social Perspective

Algorithmic fairness from a social perspective has been investigated across various domains, such as HR operations (Dastin, 2018; Juijn et al., 2023) or judicial decision-making (Larson et al., 2016). Within this body of research, algorithmic fairness is often discussed in relation to philosophical reasoning and psychology theory (Dolata et al., 2022). From a philosophical standpoint, algorithmic fairness is interpreted according to stakeholders' assumptions about the role and legitimacy of algorithmic tools, making it a contested, socially constructed, and situated concept, i.e., what is considered to be "fair" depends on who is making the judgment and according to what assumptions (Dolata et al., 2022).

From a psychological perspective, fairness perceptions are frequently studied through organisational justice theory, which views justice as a subjective experience shaped by attitudes and behaviours (Robert et al., 2020). This theory comprises multiple dimensions, including distributive justice, procedural justice, informational justice, and interpersonal justice. Distributive justice refers to the perceived fairness of decision outcomes, while procedural justice pertains to the fairness of the processes used to reach those decisions. Informational justice involves the perceived adequacy and honesty of the information provided about the decision-making process, and interpersonal justice concerns the degree of respect decision-subjects feel they are afforded throughout the process (Greenberg, 1987).

While the social perspective offers valuable insight into how algorithmic fairness is defined and evaluated by human actors in context, it often neglects the technical mechanisms driving algorithmic decisions and the interplay between human and technological elements. As a result, scholars advocate for a sociotechnical perspective that accounts for the mutual shaping of social and technical systems (Dolata et al., 2022; Dolata & Schwabe, 2024).

Sociotechnical Perspective

The sociotechnical perspective claims that the technical and social dimensions of algorithmic fairness, such as statistical fairness metrics on one hand, and individuals' contextual or cultural influences on the other, should not be considered in isolation. These dimensions should rather be understood as deeply entangled and mutually constitutive. To adequately capture the factors shaping the construction of algorithmic fairness, this perspective advocates for examining the entire sociotechnical system and the interdependencies between its technical and social components (Dolata et al., 2022; Draude et al., 2020).

The sociotechnical perspective moves beyond the traditional "human-in-the-loop" model, which assumes that humans merely oversee and control algorithms that operate independently. Instead, it conceptualises humans and algorithms as "collective moral agents" (Dolata et al., 2022, p. 765), recognising that technical and social factors continuously influence one another. Failure to consider these reciprocal influences risks selecting technical fairness criteria that appear just in isolation but are unjust when situated within their broader social context, and vice versa. Such disconnects may ultimately produce outcomes that are unfair, despite conforming to formal fairness metrics. Therefore, comprehensive assessments of algorithmic fairness must incorporate these sociotechnical interactions to ensure that both technical and social considerations are adequately addressed (Dolata et al., 2022).

6.2.2 Algorithmic Fairness in the HR Context

HR operations have long been susceptible to discrimination and bias (Demuijnck, 2009; Triana et al., 2021; Yeung et al., 2021), as HR decisions inherently involve consideration of diverse individuals and their unique characteristics (Cascio and Aguinis, 2013; Köchling & Wehner, 2020). Existing research about ADM systems in HR predominantly focuses on job candidate selection and employee performance assessments (Giermindl et al., 2022). Scholars have explored how various stakeholders perceive ADM systems by comparing fairness perceptions among decision-subjects and end-users in contexts where decisions are made solely by humans, solely by algorithms, or through a hybrid approach (Acikgoz et al., 2020; Feldkamp et al., 2023). The groups primarily studied in existing research include frontline workers (Zhou et al., 2023) and job applicants (Lavanchy et al., 2023), while fewer studies address the fairness perceptions of end-users (Feldkamp et al. 2023) or developers (Kasinidou et al., 2021; Kleanthous et al., 2022), often examining these perceptions through a social lens to understand how algorithmic fairness is constructed (Narayanan et al., 2024).

By primarily focusing on frontline workers, empirical research often overlooks the roles of those responsible for developing, implementing and using ADM systems, who are also instrumental in constructing conceptions of algorithmic fairness. Our research seeks to address this gap by amplifying the voices of these stakeholders, thereby fostering a more comprehensive understanding of the diversity in how algorithmic fairness is constructed. This, in turn, supports the development and evaluation of ADM systems that strive to be equitable for all affected parties.

6.3 Methods

To explore how algorithmic fairness is constructed by different stakeholders in practice, we conducted 30 semi-structured interviews. Interviewees involved HR professionals as well as consultants specialising in HR processes and responsible AI, and people analysts. To capture potential cultural influences, we interviewed individuals working in Australia, Germany, the United Kingdom (UK), the United States (US), Canada and New Zealand with the majority located in Germany and Australia. We included the UK, US, Canada and New Zealand as interviewees from those countries work for global software development companies or provide global AI consultant services, therefore providing specialised insights that are relevant for our research. Participants were recruited via LinkedIn through interview invitations that clearly outlined the purpose and objectives of the study. The interviews were conducted between July 2024 and March 2025, with three additional interviews conducted in August 2025 and each session lasting between 30 and 60 minutes. Interviews were audio-recorded and subsequently transcribed for analysis.

Our interview protocol was structured into distinct sections to capture a comprehensive understanding of our interviewees' backgrounds and experiences with algorithms in HR, consulting and development contexts. Initially, we explored their current and prior work experience both with and without algorithmic systems to contextualise their familiarity with and use of such technologies. Subsequently, the focus shifted to how participants employ algorithms in their professional roles or have observed their application in practice. Although most of the HR professionals interviewed had not yet implemented ADM systems, we probed their reasons for non-adoption and their perceptions regarding the use of such systems. The final section addressed interviewees' conceptualisations of algorithmic fairness. This topic was deliberately positioned at the end of the interview to enable us to examine how their understanding of algorithmic fairness is shaped by their professional experiences and reflections on ADM system use.

Our interviews revealed that the majority of HR professionals had not previously used ADM systems in their professional operations. Notably, many reported using generative AI tools, predominantly ChatGPT, to support operational tasks such as drafting job descriptions or summarising content, but not for decision-making purposes. Consequently, our discussions centred on hypothetical scenarios involving ADM applications, such as candidate selection during recruitment. We further explored the reasons behind their reluctance or hesitation to implement ADM systems. Building on these insights, we examined their conceptualisations of algorithmic fairness. In contrast, the consultants interviewed had practical experience with ADM system deployment and the implementation of governance frameworks, enabling them to draw on concrete examples and scenarios from their own practice.

We adopt a social constructionist theoretical framework, which argues that knowledge and reality are constructed through interactions with society, culture, and language. According to this perspective, knowledge and reality emerge as intertwined products of collective human activity, in contrast to objectivism, which asserts that "truth and meaning reside in their objects independently of any consciousness" (Crotty, 1998, p. 42) and assumes a singular, objective truth independent of human experience (Berger and Luckmann, 1991).

What is recognised as knowledge is contingent upon contextual, relational, and historical factors, such as cultural and community influences, that shape individuals' perceptions of what is "real." Consequently, knowledge must be understood as relative, given that moral frameworks, laws, and other social constructs vary across cultures and societies (Berger & Luckmann, 1991). This theoretical lens is particularly relevant for studying algorithmic fairness, as it allows us to explore how different stakeholders' understandings of algorithmic fairness are shaped by their unique contexts. To account for variations in our interviewees' experiences, social environments, and interactions with technology, we thus employ a social constructionist approach to grounded theory. This methodology facilitates the exploration of how, what, and why questions, enabling us to uncover the processes through which individuals construct diverse meanings of algorithmic fairness and identify the factors influencing these constructions (Berger & Luckmann, 1991; Charmaz, 2008). Furthermore, adopting a constructionist grounded theory approach promotes reflexivity throughout the research process, ensuring that we critically consider not only our interviewees' interpretations but also our own influence and positionality as researchers (Charmaz, 2006, 2008).

Our data analysis began with the identification of initial concepts through systematic coding of the interview transcripts, segmented according to our interview protocol. These segments were categorised into broad areas such as work experience, experience with ADM systems, and conceptions of algorithmic fairness. This iterative process of coding and categorisation was repeated until theoretical saturation was reached, and no new concepts emerged from the data. Subsequently, we identified five overarching thematic categories and proceeded to examine the relationships between them. This analytical phase included an exploration of how sociotechnical factors, as articulated by the interviewees, shape their understandings of algorithmic fairness (Charmaz, 2006).

6.4 Results

Using constructivist grounded analysis, we identified five key concepts that influence how algorithmic fairness is constructed by stakeholders. These concepts were subsequently organised into five broader thematic categories: algorithmic fairness as anti-discrimination, algorithmic fairness equated with transparency and explainability, algorithmic fairness as inclusivity, algorithmic fairness as objectivity and the absence of bias as well as algorithmic fairness dependent on contexts and social norms. In the following sections, we first present the specific factors shaping stakeholders' constructions of algorithmic fairness and then discuss how these factors influence the five overarching themes.

6.4.1 Aspects that Influence the Construction of AF

Laws and Regulations

A key concern consistently raised by all interviewees was the role of regulatory compliance as a major barrier to the implementation of ADM systems in HR. This issue was particularly pronounced among interviewees based in Germany. The German context presents significant challenges related to data protection, privacy, and security, governed by the European General

Data Protection Regulation (GDPR) (European Parliament and European Council, 2016) and the recently enacted EU AI Act, which designates AI systems used in employment, such as for recruitment or performance evaluation, as high-risk (European Parliament and European Council, 2024). In contrast, participants from Australia referred to compliance with the Australian Privacy Act (Australian Government, 1988) and noted the existence of voluntary frameworks, such as the AI Ethics Principles (Australian Government, n.d.) and government-issued AI guardrails (Australian Government, n.d.). However, Australia has not yet implemented binding AI-specific regulations, resulting in a comparatively less stringent regulatory landscape.

Our interviewees expressed concerns about the opacity of ADM systems, particularly regarding what data is collected and how it is processed. These uncertainties were closely linked to fears of unintentionally breaching legal requirements, with potential consequences such as regulatory fines and reputational harm. One HR consultant and recruiter stated: *"Yes, in general, the topic of data protection and a certain fear of using it correctly play a role because it always involves personal data"* (Interviewee 29, Recruiter). Thus, while regulatory frameworks are designed to promote the safe and ethical use of ADM systems, our findings suggest they may also have an unintended deterrent effect.

Cultural Influences

Cultural influences were particularly evident in discussions related to risk and change aversion. Germany was frequently described as a *"conservative country that currently weighs up the advantages and disadvantages"* (Interviewee 23, AI Consultant), indicating a generally cautious attitude toward the adoption of emerging technologies, especially in comparison to more innovation-driven countries such as the United States. Several interviewees noted that Germany is not widely known as a digitalisation pioneer, and that organisational memory of previous, often unsuccessful, digital transformation efforts continues to shape attitudes toward newer technologies. As one participant explained, German companies remain hesitant due to past implementation failures as part of the digitalisation: *"Keeping the topic of digitalisation and the digital transformation in mind, I think we, as Germany overall, haven't exactly done a great job"* (Interviewee 29, Recruiter).

Furthermore, given Germany's strong focus on engineering, ADM systems are more commonly applied in technical and industrial domains than in human-centric fields, such as HR: *"I think they're starting to realise that what has somewhat distinguished us, especially Germany, is actually happening more in the product and engineering sectors. I believe there's significantly more happening there than perhaps in other countries"* (Interviewee 23, AI Consultant). Our findings also indicate that German HR professionals appeared more reluctant to consider future adoption of ADM systems, while Australian HR professionals expressed comparatively greater openness and interest. These insights suggest that cultural differences, particularly in relation to general change aversion, play a significant role in shaping how ADM technologies are perceived and adopted across national contexts.

Moreover, our findings suggest that, compared to Australia, Germany exhibits a more structured and traditional approach to professional career development. In the German context, there appears to be an expectation that individuals follow a *"linear career path"* (Interviewee 23, AI Consultant), whereby careers are typically built within a single domain, with limited deviation across roles or industries. Several interview participants expressed concern that, while human decision-makers can recognise and appreciate the value of non-linear career trajectories, ADM systems may not. As a result, potentially well-suited candidates with diverse or unconventional career paths might be systematically excluded during the selection process. This perceived limitation raises broader concerns about diversity in hiring outcomes. If ADM systems reinforce rigid career expectations, they may inadvertently contribute to a homogenised workforce by filtering out candidates who do not conform to predefined patterns.

One interviewee noted: *"I want to have diversity by default. This means that as a result of unfairness, we end up with uniformity in the workforce within companies, which I see as a negative outcome"* (Interviewee 23, AI Consultant).

Algorithm Aversion

HR professionals with limited experience and familiarity in working with ADM systems often demonstrated a tendency toward algorithm aversion, which is defined as the reluctance to rely on algorithmic decision-making (Hannon et al., 2024). This aversion appears to be primarily driven by a lack of trust, which results from limited direct interaction with and a lack of understanding of such systems. Moreover, our interviews suggest that many HR professionals continue to rely on their intuition or "gut feeling" when making decisions, indicating a preference for human judgment over automated processes: *"Unfortunately, we have always done it this way, where we make the decisions ourselves, and delegating this responsibility is generally a problem. I believe that one of the big issues would be handing over responsibility to the decision-making power of technology"* (Interviewee 10, Recruiter).

Our findings further suggest that individuals' prior experiences as decision-subjects, rather than as end-users, can significantly shape their perceptions of ADM systems, even when later encountering them in a professional or operational context. Several HR professionals described instances in which they had been assessed by algorithmic systems. These experiences appeared to inform their attitudes toward using such technologies in HR practice. While ADM systems may offer potential efficiencies and improvements in decision-making, prior negative experiences as subjects of algorithmic evaluation often foster scepticism or distrust. For example, one HR professional recounted: *"I had to do a language exam as a certification and I went through a computer system, and I was just interacting with the system, right? I was talking with that system. I was reading to that system and I was like, what happens if I don't have the correct intonation? My accent will be different, right? So, I would probably have a* *lot of errors because of this, right? So, when I read about the algorithm's turns, I was like ohh yeah, that could be a bias"* (Interviewee 26, People Analyst). This example highlights how a perceived lack of fairness or sensitivity to individual differences, such as language variation, can reinforce algorithm aversion. These personal experiences tend to be salient and emotionally resonant, making them more memorable than neutral or even positive encounters. As a result, a single negative interaction with an ADM system may disproportionately influence an individual's long-term attitudes toward such technologies, leading to broader scepticism and resistance.

Furthermore, our findings suggest a strong correlation between lack of familiarity with ADM systems and a limited understanding of how they function. In many cases, HR professionals lacked opportunities to interact directly with ADM systems in their day-to-day operations, which constrained their ability to build informed perspectives. As a result, their understanding of algorithmic fairness was often based on second-hand knowledge rather than direct experience. Second-hand knowledge refers to information acquired indirectly, such as through media coverage, public discourse, or anecdotal accounts from others, rather than through personal engagement with the technology (Mahmud et al., 2022). These findings show that, in the absence of practical experience, narratives from external sources often fill the knowledge gap and significantly influence how fairness is understood and evaluated. This reliance on second-hand knowledge may reinforce algorithm aversion, as these external sources tend to highlight high-profile failures or risks associated with ADM systems.

Technical Fairness Considerations

Technical fairness considerations were often framed by interviewees as fairness metrics or success criteria that must be tailored to the specific context in which ADM systems are deployed. In this framing, algorithmic fairness is conceptualised in technological terms, expressed through numerical measures and surrogate indicators designed to approximate equitable outcomes. This includes selecting appropriate datasets, defining fairness constraints, and acknowledging that no single, universally accepted technical definition of algorithmic fairness exists (Morse et al., 2022). As one AI consultant explained: *"If you are to artificially impose like a fairness, like technical fairness metric to try to reallocate resources to the disadvantaged group, in the long run, that might actually end up hurting the disadvantaged group and increasing the inequality between them"* (Interviewee 25, AI Consultant).

Our findings indicate that technical fairness considerations are most often raised by stakeholders with deeper technical expertise, namely, software developers and AI consultants. In contrast, while HR professionals occasionally referred to technical concepts, such as biased training data, they were generally unable to elaborate further. This suggests their knowledge was largely based on second-hand information rather than direct, experiential engagement with ADM systems. These distinctions highlight how technical expertise shapes the depth and nature of fairness discussions, with implications for how fairness is constructed across different stakeholder groups.

6.4.2 Construction of Algorithmic Fairness

Algorithmic Fairness as Anti-Discrimination

This was the primary way in which interviewees constructed algorithmic fairness. In this view, ADM systems are considered fair if they do not disadvantage individuals based on their membership in protected groups. Algorithmic fairness is thus understood as the mitigation or absence of discriminatory or exclusionary outcomes in processes such as recruitment or performance evaluation, particularly when such outcomes are associated with protected characteristics, including gender, age, ethnicity, or national origin. This understanding also extends to concerns about disadvantaging candidates based on non-traditional or non-linear career trajectories, which may lead to the unintentional exclusion of qualified individuals and a potential loss of talent: *"On the one hand, there's the issue of actual discrimination and* *unfairness, and on the other hand, there's certainly a significant potential loss of talent simply because AI has misjudged things entirely"* (Interviewee 23, AI Consultant).

The framing of algorithmic fairness as anti-discrimination was most often articulated by stakeholders with limited familiarity or direct experience with ADM systems. This suggests that a lack of technical expertise may result in a relatively narrow construction of algorithmic fairness from a social perspective, that does not fully account for the technical mechanisms through which algorithmic bias might emerge or how fairness metrics are operationalised in practice. For these stakeholders, algorithmic fairness appeared to be conflated with general notions of fairness in human decision-making, without clearly distinguishing the specificities of fairness in ADM processes.

Moreover, stakeholders who equated fairness with anti-discrimination also tended to express strong risk aversion. Their concerns centred around compliance with legal and regulatory frameworks and the potential consequences of non-compliance, including financial penalties and reputational harm. This understanding often referenced legal protections and may be reinforced by recent regulatory developments such as the EU AI Act, which frames fairness primarily through the lens of non-discrimination but does not provide a detailed operational definition of algorithmic fairness. In the absence of technical expertise, stakeholders may rely on regulatory language as a substitute for fairness, thereby reinforcing a narrow, compliance-oriented conception of algorithmic fairness.

Algorithmic Fairness Equated with Transparency and Explainability

This conception was shared by both stakeholders with and without technical expertise, suggesting that this understanding of fairness is not solely driven by technical knowledge of ADM systems. However, HR professionals tended to articulate this perspective from a personal standpoint, expressing difficulties in understanding how ADM systems operate, which in turn led to a lack of trust and a perception that these systems were unfair. Their views were often influenced by second-hand information, largely derived from media reports highlighting negative incidents involving ADM systems, rather than from direct experience.

By contrast, interviewees with technical expertise emphasised the importance of designing ADM systems to be transparent and explainable for end-users. They viewed transparency as a crucial mechanism to build trust and, ultimately, to shape perceptions of fairness by directly addressing end-users' concerns: *"The other thing that we also do in terms of fairness is making sure that it is explained and it is available to as many people as possible, regardless of their technical background"* (Interviewee 6, Data & AI Specialist). Although many stakeholders with technical backgrounds also found it challenging to provide a precise definition of algorithmic fairness, one AI consultant noted that algorithmic fairness is not discussed *"as such in practice"* (Interviewee 23, AI Consultant), they broadly agreed that transparency and explainability form the foundation for achieving fairness. As one software developer explained: *"The only way to make sure that there's fairness is that the people who would be impacted negatively have the ability to go and see what is it that's causing the negative impact. [...] So, to me, transparency is like the prerequisite to fairness, because you can't really enforce actions without it"* (Interviewee 5, Software Developer).

Algorithmic Fairness as Inclusivity

In some cases, algorithmic fairness was equated with inclusivity. This is particularly interesting because most interviewees primarily discussed fairness in terms of algorithmic outcomes, focusing on distributive fairness (Greenberg, 1987) rather than on equitable access to ADM systems themselves. Regarding inclusivity, several examples emerged. For instance, one HR professional working in Learning and Development described their leadership assessments, evaluating individuals' suitability for leadership roles, as follows: "*The increase of neurodiversity in the workplace and actually have we considered what more inclusive design* *looks like in some of our technology experiences [...] that's really top of mind at the moment. Like, have we created a fair playing field as we assess people, because this then leads to you going into a nine-box grid: are you high potential? This is particularly something we do in our leadership program."* (Interviewee 3, HR Learning and Development Professional).

Moreover, software developers discussed inclusivity from the perspective of ADM system design: *"The other part of our fairness is that we don't believe fairness is just an algorithmic issue. It is relevant to the whole system that AI is embedded in. For example, if I am delivering this through a mobile device and it doesn't have the right user interface, and if it adversely impacts a certain group, like maybe somebody who's colourblind, that is unfair, right? But that's not AI most of the time. Companies often demonstrate through testing that their AI is fair, but you wouldn't see how many people dropped off from their system without even completing it. So that's why we thought inclusivity is very important, and that's a way we measure inclusivity."* (Interviewee 8, Software Developer).

