Developing unbiased Artificial Intelligence in Recruitment and Selection: A processual framework

A dissertation presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Management

at Massey University, Albany, Auckland,
New Zealand.

Melika Soleimani

2022

Abstract

For several generations, scientists have attempted to build enhanced intelligence into computer systems. Recently, progress in developing and implementing Artificial Intelligence (AI) has quickened. AI is now attracting the attention of business and government leaders as a potential way to optimise decisions and performance across all management levels from operational to strategic. One of the business areas where AI is being used widely is the Recruitment and Selection (R&S) process.

However, in spite of this tremendous growth in interest in AI, there is a serious lack of understanding of the potential impact of AI on human life, society and culture. One of the most significant issues is the danger of biases being built into the gathering and analysis of data and subsequent decision-making. Cognitive biases occur in algorithmic models by reflecting the implicit values of the humans involved in defining, coding, collecting, selecting or using data to train the algorithm. The biases can then be self-reinforcing using machine learning, causing AI to engage in ‘biased’ decisions. In order to use AI systems to guide managers in making effective decisions, unbiased AI is required.

This study adopted an exploratory and qualitative research design to explore potential biases in the R&S process and how cognitive biases can be mitigated in the development of AI-Recruitment Systems (AIRS). The classic grounded theory was used to guide the study design, data gathering and analysis. Thirty-nine HR managers and AI developers globally were interviewed.

The findings empirically represent the development process of AIRS, as well as technical and non-technical techniques in each stage of the process to mitigate cognitive biases. The study contributes to the theory of information system design by explaining the phase of retraining that correlates with continuous mutability in developing AI. AI is developed through retraining the machine learning models as part of the development process, which shows the

mutability of the system. The learning process over many training cycles improves the algorithms' accuracy.

This study also extends the knowledge sharing concepts by highlighting the importance of HR managers' and AI developers' cross-functional knowledge sharing to mitigate cognitive biases in developing AIRS. Knowledge sharing in developing AIRS can occur in understanding the essential criteria for each job position, preparing datasets for training ML models, testing ML models, and giving feedback, retraining, and improving ML models.

Finally, this study contributes to our understanding of the concept of AI transparency by identifying two known cognitive biases – similar-to-me bias and stereotype bias – in the R&S process that assist in assessing the ML model outcome. In addition, the AIRS process model provides a good understanding of data collection, data preparation and training and retraining the ML model and indicates the role of HR managers and AI developers to mitigate biases and their accountability for AIRS decisions.

The development process of unbiased AIRS offers significant implications for the human resource field as well as other fields/industries where AI is used today, such as the education system and insurance services, to mitigate cognitive biases in the development process of AI. In addition, this study provides information about the limitations of AI systems and educates human decision makers (i.e. HR managers) to avoid building biases into their systems in the first place.

Acknowledgements

First and foremost, I would like to thank my parents and sister. Without their support, tremendous understanding, and encouragement in my life, it would be impossible for me to study and even migrate to New Zealand and experience a new life. Thanks for believing in me.

I am deeply grateful to Prof David Pauleen for his invaluable advice, continuous support, and patience during my PhD study. I could not forget the moment you replied to my first email while you were on sabbatical on July 24th, 2017, and it got me one step closer to my goals. Throughout this journey, you have always been available and answered my emails patiently. I could not have imagined having a better supervisor and mentor for my PhD study.

I would like to express gratitude to Dr. Ali Intezari for his treasured support, which was really influential from the beginning of my PhD journey to the end. Without his support and patience, I could not have started my study. His immense knowledge and plentiful experience have encouraged me in my academic research and daily life. Thanks for always being available and caring when I needed you.

I am thankful to Dr Taskin Nazim for his guidance and feedback and the detailed questions throughout my study. My special thanks also go to Prof Jim Arrowsmith, who was a concerned and helpful ally although he came on board late.

I would also like to thank all members of the Management, Analytics, and Decision Making group, particularly Dr Kasuni Weerasinghe and Dr Hoa Nguyen, for always listening to me patiently and sharing their experiences with me.

I also appreciate the involvement of all the participants of this study, who gave their precious time to provide me with valuable insights and motivate me about the topic.

Last but most certainly not least, I would like to thank my PhD colleagues Nimeesha Odedra, Zagross Hadadian, and Patricia Loga for their support, a cherished time spent together while doing our PhDs, and in social settings. Furthermore, appreciation is given to my friends, Erin Burrell and Saba Samiei, for their help in the stage of data collection. Finally, a special thanks to my friend Shahab for his kind help and support that have made my study, work and life in New Zealand such a wonderful time.

List of Publications

- Developing unbiased AI to empower HR managers' decision-making in the recruitment and selection process. *13E2021 Doctoral Symposium*, Galway, Ireland, August 2021.
- The development process of unbiased Artificial Intelligence in Recruitment and Selection. *NZIS Doctoral Consortium*, Auckland, New Zealand, June 2021.
- Mitigating Cognitive Biases in Developing AI-Assisted Recruitment Systems: A Knowledge-Sharing Approach. *International Journal of Knowledge Management*, 18 (1), 1-18.
- Cognitive biases in developing biased Artificial Intelligence recruitment systems. *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, United States, 5-8 January 2021. IEEE Computer Society.
- Artificial intelligence and cognitive biases. *Australasian Business Ethics Network (ABEN)*, Auckland, New Zealand, 9-11 December 2018.
- Decision Making through Human Insight and Artificial Intelligence. *NZIS Doctoral Consortium*, Auckland, New Zealand, July 2018.

Table of Contents

List of Figures

List of Tables

Abbreviations

R&S = Recruitment and Selection

AI = Artificial Intelligence

AIRS = AI-Recruitment Systems

ML = Machine Learning

BI = Business Intelligence

DW = Data Warehouse

OLTP = Online Transaction Processing

OLAP = Online Analytical Processing

ERP = Enterprise Resource Planning

ETL = Extract, Transform, and Load

CRISP-DM = Cross-Industry Standard Process for Data Mining

TDSP = Team Data Science Process

DELTA = Data, Enterprise, Leadership, Targets, and Analysts

AIE = Applied Information Economics

MAD skills = Magnetic, Agile, and Deep skills

Chapter 1 Introduction

1.1 Research Background

The rapid development of communication and information technologies has placed businesses in new and different competitive situations (Akerkar, 2019). Organisations attempt to improve their expertise, which usually provides a competitive edge. According to Porter (1985), the success or failure of an organisation is rooted in the ability to make decisions.

The term ‘decision-making’ was introduced into the business world during the last century by Chester Barnard (Barnard, 1968). The concept of ‘decision-making’ led to managers thinking differently about their tasks and directing their actions toward being concise and conclusive and laid the foundation of managerial decision-making (Buchanan & O’Connell, 2006).

In many situations decision-making has been improved by managing risks, a nuanced understanding of human behaviour, and technological advances that support and mimic cognitive processes (Buchanan & O’Connell, 2006). In business today, decision-making requires managers to skilfully use a variety of resources including data, information, knowledge and wisdom (Intezari & Pauleen, 2018b) and advances in technology can help managers to access and benefit from these.

One of the business areas that have been influenced by the emergence of technology is the recruitment and selection (R&S) process. Various forms of technological developments such as online recruitment, gamification, and applicant tracking systems have made a substantial impact on the R&S process (Woods, Ahmed, Nikolaou, Costa, & Anderson, 2020). Recently, AI has gained increasing attention in R&S. According to Upadhyay and Khandelwal (2018), AI adoption was one of the most prevalent hiring trends among employers in 2018.

AI, as a form of computing, allows machines to perform functions by acting and reacting to data as inputs of algorithms (Russell & Norvig, 2010). Managers are using AI to collect and process data in almost every aspect of business from logistics to sales to human resources (Akerkar, 2019).

However, according to Pauleen, Rooney, and Intezari (2017) some problems in the social, cultural and political realms cannot be completely resolved by solely relying on data and information analyses. Additionally, Selbst, Boyd, Friedler, Venkatasubramanian, and Vertesi

(2019) believed that technical implementations have failed to “account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms” (p. 5). For example, Cooper and Abrams (2021) suggested that humans can overcome the algorithmic unfairness by devising a fair solution that comes from an understanding of societal roots of unfairness. Moreover, Teodorescu, Morse, Awwad, and Kane (2021) point to humans’ greater capability of identifying wild outliers, because they cannot be fooled as easily as AI by small changes.

Nilsson (2010) proposed that making effective decisions requires different capabilities, depending on the situation. True understanding of complex situations in a particular time can be obtained from deep insight (Intezari & Pauleen, 2013). Human insight plays a critical role in the managerial decision-making process, by incorporating important qualities such as understanding, experience, self-awareness, knowledge and problem-solving (Rosenfeld & Kraus, 2018).