Our results therefore indicate that understanding algorithmic fairness as inclusivity is not strongly influenced by stakeholders' technical expertise, as interviewees from both groups referenced inclusivity. However, those with higher technical knowledge tended to focus more on the design of the underlying algorithms, while interviewees with less technical expertise discussed inclusivity challenges based on personal or professional experiences, including environments they perceived as non-inclusive.

Algorithmic Fairness as Objectivity and the Absence of Bias

Stakeholders, especially HR professionals, emphasised that ADM systems should be objective and free from biases that might advantage or disadvantage certain individuals or groups: *"Um, it's a really hard word to define. It was spinning in my head for a little while. I think fairness to me means unbiased and I want to say objective."* (Interviewee 11, HR Business Partner).

This understanding of algorithmic fairness is closely linked to anti-discrimination, emphasising that decision-subjects should not be disadvantaged based on protected attributes or other characteristics. It also highlights that decision-makers, whether human or algorithmic, should avoid prejudice and instead base their decisions on facts rather than personal preferences: *"Making fair decisions can somehow include both a natural emotional perspective and also a very objective consideration where you say, OK, I have clear facts that you can base the decision on."* (Interviewee 28, People Analyst). Our findings therefore suggest that the construction of algorithmic fairness as objectivity and absence of bias is primarily shaped by a social perspective, lacking consideration of the technical complexities involved. Notably, this view was mainly held by HR professionals, while stakeholders with deeper technical knowledge of ADM systems did not assume that such systems can be completely objective or free from bias.

Algorithmic Fairness Dependent on Context and Social Norms

Since social contexts and norms evolve over time, there cannot be a single, fixed definition of algorithmic fairness. Therefore, the contexts in which ADM systems are applied need to be regularly reassessed to account for such changes and their potential impacts on fairness: *"So, I think we have to think very carefully about the potential impact of imposing any sort of criteria, any sort of success metric. And it's not just about fairness, it's about what KPIs we need to put in place, like what are we optimising for? And the answer can't just be one-dimensional, I guess in most of these cases."* (Interviewee 25, AI Consultant). Algorithmic fairness as context-dependent is primarily influenced by familiarity with ADM systems and considerations of technical fairness. Our findings therefore suggest a relationship between experience and the engagement with technical fairness concepts. Interviewees with greater technological experience and a deeper understanding of ADM systems, whether from direct use or second-hand knowledge, tended to discuss technical notions such as disparate impact, reflecting their better grasp of ADM system functionality.

This understanding also seems to correlate with a more general openness to innovation and change. While many HR professionals displayed change and risk aversion and questioned the value of ADM implementation, others expressed willingness to adopt ADM systems and AI technologies to improve efficiency and avoid falling behind competitors. Those showing interest in ADM systems, despite lacking direct usage experience, demonstrated a higher level of both technical and sociotechnical understanding than their more risk-averse counterparts.

This indicates that stakeholders who construct algorithmic fairness as dependent on context and social norms adopt a sociotechnical perspective, recognising that there is no one-size-fits-all definition of algorithmic fairness. Instead, they understand algorithmic fairness as context-dependent, necessitating the application of diverse technical fairness concepts tailored to different social and situational contexts.

6.5 Discussion

We identified five key aspects that shape how algorithmic fairness is constructed in practice. We then synthesised these aspects into five distinct understandings of algorithmic fairness, including algorithmic fairness as anti-discrimination, as transparency and explainability, as inclusivity, as the absence of bias and objectivity, and as dependent on context and social norms.

When differentiating between stakeholders with and without technical expertise and experience, clear contrasts emerge. HR professionals tend to view algorithmic fairness through a social lens, framing it as anti-discrimination, objectivity and absence of bias. This narrow understanding appears to be shaped largely by second-hand knowledge or limited first-hand experience with ADM systems (Bucher, 2017). In doing so, HR professionals often treated algorithmic fairness as a socially or normatively defined concept, without recognising the ways in which technological integration shapes, constrains, or reconfigures fairness in practice. Therefore, considerations of contextual factors or technical dimensions of fairness receive comparatively little attention as HR professionals did not consider the interplay between human actors and technological systems.

In contrast, stakeholders with a more technical understanding of ADM systems, specifically software developers and AI consultants, construct algorithmic fairness through a lens of contextual understanding, social norms and technical fairness considerations. In their explanations, they emphasise the interplay between social and technical factors, such as varying contexts, in which technical fairness metrics deemed fair in one context may not be fair in another (Lee, 2019). For example, changing circumstances in hiring practices or evolving diversity goals influence which technical fairness considerations should be applied. Thus, these stakeholders approach algorithmic fairness from a sociotechnical perspective, acknowledging the mutual interdependencies between social and technical aspects (Dolata et al., 2022).

Comparing these different stakeholders and their constructions of algorithmic fairness, we identify a clash between algorithmic fairness constructions. This is particularly the case when viewing algorithmic fairness as pure objectivity and viewing it as dependent on context and social norms. The construction of algorithmic fairness as objectivity and the absence of bias aims to assess decision-subjects solely based on objective facts. However, this approach neglects the specific contexts and nuances of individual decision-subjects. For instance, hiring for potential by considering candidates with non-linear career paths cannot be adequately addressed if fairness is constructed solely through an objective lens. Consequently, stakeholders with technical knowledge recognise that fairness must adapt to diverse social norms and evolving circumstances rather than relying on a one-size-fits-all objective standard.

Moreover, across stakeholder groups, constructions of fairness were predominantly outcome-oriented, privileging assessments of whether decision results were perceived as fair. By contrast, only a minority of interviewees discussed procedural fairness, emphasising transparency, explainability, or inclusivity of ADM systems.

Therefore, our findings reveal a misalignment between outcome-based and process-based conceptions of algorithmic fairness as well as between socially framed and sociotechnical understandings. This divergence helps explain why scholarly calls for a sociotechnical perspective on fairness (e.g., Dolata et al., 2022) remain challenging to implement in practice. While end-users often lack the technical expertise necessary to engage with system complexities, they nonetheless play a central role in integrating ADM systems into organisational decision-making. Addressing this gap requires sustained collaboration between technical experts and end-users to ensure that both sociotechnical dynamics and human considerations are incorporated into the design and evaluation of fair ADM systems.

6.6 Conclusion

The aim of our research was to investigate how stakeholders involved in developing, implementing, and using ADM systems within the organisational HR context construct the concept of algorithmic fairness, and which factors influence their constructions. Our analysis uncovered five distinct factors that influence different constructions of algorithmic fairness in practice. Additionally, we identified key relationships between these factors and how they shape varying understandings of algorithmic fairness.

Our research makes several important contributions. Theoretically, we expand the growing body of literature on algorithmic fairness by moving beyond the predominant focus on decision-subjects, such as frontline workers or job applicants, and their perceptions of fairness (Giermindl et al., 2022; Zhou et al., 2023). Instead, we include stakeholders responsible for developing, implementing, and using ADM systems, which are groups often overlooked in past research. Understanding their perspectives is crucial for explaining why certain decisions may be perceived as fair by some individuals but unfair by others.

Additionally, our study contributes to theory by identifying which stakeholder groups construct fairness from which perspectives, helping to reveal gaps in fairness considerations and informing strategies to improve knowledge, especially when viewing algorithmic fairness as a sociotechnical construct in Information Systems (Dolata et al., 2022).

Practically, our research offers valuable insights from stakeholders involved both directly in ADM system development and implementation, as well as potential end-users. This comprehensive understanding helps clarify the diverse meanings of fairness, acknowledging that there is no single universally accepted definition. These insights support practitioners in considering multiple fairness perspectives and enable clearer communication when discussing fairness in ADM contexts. Moreover, our findings can assist in designing and evaluating ADM systems to better align with varied understandings of algorithmic fairness. Finally, our research may support policymakers in articulating a more precise understanding of fairness in future legislation, helping to address the prevailing confusion and conflation of fairness with related but distinct concepts such as transparency, each of which requires careful conceptual and practical differentiation.

Our research has several limitations that should be acknowledged. First, while our sample size of 30 interviews is sufficient for qualitative analysis (Guest et al., 2006), it remains limited in scope. Notably, most HR professional interviewees had little to no prior experience using ADM systems in their professional roles, resulting in a general lack of familiarity with these technologies. Including HR professionals and organisations with established ADM system usage in future studies could provide more nuanced insights into how algorithmic fairness is constructed in practice. Second, our study primarily focused on HR professionals based in Australia and Germany. Expanding the geographical scope in future research could help uncover cultural or institutional differences that may influence constructions of algorithmic fairness in more nuanced ways. Finally, our study focused on those involved in developing, implementing, or using ADM systems but did not include decision-subjects, such as frontline workers or job candidates. Including these perspectives in future research could lead to richer and more context-sensitive understandings of algorithmic fairness.

6.7 References

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

Australian Government. (n.d.). *Australia's AI ethics principles*. Retrieved July 12, 2025, from https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles

Australian Government. (n.d.). *The ten guardrails*. Retrieved July 12, 2025, from https://www.industry.gov.au/publications/voluntary-ai-safety-standard/10-guardrails

Berger, P. L., & Luckmann, T. (1991). *The social construction of reality: A treatise in the sociology of knowledge*. Penguin Books.

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of facebook algorithms. *Information, Communication & Society, 20*(1), 30–44. https://doi.org/10.1080/1369118x.2016.1154086

Calvi, A., & Kotzinos, D. (2023). Enhancing AI fairness through impact assessment in the European Union: A legal and computer science perspective. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1229–1245. https://doi.org/10.1145/3593013.3594076

Cascio, W., & Aguinis, H. (2013). *Applied psychology in human resource management*. Pearson Education.

Charmaz, K. (2006). *Constructing grounded theory*. Sage Publications.

Charmaz, K. (2008). Constructionism and the grounded theory method. In J. A. Holstein & J. F. Gubrium (Eds.), *Handbook of constructionist research* (pp. 397–412). Guilford Press.

Crotty, M. (1998). *The foundations of social research*. SAGE Publications Ltd.

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

Demuijnck, G. (2009). Non-discrimination in human resources management as a moral obligation. *Journal of Business Ethics, 88*(1), 83–101. https://doi.org/10.1007/s10551-009-0100-6

Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal, 32*, 754–818.

Dolata, M., & Schwabe, G. (2024). Towards the socio-algorithmic construction of fairness: The case of automatic price-surging in ride-hailing. *International Journal of Human–Computer Interaction, 40*(1), 55–65. https://doi.org/10.1080/10447318.2023.2210887

Draude, C., Klumbyte, G., Lücking, P., & Treusch, P. (2020). Situated algorithms: A sociotechnical systemic approach to bias. *Online Information Review, 44*(2), 325–342. https://doi.org/10.1108/oir-10-2018-0332

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

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

Greenberg, J. (1987). A taxonomy of organizational justice theories. *The Academy of Management Review, 12*(1), 9–22. https://doi.org/10.2307/257990

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59–82. https://doi.org/10.1177/1525822x05279903

Hannon, O., Ciriello, R., & Gal, U. (2024). Just because we can, doesn't mean we should: Algorithm aversion as a principled resistance. *Proceedings of the Annual Hawaii International Conference on System Sciences.* Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2024.730

Hertweck, C., & Heitz, C. (2021). A systematic approach to group fairness in Automated Decision Making. *2021 8th Swiss Conference on Data Science (SDS)*, 1–6. https://doi.org/10.1109/sds51136.2021.00008

Juijn, G., Stoimenova, N., Reis, J., & Nguyen, D. (2023). Perceived algorithmic fairness using organizational justice theory: An empirical case study on algorithmic hiring. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 775–785. https://doi.org/10.1145/3600211.3604677

Kasinidou, M., Kleanthous, S., Barlas, P., & Otterbacher, J. (2021). I agree with the decision, but they didn't deserve this: Future developers' perception of fairness in algorithmic decisions. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 690–700. https://doi.org/10.1145/3442188.3445931

Kleanthous, S., Kasinidou, M., Barlas, P., & Otterbacher, J. (2022). Perception of fairness in algorithmic decisions: Future developers' perspective. *Patterns, 3*(1), 100380. https://doi.org/10.1016/j.patter.2021.100380

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

Lämmermann, L., Zwickel, A., & Markgraf, M. (2022). AI fairness at subgroup level: A structured literature review. *ECIS 2022 Proceedings*.

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). *How we analyzed the COMPAS recidivism algorithm*. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Lavanchy, M., Reichert, P., Narayanan, J., & Savani, K. (2023). Applicants' fairness perceptions of algorithm-driven hiring procedures. *Journal of Business Ethics, 188*(1), 125–150. https://doi.org/10.1007/s10551-022-05320-w

Lee, M. S. A. (2019). Context-conscious fairness in using machine learning to make decisions. *AI Matters, 5*(2), 23–29. https://doi.org/10.1145/3340470.3340477

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*, 121390. https://doi.org/10.1016/j.techfore.2021.121390

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607

Morse, L., Teodorescu, M. H. M., Awwad, Y., & Kane, G. C. (2022). Do the ends justify the means? Variation in the distributive and procedural fairness of machine learning algorithms. *Journal of Business Ethics, 181*(4), 1083–1095. https://doi.org/10.1007/s10551-021-04939-5

Narayanan, D., Nagpal, M., McGuire, J., Schweitzer, S., & De Cremer, D. (2024). Fairness perceptions of artificial intelligence: A review and path forward. *International Journal of Human–Computer Interaction, 40*(1), 4–23. https://doi.org/10.1080/10447318.2023.2210890

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/10.1111/isj.12482

Pessach, D., & Shmueli, E. (2023). A review on fairness in machine learning. *ACM Computing Surveys, 55*(3), 1–44. https://doi.org/10.1145/3494672

Privacy Act 1988 (Cth.) (1988). https://www.legislation.gov.au/C2004A03712/2019-08-13/text

Regulation (EU) 2016/679 (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Regulation (EU) 2024/1689 (2024). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689

Rieskamp, J., Hofeditz, L., Mirbabaie, M., & Stieglitz, S. (2023). Approaches to improve fairness when deploying AI-based algorithms in hiring – Using a systematic literature review to guide future research. *Proceedings of the Annual Hawaii International Conference on System Sciences.* Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2023.026

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. Human–*Computer Interaction, 35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Triana, M. D. C., Gu, P., Chapa, O., Richard, O., & Colella, A. (2021). Sixty years of discrimination and diversity research in human resource management: A review with suggestions for future research directions. *Human Resource Management, 60*(1), 145–204. https://doi.org/10.1002/hrm.22052

Yeung, D. Y., Zhou, X., & Chong, S. (2021). Perceived age discrimination in the workplace: The mediating roles of job resources and demands. *Journal of Managerial Psychology, 36*(6), 505–519. https://doi.org/10.1108/jmp-04-2020-0185

Zhou, Y., Wang, L., & Chen, W. (2023). The dark side of AI-enabled HRM on employees based on AI algorithmic features. *Journal of Organizational Change Management, 36*(7), 1222–1241. https://doi.org/10.1108/jocm-10-2022-0308

Chapter 7

Too Artificial for HR? Understanding HR Professionals' Concerns about the Adoption of Algorithmic Decision-Making Systems

Link to Thesis Narrative

In this chapter, I investigate HR professionals' concerns about the adoption of ADM systems and the elements that shape those concerns. In Chapter 4, I identified that individuals perceive ADM systems to be less fair than human decision-making. This finding provided a foundation and rationale for my investigation of the not-yet-adoption of ADM systems. In Chapter 5 and Chapter 6, I analysed how different stakeholders who are involved developing, implementing, and using ADM systems construct algorithmic fairness. Key factors that influence how algorithmic fairness is constructed involved algorithm aversion due to a lack of familiarity and understanding of how these systems operate. The research underpinning Chapter 6 identified a research constraint: the limited adoption of ADM systems by HR professionals (outlined in Chapter section 2.4). In response to this constraint, I used discussions with study participants about hypothetical, 'real-world' scenarios of ADM system use, rather than discussions about their first-hand experience of such systems. This constraint also made me curious to understand why ADM systems are not extensively used by HR professionals in practice. To find out, I decided to study the participants' concerns about adopting ADM systems in HR practice. In this chapter, I report on participants' concerns about the use of ADM systems in HR practices as well as the factors that shape these concerns. Therefore, my research findings establish the link between algorithmic fairness concerns and the use of ADM systems.

State: Submitted to the Australasian Conference on Information Systems (ACIS) 2025.

Abstract

Algorithmic decision-making (ADM) systems are increasingly integrated into Human Resource (HR) operations to enhance efficiency and mitigate human bias. Despite their potential, adoption in practice remains limited. While research has focused on job applicants or employees affected by ADM decisions, there is less understanding of how HR professionals, as key stakeholders, perceive and respond to these technologies. We explore HR professionals' perceptions and concerns about ADM adoption and how these are shaped by sociotechnical factors. Drawing on 30 semi-structured interviews with HR professionals, software developers, AI and HR consultants, we examine the reasons why ADM systems have not been adopted yet. Three immediate reasons emerge: concerns about transparency, explainability, and reliability; ethical issues, discrimination, and diversity; and worries about regulatory non-compliance. Four underlying factors shape these concerns: lack of familiarity, lack of understanding, technological change aversion, and epistemic overconfidence in human judgment. Our findings highlight sociotechnical barriers to ADM adoption and provide guidance for responsible implementation in HR.

7.1 Introduction

Algorithmic decision-making (ADM) systems are increasingly used in organisations to improve processes, making them more efficient and reliable (Kordzadeh & Ghasemaghaei, 2022). Especially in Human Resources (HR) operations, the use of ADM systems has the potential to increase efficiency, and reduce costs (Giermindl et al., 2022). For example, in recruitment ADM systems can be used to scan job applicants to reduce the workload of human recruiters (Hunkenschroer & Luetge, 2022). Their use further promises to reduce human biases in the decision-making process, which often tend to discriminate and prejudice certain groups or individuals based on a recruiter's own personal preferences or prejudices (Lee, 2018; Storm et al., 2023).

However, past examples reveal ADM systems as not entirely unbiased (Köchling & Wehner, 2020; Zhou et al., 2023). Moreover, research has shown that individuals often prefer human decision-making over ADM (Narayanan et al., 2024), even if they see that ADM systems outperform human decision-making (Dietvorst et al., 2015). This preference to rely on human rather than algorithmic decision-making is referred to as "algorithm aversion" (Jussupow et al., 2024). While there is a vast body of literature that focuses on algorithm aversion and AI adoption, with a particular focus on reasons why individuals are averse towards algorithmic decision-making, there is only limited research on why HR professionals may be averse towards using algorithmic decision-making systems (Feldkamp et al., 2023). In the context of HR, previous research predominantly focuses on algorithm aversion by individuals affected by ADM, such as job applicants (Choung et al., 2023) or employees (Acikgoz et al., 2020; Juijn et al., 2023; Lavanchy et al., 2023). Only limited research focuses on the users of ADM systems, such as HR managers (Feldkamp et al., 2023; Lacroux & Martin-Lacroux, 2022).

While the literature discusses ADM systems in HR in terms of efficiency and bias reduction, we wonder why such systems are not used more extensively in practice. Moreover, we want to understand whether HR professionals perceive ADM systems as inherently problematic, reflecting a stable form of aversion, or whether their reluctance results from uncertainty about the systems' maturity, relevance, and practical value (Marocco et al., 2024). This distinction suggests that what is often described as rejection may instead reflect a state of not-yet-adoption, meaning a temporary condition that can be overcome through usability, exposure, and institutional support (Bick et al., 2025; Cubric, 2020; Hangl et al., 2023; Horani et al., 2025). Recent developments in generative AI, such as ChatGPT, show this dynamic. Within a short period, these tools have moved from niche innovations to widespread use in personal and professional contexts (Agrawal, 2024; Bick et al., 2025), highlighting how increased accessibility, usability, and familiarity can reduce initial resistance and foster adoption. These insights likely extend to other AI systems, including ADM systems.

To understand these aspects, we investigate their use in HR from a sociotechnical perspective, which does not only focus on the technical aspects of ADM systems but on the interaction between social and technical structures (Pink et al., 2025). This will help us understand how the lack of use of ADM systems is shaped by individuals, such as their experiences and different contexts, as well as by technical aspects (Berger & Luckmann, 1991; Charmaz, 2008). We ask the following two research questions: *(1) What are the concerns about the adoption of ADM systems by HR professionals? (2) How are those concerns shaped?*

We conducted 30 semi-structured interviews with HR professionals, software developers, AI consultants, and HR consultants. We adopt social constructivist theory, which understands knowledge and reality as socially and culturally constructed through interaction and language (Berger & Luckmann, 1991). This allows us to trace how HR professionals' perceptions, concerns and decisions around ADM systems are influenced by their social environments, professional norms and prior experiences. We thus explore how individuals make sense of ADM technologies and the factors that shape their reluctance towards adoption, while being cognisant of our own role as researchers in interpreting these meanings (Charmaz, 2008).