According to Uzonwanne (2018), if a decision maker takes a step by step approach to decision-making using facts and information, a rational model of decision-making is followed. Nevertheless, sometimes decision makers employ heuristics that lead to suboptimal decisions (Tversky & Kahneman, 2013) due to cognitive bias (Bazerman & Moore, 2009).

Cognitive biases occur in algorithmic models too, reflecting the implicit values of the humans who are involved in defining, coding, collecting, selecting or using data to train the algorithm (IBM, 2018). Algorithms use past data about individuals to predict their future behaviour and make decisions based on the prediction (Marjanovic, Cecez-Kecmanovic, & Vidgen, 2021).

Various biases present in historical data will result in biases in algorithms (Marjanovic et al., 2021). Biases in learning from datasets can then be self-reinforcing using machine learning, leading AI to engage in unrestrained ‘biased’ decisions (Varshney, 2018). Machine learning is a facet of AI that relies on algorithms.

AI improves decision-making efficiency by utilising machine learning (ML) algorithms to extract patterns from very large sets of data (Rhem, 2020). Essentially, ML is the application of mathematical models to prepare training data and derive results from the datasets using statistics and mathematical models (Cormen, 2009).

AI may assist to quickly process large amounts of real-time data and calculate nonlinear influencing factors such as complex multivariate data (Kumar, 2017; PWC, 2016). However, biased AI may be harmful to organisations and/or individuals if it leads to economically suboptimal results, and any socially unacceptable scenarios that involve unfair discrimination (Vasconcelos, Cardonha, & Gonçalves, 2017). Biases can be harmful to organisations and/or individuals whether on grounds of efficiency (economically suboptimal results) and/or equity (‘unfair’) results. Evidence from the industry shows that biases can find their way into developing AI in the recruitment and selection process when training datasets are small and non-representative or “over- representative of certain groups” (Ntoutsi et al., 2020, p. 4).

For example, Amazon AI Recruiter learned from biased datasets and was biased against women as they had bias against hiring women applicants before (Kaplan & Haenlein, 2019). Moreover, biased algorithms may be created when developers are unable to formulate users' assumptions objectively or when inaccurate selection criteria are used when formulating assumptions (Tambe, Cappelli, & Yakubovich, 2019).

While AI can be an enabler for R&S (Wright & Atkinson, 2019), it is critical that HR managers and AI developers have a deep understanding of the potential biases and errors in AI in the R&S process that may lead to ineffective, poor and even harmful decisions. Hence this study aims to explain human cognitive biases that are highly likely to be embedded in AIRS during the development process.

The focus of the study is on the human cognitive biases that may affect recruitment and selection decision-makers when using AIRS. Human cognitive biases detected in algorithms might arise outside the training of the machine learning models. For example, cognitive biases are prevalent in collected data utilized for training machine learning models and the model parameters that are either obtained by training the model using real-world data or chosen by AI developers. In this research algorithmic bias in the recruitment and selection process is considered as the model biased outcomes that influence choosing certain groups of candidates.

1.2 Research Problem and Questions

AI is an area of computer science that emphasises the simulation of intelligent behaviour such as learning and reasoning in computers that can work and react like humans (Russell & Norvig, 2010). Attention to AI's application in decision-making is growing in both industry and academia and one of the business processes that are increasingly embracing AI as a decision support system is the selection phase in the Human Resources (HR) recruitment process. Even though AI is considered a decision aid for the recruitment and selection process, there are some undesirable consequences and implications of using AI.

Davenport and Michelman (2018) discussed some of these implications such as fairness and algorithmic bias, transparency and explainability, privacy and data security, trust and disclosure. Likewise, Benjamins, Barbado, and Sierra (2019) pointed to unfair biases of AI that lead to discrimination and the lack of explanations of the results of AI.

Technologists propose AI as a tool which has the ability to learn from data. This ability is the strength of AI: to extract information from data through machine learning algorithms (Akerkar, 2013) which can help managers make effective decisions. However, Krogh (2018) states that studying AI for problem-solving and decision-making in organisations is a new area of academic research which is needed to assist practitioners in approaching AI realistically. Moreover, according to Davenport and Michelman (2018) problems related to AI have not been discussed widely. They mention that businesses should try to anticipate the potential AI-related sources of social and economic harm, and as mentioned above, one source of problems is related to biases.

Scholars such as Simon (1960), Tversky and Kahneman (1974), Thaler (2000) and Bazerman and Moore (2013) have suggested that human decision-making is limited in many ways such as by bounded rationality, willpower, ethics and morality. Bounded rationality is a core concept of organisational behaviour proposed by Simon (1960) and emphasises the limitations of looking at all possible choices and their outcomes as human rationality is heavily influenced by the situation and human brainpower (Bazerman & Sezer, 2016).

Simon (1997) argued that the limitations on human rationality and power of the human brain exist due to "the disparity between the complexity of the world and the fitness of human computational capabilities, with or without computers" (p. 319). Bounded willpower reflects the fact that people sometimes make decisions that are not in their long-term interests

(Mullainathan & Thaler, 2000). A bounded ethicality situation may be characterised as one in which the central self, and the motivation that drives it, plays a significant role. Ethicality is bounded when judgements are influenced unconsciously by a particular vision of the self (Chugh, Bazerman, & Banaji, 2005).

Human decision-making boundaries and limitations lead to cognitive biases. Tversky and Kahneman (1973, 1974, 1983) postulated in a series of influential articles that bounded rationality causes individuals to use short-cut heuristics to make sound decisions under uncertain conditions. These heuristics were associated with a series of cognitive biases (Chugh et al., 2005). Bounded ethicality leads to biased decisions due to a stubborn view of being moral, competent, and deserving, and thus, unaffected by conflicts of interest. Since AI systems are designed and developed by humans, AI is prone to bias.

The risk of bias in the design of AI is due to encoding biases in datasets and algorithms. Human bias can arise in choosing a dataset that is not diverse enough in some factors such as age, race, colour, and region when developers formulate the hypothesis in algorithms (Vasconcelos, Cardonha, & Gonçalves, 2017). Therefore, it is critical that those involved in the design, implementation and use of AI are aware of and minimise the impact of the biases on the output of AI and the subsequent decisions.

Recent research studied the challenges of developing AI-assisted decision-making and point to cognitive biases as one of the key issues. Kaplan and Haenlein (2019) indicated cognitive biases that might manifest at the beginning of the development process. Shrestha, Ben-Menahem, and Krogh (2019) stated that biases can stem from the datasets with which the algorithm is trained. Similarly, Martin (2018) and Barocas and Selbst (2016) considered biased training datasets as the critical issue that leads to developing biased AI. Tambe, Cappelli, and Yakubovich (2019) explained another source of biases that might happen in designing algorithms due to not choosing the best set of variables or features.

Even though the above scholars have studied cognitive biases in AI, there is a lack of empirical studies on cognitive biases in developing AI and approaches to mitigate them. Since many believe that AI has the greatest potential to eliminate hiring bias (Polli, 2019), investigating potential biases in developing AI-Recruitment Systems and methods to mitigate them can assist in mitigating hiring bias. Additionally, Dwivedi et al. (2019) suggest that IS researchers should propose design criteria from a technology-human interaction perspective to build ideal AI systems for human decision makers.

Consequently, this exploratory research focuses on R&S due to the increasing use of AI applications in this decision-making process, identifying the potential biases and examining how the biases can be mitigated in the AI development process. Accordingly, this research addresses two main research questions:

RQ1- Which cognitive biases are more likely to be observed in recruitment and selection decisions?

RQ2- How can cognitive biases that emerged in this study be mitigated in developing AI-Recruitment Systems?

1.3 Research Objective and Significance

The objective of this study is to use empirical data to develop a theoretical model that explains approaches to mitigate cognitive biases in the development process of AI-Recruitment Systems (AIRS). This model could have a wide-ranging significance for both AI development and the R&S process, as well as decision-making in general.

The significance of this study is the contribution it makes to advancing concepts and practical impacts. First, this study contributes to the theory of information system design. The development process of unbiased AIRS shows retraining as part of the AI development process that refers to mutability in design theory. AI developers can conduct a test and retrain and change the ML models based on business use cases. Moreover, mutability in machine learning happens continuously to adjust the models to the required changes. The required changes are based on the feedback from the users of the AI-system (i.e., HR managers), new training data, or a change to the model definition.

Second, this study contributes to biases in developing AI and mitigation techniques by explaining cognitive biases and mitigation techniques in the development process of AIRS, such as having a good understanding of job position requirements to find relevant ML features, collecting enough and representative datasets, labelling and annotating datasets precisely, and monitoring AIRS constantly to detect errors and retrain the model.

Third, the AIRS development process model contributes to the knowledge sharing concept. It illustrates the importance of HR managers' and AI developers' knowledge sharing in all stages of the development process such as preparing datasets and labelling them and providing feedback to help AI developers to determine biases and improve their machine learning models.