We identified four common perceptions and concerns about ADM systems expressed by HR professionals, which directly influenced the reasons why they have not adopted ADM systems yet. These involve concerns regarding transparency, explainability and reliability of ADM systems, ethical concerns, such as biased and potentially discriminatory outcomes from such systems, general aversion towards technological change, as well as concerns about regulatory and legal compliance. To gain a deeper understanding of these concerns we further reveal underlying factors that shape these concerns, organised into three themes: the lack of familiarity with ADM systems, the understanding of their functionalities, technological change aversion as well as epistemic overconfidence in the human ability to make better decisions than ADM systems. Our analysis provides a nuanced understanding of the sociotechnical aspects influencing ADM system adoption and use in HR contexts or rather lack thereof. By examining the perspectives of HR professionals, software developers, and AI and HR consultants, we capture a comprehensive view of the barriers to ADM system adoption. These insights highlight the interplay between domain-specific knowledge, technical literacy and regulatory context in shaping adoption decisions.

The remainder of the paper is organised as follows. We first provide an overview of the literature on different types of algorithmic systems and reasons for algorithm aversion. We then explain our methodology, and how we conducted and analysed our interviews. We then discuss our findings in the form of common perceptions and concerns about ADM systems and underlying factors shaping them. We end the paper with a discussion and conclusion.

7.2 Background Literature

We review existing literature to delineate differences of algorithmic systems, showing how these differences may influence HR professionals' perceptions and concerns. We then examine the concept of algorithm aversion, focusing on the underlying factors that contribute to resistance towards ADM. Finally, we contextualise ADM systems within the HR context, drawing on prior studies to explore how these systems are commonly perceived in HR practice.

7.2.1 Algorithmic Systems

Algorithms are encoded procedures designed to transform input data into a desired output through automated processes (Mahmud et al., 2022). In AI systems algorithms provide functionality in specific ways (Saranya & Subhashini, 2023), commonly understood as "the ability of a system to identify, interpret, make inferences and learn from data to achieve predetermined organisational and societal goals" (Mikalef & Gupta, 2021, p.3).

Machine Learning and Deep Learning as subfields of AI, involve the development of mathematical inference models that can identify patterns within data to make predictions or informed decisions (Saranya & Subhashini, 2023). In today's age, organisations increasingly employ AI for algorithmic decision-making, often also referred to as ADM systems (Mahmud et al., 2022). Generative AI, another subdomain of AI, has the ability to create new content, such as images, text, or audio, in response to user prompts, by first learning from existing datasets, thus encoding patterns in AI models. This generative capability is largely enabled by deep learning architectures, which allow models to synthesise new outputs that reflect the structure and nuances of the input data (Cronin, 2024). The most common generative AI systems are chatbot systems like ChatGPT.

7.2.2 Algorithm Aversion

Algorithm aversion is the tendency of individuals to avoid relying on algorithmic decisions, despite their potential to surpass human judgment. While definitions vary, a common view characterises it as the conscious or unconscious rejection of algorithmic outputs in favour of human decisions, often prior to any actual interaction with the system (Dietvorst et al., 2015; Jussupow et al., 2024; Mahmud et al., 2022). For our study, we adopt this anticipatory form of algorithm aversion, which may influence the adoption of ADM systems. This approach aligns with our focus on HR professionals, many of whom had not yet implemented such systems in practice.

A key contributor to algorithm aversion involves individuals' beliefs about the superiority of human decision-making. Human decisions are associated with uniquely human characteristics like intuition, creativity, empathy, and adaptability, which are largely lacking from algorithmic processes, which are thus seen as rigid, impersonal, and unable to handle complex or novel situations. Despite evidence that ADM systems often produce more consistent outcomes, many individuals continue to trust their own judgment, particularly in ethically sensitive contexts (Dietvorst et al., 2015; Jauernig et al., 2022).

These beliefs are further shaped by prior experience and familiarity with ADM systems. Research shows that people with prior experience or technical knowledge tend to exhibit lower levels of algorithm aversion (Mahmud et al., 2022). By contrast, those without direct experience often form views based on algorithmic imaginaries (Bucher, 2017), drawing on media narratives or organisational assumptions rather than firsthand understanding. Additionally, older individuals are more likely to perceive ADM systems as irrelevant or ineffective, which further decreases trust and openness toward these technologies (Araujo et al., 2020; Lourenço et al., 2020).

Moreover, users are often required to place trust in systems they do not fully understand, particularly when dealing with opaque models such as neural networks (Hoff & Bashir, 2015; Omrani et al., 2022). A lack of transparency and explainability undermines cognitive trust, especially among users with limited technical background. Unlike human decision-makers, algorithms rarely offer accessible justifications for their outputs, leading to scepticism and reduced adoption. While emotional trust may be influenced by broader attitudes toward technology, cognitive trust hinges on clear, interpretable, and reliable system behaviour (Mahmud et al., 2022; Omrani et al., 2022).

Ethical concerns further exacerbate algorithm aversion. Individuals frequently question whether algorithms can make fair decisions or identify discrimination, especially given their reliance on biased training data or embedded design assumptions (Jauernig et al., 2022; Verma et al., 2021). While algorithms are often presented as neutral, many users perceive them as reinforcing existing societal inequalities due to human involvement in their development and the biases reflected in the data used to train them. Practical examples of algorithmic discrimination, such as biased hiring tools, have further eroded trust and increased scepticism toward their use, especially in sensitive contexts (Dastin, 2018). Even when ADM systems are optimised for fairness using technical metrics, users may still perceive their outputs as impersonal, rigid, or unjust. This perceived disconnect arises from a misalignment between formal definitions of fairness embedded in ADM systems and the more context-sensitive, normative understandings held by users and decision-subjects (Lee & Baykal, 2017). In ethically complex decisions such as hiring or promotion, many users prefer human decision-makers, who are believed to be capable of empathy and moral reasoning, which are qualities not attributed to algorithms (Jauernig et al., 2022; Mahmud et al., 2022).

A perceived loss of control is another key barrier to adoption. Individuals are reluctant to delegate important decisions to systems they cannot fully oversee or influence, especially in domains like HR, where decision-making is central to professional expertise (Schaap et al., 2024). When ADM systems limit users' ability to intervene or adjust outcomes, disengagement often follows. Perceived control involves not only actual system functionality but also whether users believe their input genuinely impacts results, which is often unclear due to ADM opacity (Oz et al., 2024). Furthermore, responsibility frequently shifts from users to developers or organisations, complicating accountability and increasing resistance (Jones-Jang & Park, 2022; Mahmud et al., 2023; Ram & Sheth, 1989).

7.2.3 Algorithmic Systems in the HR Context

Algorithmic systems are increasingly integrated into various HR functions. In recruitment, algorithmic tools are employed to select job candidates (Hunkenschroer & Luetge, 2022; Rieskamp et al., 2023). As part of performance assessments, such systems assess employee performance using diverse metrics, including emotional analysis (Pinheiro et al., 2017), internet browser activity (Angrave et al., 2016), social network engagement (Leicht-Deobald et al., 2019), and levels of work engagement (Burnett & Lisk, 2019). These tools can also identify areas for employee improvement (Budhwar et al., 2022; Manoharan et al., 2011). In learning and development, algorithmic systems offer personalised training recommendations aligned with employees' skill profiles and career aspirations (Tinguely et al., 2023).

The use of algorithmic systems in HR processes is commonly referred to as algorithmic management (Parent-Rocheleau & Parker, 2022). Such systems can fully automate HR processes, replacing humans entirely (Leicht-Deobald et al., 2019) or support HR professionals by automating specific tasks and thereby supporting daily operations (Parent-Rocheleau & Parker, 2022), where decision-making autonomy is shared between HR professionals and algorithmic systems (Tinguely et al., 2023). While algorithmic management offers benefits such as increased efficiency (Leicht-Deobald et al., 2019), it also raises concerns, particularly regarding ethical risks (Douglas et al., 2024) and reduced personal integrity in decision-making processes (Leicht-Deobald et al., 2019). These concerns are closely linked to the concept of algorithm aversion, as humans may view ADM systems as less fair compared to human decision-making (Lee, 2018), thereby increasing distrust and resistance to ADM system implementation (Omrani et al., 2022).

Although algorithmic systems are extensively discussed in the literature, their implementation in industry remains limited. This discrepancy between the theoretical potential of these technologies and their actual use in HR contexts highlights a significant gap in current research. While much of the existing scholarship has concentrated on specific applications, particularly in recruitment, and has frequently addressed ethical concerns (Köchling & Wehner, 2020; Narayanan et al., 2024; Rieskamp et al., 2023; Robert et al., 2020; Starke et al., 2022), there is a lack of empirical insight into the broader organisational reluctance to adopt such systems. Against this background, our research investigates the factors contributing to HR professionals' aversion to algorithmic decision-making systems, with the aim to better understand their reasons for being hesitant towards adopting ADM systems.

7.3 Method

To investigate the factors that underpin and contribute to HR professionals' reluctance to adopt AI systems in their business operations, we conducted 30 semi-structured interviews with a diverse set of stakeholders, including HR professionals, software developers, people analysts and AI as well as HR consultants. The HR professionals represented a range of functional areas, including recruitment, learning and development, as well as people and culture. The software developers and consultants were selected due to their engagement with algorithmic technologies and their close collaboration with HR departments. This selection enabled us to obtain valuable perspectives from individuals positioned outside the traditional HR domain, thereby enriching the diversity and depth of the insights obtained.

Most interview participants were based in Australia and Germany, with additional participants located in the United States (US), the United Kingdom (UK), Canada and New Zealand. This geographical selection was intentional, aimed to capture and contrast regulatory environments. For example, the selection allows comparing Germany's adherence to the EU AI Act versus Australia's relatively less developed AI-specific regulatory framework. Participants from the US, Canada and the UK were primarily consultants or software developers, included due to their global scope of work, which provided valuable insights into cross-jurisdictional implementation challenges and regulatory considerations. Interview participants were recruited through LinkedIn by outlining our research purpose and approach. The interviews were conducted between July 2024 and March 2025 with three additional interviews conducted in August 2025, lasting between 30 and 60 minutes each. All interviews were conducted online using Microsoft Teams or Zoom and were recorded and transcribed. Transcriptions were manually validated for accuracy.

The interview protocol is organised into several thematic sections. It begins by asking interview participants about their professional background and experience with AI systems, in particular ADM systems, or lack thereof. This establishes a baseline understanding of their exposure and familiarity with ADM systems. The subsequent section concentrates on the application of AI systems in the workplace, inviting participants to either describe their current usage or, for HR professionals without prior engagement, to reflect on the reasons that underlie not-yet-adoption. Interviews with software suppliers and consultants focused on their experience in implementing ADM systems, observations from deployments and their perspectives on the factors contributing to HR professionals' reluctance to adopt to ADM systems.

Our findings indicate that most HR professionals interviewed had limited prior engagement with ADM systems in their professional roles, with the notable exception of generative AI tools, mainly ChatGPT. As a result, discussions with these participants largely revolved around hypothetical use cases of ADM systems, such as recruitment, and the perceived barriers to adoption. In contrast, software developers and consultants brought practical experience from either implementing ADM systems or developing governance frameworks, which facilitated more in-depth conversations concerning specific application scenarios.

To analyse our data, we employed a social constructionist approach to grounded theory (Charmaz, 2006, 2008), This methodology supported a reflexive stance, recognising the co-construction of meaning between researchers and participants, as well as the influence of the researchers' interpretative frameworks. The analytical process began with the identification of initial codes derived from segmenting the transcripts in accordance with the interview protocol. These segments were subsequently categorised based on participants' professional backgrounds, experience with AI in general, and ADM systems in particular. Coding of interview data was conducted iteratively, with ongoing refinement through multiple cycles until theoretical saturation was reached. This process resulted in the identification of three core themes, with focused consideration of the interrelationships among these themes and the factors influencing participants' reasons for being hesitant to adopt ADM systems.

7.4 Results

Our analysis indicates that HR professionals' reluctance to adopt ADM systems is driven by various conceptions and concerns about ADM systems' use, such as lack of transparency of ADM systems or HR professionals' general aversion toward technological change. In the following, we first present those stated views and common perceptions. We then further identify factors that shape these views in the absence of any actual use or first-hand experience with relevant ADM tools. We organise these factors into four themes: (1) lack of familiarity with ADM systems, (2) lack of understanding of the underlying functionalities, (3) technological change aversion and (4) overconfidence in human judgement. Our analysis provides a better understanding of why HR professionals have not adopted ADM systems, by both drawing on their own perspectives as well as insights from software developers and consultants, a stakeholder group with stronger technical expertise that frequently works with HR professionals.

7.4.1 Reasons for Not-Yet-Adoption of ADM Systems

Transparency, Reliability and Explainability

A central concept that emerged from our analysis was the concern regarding the lack of transparency, explainability and reliability of AI systems. This was especially evident when interview participants referred to AI systems as "black boxes". For example, HR professionals raised concerns about the lack of transparency of AI systems in relation to how AI systems operate, what data is used and how data is processed: *"Or simply what we might lose, or what might ultimately happen with the data. I can imagine that, even though I haven't dealt with it much so far"* (Interview Partner 10, Recruiter). Transparency was closely linked to trust throughout the interviews. HR professionals emphasised the importance of insights into how ADM systems and their underlying algorithms work, particularly how it weighs certain criteria or makes recommendations or decisions. Interview participants were particularly hesitant to deploy these tools in processes with legal and ethical implications and trust the outcomes: *"How can I trust a black box to be fair, how can I if I haven't seen the algorithm?"* (Interview Partner 11, HR Business Partner).

Participants were also concerned about the reliability of AI-generated outcomes. As one HR professional expressed: *"How do you make sure this stuff is the right answer?"* (Interview Partner 3, HR Learning and Development Professional), indicating their concerns about hallucinations and inaccuracies in ADM systems. Additionally, several participants highlighted a general reluctance to rely on automated systems by drawing parallels to established technologies: *"Exactly, and as of today [...] I don't even trust my navigation system [...] I don't trust it enough, so I always have Google Maps running on the side because I think that's where I get the traffic updates. I just don't trust my own system enough, even though in the end it's exactly the same route"* (Interview Partner 10, Recruiter). Such remarks underscore mistrust in automated outputs, even in well-established technologies, which in turn informs scepticism toward the reliability of more complex ADM systems in HR contexts.

In addition, lack of explainability has emerged as a key barrier to ADM system adoption. Interview participants indicated that inability to understand how ADM system function or generate outcomes contributed to their decision not to use them: *"Maybe I need someone who explain the system to me, and then I would be inclined to use it"* (Interview Partner 11, Recruiter). This lack of explainability was often framed as a decisive factor for preferring human judgement over algorithmic decision-making. As one software developer working closely with HR professional noted: *"In general, people see the outcome, but it's hard* *for them to challenge it. And then when they accept it, it's hard for them to articulate why they've accepted it"* (Interview Partner 6, Data & AI Specialist).

Ethical Issues and Bias

Another major barrier to ADM system adoption among HR professionals concerns ethical considerations, particularly the issue of bias, which could lead to discriminatory and less diverse outcomes. A primary concern was the potential for ADM systems to treat decision-subjects, such as job applicants and employees, less fairly than human decision-makers would. This was primarily discussed in the context of recruitment, particularly with regard to job interviews and resume selection. One source of concern relates to bias embedded within the algorithm itself or within the datasets used to train it. This was perceived as a risk for individuals based on protected attributes, such as their gender or origin. For instance, AI-based job interviews may misinterpret or inaccurately transcribe candidates with accents, resulting in a disadvantage compared to others: "*If they may have incorrect transcription [...] or hearing the candidate incorrectly, it also cannot capture things like hand gestures or possibly their facial expressions and their tones, which could be something that would affect kind of the quality of the interview"* (Interview Partner 16, Recruiter).

Another issue raised was the presence of developer biases inadvertently embedded in the model, which can lead to skewed predictions or decisions: *"People building these models and I do see it as a risk that the outputs themselves can then be potentially biased".* (Interview Partner 7, HR Business Partner). HR professionals also raised ethical issues resulting from the mere use of such systems. A recurring topic was the possibility that job applicants might "game" the system. When job applicants were aware that algorithms are used to rank or select resumes for a role, they may deliberately tailor their applications with specific keywords that are likely to improve their ranking compared to other candidates' applications. This may result in favouring candidates not suitable for the role. Furthermore, it disadvantaged individuals who are less technologically literate. Participants noted that this practice could homogenise applications, diminishing both diversity and the quality of selected candidates: *"I think humans will naturally try and win a process, right? And so, I think the risk is that we'll get certain groups of people that understand exactly what is being looked for and will be will become excellent up job application buyers, you know or excellent performers cause they know what boxes to tick, but they may not actually be the quality candidate."* (Interview Partner 2, HR Business Partner).

A frequently discussed concern was the potential for decreased diversity when algorithms are used to select candidates, particularly if those systems are designed to identify profiles that mirror prior successful hires. In cases where such systems operate without human oversight, the risk of reduced diversity becomes even more pronounced. For instance, job applicants with "non-linear" career trajectories, while possibly highly suitable for the role, may be overlooked if they do not meet all formal requirements outlined in the job description. Such job applicants may possess other valuable attributes that an ADM system is not equipped to recognise. Consequently, ADM tools may fail to "hire for potential" due to overlooking talented candidates, thereby reinforcing homogeneity.

Compliance and Regulatory Concerns

We further found concerns about legal and regulatory compliance. Most of our interview participants were based in Australia and Germany, which are highly regulated countries with regard to data and privacy protection. In the European Union, the General Data Protection Regulation (GDPR) governs data privacy (European Parliament and European Council, 2016), while Australia operates under the Privacy Act 1988 (Cth). Both regulatory frameworks emphasise the protection of individual rights and impose strict requirements on how organisations collect, store, and process personal data. Furthermore, the regulatory landscape is evolving with the introduction of AI-specific legislation (Australian Government, 1988;

European Parliament and European Council, 2016). While Australia is in the process of developing such frameworks, the European Union has already enacted the EU AI Act, effective since August 1, 2024. This legislation adopts a risk-based approach to ensure ethical and safe AI deployment, including safeguards against discrimination and misuse (European Parliament and European Council, 2024).

In this context, HR professionals frequently discussed uncertainty around how exactly AI systems collect and handle data, which was perceived as directly increasing the risk of regulatory non-compliance. As one participant noted: *"There were concerns raised by our privacy team in regard to what kind of information we're telling candidates about how we're utilising these tools and what it means"* (Interview Partner 18, Recruiter). In several cases, privacy and compliance teams were seen as barriers to adoption, not necessarily because of opposition to ADM systems itself, but due to a lack of clarity on how to ensure compliance with emerging legislation. HR professionals expressed particular concern about adopting of ADM systems by foreign vendors due to the risk of aligning such technologies with domestic legal and regulatory requirements. The potential for unintentional violations of data protections laws was perceived as an unacceptable risk.

Additionally, the HR function was characterised as one of the most risk-averse department within organisations. This was attributed to its strong compliance focus, its perceived reputational exposure and its general caution toward innovation. As one software developer working closely with HR professionals explained: *"So as a business strategy, our number one is to avoid HR until the last minute and that's simply because they're the most risk adverse. They feel like implementations reflect on their reputation [...] there's this huge perceived reputational risk by the HR department also because they live very heavily in compliance world, they're very risk adverse"* (Interview Partner 27, Software Developer). Overall, the risk of legal non-compliance was closely linked to concerns about reputational damage. HR professionals highlighted the fear of public backlash as a critical factor. As one interview partner noted: *"Today, there is a relatively high chance when implementing a new approach that it could trigger a backlash in some newspaper article"* (Interview Partner 28, People Analytics Lead).

7.4.2 Factors Underpinning Reasons for Not-Yet-Adoption

We listed and discussed reasons provided by HR professionals for why they currently refrain from adopting ADM systems. Further analysis revealed that this is not the full picture. Triangulating these findings with the perspectives of IT developers and HR consultants revealed deeper seated factors that underpin these espoused reasons. For example, it became clear that HR professionals often argue their point without first-hand knowledge of ADM systems, yet many exhibited generally adverse attitudes towards ADM or technology-driven change more generally. We distinguish four factors: a lack of familiarity and a lack of understanding of how these technologies work, a general aversion to technology change, and finally overconfidence in human abilities vis-à-vis such systems. We argue that these factors act as barriers to adoption.

Lack of Familiarity

Our findings reveal that a limited familiarity with ADM systems, and associated regulations, significantly contribute to perceived difficulties of navigating complex regulatory environments. Without experience in using ADM systems or interpreting associated regulations, HR professionals often assume by default heightened risks that could lead to non-compliance and, in turn, potential negative consequences such as reputational damage. Interview participants noted that data privacy teams also share similar concerns, often due to limited understanding of ADM applications. In several cases, privacy and compliance teams were described as barriers to adoption, not out of opposition to ADM systems, but because they lacked clarity on how to comply with regulatory requirements. This risk aversion was confirmed by software developers who work closely with HR departments, as one explained: *"So as a business strategy, our number one is to avoid HR until the last minute and that's simply because they're the most risk adverse. They feel like implementations reflect on their reputation*" (Interview Partner 27, Software Developer).