Fourth, the process model of developing unbiased AIRS enhances the AI transparency literature by providing information on the data collection and preparation process for developing AIRS and the training and retraining process of machine learning models. This information extends the knowledge of the users of the systems (i.e., HR managers) and individuals who are affected by AIRS decisions such as job seekers. In addition, it emphasises that both AI developers and HR managers are responsible for the outcome of the AIRS decision-making.

As for the practical implications of this study, the development process of unbiased AIRS can have practical implications for the HR field as well as other fields where using AI is prevalent such as the education system and insurance services. The process model of unbiased AIRS can be used as a practical guideline for AI developers and field experts to understand the challenges of developing unbiased AI and how their collaboration leads to mitigating cognitive biases in developing AI. Moreover, the development process model of unbiased AIRS helps human decision makers in each field to understand the potential limitations of AI that lead to biased decisions.

1.4 Working Terms

This dissertation uses three terms – HR managers, AI developers, and AI-Recruitment Systems – frequently which are defined in detail for clarification and consistency:

Decision-making: In this study, the term decision-making is defined as a cognitive process of human beings that involves both intuition (System 1) and reasoning (System 2).

HR managers: The term 'HR managers' refers to people who are involved in the process of R&S decision-making, such as recruiters, HR staff and line managers.

AI developers: The term 'AI developers' refers to data scientists/data engineers, AI and Machine Learning engineers, solution architect/engineers and AI project managers who are involved in developing the Artificial Intelligence software and applications that can be used in recruitment and selection.

AI-Recruitment Systems: Throughout this thesis, the AI applications and systems that HR managers use for selecting candidates are employed for analysing CVs and references, algorithmic videos, and chatbot analysis. HR managers in New Zealand have been exposed to using AIRS for screening CVs and chatbots specifically for customer service positions in contact centres and retail stores where they receive large numbers of CVs for one position.

Cognitive bias: In this research cognitive bias refers to implicit cognitive biases in that these biases are activated involuntarily, without an individual's awareness or intentional control, and can be hard to detect. Also, cognitive bias and bias have been used interchangeably.

Algorithmic bias/biased AI: Algorithmic bias sometimes called biased AI, refers to the negative impacts of applying machine learning models while encoding the biases of their developers, datasets or the surrounding society, producing predictions or inferences that discriminate against individuals or groups of individuals in favour of others.

1.5 Structure of the Dissertation

This dissertation consists of six chapters. Chapter 1 presents an introduction to the study and Chapter 2 provides the review of the literature. Chapter 3 explains the methodology design. Chapter 4 presents the findings, and Chapter 5 discusses the findings based on the relevant literature. Finally, in Chapter 6, the conclusion, implications, and limitations of the study are discussed. The content of each chapter is explained below.

Chapter 1- Introduction

Chapter 1 presents an outline of the research problem, objectives, and research questions. In addition, the significance of this study is discussed. Then, the working terms being used in this study are provided.

Chapter 2- Literature Review

Chapter 2 provides the initial literature review before data collection (see 3.4.4.2 Literature review). The literature review consists of decision-making, Artificial Intelligence, and the recruitment and selection process literature. Decision-making is explained at organisational and individual levels, along with the cognitive perspective on decision-making (cognitive systems), and cognitive biases in decision-making. Artificial Intelligence literature is reviewed including the definitions of AI, the process models of AI, as well as cognitive biases in AI. Then the recruitment and selection decision-making literature is discussed.

Chapter 3- Methodology

Chapter 3 explains the research methodology and discusses the reasons for choosing grounded theory as well as the suitability of the classic grounded theory for this study. Grounded theory and its key components are discussed. Then, data collection and coding procedures are depicted.

Chapter 4- Findings

Chapter 4 discusses the coding process and the interpretation of the data. This chapter explains the codes and categories of the development process of unbiased AI-Recruitment Systems supported by comments from informants.

Chapter 5- Discussion

Chapter 5 discusses the findings by explaining the development process of unbiased AI-Recruitment Systems and the HR managers/AI developers' collaboration to mitigate cognitive biases in each phase of the development process of AIRS. The key components of the emergent AIRS process model and interrelationships are discussed following classic grounded theory methods. Literature relevant to the findings which was not included in the original literature review is included in this chapter.

Chapter 6- Conclusion, Implications, and Contribution

Chapter 6 presents an overview of the research project as well as the contribution and implications of the findings for the literature and practitioners. Following that, the limitations of the study are discussed, and some directions for further research are suggested.

1.6 Summary

This chapter presented the main research background, research problems, and objectives along with the research questions of the study. The significance of this study was also explained. As part of this chapter, the structure of this study was presented. In the next chapter, the theoretical foundations of this study will be discussed.

Chapter 2 Literature Review

This chapter presents the initial literature review on decision-making, Artificial Intelligence (AI) as a decision support system, and the R&S process as a specific decision-making context as well as cognitive biases in both R&S and the development process of AI. The literature review was done twice due to the adopted methodology (classic grounded theory). The initial literature review took place while completing the PhD proposal required for PhD admission and the second review was done during the data interpretation and after the theory began to emerge and when developing the discussion and implications and contributions chapters (Chapter 5 and Chapter 6).

This chapter first reviews the literature on decision-making in organisations. Next, the literature on cognitive perspectives on decision-making and the notion of cognitive biases are discussed. Then, the decision support methods and tools with a focus on AI, its components and the development process of AI as well as cognitive biases in software engineering and AI are presented. Following that, the R&S process, cognitive biases in this process and AI in R&S are explained.

2.1 Decision-making

To manage organisations in the contemporary business context, managers need to employ highly effective decision-making processes. This section reviews the literature on decision-making processes within the organisational context.

2.1.1 Decision-making in organisations

From an organisational perspective Mintzberg, Raisinghani and Theoret (1976) defined decision-making as a process of committing to a particular course of action. Good decisions are made through developing an appropriate decision-making process at all organisational levels. There are three levels of organisational decisions and processes: operational, tactical and strategic (Harrington & Ottenbacher, 2009).

Decisions that are related to daily operations in companies are operational (Hitt, Ireland, & Hoskisson, 1999). Hitt, Ireland, and Hoskisson (1999) defined tactical decisions as those that organisations decide to implement to achieve the organisational goals. Tactical decisions are well-structured and routine, requiring few resources (Kline, 1994). Strategic decisions are those decisions that will have significant and often long term impacts on the organisations and multiple stakeholders. Stakeholders, who can be involved in strategic decisions, are employees, managers, stockholders, the business community and society (Intezari & Pauleen, 2018a).

Decision-making is central to management activities (Intezari & Pauleen, 2018a). Effective management is the primary and critical factor of organisational capability and affects the performance of the whole organisation (Harrison & Pelletier, 1998). As decision-making is connected to management activities, the effectiveness of decisions has an impact on management and therefore organisational effectiveness. According to Eisenhardt and Zbaracki (1992), decision effectiveness is related to the extent that a decision leads to desired outcomes.

Decision-making is defined as a sequence of activities and the classic decision-making scholars have had different perspectives on the decision-making process (Simon, 1960; Mintzberg, Raisinghani, & Théorêt 1976; Drucker, 1967; Hogarth 1980). The different perspectives are related to the order of the functions in the decision-making process, and whether it is a linear or non-linear process (Intezari & Pauleen, 2018a).

Scholars such as Simon (1967) and Hogarth (1980) described how humans make decisions; their approach to decision-making is not a linear process. Therefore, they take the naturalist approach to decision-making (Messick & Bazerman, 1996). Naturalistic decision-making is related to understanding how people make decisions in the real world (G. Klein, 2015). Other scholars such as Drucker (1967) and Mintzberg et al. (1967) provided a structured and sequenced approach to decision-making. In their view, to achieve optimal outcomes, following the stages of the decision-making process and having a rational approach to decision-making are required (Betsch & Held, 2012).

Even though Mintzberg proposed his early model as prescriptive planning, in the late 1980s and early 1990s he was among a group of scholars such as Chaffee (1985), Burgelman (1991), Pettigrew (1992), and Van de Ven (1992) who introduced new perspectives based on politics, sociology and organisation theory (Booth, 1998). The new perspective was a departure from prescriptive planning and design schools and emphasises understanding the role of contextual factors, values, cultures, and politics within strategic thinking (Ezzamel & Willmott, 2004).

The rationalist approach to decision-making can fit normative and prescriptive models of decision-making (Intezari & Pauleen, 2018a). Normative and prescriptive models advise how decision makers should make decisions through following the decision-making processes (Dillion, 2014). According to Bell, Raiffa and Tversky (1988), normative models explain consistent decision procedures as practiced in situ. They argue that prescriptive models describe how to train decision makers to make good decisions. In contrast, the naturalist models are basically descriptive (Intezari & Pauleen, 2018a). Descriptive models focus on the thinking process and come from the organisational behaviour field (McFall, 2015). These models emphasise the decision makers' cognitive and behavioural processes in their natural environment (Intezari & Pauleen, 2018a).

The naturalistic theory is based on Simon's (1947) classic book on administrative behaviour. He emphasised the individual as a decision maker for organisations. Simon proposed the notions of bounded rationality and satisficing based on his observations of behaviour in organisations (Beach & Connolly, 2005). According to Simon (1960), decision makers have limited cognitive capacity and individual judgement is bounded in making rational decisions. Hence, decision-making can be better understood through explaining actual decisions rather than following prescriptive decision analysis (Bazerman & Moore, 2009).

Although a classification of rational or non-rational processes can lead to a better understanding of the core components of decision-making situations, it does not represent how management decisions are actually made, since managers confront a variety of complex decision situations that are too complicated to be easily resolved by either following a pre-defined set of procedures or relying entirely one's intuition (Intezari & Pauleen, 2019).

Decision-making is based on the two cognitive systems: System 1 and System 2. These two systems represent the intuitive and rational systems (Kahneman, 2003). As these two systems are associated with the cognitive decision-making process, in the following section the cognitive perspective on decision-making and the two systems of cognitive processing are discussed.

2.2 Cognitive Perspective on Decision-Making

Decision-making is one of the fundamental cognitive processes of the human brain (Wang, Liu, & Wang, 2003). According to Galotti (2002), decision-making is defined as a mental process resulting in choosing among alternatives. The mental processes involved are perceiving, sensing, analysing, problem-solving, weighing a decision, simulating, learning, expressing, and responding through thought, experience, and the senses (Kumar, 2017).

Öllinger, Jones and Knoblich (2008) argued that mental set and insight are two basic processes involved in problem-solving. They define a mental set as the tendency to solve a specific class of problems based on fixed solutions to similar problems. However, insight in problem-solving is seen in unconventional solutions that require a sudden, unconscious and unintended process. When a decision maker infers the solution without being able to give reasons, this phenomenon is called intuition. Intuition and insight are related to non-analytical and/or non-rational mental functioning (Zander, Öllinger, & Volz, 2016).

The key differences between rational and non-rational mental functioning are conceptualised by Kahneman (2003) through advocating a dual-system view of human thinking. The importance of this dual process that accounts for human reasoning and related higher cognitive processes such as judgement and decision-making has been recognised by cognitive scholars such as Evans (1984, 1989), Schneider and Shiffrin (1977), Shiffrin and Schneider (1984) and social psychology scholars such as Chaiken (1980) and Petty and Cacioppo (1986).

2.2.1 Two Cognitive Systems

Two cognitive systems are involved in human decision-making. As shown in Figure 2-1, Kahneman (2003) introduced these two main approaches – intuition (System 1) and reasoning (System 2) – to decision-making. The concept of intuition is connected to perception, includes emotions and experience, and occurs automatically and unconsciously (Dane, Rockmann, & Pratt, 2012; Sadler-Smith & Shefy, 2004).

According to Schneider and Shiffrin (1977), automatic processing is defined as the activation of a learned sequence of nodes in long-term memory that is activated automatically in response to appropriate inputs without subject control and attention. Automatic processes operate through associative connections and computation, and humans are only conscious of the results of computation and not the process (Lockton, 2012). Since System 1 is automatic and it is beyond conscious awareness and control, humans cut through the large amounts of information rapidly and effortlessly (Hodgkinson & Sadler-Smith, 2018).

In contrast, System 2 is based on logical reasoning to minimise the effect of emotions on the decision-making (Sadler-Smith & Shefy, 2004). It is the more rational side of the thinking process (Lockton, 2012). As it is more likely to be consciously monitored and deliberately controlled, it entails detailed analysis (Hodgkinson & Sadler-Smith, 2018) by considering all aspects of a problem and the interrelationships among elements in a problem (Intezari & Pauleen, 2018a). System 2 processing requires attention and is controlled by the subject (Schneider & Shiffrin, 1977). The System 2 processes are rule-based and depend heavily on working memory (Stanovich & West, 2000). The term 'working memory' refers to a brain function for short-term maintenance and information processing to perform complex cognitive tasks such as language comprehension, learning, and reasoning (Baddeley, 1992).

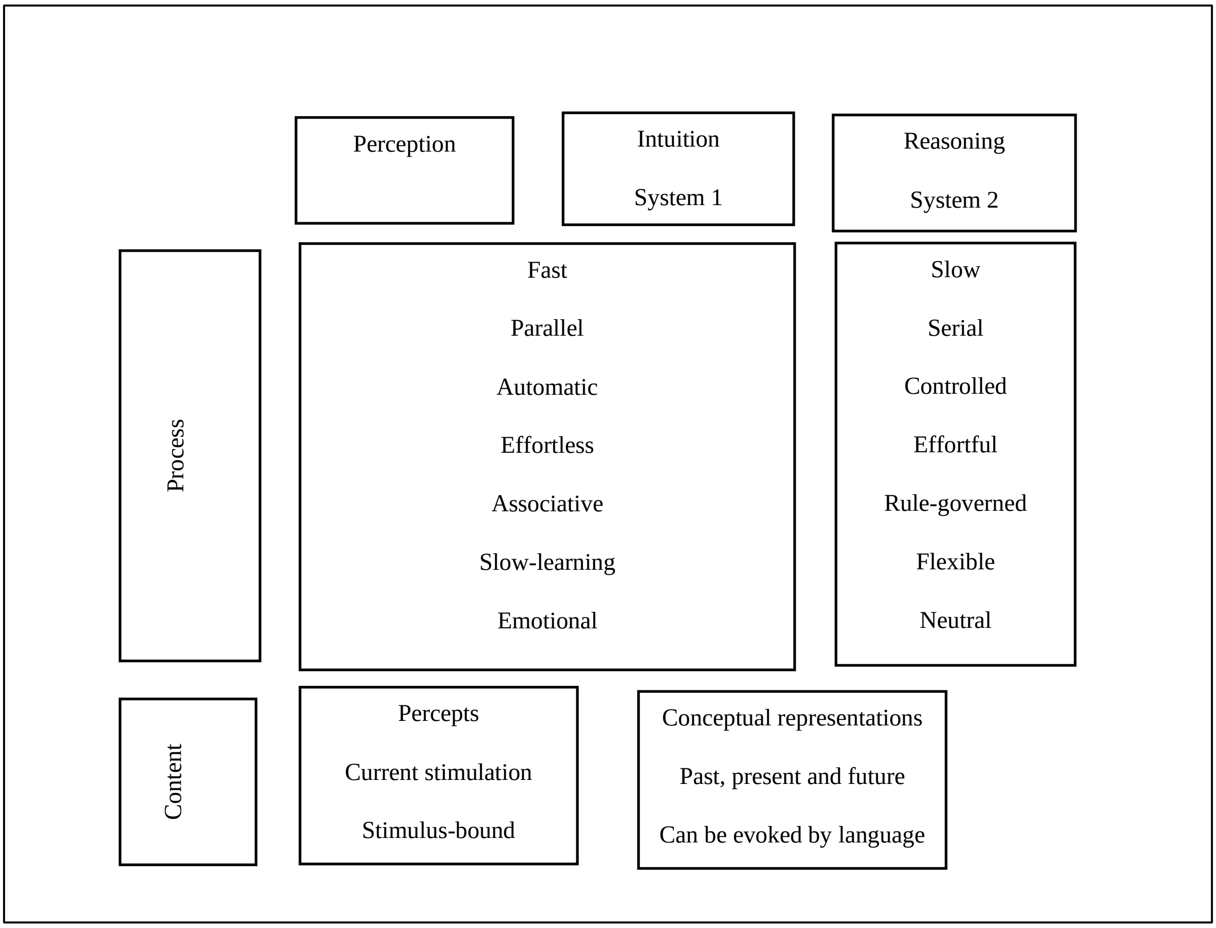

Figure 2-1 Process and Content in Two Cognitive Systems (Kahneman, 2003, p. 698)

Although some scholars propose that using rationality can lead to performing tasks efficiently (Sadler-Smith & Shefy, 2004), others (Evans, 1977; Shiffrin and Schneider, 1984; Chaiken, 1980; Petty and Cacioppo, 1986) indicate the importance of the ability to cut through details with less cognitive effort in some decision-making situations (Hodgkinson & Sadler-Smith, 2018), i.e. applying intuition to make decisions in that situation. Non-rationality and rationality are complementary to the extent that they fit the demands of particular decision-making situations (Sadler-Smith & Shefy, 2004).