Our results further indicate that lack of familiarity also underpins HR professionals' ethical and reliability concerns. Without prior experience with such systems, many HR professionals distrust their outputs, fearing decisions that might be inconsistent with their own judgement, or that are feared to be unethical by default. These apprehensions appear largely shaped by second-hand knowledge, often informed by public discourse or media coverage rather than direct engagement. As a result, their understanding of potential risks is frequently based on speculation or "algorithmic imagination" (Bucher, 2017) rather than sufficient grasp of the technical processes that underpin these systems.

Interestingly, many interview partners reported using generative AI tools, such as ChatGPT, in both personal and professional contexts. HR professionals described using ChatGPT for example to generate job descriptions and summarise content. This suggests that familiarity can influence adoption. Our findings indicate that concerns related to risks may be mitigated through direct exposure and practical engagement. When ADM tools are perceived as accessible and embedded in routine practice, HR professionals demonstrate greater willingness to adopt them. Increased familiarity with specific tools, therefore, appears to reduce hesitation and foster openness to ADM system adoption in the HR function.

Lack of Understanding

Related to lack of familiarity is lack of understanding of ADM systems. Again, without direct exposure to ADM systems, HR professionals had little opportunity to develop firsthand knowledge through experience. As a result, their understanding relies on second-hand information and common (mis)conceptions, gleaned from media discourse and public narratives.

This was evident in frequent references to widely publicised cases, such as the Amazon Hiring Algorithm (Dastin, 2018), and in the rather generalised way in which they addressed ethical issues, often without detailed technical grounding on how biases might arise. While many emphasised the reputational risks of unethical ADM systems, few could explain the mechanisms behind such risks.

By contrast, software developers and AI consultants, stakeholder groups with greater technical literacy, expressed concerns that were more technically informed and rooted in professional experience. Their focus was on the technical mechanisms that can generate bias, particularly the role of data quality in shaping fairness. Developers highlighted that inadequate or biased datasets can impair system performance and perpetuate historical inequities, producing outputs that reinforce discrimination and reduce workforce diversity. Such deficiencies pose both ethical and operational challenges.

HR professionals' limited understanding of ADM systems was also evident in their concerns about legal and regulatory compliance, particularly the potential reputational and financial consequences of non-compliance. Our findings suggest that both lack of familiarity with, and lack of comprehension of, ADM systems contribute substantially to the perceived complexity of navigating regulatory requirements. This was reflected in interviewees' views that a deeper understanding of system operations would enhance their confidence in adoption and reduce compliance-related apprehensions: *"Maybe one would need the system explained to me, and then I would be inclined to use it"* (Interview Partner 11, Recruiter).

Interestingly, HR professionals perceived generative AI systems such as ChatGPT as more accessible and user-friendly. This was attributed, rightly or wrongly, to a perceived transparency of data sources and interactive functionality that enables users to question outputs and request clarification. Although lacking detailed technical knowledge, participants reported a sense of agency in evaluating ChatGPT outputs. As one HR professional explained: "*But if I read a theme and I go oh, that doesn't make a lot of sense to me, that's a weird one that's unexpected. As a result, I can actually interrogate the model, and I can say can you show me the comments that you used to summarise this theme and it'll spit back out some specific comments and then I can go Alright, I understand how the AI came to that theme*" (Interview Partner 7, HR Business Partner). This interactivity, and the ability to converse in natural language, promoted a sense of control, allowing assessment without advanced technical expertise.

Moreover, reliance on ChatGPT's outputs was not widely viewed as ethically risky or reputationally damaging. HR professionals stressed their ability to retain decision-making authority, selectively accepting or rejecting suggestions. This perceived autonomy promotes greater comfort and trust in generative AI than in other algorithmic systems. In sum, distrust in ADM systems appears driven by both a fundamental knowledge gap and limited opportunities to interrogate outputs, which are constraints that impede critical evaluation and reinforce adoption hesitancy.

Technological Change Aversion

Our findings suggest that apprehension toward ADM systems is also to a good part driven by general resistance to technological change. This resistance was most apparent in interviewees' discussions of ADM systems, often framed through justifications for retaining manual processes perceived as more effective and reliable. Participants with extensive industry experience, particularly those with over two decades in HR roles, emphasised that ADM systems had never been necessary to achieve successful outcomes: *"I've, you know, been doing this for 20 years. I've never needed help from the computer before"* (Interview Partner 11, Recruiter). Several participants also framed this aversion in generational terms, noting that key decision-makers often belong to generations less familiar with and more cautious about emerging technologies, given the perceived adequacy of existing manual processes. As one interviewee observed: *"A lot of the decision makers are of the generation that is concerned and scared about the implementation of new technologies. I'm optimistic that a lot of Gen Z people coming through the workforce in the coming years will be open to the use of AI technologies in the recruitment process"* (Interview Partner 7, HR Business Partner).

Overall, the findings suggest that resistance to ADM system adoption in HR is shaped by a combination of ingrained reliance on established practices, generational differences in technological openness, and the perceived adequacy of existing manual processes.

Overconfidence in Human Decision-Making

The fourth theme emerging from our interviews is HR professionals' belief in the superiority of human judgement over ADM systems in performing core aspects of their work. Our analysis suggests that this epistemic overconfidence is in turn driven by a combination of their limited understanding, lack of familiarity, and broader resistance to technological change. Because HR professionals have direct access to their own reasoning processes and a, at least perceived, clear awareness of how they reach decisions, they place greater trust in their own judgement. By contrast, they view ADM systems as opaque, due to limited transparency regarding data inputs and underlying algorithmic processes. Many participants also emphasised the restricted opportunities to interrogate or challenge AI-generated outputs.

The lack of familiarity and understanding reinforces epistemic overconfidence in two ways. First, limited or no direct exposure to ADM systems increases distrust, making HR professionals reluctant to rely on them. Second, their understanding of ADM system often shaped by second-hand knowledge, further discouraging adoption. This scepticism is compounded by strong confidence in long-standing HR practices, perceived as reliably effective for decades without ADM systems, thereby reinforcing a preference for human judgement. However, our analysis indicates that perceptions of human superiority are not solely a function of distrust or unfamiliarity but reflect a deeper conviction that human expertise, particularly in contexts demanding nuanced judgement and interpersonal sensitivity, cannot be replicated by ADM systems. This belief is grounded in a strong professional identity that values discretion and relational understanding, especially in recruitment.

It was most evident in discussions on diversity in hiring, where participants expressed concerns that ADM, such as shortlisting candidates, could reduce diversity by failing to recognise non-linear or unconventional career paths. HR professionals emphasised their ability to identify potential and assess suitability beyond rigid job criteria, recognising transferable skills and qualities not apparent in formal qualifications or keyword-based data. Additional concerns centred on the perceived susceptibility of ADM systems to manipulation, such as applicants strategically inserting keywords to optimise algorithmic selection. By contrast, human evaluators were seen as less vulnerable to such tactics, owing to their capacity for holistic assessment.

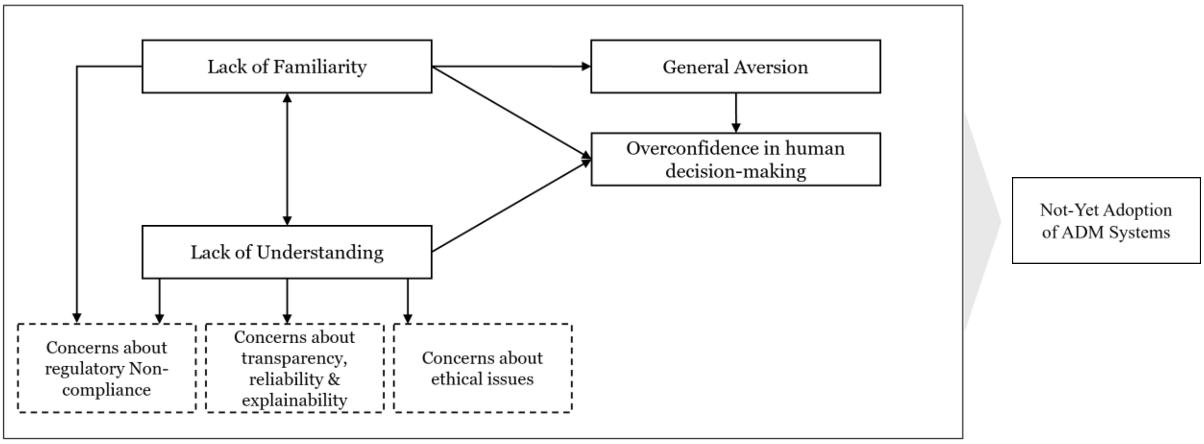
7.5 Discussion

Drawing on 30 semi-structured interviews with HR professionals, software developers, and both AI and HR consultants, we identified three overt reasons for HR professionals' reluctance to adopt ADM systems. We further revealed underlying factors that shape these concerns, organised into four themes: the lack of familiarity with ADM systems, the understanding of their functionalities, technological change aversion as well as epistemic overconfidence in the human ability to make better decisions than ADM systems as shown in figure 7.1. Overall, our findings reveal a consistent set of concerns among HR professionals regarding the integration of ADM systems into their daily operations, in contrast to the growing enthusiasm for such technologies in both academic and industry discourse.

A central concern among HR professionals was the perception of ADM systems as opaque or unexplainable, often framed as the "black box" problem (Douglas et al., 2024). Many viewed ADM tools as lacking transparency and reliability, thereby undermining trust in their outputs (Omrani et al., 2022). This distrust was amplified by a general aversion to risk and resistance to change, with perceived risk emerging as a key barrier to adopting new technologies, particularly when potential legal, ethical, or reputational consequences were involved. Several interviewees highlighted fears of non-compliance with laws and regulations, which they believed could result in financial penalties or reputational damage.

Importantly, our findings show that concerns about ADM systems are interrelated and mutually reinforcing. Lack of familiarity limits understanding, as HR professionals lacked exposure and relied predominantly on second-hand knowledge. This lack of experiential learning further entrenched their confidence in traditional decision-making approaches. In turn, this overconfidence in human judgment, despite its well-documented biases (Hunt et al., 2024; Thomas a Reimann, 2023), diminished their motivation to adopt ADM systems (Mahmud et al., 2022; Sutherland et al., 2016). Moreover, while the identification of risks associated with ADM systems is driven by second-hand knowledge, such as ethical concerns or reputational damage, the lack of familiarity drives HR professionals to fear these risks materialising. As a result, HR professionals are reluctant to adopt ADM systems at all, even when such tools have the potential to improve decisions and efficiency (Jabagi et al., 2025). Their unwillingness to take risk stems from limited exposure to ADM processes and insufficient confidence in navigating associated regulatory requirements.

Figure 7.1: Reasons for Not-Yet-Adoption of ADM Systems



Interestingly, this contrasted with HR professionals' attitudes toward generative AI systems, such as ChatGPT, which many reported using in personal and professional contexts. Its intuitive interface, human-like responsiveness, and capacity for follow-up questions were cited as key factors driving adoption. Interestingly, it was not evident that these professionals had any better understanding of how ChatGPT worked than any of the HR-related ADM systems.

Moreover, the adoption of ADM systems has been more favourable when they are employed to support decision-making rather than to make decisions entirely without human input. This tendency reflects a broader human reluctance to lose control over processes for which they hold responsibility (Mahmud et al., 2022). This is reflected in our findings, where HR professionals use ChatGPT as support but maintaining their own decision-making authority.

Our findings indicate that the limited adoption of ADM in HR is driven by technological constraints and by psychological, organisational, and regulatory factors. The preference for ChatGPT suggests that greater familiarity and perceived control can increase openness to other algorithmic tools. This insight underscores the need for user-centred design, enhanced explainability, and robust governance frameworks when introducing ADM systems in organisational contexts.

Moreover, when we look at the case of ChatGPT, we can see a rapid uptake of generative AI tools, which illustrates that familiarity, accessibility, and usability significantly influence adoption (Agrawal, 2024; Bick et al., 2025). Prior research in HR contexts is limited, leaving open questions about why ADM systems have not been widely adopted and whether HR professionals view these systems as inherently problematic or simply as not yet sufficiently mature, relevant, or useful (Marocco et al., 2024). This distinction suggests that what is often described as resistance may instead reflect a state of not-yet-adoption, a temporary condition shaped by evolving knowledge, experience, and organisational context. Importantly, barriers such as distrust, concerns about transparency, or perceived ethical risks are not fixed. They can diminish as individuals gain exposure and positive experience with AI systems (Marocco et al., 2024; Kar & Kushwaha, 2023). Therefore, the adoption trajectory observed with ChatGPT may imply that ADM systems could similarly experience increased uptake over time as HR professionals gain familiarity and hands-on experience, highlighting the need to consider non-adoption as potentially transitional rather than permanent.

We acknowledge several limitations of our research. First, our sample of 30 interviews, while sufficient in size for our qualitative analysis (Guest et al., 2006), remains limited in its scope. Including HR professionals and organisations already using ADM systems could provide more nuanced insights into both barriers and enablers of adoption. However, it has been surprisingly difficult to find a broader base of adopted system in the market in the jurisdictions included in this research. Second, our study focused primarily on HR professionals based in Australia and Germany. Expanding the geographical scope in future research could help identify whether cultural or institutional differences influence attitudes toward ADM technologies. Moreover, our sample encompassed a diverse range of HR professionals, including those primarily working in recruitment, as well as individuals working in learning and development and HR business partners. We acknowledge that ADM systems may be applied differently across various HR functions and contexts. A more evenly distributed sample representing professionals from varied areas could potentially provide deeper insights into how concerns regarding ADM adoption vary depending on specific functional roles.

7.6 Conclusion

This research investigates the concerns surrounding the adoption of ADM systems and the themes that shape these concerns within the organisational HR context. We conducted 30 semi-structured interviews with HR professionals from diverse functional areas, including recruitment and learning and development, as well as with software developers and AI as well as HR consultants, primarily based in Australia and Germany.

Our analysis identified three concerns commonly perceived by HR professionals and four underlying factors that shape these concerns, which collectively shape resistance to ADM system adoption. Our findings reveal that the main concern involves the uncertainty about transparency, explainability and reliability of ADM systems. These concerns contribute to ethical apprehensions and a general mistrust of ADM systems. Additionally, a prevailing risk aversion among HR practitioners manifests in a reluctance to alter established processes, particularly those perceived as effective in the past. This resistance is further exacerbated by fears of non-compliance with legal and regulatory frameworks, coupled with apprehensions about ultimate accountability.

Our research makes several important contributions. To the growing body of literature on algorithm aversion and the application of AI in HRM, we contribute by identifying interconnected concerns as central barriers to ADM adoption. We further advance theory by identifying new concerns not widely discussed in the literature, such as the fear that ADM systems may reduce workforce diversity by prioritising keyword matching, thereby not being able to “identify for potential” and the risk that applicants may "play" the system and thereby gaining advantages in the recruitment process.

Practically, our findings offer insights for both developers and regulators. For practitioners, our study provides a foundation for designing AI systems and implementation strategies that address end-user concerns, particularly through improved transparency, explainability, and change management processes. For policymakers and regulators, our research underscores the importance of clear communication around legal compliance and ethical considerations to support responsible AI adoption in HR. Ultimately, our study reaffirms that effective change management practices are critical components in the successful integration of AI technologies into the HR function.

7.7 References

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

Agrawal, K. P. (2024). Towards adoption of generative AI in organizational settings. Journal of Computer Information Systems, 64(5), 636-651. https://doi.org/10.1080/08874417.2023.2240744

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal, 26*(1), 1–11. https://doi.org/10.1111/1748-8583.12090

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*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Berger, P. L., & Luckmann, T. (1991). *The social construction of reality: A treatise in the sociology of knowledge*. Penguin Books.

Bick, A., Blandin, A., & Deming, D. J. (2025). The rapid adoption of generative AI (NBER Working Paper No. 32966). National Bureau of Economic Research. https://doi.org/10.3386/w32966

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society, 20*(1), 30–44. https://doi.org/10.1080/1369118x.2016.1154086

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

Burnett, J. R., & Lisk, T. C. (2019). The future of employee engagement: Real-time monitoring and digital tools for engaging a workforce. *International Studies of Management & Organization, 49*(1), 108–119. https://doi.org/10.1080/00208825.2019.1565097

Charmaz, K. (2006). *Constructing grounded theory*. Sage Publications.

Charmaz, K. (2008). Constructionism and the grounded theory method. In J. A. Holstein & J. F. Gubrium (Eds.), *Handbook of constructionist research* (pp. 397–412). Guilford Press.

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, 40*(22), 7451–7468. https://doi.org/10.1080/10447318.2023.2266244

Cronin, I. (2024). *Understanding generative AI business applications: A guide to technical principles and real-world applications*. Apress. https://doi.org/10.1007/979-8-8688-0282-9

Cubric, M. (2020). Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. *Technology in Society, 62*, 101257. https://doi.org/10.1016/j.techsoc.2020.101257

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women.* https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

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

Douglas, D. M., Lacey, J., & Howard, D. (2024). Ethical risk for AI. *AI and Ethics, 5*(3), 2189–2203. https://doi.org/10.1007/s43681-024-00549-9

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

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

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59–82. https://doi.org/10.1177/1525822x05279903

Hangl, J., Krause, S., & Behrens, V. J. (2023). Drivers, barriers, and social considerations for AI adoption in SCM. *Technology in Society, 74*, 10229. https://doi.org/10.1016/j.techsoc.2023.102299

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Horani, O. M., Al-Adwan, A. S., Yaseen, H., Hmoud, H., Al-Rahmi, W. M., & Alkhalifah, A. (2025). The critical determinants impacting artificial intelligence adoption at the organizational level. *Information Development, 41*(3), 1055-1079. https://doi.org/10.1177/02666669231166889

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

Hunt, A. K., Wang, J., Alizadeh, A., & Pucelj, M. (2024). Advancing a theoretical framework for exploring heuristics and biases within HR decision-making contexts. *Personnel Review, 53*(7), 1823–1841. https://doi.org/10.1108/pr-03-2023-0192

Jabagi, N., Croteau, A.-M., Audebrand, L. K., & Marsan, J. (2025). Do algorithms play fair? Analysing the perceived fairness of HR-decisions made by algorithms and their impacts on gig-workers. *The International Journal of Human Resource Management, 36*(2), 235–274. https://doi.org/10.1080/09585192.2024.2441448

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

Jones-Jang, S. M., & Park, Y. J. (2022). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication, 28*(1). https://doi.org/10.1093/jcmc/zmac029

Juijn, G., Stoimenova, N., Reis, J., & Nguyen, D. (2023). Perceived algorithmic fairness using organizational justice theory: An empirical case study on algorithmic hiring. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 775–785. https://doi.org/10.1145/3600211.3604677

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

Kar, A. K. & Kushwaha, A. K. (2023). Facilitators and barriers of artificial intelligence adoption in business – insights from opinions using big data analytics. *Information System Frontiers, 25*, 1351-1374. https://doi.org/10.1007/s10796-021-10219-4

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

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems, 31*(3), 388–409. https://doi.org/10.1080/0960085x.2021.1927212

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*. https://doi.org/10.3389/fpsyg.2022.895997

Lavanchy, M., Reichert, P., Narayanan, J., & Savani, K. (2023). Applicants' fairness perceptions of algorithm-driven hiring procedures. *Journal of Business Ethics, 188*(1), 125–150. https://doi.org/10.1007/s10551-022-05320-w

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

Lee, M. K., & Baykal, S. (2017). Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1035–1048. https://doi.org/10.1145/2998181.2998230

Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics, 160*(2), 377–392. https://doi.org/10.1007/s10551-019-04204-w

Lourenço, C. J. S., Dellaert, B. G. C., & Donkers, B. (2020). Whose algorithm says so: The relationships between type of firm, perceptions of trust and expertise, and the acceptance of financial robo-advice. *Journal of Interactive Marketing, 49*(1), 107–124. https://doi.org/10.1016/j.intmar.2019.10.003

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*, 121390. https://doi.org/10.1016/j.techfore.2021.121390

Mahmud, H., Islam, A. K. M. N., & Mitra, R. K. (2023). What drives managers towards algorithm aversion and how to overcome it? Mitigating the impact of innovation resistance through technology readiness. *Technological Forecasting and Social Change, 193*, 122641. https://doi.org/10.1016/j.techfore.2023.122641

Manoharan, T. R., Muralidharan, C., & Deshmukh, S. G. (2011). An integrated fuzzy multi-attribute decision-making model for employees' performance appraisal. *The International Journal of Human Resource Management, 22*(3), 722–745. https://doi.org/10.1080/09585192.2011.543763

Marocco, S., Barbieri, B., & Talamo, A. (2024). Exploring facilitators and barries to managers' adoption of AI-based systems in decision-making: A systematic review. *AI, 5*(4), 2538-2567. https://doi.org/10.3390/ai5040123

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management, 58*(3), 103434. https://doi.org/10.1016/j.im.2021.103434

Narayanan, D., Nagpal, M., McGuire, J., Schweitzer, S., & De Cremer, D. (2024). Fairness perceptions of artificial intelligence: A review and path forward. *International Journal* *of Human–Computer Interaction, 40*(1), 4–23. https://doi.org/10.1080/10447318.2023.2210890

Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F., & Agreda, S. G. (2022). To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts. *Technological Forecasting and Social Change, 181*, 121763. https://doi.org/10.1016/j.techfore.2022.121763

Oz, B., Karran, A., Boasen, J., Coursaris, C., & Léger, P.-M. (2024). The role of user control in enhancing human-AI collaboration effectiveness: Insights from a pilot study. In F. D. Davis, R. Riedl, J. V. Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 185–193). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-58396-4

Parent-Rocheleau, X., & Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resource Management Review, 32*(3), 100838. https://doi.org/10.1016/j.hrmr.2021.100838

Pinheiro, P. G., Ramos, J. J. G., Donizete, V. L., Picanco, P., & De Oliveira, G. H. (2017). Workplace emotion monitoring – An emotion-oriented system hidden behind a receptionist robot. In D. Zhang & B. Wei (Eds.), *Mechatronics and robotics engineering for advanced and intelligent manufacturing* (pp. 407-420). Springer International Publishing. https://doi.org/10.1007/978-3-319-33581-0

Pink, S., Quilty, E., Grundy, J., & Hoda, R. (2025). Trust, artificial intelligence and software practitioners: An interdisciplinary agenda. *AI & SOCIETY, 40*(2), 639–652. https://doi.org/10.1007/s00146-024-01882-7

Privacy Act 1988 (Cth.) (1988). https://www.legislation.gov.au/C2004A03712/2019-08-13/text

Ram, S., & Sheth, J. N. (1989). Consumer resistance to innovations: The Marketing Problem and its solutions. *Journal of Consumer Marketing, 6*(2), 5–14. https://doi.org/10.1108/eum0000000002542

Regulation (EU) 2016/679 (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Regulation (EU) 2024/1689 (2024). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689

Rieskamp, J., Hofeditz, L., Mirbabaie, M., & Stieglitz, S. (2023). Approaches to improve fairness when deploying AI-based algorithms in hiring – Using a systematic literature review to guide future research. *Proceedings of the Annual Hawaii International Conference on System Sciences.* Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2023.026

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human–Computer Interaction, 35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Saranya, A., & Subhashini, R. (2023). A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal, 7*, 100230. https://doi.org/10.1016/j.dajour.2023.100230

Schaap, G., Bosse, T., & Hendriks Vettehen, P. (2024). The ABC of algorithmic aversion: Not agent, but benefits and control determine the acceptance of automated decision-making. *AI & SOCIETY, 39*(4), 1947–1960. https://doi.org/10.1007/s00146-023-01649-6

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society, 9*(2). https://doi.org/10.1177/20539517221115189

Storm, K. I. L., Reiss, L. K., Guenther, E. A., Clar-Novak, M., & Muhr, S. L. (2023). Unconscious bias in the HRM literature: Towards a critical-reflexive approach. *Human Resource Management Review, 33*(3), 100969. https://doi.org/10.1016/j.hrmr.2023.100969

Sutherland, S. C., Harteveld, C., & Young, M. E. (2016). Effects of the advisor and environment on requesting and complying with automated advice. *ACM Transactions on Interactive Intelligent Systems, 6*(4), 1–36. https://doi.org/10.1145/2905370

Thomas, O., & Reimann, O. (2023). The bias blind spot among HR employees in hiring decisions. *German Journal of Human Resource Management: Zeitschrift Für Personalforschung, 37*(1), 5–22. https://doi.org/10.1177/23970022221094523

Tinguely, P. N., Lee, J., & He, V. F. (2023). Designing human resource management systems in the age of AI. *Journal of Organization Design, 12*(4), 263–269. https://doi.org/10.1007/s41469-023-00153-x

Verma, S., Ernst, M., & Just, R. (2021). *Removing biased data to improve fairness and accuracy* (No. arXiv:2102.03054). arXiv. https://doi.org/10.48550/arXiv.2102.03054

Zhou, Y., Wang, L., & Chen, W. (2023). The dark side of AI-enabled HRM on employees based on AI algorithmic features. *Journal of Organizational Change Management, 36*(7), 1222–1241. https://doi.org/10.1108/jocm-10-2022-0308

Chapter 8

Discussion and Synthesis

8.1 Introduction

The objective of this thesis is to provide insights into the construction of algorithmic fairness among various stakeholders in the HR context. To achieve this objective, I deliberately chose to focus on stakeholders involved in developing, implementing, and using ADM systems because frontline employees' or job applicants' roles and perceptions of algorithmic fairness in decision-making processes have already been extensively explored in prior literature (e.g., Acikgoz et al., 2020; Lavanchy et al., 2023; Newman et al., 2020; Ochmann et al., 2024).

This thesis addresses the central research question: "*How do stakeholders in the HR context construct algorithmic fairness?*" Each component of my research engages with this question by exploring distinct aspects of algorithmic fairness and the elements shaping its construction from the perspective of different stakeholders, including software developers and HR professionals. This research further focuses on the consequences of negative perceptions of algorithms, including concerns around algorithmic fairness, that lead HR professionals to currently resist ADM system implementation.

In Chapter 4, I identified the ways in which literature constructs algorithmic fairness depending on the stakeholder perspective that is taken in each publication. In doing so, I show that different stakeholders focus on different aspects of algorithmic fairness. I show that algorithmic fairness is predominantly examined from managerial and developer perspectives, with employees and other decision-subjects treated instrumentally. This finding is important because it highlights that without understanding how those developing, implementing, and using ADM systems perceive fairness, research cannot fully capture the multi-stakeholder nature of HR decision-making. Manuscript 2 (Chapter 5) extends this analysis by examining how software suppliers publicly represent algorithmic fairness, showing that fairness is acknowledged in websites but often without clarity or full coverage across HR processes. This is an important finding because it demonstrates that even in publicly facing communications, clarity and transparency around fairness in ADM systems are limited, which may influence stakeholder expectations and understanding in practice. Together, Manuscript 1 and Manuscript 2 provide a foundation for empirical investigation to link public discourse with the lived perceptions of stakeholders. Manuscript 3 (Chapter 6) investigates how different stakeholders, HR professionals, software developers, and consultants, construct algorithmic fairness in practice, showing that technical knowledge, social norms, and cultural context shape interpretations, and that only a minority adopt a fully sociotechnical perspective. Shared values, such as objectivity and absence of bias, emerged across groups, yet conflicting interpretations persist, showing the uneven integration of technical and social dimensions. This is an important finding as it reveals how fragmented and uneven understandings of fairness across stakeholder groups are. Moreover, it provides a foundation for my subsequent research in Manuscript 4 (Chapter 7) by demonstrating how a perceived lack of fairness among other factors can influence the successful adoption of ADM systems in HR in practice (Ochmann et al., 2024).

Manuscript 4 (Chapter 7) further examines the consequences of these constructions, specifically focusing on the not-yet-adoption of ADM systems. Here, HR professionals' concerns about transparency, explainability, ethical risks, and lack of familiarity were shown to reinforce one another (Jauernig et al., 2022; Mahmud et al., 2022) creating practical barriers to adoption despite the theoretical benefits of these systems. This finding is important because it highlights that limited adoption may reflect a state of not-yet-adoption rather than inherent resistance, suggesting that targeted exposure, training, and improved communication around those concerns could align perspectives and facilitate responsible integration of ADM systems.

In summary, my research sheds light on algorithmic fairness and the reasons why ADM has not yet been extensively implemented in HR operations. In this chapter, I summarise how my research enhances our understanding of algorithmic fairness and the reasons for the limited implementation of ADM systems in HR operations. I discuss the overall findings of my thesis and synthesise the theoretical and practical contributions of the articles included in it. Finally, I acknowledge the limitations of each individual study and propose directions for future research into algorithmic fairness and the persistent challenges surrounding the adoption of ADM systems in HR.

8.2 Summary of Research Questions and Findings

This thesis investigated nine research questions, as outlined in Table 8.1. The following section summarises the interdependencies among the findings and illustrates how the outcomes of each research question inform and connect with one another to form an integrated body of insights.

Table 8.1: Summary of Research Questions

Chapter	Research Question	Research Question
Chapter 1: Introduction	*RQ 1*	*"How do stakeholders designing, implementing, and using algorithmic decision-making systems construct algorithmic fairness in the organisational HR context?"*
Chapter 4: Manuscript 1	*RQ 2*	*"What stakeholder groups are considered in existing research?"*
Chapter 4: Manuscript 1	*RQ 3*	*"How does existing research construct algorithmic fairness from different stakeholder perspectives in the organisational context?"*
Chapter 5: Manuscript 2	*RQ 4*	*How is fairness implicitly or explicitly constructed by People Analytics vendors in practice?"*

Chapter	Research Question	Research Question
Chapter 5: Manuscript 2	*RQ 5*	*"How are the employment lifecycle phases impacted by fairness?"*
Chapter 6: Manuscript 3	*RQ 6*	*"What factors are considered in the construction of algorithmic fairness by stakeholders developing, implementing, and using algorithmic decision-making systems in the organisational HR context?"*
Chapter 6: Manuscript 3	*RQ 7*	*"How do these factors construct different understandings of algorithmic fairness?"*
Chapter 7: Manuscript 4	*RQ 8*	*"What are concerns about the adoption of algorithmic decision-making systems by HR professionals?"*
Chapter 7: Manuscript 4	*RQ 9*	*"How are those concerns shaped?"*

RQ 1: Construction of algorithmic fairness by different stakeholders in the HR context

Overall, this thesis is guided by the question *"How do stakeholders developing, implementing, and using algorithmic decision-making systems construct algorithmic fairness in the organisational HR context"* Throughout this thesis, different aspects of the question were investigated to answer RQ 1 as summarised in Table 8.2 and outlined in the sections that follow.

Table 8.2: Summary of Findings

Manuscript	Findings	Significance of Findings
Manuscript 1 (Chapter 4): Literature Review	Algorithmic fairness is predominantly studied from managerial and developer perspectives, whereas employees and other decision-subjects are often treated instrumentally rather than as primary stakeholders.	Highlights a gap in the literature: limited focus on understanding the construction of algorithmic fairness by stakeholders developing, implementing, and using ADM systems, justifying the need for multi-stakeholder investigation.

Manuscript	Findings	Significance of Findings
	Different stakeholders construct fairness based on distinct factors: developers integrate technical constraints and legal considerations; managers focus on organisational relevance by relying on decision-subjects' perceptions.	Emphasises the importance of a sociotechnical perspective that integrates both social and technical dimensions when studying algorithmic fairness.
	No single, universal construction of fairness exists. While there are overlapping elements, interpretations diverge across stakeholders.	Highlights that algorithmic fairness is context-dependent and socially constructed, reinforcing the need to explore tensions and misalignments between perspectives.
Manuscript 2 (Chapter 5): Website Analysis	Software suppliers construct algorithmic fairness explicitly (via references to fairness in outcomes/processes) or implicitly (using terms such as absence of bias, objectivity, equality, equity, ethical decision-making).	Shows that algorithmic fairness is acknowledged in public-facing materials but often without clear definition or context, indicating limited transparency.
	Algorithmic fairness is most frequently addressed in the recruitment and selection phase, less so in other phases like learning and development or compensation, despite interconnections (e.g., performance affecting rewards).	Suggests that software suppliers focus on high-impact HR processes, but partial coverage may lead to incomplete fairness considerations across the employment lifecycle.
	The findings further provided a foundation for subsequent semi-structured interviews with software developers, HR professionals, people analysts and consultants, to explore factors shaping fairness constructions in practice.	Establishes a link between public discourse and empirical investigation, contextualising how stakeholders' perceptions align with, or diverge from, vendor messaging.
Manuscript 3 (Chapter 6): Construction of Algorithmic Fairness	Multiple factors influence how stakeholders construct algorithmic fairness: laws and regulations, cultural influences, algorithm	Demonstrates that algorithmic fairness is shaped by both social and technical factors, highlighting the need for a sociotechnical perspective.

Manuscript	Findings	Significance of Findings
	aversion, and technical fairness considerations.	
	Stakeholders with limited technical expertise conceptualise algorithmic fairness predominantly through a social lens, while those with technical expertise integrate technical dimensions. Only a few stakeholders conceptualise algorithmic fairness through a sociotechnical lens.	Highlights uneven integration of social and technical considerations and the limited adoption of fully sociotechnical approaches in practice.
	Shared interpretations emerge across public-facing discourse as part of Manuscript 2 and stakeholder interviews, e.g., algorithmic fairness as objectivity and absence of bias.	Shows that certain algorithmic fairness concepts are consistently valued, providing points of alignment across stakeholder perspectives.
	Lack of familiarity and technical knowledge leads to socially grounded, sometimes conflicting interpretations of algorithmic fairness.	Explains why stakeholder understandings diverge, motivating further investigation into concerns and not-yet-adoption of ADM systems in HR, as outlined in Manuscript 4.
	Tension exists between algorithmic fairness as objective vs. fairness as context-dependent and socially constructed, with the latter incorporating social norms and situational context.	Indicates that purely objective definitions of algorithmic fairness may be insufficient; emphasises the importance of context-sensitive, nuanced approaches.
Manuscript 4 (Chapter 7): Not-Yet-Adoption of Algorithmic Decision-Making Systems	HR professionals demonstrate a reluctance to adopt ADM systems in their operational practices.	Highlights a practical gap between using ADM systems in theory and real-world adoption in HR contexts.
	Concerns driving not-yet-adoption include transparency, explainability, reliability, ethical risks, resistance to technological change, and legal/regulatory compliance.	Identifies critical barriers that must be addressed for effective algorithmic decision-making system implementation.

Manuscript	Findings	Significance of Findings
	Underlying factors shaping these concerns include lack of familiarity and understanding, reliance on second-hand knowledge, and limited hands-on experience.	Shows that insufficient exposure and training amplify ethical and practical apprehensions, impeding adoption.
	Risk perceptions are mutually reinforcing, i.e., unfamiliarity increases perceived ethical and reputational risks.	Explains why HR professionals remain hesitant even when algorithmic decision-making systems offer efficiency gains; emphasises the role of perceived risk.
	Findings illustrate the interplay between algorithmic fairness and real-world adoption, showing that algorithmic fairness influences HR decision-making regarding algorithmic decision-making.	Provides an understanding of how concerns about algorithmic fairness affect technology integration, which informs strategies to address adoption barriers.

RQ 2 & RQ 3: Construction of algorithmic fairness by literature

In Chapter 2, I outlined my epistemological approach towards the social construction of algorithmic fairness, whereby algorithmic fairness can be understood as a complex social phenomenon, the meaning of which is mediated by culture, context, and power relations (Berger & Luckmann, 1991). Taking this epistemological stance, I conducted an organising literature review (Leidner, 2016) to address RQ 2, *"What stakeholder groups are considered in existing research?"* and RQ 3, *"How does existing research construct algorithmic fairness from different stakeholder perspectives in the organisational context?"*. The purpose of this literature review was to broaden the field's understanding of how algorithmic fairness has previously been investigated (Chapter 4). Analysing different literature reviews that focus on algorithmic fairness, I found that most literature reviews only focus on one particular stakeholder group and do not compare algorithmic fairness across different stakeholder groups (e.g., Narayanan et al., 2024). Otherwise, they investigate algorithmic fairness without focusing on a particular group (e.g., Hunkenschroer & Luetge, 2022; Lämmermann et al., 2022; Rieskamp et al., 2023). In this context, my aim was to map existing research on algorithmic fairness within the literature to different stakeholder groups (Webster & Watson, 2002), and thereby address RQ 2. My research aim was also to organise current literature into those different stakeholder groups to synthesise and organise the broad body of literature about algorithmic fairness, and thereby improve the comprehensibility of this topic (Leidner, 2016), thereby addressing RQ 3. I organised the construction of fairness based on the stakeholder perspective that is taken in the respective publication. I identified that most of the literature on algorithmic fairness takes a managerial or a developer's perspective. I observed that algorithmic fairness has been constructed either directly, by analysing the understanding of algorithmic fairness by decision-makers (Feldkamp et al., 2023) or software developers (Kasinidou et al., 2021; Kleanthous et al., 2022), or indirectly, by exploring algorithmic fairness implications for HR managers or software developers based on how individuals subject to algorithmic decisions perceive fairness (e.g., Lavanchy et al., 2023). This finding indicates that, within the literature, decision-subjects, such as employees, are often used instrumentally to infer implications for other stakeholder groups rather than being examined as primary stakeholders in their own right. Furthermore, my analysis revealed that different stakeholder perspectives incorporated distinct elements in their construction of algorithmic fairness. For instance, the developer perspective was shaped not only by developers' subjective interpretations of fairness but also by technical constraints and legal considerations. This finding underscores the importance of adopting a sociotechnical lens when examining algorithmic fairness, to integrate social and technical dimensions rather than treat them in isolation (Dolata et al., 2022).

My analysis also demonstrated that the diverse lenses adopted in the literature highlight the absence of a one-size-fits-all approach to constructing algorithmic fairness across stakeholder groups. Different stakeholders incorporate varying elements into their fairness frameworks and, while some elements appear across multiple groups, the interpretations of these elements and the emphasis given to them often diverge. In this regard, Chapter 4 serves as a foundation for my empirical investigations because it maps the stakeholder groups identified in the literature as relevant to the construction of algorithmic fairness. Chapter 4 also establishes a basis for the examination of algorithmic fairness from a practical perspective, thereby enabling an exploration of the factors shaping stakeholder understandings of algorithmic fairness and the potential tensions or misalignments that may arise between those understandings.

RQ4 & RQ5: Construction of Algorithmic Fairness by Software Suppliers Through Their Websites

Building on the knowledge I obtained through the literature review, I investigated software suppliers, in particular people analytics vendors, to obtain a better understanding of whether and how algorithmic fairness is discussed in practice by software suppliers. In particular, I investigated how this stakeholder group publicly constructs and communicates notion of algorithmic fairness on their websites. This research addressed RQ 4: *"How is fairness implicitly or explicitly constructed by People Analytics vendors in practice?"* I wanted to understand which phases of the employment lifecycle were given particular focus with regard to algorithmic fairness, thereby addressing RQ 5: *"How are the employment lifecycle phases impacted by fairness?"* To answer both questions, and in order to include systems that are able to support decision-making and prediction-making processes across different HR operations, I conducted a thematic analysis of 47 people analytics vendors' websites that offer trend analysis and/or predictive modelling software (available in Australia). My results indicated that algorithmic fairness was constructed either explicitly, through references to fairness in outcomes or processes, or implicitly, via a range of related terms. These terms included fairness framed as the absence of bias, as objectivity and limited subjectivity, as equality and equity, and as fairness achieved through ethical decision-making. However, even in those instances where fairness was explicitly referenced, its meaning was rarely clarified within the relevant context. Notably, fairness was never specified in terms of fairness for whom. Furthermore, none of the websites examined provided an explanation of how their software or underlying algorithms produced fair outcomes or processes, which reflected a lack of transparency and explainability (Hunkenschroer & Luetge, 2022). In addition, by mapping how algorithmic fairness is constructed across different phases of the employment lifecycle, I observed that algorithmic fairness was most frequently discussed in relation to the recruitment and selection phase. This frequency during the recruitment and selection phase suggests a recognition that unfair hiring decisions can have a more significant impacts on individuals' lives than unfair decisions during other phases, such as learning and development (Kordzadeh & Ghasemaghaei, 2022). However, I also caution against viewing these lifecycle phases in isolation. For example, while fairness appears to play a central role in performance management, it is mentioned less frequently in relation to compensation and rewards. However, given that performance evaluations often influence compensation outcomes (Parent-Rocheleau & Parker, 2022), considerations of fairness in performance management inevitably have implications for fairness in compensation. Therefore, if software suppliers address fairness in one phase, such as performance management, they must, either implicitly or explicitly, account for fairness in interconnected phases such as compensation and rewards.

This study served as the foundation for the two subsequent empirical investigations, during which I conducted semi-structured interviews (Walsham, 1995) to gather first-hand insights from various stakeholders regarding the factors that shape their constructions of algorithmic fairness.

RQ 6 & RQ 7: Construction of Algorithmic Fairness by Stakeholders Developing, Implementing, and Using ADM Systems

Chapter 6 reported on the semi-structured interviews I conducted with 30 stakeholders who develop, implement, and use ADM systems. The aim of this study was to understand the factors that influence the construction of algorithmic fairness, thereby addressing RQ 6: *"What factors are considered in the construction of algorithmic fairness by stakeholders developing, implementing, and using algorithmic decision-making systems in the organisational HR context?"* My intent was also to understand how these stakeholders construct fairness based on these different factors, thereby addressing RQ 7: *"How do these factors construct different understandings of algorithmic fairness?"* This study identified several different factors that influence how algorithmic fairness is constructed, including laws and regulations, cultural influences, algorithm aversion, and considerations of technical fairness. I also explored how algorithmic fairness was spoken about during my interviews and analysed the factors shaping the varied understandings of algorithmic fairness. These understandings included fairness as: anti-discrimination, transparency and explainability, inclusivity, objectivity and the absence of bias, and, lastly, as dependent on social norms. Clear overlaps emerged between these findings and those in Chapter 5, which analysed how software developers publicly frame algorithmic fairness on their websites. Notably, conceptions such as algorithmic fairness as objectivity and the absence of bias appeared consistently across both public-facing discourse and stakeholder interviews, suggesting shared interpretations of algorithmic fairness.