Fitting the demands of the decision-making situation refers to the level of processing of the information, which is related to memory research (Craik & Lockhart, 1972). According to Hodgkinson and Sadler-Smith (2018), there are two different but complementary ways of information processing: automatic and controlled. Automatic processing enables individuals to reduce large amounts of information quickly and easily. Controlled processing represents a deeper level of processing and involves detailed analysis.

To conclude, based on the information processing constraints from cognitive limitations of humans (Simon, 1960), two cognitive capabilities – analytical and intuitive – are required to overcome information overload and decision-making under uncertainty (Hodgkinson & Sadler-Smith, 2018). However, intuitive and analytical skills can be distorted by cognitive biases (Bazerman & Moore, 2009; Kahneman, Lovallo, & Sibony, 2011) that leads to making suboptimal decisions (Tversky & Kahneman, 2013). Cognitive biases involved in decision-making are discussed in the following section.

2.2.2 Cognitive Biases in Decision-Making

Two key cognitive biases – teleology and essentialism – constrain biological reasoning and both can be thought of as biases associated with intuitive reasoning (System 1) modes within the dual process framework (Evans & Rosengren, 2018). These intuitive cognitive biases may hinder understanding of evolution (Evans, 2001) as well as other counterintuitive scientific concepts (Bloom & Weisberg, 2007).

Teleological reasoning is associated with both cognitive and social psychological factors (Scott, 2021). Psychological researchers defined teleological reasoning as a “developmentally persistent cognitive default” (Kelemen, Rottman, & Seston, 2013, p. 1057). Humans tend to explain mysterious phenomena by referring to their default functions and purposes (Kelemen et al., 2013). A belief that a hurricane is intended to punish a sinful populace can be an example of teleological bias (Scott, 2021).

An essentialist bias happens when individuals rapidly categorise things in the world. A quick classification could be crucial for survival: for example, in the case of distinguishing dangerous from non-dangerous animals (Evans & Rosengren, 2018). Even though essentialism can sometimes facilitate reasoning, it can affect decision-making. An essentialist view can be grounded in genetic or social deterministic theories (Nürnberger, Nerb, Schmitz, Keller, & Sütterlin, 2016).

The belief in genetic determinism refers to a person's behaviour that is determined by a biological basis (Keller, 2005). The belief in social determinism is the reflection of social factors such as upbringing, socialisation, and social background which shape individuals' fundamental characteristics permanently and profoundly (Rangel & Keller, 2011). For example, an essentialist belief that women are biologically more nurturing than men leads to gender-based division of labour (Brescoll, Uhlmann, & Newman, 2013).

Cognitive bias is also a systematic error in thinking and reasoning (System 2) (Leighton, 2010). Some scholars have worked on cognitive biases by identifying types of heuristics (Tversky & Kahneman, 1974; Bazerman & Moore, 2013; Das & Teng, 1999), outlining a list of biases in the stages of information processing (Hogarth, 1980) or gathering, as well as generating and evaluating data (Haley & Stumpf, 1989). Table 2-1 illustrates potential cognitive biases in decision-making, including definitions and effects on decision-making.

Table 2-1-Cognitive biases in decision-making and definitions

Cognitive Biases	**Definitions**	**Effects on decision-making**
• Availability (Tversky & Kahneman, 1974), (Schwenk, 1982), (Hodgkinson, 2001), (Bazerman & Moore, 2013), (Blumenthal-Barby, 2016) • Accessibility (Haley & Stumpf, 1989) • Fundamental attribution error (Nisbett & Ross, 1980)	Availability biases prevailed when participants focused heavily on the value-laden or emotional information even when others presented more objective information (Haley & Stumpf, 1989, p. 490).	"Judgements of the probability of easily recalled events are distorted" (Hodgkinson, 2001, p. 7).

• Representativeness (Tversky & Kahneman, 1974), (Schwenk, 1982), (Bazerman & Moore, 2013) • Salience of some information, vividness (Haley & Stumpf,1989) • Affect heuristic, representative heuristic (Blumenthal-Barby, 2016)	"Probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B . . . [and] not influenced by factors that should affect judgments . . . prior probability outcomes . . . sample size . . . chance . . . predictability . . . validity . . ." (Blumenthal-Barby, 2016, p. 6).	"Engaging decision makers in behaviour such as race discrimination" (Bazerman & Moore, 2013, p. 9).
• Confirmation (Bazerman & Moore, 2013), (Blumenthal-Barby, 2016)	"The tendency to perceive more support for [one's] beliefs than actually exists in the evidence at hand" (Blumenthal-Barby, 2016, p. 6)	"Interpreting evidence to support the conclusions of a given statement and hypothesis" (Bazerman & Moore, 2013, p. 10)
• Adjustment and anchoring (Tversky & Kahneman, 1974), (Haley & Stumpf, 1989), (Blumenthal-Barby, 2016) • Anchoring as a reason for confirmation bias (Bazerman & Moore, 2013)	Different starting points yield different estimates that are biased toward the initial values (Tversky & Kahneman, 1974).	"Decision makers do not search for additional information or ignore qualitative information presented to them by others" (Haley & Stumpf, 1989, p. 490).
• Perseverance (Haley & Stumpf, 1989), (Blumenthal-Barby, 2016)	Individuals enact perseverance heuristics in data gathering when they adhere to their prior beliefs and ignore subsequent disconfirming evidence (Blumenthal-Barby, 2016, p. 6).	Ignoring when "others introduced disconfirming or contradictory information" (Haley & Stumpf, 1989, p. 490).
• Positivity (Haley & Stumpf, 1989) • Wishful thinking (Hodgkinson, 2001)	Some managers' preferences for holistic information and their nonlinear, problem constructions would seem to encourage positivity biases (Blumenthal-Barby, 2016, p. 6).	"Probabilities of desired outcomes are judged to be inappropriately high" (Hodgkinson, 2001, p. 7).
• Social desirability (Nutt , 1986), (Haley & Stumpf, 1989), (Callegaro, 2011)	Some managers place importance on interpersonal relations and social approval. Social-desirability biases have less cognitive and more motivational origins (Haley & Stumpf, 1989).	"Reporting an answer in a way they deem to be more socially acceptable than would be their 'true' answer" (Callegaro, 2011, p. 2).
• Illusory correlation (Haley & Stumpf, 1989), (Hodgkinson, 2001)	This bias causes individuals to build up erroneous linkages around salient events. Uncertain situations may hinder some managers from forming these causal connections (Haley & Stumpf, 1989).	"Encouraging the belief that unrelated variables are correlated" (Hodgkinson, 2001, p. 7)
• Ambiguity aversion (Osmont, Cassotti,	"The display of preferences for known or certain probabilities over uncertain	"Avoiding ambiguous options and human choices depends on the

Agogué, Houdé, & Moutier, 2015), (Blumenthal-Barby, 2016)	probabilities regardless of actual benefits" (Blumenthal- Barby, 2016, p. 6).	option's presentation" (Osmont, Cassotti, Agogué, Houdé, & Moutier, 2015, p. 572).
• Bandwagon effect (Blumenthal- Barby, 2016)	"An accelerating diffusion through a group or population of a pattern of behaviour, the probability of any individual adopting it increasing with the proportion who have already done so" (Blumenthal- Barby, 2016, p. 6).	"Adoption without assessment" (Secchi & Bardone, 2009, p.5); "Disability to exercise the decision maker mindfulness, i.e. the level of attention and awareness typical of active information processing" (Gardner & Avolio, 2003, p. 58).
• Commission bias (Blumenthal-Barby, 2016)	"Tendency toward action rather than inaction" (Blumenthal-Barby, 2016, p. 6)	"These seem to relate to a functional reason for associating action with responsibility in that inaction produces less material evidence of wrong doing" (Feldman, Kutscher, & Yay, 2018, p. 7).
• Decoy effect (Felfernig 2014), (Blumenthal-Barby, 2016)	"The addition of such [asymmetrically dominated] alternatives increases the share of the item that dominates it" (Blumenthal-Barby, 2016, p. 6).	"Increasing the selection share of specific target items" (Felfernig, 2014, p. 2)
• Default bias or status quo/Sunk-cost bias (Arkes & Blumer, 1985), (Blumenthal-Barby, 2016)	"Individuals have a strong tendency to remain at the status quo, because the disadvantages of leaving it loom larger than advantages" (Blumenthal-Barby, 2016).	"To continue an endeavour once an investment in money, effort, or time has been made" (Arkes & Blumer, 1985, p. 124).
• Frequency/percentage framing effect (Kahneman & Tversky , 1979), (Blumenthal-Barby, 2016)	"Frequency scales generally . . . lead to higher perceived risk" (Blumenthal-Barby, 2016, p. 6).	"Implementing different behaviours when facing to the multiple choice that expressing the same meaning" (Li & Ling, 2015, p. 96).
• Impact bias (Blumenthal-Barby, 2016), (Hodgkinson, 2001) • Hindsight bias and the curse of knowledge as a reason of confirmation bias (Bazerman & Moore, 2013)	"Failure to anticipate our remarkable ability to adapt to new states. People tend to overestimate the long-term impact of both positive events . . . and negative events" (Blumenthal-Barby, 2016, p. 6).	"Quickly dismiss the possibility that things could have turned out differently than they did" (Bazerman & Moore, 2013, p. 10).
• Loss/gain framing bias or loss aversion bias (Blumenthal-Barby, 2016), (Bouteska & Regaieg, 2018)	"Losses loom larger than corresponding gains" (Blumenthal-Barby, 2016, p. 6).	"Do not value the gain and loss in the same way" (Bouteska & Regaieg, 2018, p. 4).