This study aimed to examine algorithmic fairness from a sociotechnical perspective, to understand the interrelation between social and technical factors that shape its construction (Dolata et al., 2022). However, my findings indicate that most stakeholders, particularly those with limited experience or knowledge of ADM systems, tended to conceptualise algorithmic fairness predominantly through a social lens. Their understanding often centred on perceptions of fairness in outcomes or procedures, aligning with principles of distributive and procedural justice as outlined in organisational justice theory (Greenberg, 2003; Dolata et al., 2022). In contrast, only those stakeholders with some degree of technical expertise – typically software developers, AI consultants, or HR professionals with a technical background – incorporated technical dimensions of fairness into their constructions. This disparity highlights the uneven integration of technical and social considerations across stakeholder groups and suggests that a truly sociotechnical understanding of algorithmic fairness remains limited in practice.

My analysis revealed a tension between differing constructions of algorithmic fairness, specifically between fairness understood as objectivity, as also identified in Chapter 5, and fairness viewed as context-dependent and shaped by social norms. The former construction frames algorithmic fairness in terms of the objective assessment of decision-subjects, emphasising neutrality and the elimination of bias. However, this perspective often overlooks the contextual nuances and lived realities of individual cases. For example, evaluating candidates for potential, such as those with non-linear career trajectories (Hunkenschroer & Luetge, 2022), cannot be fully captured within a purely objective framework. In contrast, stakeholders with technical expertise in ADM systems tended to adopt a more contextual and sociotechnical conception of fairness. This conception acknowledges that fairness must be responsive to diverse social contexts and dynamic normative expectations, rather than relying on static, one-size-fits-all definitions rooted solely in objectivity (Dolata et al., 2022).

The study reported in Chapter 6 highlighted the complexity, and multiplicity, of algorithmic fairness constructions across different stakeholder groups. The study found that these constructions were driven by various factors, which mainly consisted of a lack of understanding of, and familiarity with, ADM systems. A lack of technical knowledge about, and experience with, ADM systems often lead to socially grounded, and sometimes conflicting, interpretations of fairness. These findings motivated the decision to undertake a deeper exploration of stakeholder concerns regarding the use of algorithmic systems. As a result, this study laid the groundwork for the research reported on in Chapter 7. This research I examined the underlying reasons and concerns driving the not-yet-adoption of ADM systems within the organisational HR context.

RQ 8 & RQ 9: Non-Adoption of ADM Systems in the Organisational HR Context

During the interviews reported on in Chapter 6, which explored how algorithmic fairness is constructed and the factors shaping these constructions, I observed a notable reluctance among HR professionals to adopt ADM systems within their operational practices. Although this line of inquiry was not part of the original research design at the outset of my research journey, I chose to remain open to emerging areas of investigation that surfaced during the empirical process (Cloutier, 2024). This observation prompted a shift in focus towards understanding the underlying reasons for the not-yet-adoption of ADM systems in HR, leading to the formulation of Research Question 8: *"What are the concerns about the adoption of algorithmic decision-making systems by HR professionals?"* In addressing RQ 8, my analysis revealed several key concerns driving reluctance towards adoption, including issues related to the transparency, explainability, and reliability of algorithmic systems; ethical concerns, such as the potential for algorithms to produce unfair outcomes or perpetuate discrimination; a general resistance to technological change; and apprehensions regarding legal and regulatory compliance. RQ 9 *"How are those concerns shaped?"*, guided my investigation into the underlying factors influencing these concerns. My findings indicated that a lack of familiarity and understanding of algorithmic systems significantly contributes to ethical apprehensions. Specifically, concerns that ADM may unfairly disadvantage certain individuals or groups, thereby potentially reducing diversity within the workforce, were particularly pronounced among participants with limited technical knowledge.

These findings demonstrate that concerns surrounding ADM systems are deeply interrelated and mutually reinforcing. One pattern clearly reflected in the interviews was that a lack of familiarity with ADM systems often results in limited understanding because HR professionals predominantly rely on second-hand knowledge (Bucher, 2017). This absence of direct, experiential learning reinforces the HR professionals' confidence in traditional decision-making processes. Consequently, an overreliance on human judgment, despite its well-documented susceptibility to bias (Hunt et al., 2024; Thomas & Reimann, 2023), further diminishes the HR professionals' motivation to engage with or adopt ADM systems (Mahmud et al., 2022; Sutherland et al., 2016).

Another finding from this study is that risk perceptions associated with ADM systems, such as ethical concerns, or concerns relating to reputational damage, are often shaped by indirect sources and are amplified by a lack of hands-on experience. Being unfamiliar with these systems heightened participants' fears that these risks will materialise, thereby reinforcing their aversion to adoption. Even when ADM systems present clear opportunities for improving efficiency (Hunkenschroer & Luetge, 2022), HR professionals remain hesitant to engage with them due to a perceived inability to manage the risks of using these systems effectively. This reluctance is compounded by the participants' limited exposure to ADM systems and their lack of confidence in navigating regulatory and compliance frameworks.

Chapter 7 built directly upon the preceding analyses by deepening the exploration of stakeholder perspectives on ADM. While earlier chapters focused on the construction of algorithmic fairness in public-facing communications by software developers and in different stakeholder understandings, Chapter 7 shifted attention to the practical implications of these perceptions by examining the factors contributing to the not-yet-adoption of ADM systems by

HR professionals. This study complements the theoretical and empirical insights on algorithmic fairness by highlighting how limited familiarity and reliance on second-hand knowledge shape ethical concerns, risk perceptions, and resistance to technological change.

Together, all findings provide a cohesive narrative that connects the constructions of fairness to real-world challenges in implementing ADM systems, thus offering a comprehensive understanding of why algorithmic fairness is constructed in varied ways and how these constructions impact the adoption and integration of ADM systems. This integrated perspective informs the concluding sections of this chapter, which discuss the theoretical and practical contributions of this research, its limitations, and the potential future research directions to address these limitations.

8.3 Overall Discussion of Findings

Examining the findings from the different studies reported on in this thesis, it becomes evident that algorithmic fairness is constructed in varied and often inconsistent ways across stakeholder groups. Each chapter provided a different lens through which to view this phenomenon, including how algorithmic fairness is conceptualised in academic literature, how it is presented by software suppliers on their websites, how it is understood by stakeholders in practice, and, finally, the concerns expressed by HR professionals regarding the use of ADM systems.

Across all four studies, a key overarching observation was that stakeholders often found it challenging to articulate what fairness means to them in the context of ADM systems. This observation was particularly clear in the interviews conducted in Chapter 6, where most participants, regardless of their role, found it difficult to define fairness clearly or consistently without relying on other ethical terms (Dolata et al., 2022; Starke et al., 2022). This difficulty was mirrored in the findings from the website analysis in Chapter 5, whereby software suppliers often referenced "fairness" without clarifying what fairness entails or how it is achieved in their systems. Even in cases where fairness was discussed explicitly, such as “fair recruitment” or “bias-free decision-making,” the underlying assumptions or mechanisms were rarely explained. This ambiguity around fairness suggests that algorithmic fairness is not a fixed or universally agreed-upon concept (Dolata et al., 2022; Kleanthous et al., 2022; Mulligan et al., 2019). Instead, it is constructed through related ethical ideas, such as transparency, objectivity, explainability, non-discrimination, and bias mitigation, which serve as proxies for fairness (Starke et al., 2022). These related terms often functioned as indirect articulations of algorithmic fairness, allowing stakeholders to express ethical intentions without necessarily defining fairness itself.

The specific way in which algorithmic fairness was constructed varied depending on the stakeholder’s role, experience, and technical knowledge. For example, in the interviews reported on in Chapter 6, software developers and technically skilled consultants tended to take a sociotechnical view of fairness. They were more likely to discuss fairness as a balance between technical notions of algorithmic fairness and contextual or social considerations. This included for example, fairness across demographic groups or compliance with ethical norms (Dolata et al., 2022; Draude et al., 2020). In contrast, HR professionals, especially those with limited technical backgrounds, framed algorithmic fairness using socially grounded and role-relevant terms. Their constructions of algorithmic fairness were largely focused on fairness as non-discrimination, equality of opportunity, and respecting individual differences, which are concepts that are already embedded in existing HR practices and legal frameworks (Feldkamp et al., 2023). This finding is important because it highlights how existing professional norms shape how algorithmic fairness is understood, often limiting the integration of more complex or technical fairness concepts into HR discourse. Interestingly, the literature review (Chapter 4) revealed a similar pattern. Much of the academic work on algorithmic fairness is written either from a technical perspective, such as computer science (e.g., Friedler et al., 2019; Verma

& Rubin, 2018), or from a social perspective, such as organisational justice theory (e.g., Juijn et al., 2023; Narayanan et al., 2024). Very few studies compared or integrated these perspectives into a sociotechnical perspective (Dolata et al., 2022; Dolata & Schwabe, 2024; Draude et al., 2020). These findings further reinforce the idea that fairness is not only multi-dimensional but also stakeholder-contingent, and that understanding fairness requires us to pay close attention to whose perspective is being considered.

The misalignment between stakeholder perspectives on algorithmic fairness can create significant barriers to adoption. Although all interviewed stakeholders value fairness, they do not always share the same vision of what it entails, how it should be operationalised, or who should be responsible for it. This divergence reflects the broader theme across all the studies, which is a disparity between ethical intentions and actual practice. While algorithmic fairness is widely acknowledged as important, it remains inconsistently defined and communicated. Without explicit alignment across perspectives, these differences can generate tension, particularly in cross-functional settings where technical teams and HR departments must collaborate (Corrêa et al., 2025).

The lack of clarity among different stakeholder groups about the meaning of algorithmic fairness, and the difficulties regarding explaining this meaning, contribute to a broader issue that became particularly apparent during the research into HR professionals' reluctance to adopt ADM systems in HR practice (Chapter 7). Many of the research participants expressed concerns about the reliability, transparency, and ethical ramifications of ADM systems (Mahmud et al., 2022). Their hesitancy was often shaped by limited familiarity and second-hand knowledge, rather than direct experience (Bucher, 2017; Mahmud et al., 2022). These findings reinforced those reported in earlier chapters, which showed a gap between intention and understanding.

In sum, this thesis shows that algorithmic fairness in HR is not a single, shared concept but a set of overlapping, fragmented, and role-dependent constructions. Across literature, public communication, and practice, fairness is framed in different ways, and often not clearly defined. While technical stakeholders may approach fairness through measurable attributes and models, HR professionals often rely on social values and legal obligations, creating gaps in expectations, understanding, and implementation. There is, consequently, a need for more interdisciplinary dialogue, shared frameworks, and transparent communication with regard to the design, implementation as well as the governance of ADM systems in HR. I discuss this need in Chapter section 8.5.

8.4 Theoretical and Practical Contributions

This thesis addresses the need to investigate how different stakeholders construct algorithmic fairness by viewing algorithmic fairness as part of a sociotechnical system, wherein technical elements and social elements work together in joint optimisation (Dolata et al., 2022). This thesis further provides answers to the question of why ADM systems have not been extensively implemented in practice, and finds clear evidence of algorithm aversion among HR professionals (Hannon et al., 2024). My thesis specifically focused on stakeholders developing, using, and implementing ADM systems, and aimed to advance the theoretical and practical understanding of algorithmic fairness and the reasons for the not-yet-adoption of ADM systems in practice. The four manuscripts presented as part of this thesis contribute to a better understanding of these matters by providing novel insights into the social construction of algorithmic fairness and into stakeholder concerns about adoption of ADM systems in practice.

8.4.1 Theoretical Contributions

This thesis provides several theoretical contributions, which are synthesised and organised in the subsequent sections according to the thematic focus of the contribution.

Structuring the Multidisciplinary Landscape of Algorithmic Fairness

Given the expanding body of literature on algorithmic fairness, across diverse domains and stakeholder groups, the organising literature review in Chapter 4 contributes to theory by synthesising and clarifying this broad field to make it more comprehensible (Leidner, 2016). The chapter identifies and categorises stakeholder groups discussed in the organisational HR context and systematically compares how algorithmic fairness is constructed from these differing perspectives. This work provides a nuanced understanding of the factors shaping conceptions of algorithmic fairness and how these depend upon the stakeholder lens taken. While decision-subjects are frequently discussed in the literature, this focus often serves to infer implications for managers or software developers by examining what individuals affected by algorithmic decisions perceive as fair or unfair. By comparatively analysing constructions of fairness across different stakeholder groups, my research addresses a significant gap in the literature, which tends to examine primarily decision-subjects (e.g., Acikgoz et al., 2020; Lavanchy et al., 2023; Parent-Rocheleau & Parker, 2022) to inform other stakeholders' perspectives, such as managers or developers. I also demonstrate that existing literature reviews on algorithmic fairness typically adopt either a social lens, grounded in organisational justice theory (e.g., Ochmann et al., 2024) or a purely technical perspective (Barocas & Selbst, 2016; Lämmermann et al., 2022). By adopting a sociotechnical stance and incorporating multiple stakeholder viewpoints, my review bridges this divide and highlights the critical importance of recognising diverse constructions of algorithmic fairness.

Expanding Theoretical Understanding of Stakeholder Perspectives

Chapters 5 and 6 together contribute to theory by broadening the focus of algorithmic fairness research beyond decision-subjects to include stakeholders who actively shape, operationalise, and communicate fairness in algorithmic systems. Through interviews and a website analysis, I illuminate the perspectives of actors involved in the development, implementation, and use of ADM systems. The analysis of software suppliers' websites, as outlined in Chapter 5, demonstrates how these critical stakeholders, whose design choices directly influence algorithmic functionality (Kasinidou et al., 2021; Kleanthous et al., 2022), publicly construct notions of fairness, frame them for external audiences, and potentially shape societal expectations. The interview analysis in Chapter 6 extends this work by revealing how fairness considerations are embedded in practice and interpreted across organisational contexts. This study highlights gaps between public claims, internal priorities, and the lived realities of ADM system use. Together, these findings offer a rich and multi-layered account of how different stakeholders construct fairness, and they reveal asymmetries in understandings that can inform both theory and practice.

Providing Domain-Specific Insights into Algorithm Aversion in HR

In Chapter 7, I examined the concerns and reasons for the not-yet-adoption of ADM systems by HR professionals. This study extends the existing body of literature on algorithm aversion, which typically addresses general algorithm aversion without a domain-specific focus (e.g., Mahmud et al., 2022). By concentrating on HR professionals, my study contributes to theory by offering domain-specific insights into the particular concerns held by this stakeholder group. I identified several reasons uniquely pertinent to the HR context, including apprehensions about discrimination by ADM systems, which may result in talent loss and a consequent reduction in workforce diversity. My findings also indicate that the type of algorithmic system influences levels of algorithm aversion, a factor insufficiently explored in prior research. Together, these contributions provide a nuanced understanding of the barriers to ADM system adoption in HR and enrich the growing discourse on the application of these systems in the HR context.

8.4.2 Practical Contributions

This thesis also provides several practical contributions, which are synthesised and organised according to the thematic focus of the contributions. In the subsequent sections I present these practical contributions in detail.

Informing the Design and Evaluation of Fair ADM Systems

The literature review presented in Chapter 4 identifies the various aspects of algorithmic fairness considered by different stakeholders, including regulatory implications, technical notions of fairness, and social fairness perceptions. I argue that this understanding is crucial because it informs the development and evaluation of ADM systems with respect to algorithmic fairness. Specifically, my findings provide valuable guidance on designing fair algorithms by highlighting the differing emphases that stakeholders place on aspects of algorithmic fairness. Moreover, the findings of the interview analysis in Chapter 5 contribute to the design and evaluation of ADM systems by enabling better alignment with varied stakeholder understandings of algorithmic fairness. My findings further offer actionable insights for developers and organisations, as can be drawn from my study on the not-yet-adoption of ADM systems, detailed in Chapter 7. This study provides a foundational basis for designing ADM systems and implementation strategies that directly address end-user concerns, particularly by enhancing transparency, explainability, and change management processes.

Supporting Policymakers in Communicating and Operationalising Fairness

For policymakers and regulators, my research emphasises the necessity of clear communication regarding legal compliance and ethical considerations, to facilitate responsible ADM system adoption within HR. Ultimately, this research reaffirms that effective change management practices constitute critical components for the successful integration of AI technologies into HR functions, as discussed in Chapter 7. Moreover, the insights from Chapters 4, 5, and 6 could assist policymakers and regulators to recognise the influence of diverse stakeholder perspectives on the conceptualisation and implementation of algorithmic fairness. Therefore, my research can offer support to policymakers who wish to articulate a more precise and nuanced conceptualisation of fairness in future legislation, thereby helping them to mitigate the prevailing confusion and conflation with related yet distinct concepts such as transparency, each of which demands careful conceptual and practical differentiation.

Facilitating Cross-Stakeholder Communication on Algorithmic Fairness

The thesis findings derived from the interview analysis in Chapter 6 provide valuable insights from stakeholders involved both directly in the development and implementation of ADM systems, as well as from potential end-users. This comprehensive understanding highlights the diverse interpretations of algorithmic fairness, and recognises that no single, universally accepted definition exists. These insights can offer assistance to practitioners who wish to acknowledge multiple fairness perspectives and facilitate clearer communication when addressing fairness within ADM contexts.

Enhancing Explainability as a Driver of Fairness Perceptions

The thesis findings highlight the importance of explainability in ADM systems. This matter was emphasised in the interview studies (Chapter 6), which revealed that stakeholders often equate algorithmic fairness with a lack of explainability and transparency. Furthermore, my investigation into the not-yet-adoption of these systems (Chapter 7) demonstrated that one of the primary reasons for their limited use is HR professionals' insufficient understanding of how these systems function, which leads to a lack of trust. Consequently, increased practical engagement, whether by software suppliers or organisations developing in-house ADM systems, is essential. This engagement must highlight the important role that explainability plays in ensuring perceptions of fairness and facilitating the adoption of ADM systems.

8.5 Limitations and Future Research

While this thesis offers novel insights into how various stakeholders understand and approach algorithmic fairness, several limitations must be acknowledged. While these limitations reflect the constraints inherent in the research design and context, they also highlight valuable opportunities for future investigation.

First, the organising literature review undertaken for this thesis (Chapter 4) relied on inductive thematic development, which, while effective in capturing emerging patterns and diverse perspectives, may limit conceptual clarity and theoretical coherence. Compared to systematic or theory-driven reviews, organising reviews are often regarded as making a more modest theoretical contribution because they lack the guiding structure provided by an explicit theoretical model (Leidner, 2016). To advance the field, future research could undertake a systematic or theory-driven literature review, enabling greater analytical precision and more direct contributions to theory building. This approach would facilitate a mapping of the evolution of fairness-related concepts across disciplines, such as law, computer science, and ethics, and help synthesise cross-domain insights into algorithmic fairness, which my study does not encompass.

Second, while my research engaged with fairness-related statements and claims made by software suppliers on their websites, it did not delve deeply into the meanings behind those statements and claims. Often, terms like “fair,” “transparent,” or “ethical” are employed rhetorically in marketing materials, with little explanation or operational detail. It remains unclear whether these claims reflect internal commitments or are strategic responses to current public discourse. Therefore, it is difficult to make generalisation based on my findings. Instead, future research could conduct content analyses of corporate communication, followed by interviews with software developers working at those organisations to examine how fairness is defined and operationalised in practice. This research could contrast external-facing narratives with internal practices and shed light on possible gaps or contradictions.

Third, this research did not involve interview participants who were software providers working at the software supplier companies that were part of the thematic analysis of software providers' websites. Consequently, the research missed an opportunity to triangulate publicly available fairness discourse with internal organisational perspectives. While this omission was largely due to access constraints, future studies could aim to include software vendors in a multi-stakeholder design, and to compare how algorithmic fairness is constructed across developers, HR professionals, and end-users. Methodologically, this further research might involve embedded case studies or ethnographic research within firms who develop ADM systems.

Fourth, the research design for this thesis employed a small-n qualitative design compared to larger datasets (Tsoukas, 2018). This approach was driven by limited access to individuals with direct experience of ADM systems and the sensitive nature of fairness discussions within HR, where reputational risks may discourage disclosure, particularly regarding ethically sensitive applications, such as recruitment. Consequently, the research prioritised in-depth analysis over broad representativeness, which resulted in rich, context-specific insights. However, this approach limits the generalisability of the research findings. Future research should consider large-scale surveys, or mixed method approaches to validate and extend these results across larger populations, and to identify statistically robust trends.

A fifth limitation relates to geographic scope. The majority of interviews were conducted with professionals located in Australia, which limits the insights into the influences of culture. Moreover, the participant sample was unevenly distributed, with a higher number of Australian participants due to challenges in recruiting German HR professionals. While some of these findings were validated through discussions with Germany-based scholars in the same research domain, this validation cannot fully compensate for the lack of broader international representation. Future research should expand to include participants from other regions, particularly countries in Asia, Africa, North America, and South America, where cultural, institutional, and regulatory environments may differ significantly. Such comparative work could examine how culturally embedded notions of fairness shape expectations of ADM systems, and how national regulatory regimes influence their design and deployment.

Sixth, and in relation to HR functions, while this research aimed to investigate a range of operational domains, including recruitment, performance appraisal, and compensation, most interview data came from professionals involved in recruitment and selection. Consequently, the research findings may not be fully representative of fairness considerations in other HR contexts. Future research could explore fairness perceptions across a wider spectrum of HR tasks, investigating, for instance, how algorithmic tools shape decisions about employee development, promotion, or termination. This future work could also involve longitudinal research to track how perceptions of fairness evolve as organisations implement ADM tools over time.

A seventh limitation concerns the limited direct experience that many HR professionals in the study had with ADM systems. While this lack of participant experience provided an insight into expectations and concerns prior to adoption, it limited the ability to capture post-implementation experiences. To better understand how algorithmic fairness is negotiated in day-to-day use, future research should include HR professionals and organisations with mature ADM practices. This work could involve real-time observation of decision-making processes or retrospective accounts of system deployment and adaptation.

Eighth, many interview participants, as well as researchers consulted during the study, confirmed the difficulty of accessing actual ADM users within organisations. This difficulty supported the decision to engage potential end-users through hypothetical scenarios. Although this refined approach provided useful insights into stakeholder expectations, it also pointed to the need for creative research methods, such as simulations, that allow stakeholders to interact with ADM systems in controlled settings. These methods could result in more valid data while still preserving participant confidentiality and comfort.

Ninth, the research findings suggest that individuals without technical backgrounds often understand algorithmic fairness predominantly through a social or ethical perspective, rather than from a technical or statistical perspective. This difference highlights the importance of developing a sociotechnical perspective on fairness that is accessible to non-expert users. Future research could explore how to design interfaces, training materials, and governance mechanisms, that translate complex technical fairness metrics into concepts that resonate with everyday users. Participatory design approaches may be particularly useful here because they could enable users to co-construct meanings of fairness alongside system developers and policymakers.

Finally, the timing of this research is an important consideration for limitations and future work. Recruitment for interview participants began in 2024, when generative AI tools like ChatGPT were just emerging, and ADM systems were primarily used in large, tech-savvy organisations such as Amazon (Dastin, 2018). Many smaller or less technologically advanced companies had not yet adopted these systems, which limited access to participants with direct experience. Since then, the landscape has shifted rapidly, with widespread adoption of generative AI and the introduction of new AI systems (Agrawal, 2024; Bick et al., 2025). These developments highlight the need for future research to investigate whether the current limited adoption of ADM systems in HR reflects a not-yet-adoption phase that can be overcome through usability, exposure, and institutional support (Bick et al., 2025; Cubric, 2020; Hangl et al., 2023; Horani et al., 2025). Algorithm aversion literature (Mahmud et al., 2022; Lee, 2018; Ochmann et al., 2024) and broader AI adoption studies (Hangl et al., 2023; Horani et al., 2025; Khanfar et al., 2026) suggest that perceived barriers such as distrust, transparency, or uncertainty are not fixed and can change as users gain experience and familiarity (Marocco et al., 2024; Kar & Kushwaha, 2023). Moreover, prior exposure to similar technologies can lower adoption barriers and increase the likelihood of adopting other AI systems over time (Horowitz et al., 2024). Investigating these dynamics in HR-specific contexts will help clarify how evolving AI capabilities and organisational exposure shape adoption trajectories and inform strategies for integrating ADM systems effectively.

In sum, while this research offers valuable insights into the contested and multifaceted nature of algorithmic fairness in HR, it also opens multiple pathways for deeper and broader inquiry. Addressing these limitations will not only strengthen the empirical foundations of the field but also contribute to more inclusive, accountable, and socially meaningful ADM systems.

8.6 Concluding Remarks

This thesis, including its constituent publications, has investigated the construction of algorithmic fairness and the factors contributing to the not-yet--adoption of ADM systems within the organisational HR context. Grounded in a relational onto-epistemological worldview, the research elucidates how algorithmic fairness is shaped by diverse factors, which are contingent upon the stakeholder perspectives considered. First, I provided an organising analysis of existing literature to reveal how algorithmic fairness is conceptualised differently across stakeholder groups, and I emphasised that perceptions of decision-subjects significantly influence managerial and developer interpretations of fairness. Second, I explored how software suppliers publicly construct algorithmic fairness on their websites, and I identified the employment lifecycle phases during which fairness is predominantly foregrounded. Third, I examined how stakeholders involved in the development, implementation, and use of ADM systems construct algorithmic fairness in practice within the HR domain. Fourth, I analysed the underlying reasons for the limited adoption of these systems, with algorithmic fairness emerging as a critical concern. Collectively, these contributions advance scholarly understandings of algorithmic fairness and its implications for the adoption of ADM in HR practices. It is my hope that this research inspires future scholars to critically interrogate the proliferation and sociotechnical dynamics of algorithmic systems in organisational settings, including HR contexts.

References

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

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

Adams, J. S. (1965). Inequity in social exchange. *Advances in Experimental Social Psychology*, *2*, 267–299.

Afiouni, R. (2019). Organizational learning in the rise of machine learning. *ICIS 2019 Proceedings*. ICIS, Munich, Germany.

Agarwal, A., Beygelzimer, A., Dudík, M., Langford, J., & Wallach, H. (2018). *A reductions approach to fair classification* (No. arXiv:1803.02453). arXiv. http://arxiv.org/abs/1803.02453

Agrawal, K. P. (2024). Towards adoption of generative AI in organizational settings. *Journal of Computer Information Systems, 64*(5), 636-651. https://doi.org/10.1080/08874417.2023.2240744

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, *60*, 102387. https://doi.org/10.1016/j.ijinfomgt.2021.102387

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal*, *26*(1), 1–11. https://doi.org/10.1111/1748-8583.12090

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

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*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Arif, I., Aslam, W., & Hwang, Y. (2020). Barriers in adoption of internet banking: A structural equation modeling – Neural network approach. *Technology in Society*, *61*, 101231. https://doi.org/10.1016/j.techsoc.2020.101231

Bankins, S., Formosa, P., Griep, Y., & Richards, D. (2022). AI decision making with dignity? Contrasting workers' justice perceptions of human and AI decision making in a human resource management context. *Information Systems Frontiers*, *24*(3), 857–875. https://doi.org/10.1007/s10796-021-10223-8

Bao, W., Lianju, N., & Yue, K. (2019). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications*, *128*, 301–315. https://doi.org/10.1016/j.eswa.2019.02.033

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*, 671–732. https://doi.org/10.2139/ssrn.2477899

Berger, P. L., & Luckmann, T. (1991). *The social construction of reality: A treatise in the sociology of knowledge*. Penguin Books.

Bick, A., Blandin, A., & Deming, D. J. (2025). *The rapid adoption of generative AI* (NBER Working Paper No. 32966). National Bureau of Economic Research. https://doi.org/10.3386/w32966

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research*, *81*, 1–11. http://arxiv.org/abs/1712.03586
Bogen, M. (2019). All the ways hiring algorithms can introduce bias. *Harvard Business Review*. https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias
Bostrom, R. P., & Heinen, J. S. (1977). MIS problems and failures: A socio-technical perspective. Part I: The causes. *MIS Quarterly*, *1*(3), 17. https://doi.org/10.2307/248710
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44. https://doi.org/10.1080/1369118x.2016.1154086
Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., Boselie, P., Lee Cooke, F., Decker, S., DeNisi, A., Dey, P. K., Guest, D., Knoblich, A. J., Malik, A., Paauwe, J., Papagiannidis, S., Patel, C., Pereira, V., Ren, S., Rogelberg, S., Saunders, M. N. K., Tung, R. L. & Varma, 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
Burnett, J. R., & Lisk, T. C. (2019). The future of employee engagement: Real-time monitoring and digital tools for engaging a workforce. *International Studies of Management & Organization*, *49*(1), 108–119. https://doi.org/10.1080/00208825.2019.1565097
Burton, J. W., Stein, M., & 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
Capterra. (n.d.). *Find the right software with Capterra Australia*. Retrieved 20 July 2025, from https://www.capterra.com.au/
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
Cecez-Kecmanovic, D., Galliers, R. D., Henfridsson, O., Newell, S., & Vidgen, R. (2014). The sociomaterialty of information systems: Current status, future directions. *MIS Quarterly*, *38*(3), 809–830. https://doi.org/10.25300/misq/2014/38:3.3
Charmaz, K. (2006). *Constructing grounded theory*. Sage Publications.
Charmaz, K. (2008). Constructionism and the grounded theory method. In J. A. Holstein & J. F. Gubrium (Eds), *Handbook of constructionist research* (pp. 397–412). Guilford Press.
Charmaz, K. (2020). "With constructivist grounded theory you can't hide": Social justice research and critical inquiry in the public sphere. *Qualitative Inquiry*, *26*(2), 165–176. https://doi.org/10.1177/1077800419879081
Chatterjee, S., Sarker, S., Lee, M. J., Xiao, X., & Elbanna, A. (2021). A possible conceptualization of the information systems (IS) artifact: A general systems theory perspective. *Information Systems Journal*, *31*(4), 550–578. https://doi.org/10.1111/isj.12320
Chen, Chiang, & Storey. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165. https://doi.org/10.2307/41703503
Cheng, H.-F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019). Explaining decision-making algorithms through UI: Strategies to help non-expert

stakeholders. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300789
Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153–163. https://doi.org/10.1089/big.2016.0047
Cloutier, C. (2024). Strategies for generating deliberately emergent qualitative research designs. *The Journal of Applied Behavioral Science*, *60*(2), 358–380. https://doi.org/10.1177/00218863241235417
Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383
Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797–806. https://doi.org/10.1145/3097983.3098095
Corrêa, N. K., Santos, J. W., Galvão, C., Pasetti, M., Schiavon, D., Naqvi, F., Hossain, R., & Oliveira, N. D. (2025). Crossing the principle–practice gap in AI ethics with ethical problem-solving. *AI and Ethics*, *5*(2), 1271–1288. https://doi.org/10.1007/s43681-024-00469-8
Crotty, M. (1998). *Foundations of social research: Meaning and perspectives in the research process*. Routledge.
Cubric, M. (2020). Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. *Technology in Society, 62*, 101257. https://doi.org/10.1016/j.techsoc.2020.101257
Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/
Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. https://doi.org/ 10.1037/0003-066X.34.7.571
Demuijnck, G. (2009). Non-discrimination in human resources management as a moral obligation. *Journal of Business Ethics*, *88*(1), 83–101. https://doi.org/10.1007/s10551-009-0100-6
Denham, M. A., & Onwuegbuzie, A. J. (2013). Beyond words: Using nonverbal communication data in research to enhance thick description and interpretation. *International Journal of Qualitative Methods*, *12*(1), 670–696. https://doi.org/10.1177/160940691301200137
Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y. K., Mäntymäki, M., & Pappas, I. O. (2023). Artificial intelligence (AI) and information systems: Perspectives to responsible AI. *Information Systems Frontiers*, *25*(1), 1–7. https://doi.org/10.1007/s10796-022-10365-3
Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science*, *31*(10), 1302–1314. https://doi.org/10.1177/0956797620948841
Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033
Dodge, J., Liao, Q. V., Zhang, Y., Bellamy, R. K. E., & Dugan, C. (2019). Explaining models: An empirical study of how explanations impact fairness judgment. *Proceedings of the*

24th International Conference on Intelligent User Interfaces, 275–285. https://doi.org/10.1145/3301275.3302310

Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal*, *32*(4), 754–818. https://doi.org/10.1111/isj.12370

Dolata, M., & Schwabe, G. (2024). Towards the socio-algorithmic construction of fairness: The case of automatic price-surging in ride-hailing. *International Journal of Human–Computer Interaction*, *40*(1), 55–65. https://doi.org/10.1080/10447318.2023.2210887

Douglas, D. M., Lacey, J., & Howard, D. (2024). Ethical risk for AI. *AI and Ethics*, *5*(3), 2189–2203. https://doi.org/10.1007/s43681-024-00549-9

Draude, C., Klumbyte, G., Lücking, P., & Treusch, P. (2020). Situated algorithms: A sociotechnical systemic approach to bias. *Online Information Review*, *44*(2), 325–342. https://doi.org/10.1108/oir-10-2018-0332

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *ITCS '12: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ITCS, Cambridge, MA. https://doi.org/10.1145/2090236.2090255

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors*, *44*(1), 79–94. https://doi.org/10.1518/0018720024494856

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment*, *50*(3), 387–395. https://doi.org/10.1207/s15327752jpa5003_8

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers*, *24*(5), 1709–1734. https://doi.org/10.1007/s10796-021-10186-w

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

Fazelpour, S., & Lipton, Z. C. (2020). Algorithmic fairness from a non-ideal perspective. *AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 57–63. https://doi.org/10.1145/3375627.3375828

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

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015). Certifying and removing disparate impact. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 259–268. https://doi.org/10.1145/2783258.2783311

Festinger, L. (1957). *A theory of cognitive dissonance*. Row, Peterson.

Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E. P., & Roth, D. (2019). A comparative study of fairness-enhancing interventions in machine learning. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 329–338. https://doi.org/10.1145/3287560.3287589

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance*, *77*(1), 5–47. https://doi.org/10.1111/jofi.13090

Gal, U., Riemer, K., Harth-Kitzerow, C., Aboud, C., & Briggs, S. (2019). *People analytics: Using data and algorithms to shape the employee experience*.

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

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

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

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

Glymour, B., & Herington, J. (2019). Measuring the biases that matter: The ethical and casual foundations for measures of fairness in algorithms. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 269–278. https://doi.org/10.1145/3287560.3287573

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

Green, B. (2022). Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology*, *35*(4), 90. https://doi.org/10.1007/s13347-022-00584-6

Grgić-Hlača, N., Zafar, M. B., Gummadi, K. P., & Weller, A. (2018). Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). https://doi.org/10.1609/aaai.v32i1.11296

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

Haas, C. (2019). The price of fairness – A framework to explore trade-offs in algorithmic fairness. *ICIS 2019 Proceedings*. ICIS, Munich, Germany.

Hangl, J., Krause, S., & Behrens, V. J. (2023). Drivers, barriers, and social considerations for AI adoption in SCM. *Technology in Society, 74*, 10229. https://doi.org/10.1016/j.techsoc.2023.102299

Hannon, O., Ciriello, R., & Gal, U. (2024). Just because we can, doesn’t mean we should: Algorithm aversion as a principled resistance. *Proceedings of the Annual Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2024.730

Hardt, M., Price, E., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *30th Conference on Neural Information Processing Systems*.

Hart, C. (1998). *Doing a literature review: Releasing the social science research imagination*. Sage Publications.

Hazelrigg, L. E. (1986). Is there a choice between ‘constructionism’ and ‘objectivism’? *Social Problems*, *33*(6), S1–S13. https://doi.org/10.2307/800671

Helberger, N., Araujo, T., & de Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Computer Law & Security Review*, *39*, 105456. https://doi.org/10.1016/j.clsr.2020.105456

Herm, L.-V., Heinrich, K., Wanner, J., & Janiesch, C. (2023). Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. *International Journal of Information Management*, *69*, 102538. https://doi.org/10.1016/j.ijinfomgt.2022.102538

Heyder, T., Passlack, N., & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. *The Journal of Strategic Information Systems*, *32*(3), 101772. https://doi.org/10.1016/j.jsis.2023.101772

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, *1*(3), 333–342. https://doi.org/10.1111/j.17549434.2008.00058.x

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Holm, S. (2023). Egalitarianism and algorithmic fairness. *Philosophy & Technology*, *36*(1), 6. https://doi.org/10.1007/s13347-023-00607-w

Holstein, K., Wortman Vaughan, J., Daumé, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What so industry practitioners need? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3290605.3300830

Horani, O. M., Al-Adwan, A. S., Yaseen, H., Hmoud, H., Al-Rahmi, W. M., & Alkhalifah, A. (2025). The critical determinants impacting artificial intelligence adoption at the organizational level. *Information Development, 41*(3), 1055-1079. https://doi.org/10.1177/02666669231166889

Horowitz, M. C., Kahn, L., Macdonald, J., & Schneider, J. (2024). Adopting AI: how familiarity breeds both trust and contempt. *AI & Society, 39*, 1721-1735. https://doi.org/10.1007/s00146-023-01666-5

Hoshino, R., Slobodin, A., & Bernoudy, W. (2018). An automated employee timetabling system for small businesses. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). https://doi.org/10.1609/aaai.v32i1.11383

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

Hunt, A. K., Wang, J., Alizadeh, A., & Pucelj, M. (2024). Advancing a theoretical framework for exploring heuristics and biases within HR decision-making contexts. *Personnel Review*, *53*(7), 1823–1841. https://doi.org/10.1108/pr-03-2023-0192

Irvine, A., Drew, P., & Sainsbury, R. (2012). 'Am I not answering your questions properly?' Clarification, adequacy and responsiveness in semi-structured telephone and face-to-face interviews. *Qualitative Research*, *13*(1), 87–106. https://doi.org/10.1177/1468794112439086

Jabagi, N., Croteau, A.-M., Audebrand, L. K., & Marsan, J. (2019). Gig-workers' motivation: Thinking beyond carrots and sticks. *Journal of Managerial Psychology*, *34*(4), 192–213. https://doi.org/10.1108/jmp-06-2018-0255

Jabagi, N., Croteau, A.-M., Audebrand, L. K., & Marsan, J. (2025). Do algorithms play fair? Analysing the perceived fairness of HR-decisions made by algorithms and their impacts on gig-workers. *The International Journal of Human Resource Management*, *36*(2), 235–274. https://doi.org/10.1080/09585192.2024.2441448

Jacobs, A. Z., & Wallach, H. (2021). Measurement and fairness. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 375–385. https://doi.org/10.1145/3442188.3445901

Jago, A. S. (2019). Algorithms and authenticity. *Academy of Management Discoveries*, *5*(1), 38–56. https://doi.org/10.5465/amd.2017.0002

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

Jo, T. (2021). *Machine learning foundations: Supervised, unsupervised, and advanced learning*. Springer International Publishing. https://doi.org/10.1007/978-3-030-65900-4

Jones-Jang, S. M., & Park, Y. J. (2022). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication*, *28*(1). https://doi.org/10.1093/jcmc/zmac029

Joshi, K. (1989). The measurement of fairness or equity perceptions of management information systems users. *MIS Quarterly*, *13*(3), 343. https://doi.org/10.2307/249010

Juijn, G., Stoimenova, N., Reis, J., & Nguyen, D. (2023). Perceived algorithmic fairness using organizational justice theory: An empirical case study on algorithmic hiring. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 775–785. https://doi.org/10.1145/3600211.3604677

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

Kar, A. K. & Kushwaha, A. K. (2023). Facilitators and barriers of artificial intelligence adoption in business – insights from opinions using big data analytics. *Information System Frontiers, 25*, 1351-1374. https://doi.org/10.1007/s10796-021-10219-4

Kasinidou, M., Kleanthous, S., Barlas, P., & Otterbacher, J. (2021). I agree with the decision, but they didn’t deserve this: Future developers’ perception of fairness in algorithmic decisions. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 690–700. https://doi.org/10.1145/3442188.3445931

Kasirzadeh, A. (2022). Algorithmic fairness and structural injustice: Insights from feminist political philosophy. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 349–356. https://doi.org/10.1145/3514094.3534188

Kelan, E. K. (2024). Algorithmic inclusion: Shaping the predictive algorithms of artificial intelligence in hiring. *Human Resource Management Journal*, *34*(3), 694–707. https://doi.org/10.1111/1748-8583.12511

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

Khanfar, A. A., Mavi, R. K., Iranmanesh, M., & Gengatharen, D. (2026). Determinants of artificial intelligence adoption: research themes and future directions. *Information Technology and Management, 27*, 31-51. https://doi.org/10.1007/s10799-024-00435-0

Kleanthous, S., Kasinidou, M., Barlas, P., & Otterbacher, J. (2022). Perception of fairness in algorithmic decisions: Future developers’ perspective. *Patterns*, *3*(1), 100380. https://doi.org/10.1016/j.patter.2021.100380

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2021). Highly accurate, but still discriminatory: A fairness evaluation of algorithmic video analysis in the recruitment context. *Business & Information Systems Engineering*, *63*(1), 39–54. https://doi.org/10.1007/s12599-020-00673-w

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

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085x.2021.1927212

Kozodoi, N., Jacob, J., & Lessmann, S. (2022). Fairness in credit scoring: Assessment, implementation and profit implications. *European Journal of Operational Research*, *297*, 1083–1094. https://doi.org/10.1016/j.ejor.2021.06.023

Lämmermann, L., Zwickel, A., & Markgraf, M. (2022). AI fairness at subgroup level – A structured literature Review. *ECIS 2022 Proceedings*. ECIS, Timișoara, Romania.