• Omission bias (Blumenthal-Barby, 2016), (Feldman, Kutscher, & Yay, 2018)	"Judge harmful commissions as worse than the corresponding omissions" (Blumenthal-Barby, 2016, p. 6)	"Possibility of negative outcomes and/or harm; Outcome not yet known" (Feldman, Kutscher, & Yay, 2018, p. 26)
• Optimism bias or optimistic/overconfidence (Sharot, 2011), (Bazerman & Moore, 2013), (Blumenthal-Barby, 2016), (Bouteska & Regaieg, 2018) • Illusion of control (Hodgkinson, 2001) • Illusion of manageability (Das & Teng, 1999)	Overestimation of personal control over outcomes	"Mispredict future occurrence" (Sharot, 2011, p. 945)
• Order effects or primacy/recency bias (Blumenthal-Barby, 2016)	"Information presented at the beginning or end of a series is remembered and chosen more often than information presented in the middle of the series" (Blumenthal-Barby, 2016, p. 6).	"The recall of items in the middle of the list is generally poor" (Scanlan, 2011, p. 2).
• Outcome bias (Blumenthal-Barby, 2016), (Gino, 2016)	"Allowing a prior event or decision outcome to influence subsequent independent decisions" (Blumenthal-Barby, 2016, p. 6)	"More likely to neglect intentions and overweight outcomes" (Gino, 2016).
• Relative risk bias (Blumenthal-Barby, 2016), (Siegerink & Rohmann, 2018)	"A stronger inclination to [act] when presented with the relative . . . risk than when presented with the same [information] described in terms of the absolute . . . risk" (Blumenthal-Barby, 2016, p. 6)	"Measurement error and misclassification" (Spiegelman & Valanis, 1998, p.406); "Relative risk estimates do not always convey the necessary context for a meaningful interpretation of the data" (Siegerink & Rohmann, 2018, p. 653).
• Selective perception (Hodgkinson, 2001) • Prior hypotheses and focusing on limited targets (Das & Teng, 1999)	"Expectations may bias observations of variables relevant to strategy" (Hodgkinson, 2001, p. 7).	Decision makers "may have prior perceptions about the relationships of salient variables, so that they might overlook information and evidence that may prove the opposite" (Das & Teng, 1999, p. 762).
• Conservatism (Hodgkinson, 2001), (Pompian, 2011)	"People cling to their prior views or forecasts at the expense of acknowledging new information" (Pompian, 2011, p. 63).	"Failure to revise sufficiently forecasts based on new information " (Hodgkinson, 2001, p. 7)

• Law of small numbers (Hodgkinson, 2001) • Exposure to limited alternatives (Das & Teng, 1999)	Overestimation of the degree to which small samples are representative of populations (Hodgkinson, 2001, p. 7)	"Decision makers are found to adopt sequential attention to alternatives and to use intuition to supplement rational analysis" (Das & Teng, 1999, p. 762).
• Regression bias (Hodgkinson, 2001), (Bazerman & Moore, 2013)	"Individuals tend to ignore the fact that extreme events tend to regress to the mean on subsequent trials" (Bazerman & Moore 2013, p. 58).	"Failure to allow for regression to the mean" (Hodgkinson, 2001, p. 7)
• Logical reconstruction (Hodgkinson, 2001), (Hertwig, Fanselow, & Hoffrage, 2003)	"'Logical' reconstruction of events which cannot be accurately recalled" (Hodgkinson, 2001, p. 7)	"Evaluate the appropriateness of ex ante behaviour that resulted in bad or good ex post outcomes" (Hertwig, Fanselow, & Hoffrage, 2003, p. 357)
• Insensitivity to outcome probabilities (Das & Teng, 1999)	Decision makers "tend to be influenced more by the value of possible outcomes than by the magnitude of the probabilities" (Das & Teng, 1999, p. 762).	"Managers are more likely to use a few key values to describe a situation and they see problems as unique" (Das & Teng, 1999, p. 762).
• Functional-fixedness (Nutt, 1986), (Haley & Stumpf, 1989)	Anchoring may lead to functional-fixedness output biases. Functional fixedness biases can reflect excessive reliance on certain problem-solving methods. Some managers identify unusual aspects of non-preferential alternatives (Nutt. 1986).	"Functional-fixedness biases arise when standard routes prove dysfunctional for final problems" (Haley & Stumpf, 1989, p. 490).

Kahneman, Lovallo and Sibony (2011) argued that it is obvious to executives that cognitive biases can influence reasoning in business decisions. Consequently, cognitive biases affect the quality of business decisions at either the individual or the organisational level. By providing statistical and probability functions and offering multiple choices, decision support tools and methods can help reduce individual biases (Bhatt & Zaveri, 2002). In the following section decision support methods and tools are explained in detail.

2.3 Decision support methods and tools

There are a variety of tools and techniques that support managers to gain insight into the decision problems (Ragsdale, 2007). To support the decision-making process, relevant information should be integrated from various sources (Tvrdíková, 2007). Information is the raw input for decision-making, and the application and use of information represents its value (Tripathi, 2011). Managers often engage in addressing multidimensional, multi-layered, and emergent problems (Intezari & Pauleen, 2018a). To solve a problem, potential solutions should be found through framing the problem and forming assumptions about it. Then, the solutions are evaluated and the best one is chosen. The more complex the search or evaluation, the more "computational decision support tools" are required (Bonabeau, 2003, p. 121). Based on the degree of complexity of the search and evaluation of the solutions, four categories of problems are shown in Table 2-2.

For problems where searching and evaluating the options are complex, decision makers may rely on their intuitive decisions. For example, a fire-fighter making decisions in a burning building or a soldier on a battlefield (Bonabeau, 2003). However, Bonabeau (2003) suggests that in the business decision-making context, with more options to consider, more data to weigh, and more unusual challenges to face, it is important to rely more on analysis as intuition is not a method of assessing complexity, but of ignoring it.

In some situations, a decision maker has a few solutions; however, each solution has many consequesnce and there is no clear way to measure the merits of the solutions. To evaluate those solutions, some decision tools such as spreadsheet modelling are applied (Ragsdale, 2007; Bonabeau, 2003). For example, a pharmaceutical manufacturer needs to determine whether a new distribution centre can be built or warehouse operations can be outsourced. The spreadsheet modelling assists the manufacturer in analysing the cost and service implications and how this will be done (Smith, 2003).

In some cases, searching for options is complex but evaluating them is not complicated in that some decision techniques assist decision makers in evaluating large quantities of options with simple and understandable models. Decision techniques such as classification, association, and clustering in data mining are helpful to reduce the complexity of the search space (Chen & Huang, 2013). For example, data mining techniques can analyse various options from customers' data to predict the credit score and rank it to avoid future risks of giving loans to people who cannot repay them (Sudhamathy, 2016).

In more complex situations where many potential solutions exist and evaluating the consequences of options is complex, sophisticated tools (Bonabeau, 2003) such as Decision Support Systems (DSS), Business Intelligence (BI), and AI are required.

Table 2-2 Problem categories (Bonabeau, 2003, p. 121)

Evaluation		Search: Simple	Search: Complex
	Complex	Few options, Complex consequences	Many options, Complex consequences
	Simple	Few options Simple consequences	Many options Simple consequences

To make effective decisions, managers must be able to overcome the complexity both in searching for many solutions and evaluating the solutions that are dependent on each other. HR decisions such as choosing a suitable candidate involve choosing in the face of uncertainty (Strohmeier, 2020). Myriad sophisticated tools have been used to do so, such as DSS (Bonabeau, 2003). These systems appeared toward the end of the 1960s, to augment the decision-making process in organisations (Vizecky & El-Gayar, 2011). Since the early 1970s, the data needed to support decision-making and DSSs have existed as the first generation of decision support data management to analyse data for making informative decisions (Watson & Marjanovic, 2012). A DSS can be defined as computer software that facilitates and accepts inputs of a large number of facts and methods to generate meaningful comparisons, graphs, and trends that can improve the decision-making capabilities of decision makers (Bhatt & Zaveri, 2002).

Over time, new generations of DSSs have emerged. During the early 1990s, Enterprise Resource Planning (ERP) systems made data available across organisations, enabling Business Intelligence (BI) and supporting managers' decision-making (Bumblauskas, Herb, Bumblauskas, & Igou; 2017). Then, as part of BI, business analytics became the key analytical component (Davenport, 2006). Data management and data warehousing are considered as the core elements of BI and analytics. A data warehouse (DW) is a repository that stores data for analytical processing and decision-making (Santos & Bernardino, 2008). A DW integrates historical data and the data gathered from different operational sources (OLTP: Online Transaction Processing) into a central repository that can be accessed by analytical applications (OLAP: Online Analytical Processing) with varying requirements. This process of extracting data from different sources, transforming it and loading it into a DW is called the ETL (Extract, Transform, Load) process (Chen et al., 2012).

With the advent of big data in 2010, big data analytics have been used to describe the datasets that are large (from terabytes to exabytes), vary in data types (structured to unstructured), and datasets that are generated and analysed quickly. Thus, traditional analytics tools are often inadequate and advanced technology such as AI is required to collect and process big data (Chen et al., 2012; Intezari & Gressel, 2017).

2.4 Artificial Intelligence

According to Akerkar (2019), the computational power of AI and the amount of data which AI can collect and analyse is the most significant difference compared to other decision tools and techniques. He argues that for businesses to have a competitive advantage, AI is necessary to promote automation, cost reduction and intelligent decision-making. As mentioned above, AI system developers are susceptible to cognitive biases. The effects of cognitive biases on AI-assisted decision-making are the focus of this research. In the following paragraphs the definitions of AI, components of AI, and specified cognitive biases in software engineering are explained.

2.4.1 Artificial Intelligence Definitions

The term AI was coined by John McCarthy in 1956. However, work had been started during the Second World War by Alan Turing, an English mathematician and computer scientist (Russell & Norvig, 2010). The collaboration of scholars such as Herbert Simon, Allen Newell, Claude Shannon, Nathaniel Rochester and other researchers from Carnegie Tech, together with research by Marvin Minsky at the Massachusetts Institute of Technology and John McCarthy of Stanford, developed early computer models of human cognition (Nilsson, 2010; Bonabeau, 2003). AI has been defined by scholars and some of the definitions that were explored during the literature review are provided in Table 2-3.

Table 2-3- Artificial Intelligence Definitions

Scholars	Definitions of AI
McCarthy (1956)	"The science and engineering of making intelligent machines" (McCarthy, 2007, p. 2)
Bellman (1978)	"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning " (as cited in Russell & Norvig, 2010, p. 2)
Haugeland (1985)	"The exciting new effort to make computers think ... machines with minds, in the full and literal sense" (as cited in Russell & Norvig, 2010, p. 2)
Charniak and McDermott (1985)	"The study of mental faculties through the use of computational models" (as cited in Russell & Norvig, 2010, p. 2)
Schank (1987)	"The question of what AI is all about probably doesn't have just one answer. What AI is depends heavily on the goals of the researchers involved, and any definition of AI is dependent upon the methods that are being employed in building AI models " (p. 59)
Kurzweil (1990)	"The art of creating machines that perform functions that require intelligence when performed by people" (Russell & Norvig, 2010, p. 2)
Rich & Knight, (1991)	"The study of how to make computers do things at which, at the moment, people are better" (p. 3)
Winston (1992)	"The study of the computations that make it possible to perceive, reason, and act" (p. 5)
Ginsberg (1993)	"AI is the enterprise of constructing an intelligent artefact, tasks demonstrating intelligence are those that people do well" (p. 3) and "artefact is an inorganic object, taking an artefact to be what Allen Newell and Herbert Simon have called a physical-symbol system" (p. 8)
Poole, Mackworth, and Goebel , (1998)	"Computational Intelligence is the study of the design of intelligent agents" (p. 1)
Nilsson and Nilsson (1998)	"AI is concerned with intelligent behaviour in artefacts" (p. 1)
Nilsson (2010)	"Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment." (p. 13)
Boden (2016)	"Artificial intelligence (AI) seeks to make computers do the sorts of things that minds can do" (p. 1)

Newell and Simon (1972) argued that AI is the science of knowledge representation and reasoning. These two parts, which belong to the cognitive side, are fundamental parts of AI (Russell & Norvig, 2010). According to Kumar (2017) AI is associated with cognitive computing to "mimic the way [the] human mind works" (p. 30). He argues that AI can think using learning already acquired, solve the problem better through extending its knowledge, identify new problem areas and find solutions. Problem finding and problem-solving are the two phases of the decision-making process (Intezari & Pauleen, 2018a) so there can be an intersection between AI and decision-making. As decision-making is related to reasoning, the possible definitions of AI that point to cognitive processes and reasoning can be taken into consideration (Pomerol, 1996).

Russell and Norvig (2010) divided eight definitions of AI into four categories based on thought process and reasoning (the definitions at the top of Table 2-4), behaviour (the definitions at the bottom of Table 2-4), human performance (the definitions on the left side of Table 2-4) and ideal performance (the definitions on the right side of Table 2-4). The four categories are: thinking humanly, thinking rationally, acting humanly and acting rationally.

Thinking humanly is related to the cognitive modelling approach of AI: activities that incorporate human thinking such as decision-making and problem-solving. Thinking rationally, called the laws of thought approach, is the study of using computational models for perceiving, reasoning and acting. Acting humanly means performing functions performed by humans, which require intelligence and which humans are better than computer programs at performing. A machine that successfully passes the Turing Test (being indistinguishable from a human) should be considered an intelligent machine. Acting rationally is associated with the study of intelligent behaviour in agents. The definition of acting rationally represents the rational agent approach. The first two definitions – thinking humanly (cognitive modelling) and thinking rationally (laws of thought) – explain the cognitive processes and reasoning.

Table 2-4- Some definitions of Artificial Intelligence, organised into four categories (Russell & Norvig, 2010, p. 2)

Thinking Humanly “The exciting new effort to make computers think ... machines with minds, in the full and literal sense.” (Haugeland, 1985) “[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ...” (Bellman, 1978)	**Thinking Rationally** “The study of mental faculties through the use of computational models.” (Charniak and McDermott, 1985) “The study of the computations that make it possible to perceive, reason, and act.” (Winston, 1992)
Acting Humanly “The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990) “The study of how to make computers do things at which, at the moment, people are better.” (Rich and Knight, 1991)	**Acting Rationally** “Computational Intelligence is the study of the design of intelligent agents.” (Poole et al., 1998) “AI . . . is concerned with intelligent behavior in artifacts.” (Nilsson, 1998)

2.4.1.1 Thinking humanly: Cognitive modelling

Haugeland (1985) proposed that human thinking and machine computing are the same. The computational framework is the basis of AI and "the right kind of computational structure suffices for the possession of a mind, and for the possession of a wide variety of mental properties" (Chalmers, 2011, p. 326). Haugeland (1981) mentioned that the computational conception might need supplementation to produce a complete understanding of thinking, such as consciousness, intentionality, and the capacity to care.

Furthermore, Bellman (1978) discussed whether computers can think. He defined three words: computer, can and think. In his definition, computer means "a commercially available digital computer" (p. 12). The meaning of "think" is relevant to the performance of activities representing human thinking such as decision-making, problem-solving, learning, creating, and game playing. He defined "can" as the ability to "use standard programming methods that do not require a high level mathematical expertise" (p. 13) and do these tasks in an appropriate time frame.

Based on these two definitions of AI, the combination of computerised models and experimental techniques from psychology should be considered to model human problem-solving. The definition of AI as cognitive modelling determines how humans think and expresses the thinking process in a computerised model (Wilson & Keil, 1999; Russell & Norvig, 2010).

2.4.1.2 Thinking rationally: Laws of thought

Charniak and McDermott (1985) discussed the logic-based representation scheme. Logic is defined as the statements about all kinds of items in the world and the relations among them (Russell & Norvig, 2010; Davis, Shrobe, & Szolovits, 1993). Winston (1992) argued that it is possible for computers to reason and perceive by representing and using knowledge. He explains that using knowledge requires application of reasoning methods. However, Russell and Norvig (2010) saw two obstacles to developing computational reasoning systems. First, codifying informal knowledge through using logical notation is not easy. Second, solving a problem in practice is not the same as solving the problem in principle. Based on the two critiques, developing computational reasoning systems is complicated.