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). *How we analyzed the COMPAS recidivism algorithm*. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Lavanchy, M., Reichert, P., Narayanan, J., & Savani, K. (2023). Applicants' fairness perceptions of algorithm-driven hiring procedures. *Journal of Business Ethics*, *188*(1), 125–150. https://doi.org/10.1007/s10551-022-05320-w

Lee, A. S. (2004). Thinking about social theory and philosophy for information systems. In J. Mingers & L. Willcocks (Eds), *Social theory and philosophy for information systems* (pp. 1–26). John Wiley & Sons.

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

Lee, M. K., & Baykal, S. (2017). Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1035–1048. https://doi.org/10.1145/2998181.2998230

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–26. https://doi.org/10.1145/3359284

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1603–1612. https://doi.org/10.1145/2702123.2702548

Leewis, S., & Smit, K. (2023). *What other factors might impact building trust in government decisions based on decision support systems, except for transparency and explainability?*

Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics*, *160*(2), 377–392. https://doi.org/10.1007/s10551-019-04204-w

Leidner, D. E. (2016). Review and theory symbiosis: An introspective retrospective. *Journal of the Association for Information Systems*, *19*(06), 552–567. https://doi.org/10.17705/1jais.00501

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: The International Journal of an Emerging Transdiscipline*, *9*, 181–212. https://doi.org/10.28945/479

Li, H., Sarathy, R., Zhang, J., & Luo, X. (2014). Exploring the effects of organizational justice, personal ethics and sanction on internet use policy compliance: Determinants of IUP compliance. *Information Systems Journal*, *24*(6), 479–502. https://doi.org/10.1111/isj.12037

Lourenço, C. J. S., Dellaert, B. G. C., & Donkers, B. (2020). Whose algorithm says so: The relationships between type of firm, perceptions of trust and expertise, and the acceptance of financial robo-advice. *Journal of Interactive Marketing*, *49*(1), 107–124. https://doi.org/10.1016/j.intmar.2019.10.003

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, mksc.2019.1192. https://doi.org/10.1287/mksc.2019.1192

Maasland, C., & Weißmüller, 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*, 779028. https://doi.org/10.3389/fpsyg.2022.779028

Madaio, M., Egede, L., Subramonyam, H., Vaughan, J. W., & Wallach, H. (2022). *Assessing the fairness of AI systems: AI practitioners' processes, challenges, and needs for support* (No. arXiv:2112.05675). arXiv. http://arxiv.org/abs/2112.05675

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

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*, 121390. https://doi.org/10.1016/j.techfore.2021.121390

Mahmud, H., Islam, A. K. M. N., & Mitra, R. K. (2023). What drives managers towards algorithm aversion and how to overcome it? Mitigating the impact of innovation resistance through technology readiness. *Technological Forecasting and Social Change*, *193*, 122641. https://doi.org/10.1016/j.techfore.2023.122641

Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, *120*, 262–273. https://doi.org/10.1016/j.jbusres.2020.07.045

Manoharan, T. R., Muralidharan, C., & Deshmukh, S. G. (2011). An integrated fuzzy multi-attribute decision-making model for employees' performance appraisal. *The International Journal of Human Resource Management*, *22*(3), 722–745. https://doi.org/10.1080/09585192.2011.543763

Mara, L.-C., Ginieis, M., & Brunet-Icart, I. (2021). Strategies for coping with LGBT discrimination at work: A systematic literature review. *Sexuality Research and Social Policy*, *18*(2), 339–354. https://doi.org/10.1007/s13178-020-00462-w

Marocco, S., Barbieri, B., & Talamo, A. (2024). Exploring facilitators and barries to managers' adoption of AI-based systems in decision-making: A systematic review. *AI, 5*(4), 2538-2567. https://doi.org/10.3390/ai5040123

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, *160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

McHugh, D., Shaw, S., Moore, T. R., Ye, L. Z., Romero-Masters, P., & Halverson, R. (2020). Uncovering themes in personalized learning: Using natural language processing to analyze school interviews. *Journal of Research on Technology in Education*, *52*(3), 391–402. https://doi.org/10.1080/15391523.2020.1752337

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607

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

Möhlmann, M., Zalmanson, L., Hendridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *MIS Quarterly*, *45*(4), 1999–2922. https:// doi.org/10.25300/misq/2021/15333

Mulligan, D. K., Kroll, J. A., Kohli, N., & Wong, R. Y. (2019). This thing called fairness: Disciplinary confusion realizing a value in technology. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–36. https://doi.org/10.1145/3359221

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, *17*(1), 2–26. https://doi.org/10.1016/j.infoandorg.2006.11.001

Namvar, M., Intezari, A., Akhlaghpour, S., & Brienza, J. P. (2023). Beyond effective use: Integrating wise reasoning in machine learning development. *International Journal of Information Management*, *69*, 102566. https://doi.org/10.1016/j.ijinfomgt.2022.102566

Narayanan, D., Nagpal, M., McGuire, J., Schweitzer, S., & De Cremer, D. (2024). Fairness perceptions of artificial intelligence: A review and path forward. *International Journal of Human–Computer Interaction*, *40*(1), 4–23. https://doi.org/10.1080/10447318.2023.2210890

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, *160*, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

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/10.1111/isj.12482

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, *37*. https://doi.org/10.17705/1cais.03743

Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F., & Agreda, S. G. (2022). To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts. *Technological Forecasting and Social Change*, *181*, 121763. https://doi.org/10.1016/j.techfore.2022.121763

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown/Archtetype.

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, *28*(9), 1435–1448. https://doi.org/10.1177/0170840607081138

Orlikowski, W. J., & Scott, S. V. (2015). Exploring material-discursive practices. *Journal of Management Studies*, *52*(5), 697–705. https://doi.org/10.1111/joms.12114

Ötting, S. K., & Maier, G. W. (2018). The importance of procedural justice in human–machine interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior*, *89*, 27–39. https://doi.org/10.1016/j.chb.2018.07.022

Panarese, P., Grasso, M. M., & Solinas, C. (2025). Algorithmic bias, fairness, and inclusivity: A multilevel framework for justice-oriented AI. *AI & SOCIETY*. https://doi.org/10.1007/s00146-025-02451-2

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parent-Rocheleau, X., & Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resource Management Review*, *32*(3), 100838. https://doi.org/10.1016/j.hrmr.2021.100838

Pazner, E. A., & Schmeidler, D. (1978). Egalitarian equivalent allocations: A new concept of economic equity. *The Quarterly Journal of Economics*, *92*(4), 671. https://doi.org/10.2307/1883182

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, *14*(3), 399–441. https://doi.org/10.1177/030631284014003004

Pinheiro, P. G., Ramos, J. J. G., Donizete, V. L., Picanco, P., & De Oliveira, G. H. (2017). Workplace emotion monitoring – An emotion-oriented system hidden behind a receptionist robot. In D. Zhang & B. Wei (Eds), *Mechatronics and robotics engineering for advanced and intelligent manufacturing*. Springer International Publishing. https://doi.org/10.1007/978-3-319-33581-0

Piorkowski, D., Park, S., Wang, A. Y., Wang, D., Muller, M., & Portnoy, F. (2021). How AI developers overcome communication challenges in a multidisciplinary team: A case study. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–25. https://doi.org/10.1145/3449205

Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). *On fairness and calibration* (No. arXiv:1709.02012). arXiv. http://arxiv.org/abs/1709.02012

Prahl, A., & Van Swol, L. (2021). Out with the humans, in with the machines? Investigating the behavioral and psychological effects of replacing human advisors with a machine. *Human-Machine Communication*, *2*, 209–234. https://doi.org/10.30658/hmc.2.11

Pritchard, G., Vines, J., Briggs, P., Thomas, L., & Olivier, P. (2014). Digitally driven: How location based services impact the work practices of London bus drivers. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3617–3626. https://doi.org/10.1145/2556288.2557156

Raghavan, M. (2024). What should we do when our ideas of fairness conflict? *Communications of the ACM*, *67*(1), 88–97. https://doi.org/10.1145/3587930

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, *20*(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8

Ram, S., & Sheth, J. N. (1989). Consumer resistance to innovations: The marketing problem and its solutions. *Journal of Consumer Marketing*, *6*(2), 5–14. https://doi.org/10.1108/eum0000000002542

Rani, U., & Furrer, M. (2021). Digital labour platforms and new forms of flexible work in developing countries: Algorithmic management of work and workers. *Competition & Change*, *25*(2), 212–236. https://doi.org/10.1177/1024529420905187

Rawls, J. (1958). Justice as fairness. *The Philosophical Review*, *67*(2), 164–194.

Rawls, J., & Kelly, E. (2003). *Justice as fairness: A restatement* (3rd edn). Havard University Press.

Rieskamp, J., Hofeditz, L., Mirbabaie, M., & Stieglitz, S. (2023). Approaches to improve fairness when deploying AI-based algorithms in hiring – Using a systematic literature review to guide future research. *Proceedings of the Annual Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2023.026

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human–Computer Interaction*, *35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Sarker, I. H. (2022). AI-based modelling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science*, *3*(2), 158. https://doi.org/10.1007/s42979-022-01043-x

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, *43*(3), 695–719. https://doi.org/10.25300/MISQ/2019/13747

Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control. *Ethics and Information Technology*, *24*(1), 4. https://doi.org/10.1007/s10676-022-09624-3

Saxena, N., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D., & Liu, Y. (2019). *How do fairness definitions fare? Examining public attitudes towards algorithmic definitions of fairness* (No. arXiv:1811.03654). arXiv. http://arxiv.org/abs/1811.03654

Schaap, G., Bosse, T., & Hendriks Vettehen, P. (2024). The ABC of algorithmic aversion: Not agent, but benefits and control determine the acceptance of automated decision-making. *AI & SOCIETY*, *39*(4), 1947–1960. https://doi.org/10.1007/s00146-023-01649-6

Schulenberg, K., Hauptman, A. I., Schlesener, E. A., Watkins, H., & Freeman, G. (2023). "I felt like I wasn't really meant to be there": Understanding women's perceptions of gender in approaching AI design & development. *Proceedings of the 56th Hawaii International Conference on System Sciences*, 175–184.

Scott, S. V., & Orlikowski, W. J. (2014). Entanglements in practice: Performing anonymity through social media. *MIS Quarterly*, *38*(3), 873–893. https://doi.org/10.25300/MISQ/2014/38.3.11

Segre, S. (2016). Social constructionism as a sociological approach. *Human Studies*, *39*, 93–99. https://doi.org/10.1007/s10746-016-9393-5

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. https://doi.org/10.1145/3287560.3287598

Shank, D. B., & DeSanti, A. (2018). Attributions of morality and mind to artificial intelligence after real-world moral violations. *Computers in Human Behavior*, *86*, 401–411. https://doi.org/10.1016/j.chb.2018.05.014

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, *74*, 102700. https://doi.org/10.1016/j.ijinfomgt.2023.102700

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, *9*(2). https://doi.org/10.1177/20539517221115189

Stowers, K., Kasdaglis, N., Rupp, M., Chen, J., Barber, D., & Barnes, M. (2017). Insights into human-agent teaming: Intelligent agent transparency and uncertainty. In P. Savage-Knepshield & J. Chen (Eds), *Adavances in human factors in robots and unmanned systems* (pp. 149–160). Springer International Publishing.

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

Sullivan, R., Veen, A. & Riemer, K. (2024). Furthering engaged algorithmic management research: Surfacing foundational positions through a hermeneutic literature analysis. I*nformation and Organization,* *34*(4), 100528. https://doi.org/10.1016/j.infoandorg.2024.100528

Sun, W., Nasraoui, O., & Shafto, P. (2020). Evolution and impact of bias in human and machine learning algorithm interaction. *PLOS ONE*, *15*(8), e0235502. https://doi.org/10.1371/journal.pone.0235502

Sutherland, S. C., Harteveld, C., & Young, M. E. (2016). Effects of the advisor and environment on requesting and complying with automated advice. *ACM Transactions on Interactive Intelligent Systems*, *6*(4), 1–36. https://doi.org/10.1145/2905370
Tarafdar, M., Page, X., & Marabelli, M. (2022). Algorithms as co-workers: Human algorithm role interaction in algorithmic work. *Information Systems Journal*, *33*(2), 232–267. https://doi.org/10.1111/isj.12389
Teodorescu, M., Morse, L., West Virginia University, Awwad, Y., Center for Complex Systems, King Abdulaziz City for Science & Technology, Massachusetts Institute of Technology, Kane, G., & Boston College. (2021). Failures of fairness in automation require a deeper understanding of human-ML augmentation. *MIS Quarterly*, *45*(3), 1483–1500. https://doi.org/10.25300/MISQ/2021/16535
Thomas, O., & Reimann, O. (2023). The bias blind spot among HR employees in hiring decisions. *German Journal of Human Resource Management: Zeitschrift Für Personalforschung*, *37*(1), 5–22. https://doi.org/10.1177/23970022221094523
Tinguely, P. N., Lee, J., & He, V. F. (2023). Designing human resource management systems in the age of AI. *Journal of Organization Design*, *12*(4), 263–269. https://doi.org/10.1007/s41469-023-00153-x
Triana, M. D. C., Gu, P., Chapa, O., Richard, O., & Colella, A. (2021). Sixty years of discrimination and diversity research in human resource management: A review with suggestions for future research directions. *Human Resource Management*, *60*(1), 145–204. https://doi.org/10.1002/hrm.22052
Tsoukas, H. (2018). *Philosophical organization theory* (1st edn). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198794547.001.0001
Turel, O., & Kalhan, S. (2023). Prejudiced against the machine? Implicit associations and the transience of algorithm aversion. *MIS Quarterly*, *47*(4), 1369–1394. https://doi.org/10.25300/MISQ/2022/17961
UK Equality Act (2010). https://www.legislation.gov.uk/ukpga/2010/15/contents
Upadhyay, A. K., & Khandelwal, K. (2018). Applying artificial intelligence: Implications for recruitment. *Strategic HR Review*, *17*(5), 255–258. https://doi.org/10.1108/shr-07-2018-0051
van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly*, *45*(3).
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
Vanhée, L., & Borit, M. (2022). Viewpoint: Ethical by designer - How to grow ethical designers of artificial intelligence. *Journal of Artificial Intelligence Research*, *73*, 619–631. https://doi.org/10.1613/jair.1.13135
Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3174014
Verma, S., Ernst, M., & Just, R. (2021). *Removing biased data to improve fairness and accuracy* (No. arXiv:2102.03054). arXiv. https://doi.org/10.48550/arXiv.2102.03054
Verma, S., & Rubin, J. (2018). Fairness definitions explained. *Proceedings of the International Workshop on Software Fairness*, 1–7. https://doi.org/10.1145/3194770.3194776
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. *The International Journal of Human Resource Management*, *33*(6), 1237–1266. https://doi.org/10.1080/09585192.2020.1871398

Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems*, *4*(2), 74–81. https://doi.org/10.1057/ejis.1995.9

Wang, H., Huang, J., & Zhang, Z. (2019). The impact of deep learning on organizational agility. *ICIS 2019 Proceedings*. ICIS, Munich, Germany.

Webster, J., & Watson, R. T. W. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), xiii–xxiii.

Wong, P.-H. (2020). Democratizing algorithmic fairness. *Philosophy & Technology*, *33*(2), 225–244. https://doi.org/10.1007/s13347-019-00355-w

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment and Society*, *33*(1), 56–75. https://doi.org/10.1177/0950017018785616

Wu, C.-F., Huang, S.-C., Chiou, C.-C., & Wang, Y.-M. (2021). A predictive intelligence system of credit scoring based on deep multiple kernel learning. *Applied Soft Computing*, *111*, 107668. https://doi.org/10.1016/j.asoc.2021.107668

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, *135*, 364–381. https://doi.org/10.1016/j.future.2022.05.014

Xivuri, K., & Twinomurinzi, H. (2023). How AI developers can assure algorithmic fairness. *Discover Artificial Intelligence*, *3*(1), 27. https://doi.org/10.1007/s44163-023-00074-4

Yeung, D. Y., Zhou, X., & Chong, S. (2021). Perceived age discrimination in the workplace: The mediating roles of job resources and demands. *Journal of Managerial Psychology*, *36*(6), 505–519. https://doi.org/10.1108/jmp-04-2020-0185

Zhao, H. (2020). Explicating the social constructionist perspective on crisis communication and crisis management research: A review of communication and business journals. *Journal of Public Relations Research*, *32*(3–4), 98–119. https://doi.org/10.1080/1062726x.2020.1802732

Zhou, Y., Wang, L., & Chen, W. (2023). The dark side of AI-enabled HRM on employees based on AI algorithmic features. *Journal of Organizational Change Management*, *36*(7), 1222–1241. https://doi.org/10.1108/jocm-10-2022-0308

Appendices

Appendix 2A: Interview Details

ID	Role	Country
Interviewee 1	Head of Talent Acquisition	Australia
Interviewee 2	HR Business Partner	Australia
Interviewee 3	HR Learning & Development Professional	Australia
Interviewee 4	Recruiter	Australia
Interviewee 5	Software Developer	Canada
Interviewee 6	Data & AI Specialist	Australia
Interviewee 7	People & Culture Business Partner	Australia
Interviewee 8	Data & AI Specialist	Australia
Interviewee 9	Recruiter	Australia
Interviewee 10	Recruiter	Germany
Interviewee 11	HR Business Partner	Australia
Interviewee 12	Recruiter	Germany
Interviewee 13	Recruiter	NZ
Interviewee 14	People Analyst	Australia
Interviewee 15	Recruiter	Australia
Interviewee 16	Recruiter	Australia
Interviewee 17	HR Technology Consultant	US
Interviewee 18	Recruiter	Australia
Interviewee 19	People Analyst	Germany
Interviewee 20	Recruiter	Australia
Interviewee 21	Recruiter	Germany
Interviewee 22	HR Consultant	Australia
Interviewee 23	Responsible AI Consultant	Germany
Interviewee 24	Lead People Scientist	Australia
Interviewee 25	AI Consultant	UK
Interviewee 26	People Analytics Consultant	Australia
Interviewee 27	Software Developer	Australia
Interviewee 28	People Analyst/Data Scientist	Germany
Interviewee 29	Recruiter	Germany
Interviewee 30	Recruiter	Australia

Appendix 2B: Draft Interview Questions

HR Managers

Introduction

- Tell me about your work?
- How long have you been working in your current position?

Understanding of algorithmic fairness

- What is fairness for you?
- What does fairness mean to you?
- How do you define algorithmic fairness?

Use of Algorithmic Decision-Making Systems

- Are algorithms used in your work?

If yes:

- Tell me about the processes where algorithms are used and how they are used?
- How much do you interact with the algorithm?
- How are you affected by the algorithm?
- Have there ever been discussions about algorithmic fairness in your organisations?
- Is algorithmic fairness considered in any training modules?
- Do you have any influence on the algorithmic decision-making? If so, how can you influence the decision-making process?
- Have you ever used an algorithm in a way that you were not supposed to use it?
- Has an employee ever expressed concerns about the use of algorithms? If so, how were these concerns resolved?
- Did your understanding of fairness change through the use of technology?

If no:

- Have you thought about using algorithmic decision-making systems?
- What concerns you about the use?
- Did you think your understanding of fairness would change through the use of technology?

Software Developers

Introduction

- Tell me about your work?
- How long have you been working in your current position?
- What kind of algorithms have you developed in the past that can be applied in the organisational employment context?

Understanding of algorithmic fairness

- What is fairness for you?
- What does fairness mean to you?
- How do you define algorithmic fairness?
- Did you think your understanding of fairness would change through the use of technology?

Enactment of algorithmic fairness

- Tell me about your role and your responsibilities in an engagement with an organisation?
- Do you collect sensitive data of the employees?
- Do you think this improves algorithmic fairness? If yes, why?
- Have there ever been discussions about algorithmic fairness with an organisation?

- Prior to an implementation or roll-out of an algorithm:
  - What are the steps that need to be taken to ensure that everyone, who uses the algorithm, is trained well enough?
  - Do meetings regarding the implementation of the algorithms contain discussions about algorithmic fairness?
- Have you ever had a client that used an algorithm in a way that they were not supposed to use it?
- What do you think are the concerns of HR professionals with regarding to the implementation of algorithmic decision-making systems and have you experienced those in practice?

Consultants

Introduction

- Tell me about your work? What do you do in your job on a daily basis?
- How long have you been working in your current position?

Understanding of algorithmic fairness

- What is fairness for you? What does it mean to you?
- How would you define algorithmic fairness?

Enactment of algorithmic fairness

- How are algorithms used in the employment context? What have you seen in practice? What is your experience in practice?
- Are or were there discussions about algorithmic fairness in your organisations or in organisations you worked with? How does it work?

- How does the process of selling the software work? Do organisations who buy it care about fairness? Is it discussed?
- What are the governance processes around it?
- Who is responsible?
- Do you the clients have any influence on the algorithmic decision-making? If so, how can they influence the decision-making process?
- What are the challenges in practice?
- What needs to be done to overcome these challenges?
- Why do you think everyone talks about it but nobody uses it?
- Did your understanding of fairness change through the use of technology?
- What do you think are the concerns of HR professionals with regarding to the implementation of algorithmic decision-making systems and have you experienced those in practice?