2.4.2 Components of Artificial Intelligence

According to Kumar (2017), AI has its roots in human cognitive studies, and the concept of learning from experience – "thinking" – was the missing notion in the previous decision tools such as BI. As mentioned above, Bellman (1978) defined the term "think" as doing activities that are related to the process of human thinking such as problem-solving. The core part of problem-solving in AI is knowledge representation (Russell & Norvig, 2010). Knowledge representation is intertwined with thinking by reasoning (Davis et al., 1993). Therefore, researchers in AI have focused on the components of intelligence such as knowledge representation, reasoning and learning. Figure 2-2 illustrates these components of AI that are involved in problem-solving.

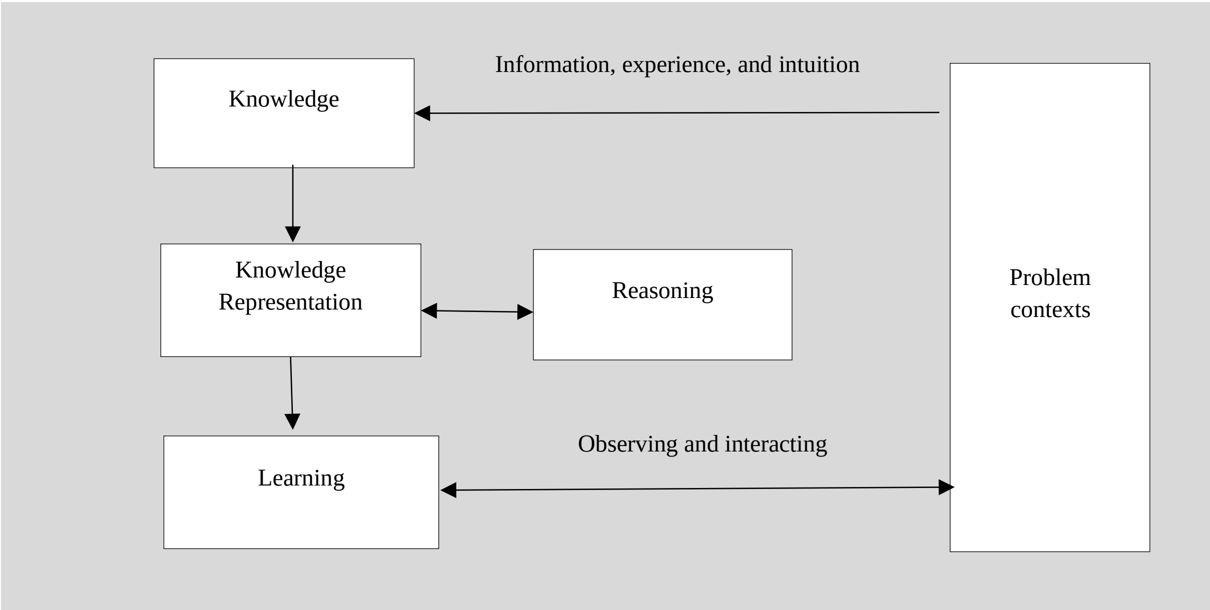

Figure 2-2 Components of Artificial Intelligence involved in problem-solving

2.4.2.1 Knowledge Representation

Knowledge representation is a way of human expression that is considered a language for expressing things about the world. Knowledge representation is one of the most familiar concepts in AI and is a "medium for efficient computation" (Davis et al., 1993, p. 26). There are several knowledge representation models such as logical representation, ontology, semantic networks, scripts and frames. Hence, knowledge representation is the way of encoding knowledge in AI (Russell & Norvig, 2010; Luger, 2004).

Algorithms for hiring offer recruiters new insights into candidates' profiles (Tambe, Cappelli, & Yakubovich, 2019) as a new approach to the production of knowledge that is called machine knowledge production. The purpose of machine knowledge production is to create knowledge from data by applying ML techniques (Bonde Thylstrup, Flyverbom, & Helles, 2019).

2.4.2.2 Knowledge

According to Roberts and Armitage (2008), knowledge includes information with awareness and understanding. They indicate that although knowledge is more than information, knowledge creation is itself dependent on information. Intezari and Pauleen (2018a) argued that there are other sources of knowledge such as a priori knowledge, experience, and intuition. Knowledge can be classified in several ways such as procedural or declarative, and tacit or explicit.

Polanyi (1958) divided knowledge into tacit and explicit types. Tacit knowledge is about what we know and are able to do, but it is difficult to formalise and describe analytically. Explicit knowledge is knowledge that is transmittable in writing, speech, or drawings (Nonaka, 1994; Intezari & Pauleen, 2018a). Explicit knowledge can be encoded (Akerkar, 2019). Encoding knowledge requires formats that the computer can both read and understand (Lele, 2018).

Procedural and declarative knowledge have been defined by cognitive scientists and AI researchers (Nilsson, 2010). Procedural knowledge explains how things work (Robillard, 1999). For example, performing a skilled action such as hitting a tennis ball involves procedural knowledge (Nilsson, 2010). Declarative knowledge describes what things are (Robillard, 1999). Nilsson (2010) states that “any knowledge represented by a declarative sentence is called declarative” (p. 236).

Thus, when a person answers a question such as “how old are you?”, the answer, “I am twenty-four years old”, is a declarative sentence. He mentions that in order to use procedural knowledge in AI programs, it is represented directly in the programs, while to codify declarative knowledge, symbolic structures are required. Representing knowledge cannot be useful unless the AI system has the ability to reason (Shapiro, 1992).

2.4.2.3 Reasoning

Reasoning is the process between two points of attending to relevant and meaningful information and deriving one or two conclusions (Khemlani, 2018). Reasoning, as a part of knowledge representation, goes on internally rather than encountering existing things in the real world (Davis et al., 1993). However, describing the complexity of the natural world is difficult and inevitably at least some limitless complexity of the world will be omitted. Therefore, some conclusions and inferences will be incorrect or imperfect and will affect representations; if representations are imperfect, they are a source of error (Davis et al., 1993).

In order to approach human cognition, AI systems extract critical information from large sets of structured and unstructured data by reasoning. The majority of machine learning-based algorithms excel at pattern recognition based on associations rather than causation (Tambe et al., 2019). Pearl (2018) proposed causal reasoning as a missing mathematical mechanism for gaining computational insight into meaning grounding. Replacing reasoning by association with causal reasoning leads to inferring causes from observed behaviours (Bishop, 2021).

Bishop (2021) gave examples of an AI in a scoring process discriminating on gender or a chatbot showing racist behaviour due to reasoning by association. Tambe, Cappelli, and Yakubovich (2019) further explained that causal reasoning helps HR decision makers to focus their attention on the relevant features and behaviours, reduces data management costs, and meets both the AI algorithm requirements of fairness and explainability.

In addition to the ability of reasoning to solve a problem, learning from past experience is a way of problem-solving. In order to mimic human beings, AI systems should have learning capabilities (Kumar, 2017).

2.4.2.4 Learning

Learning is the process of gaining knowledge through identifying, remembering and comprehending things (Banerjee et al., 2015). The capability of AI to learn separates it from other intelligent automation. It is able to learn and address the problems on its own by observing and interacting with the world (Russell & Norvig, 2010). Learning methods depend on domain-specific abstraction, similarities between different states and actions which are expected to be developed by a human expert (Rosenfeld & Kraus, 2018).

2.4.3 Machine Learning

Machine Learning (ML) is defined as a computer recognising patterns without explicit programming (Clark, 2018). To intelligently analyse data and develop the corresponding real-world applications, ML algorithms are the key. ML algorithms can be classified into four major categories: supervised, unsupervised, semi-supervised, and reinforcement learning (Mohammed, Khan, & Bashie, 2016).

Supervised learning involves determining certain goals that are to be reached from a certain set of inputs and a learning function that maps an input to an output based on sample input-output pairs (Sarker, Kayes, Badsha, Alqahtani, Watters, & Ng 2020). To infer the function, a collection of labelled training datasets are used (Han & Kamber, 2011). The common supervised tasks are classification and regression. Classification helps separate the data and regression assists in fitting the data. An example of supervised learning is the prediction of text class labels or sentiments from text, such as a tweet or a product review (Sarker, 2021).

In unsupervised learning, a dataset is analysed without human intervention (Han & Kamber, 2011). This method is commonly used to extract generative features, identify meaningful trends and structures, group results, and explore patterns in data. Typically, unsupervised learning tasks include clustering, density estimation, feature learning, dimensionality reduction, and anomaly detection (Sarker, 2021).

Semi-supervised learning can be defined as a hybrid of supervised and unsupervised learning, since it uses both labelled and unlabelled data (Sarker et al., 2020). Semi-supervised learning is useful in several contexts where labelled data may be rare and unlabelled data are numerous (Mohammed et al., 2016). Thus, the goal of a semi-supervised learning model is to predict more accurately than it would be able to with just the labelled data. Some examples of semi-supervised learning applications are machine translation, fraud detection, labelling data, and text classification (Sarker, 2021).

A reinforcement learning algorithm, also known as environment-driven learning, is a method of automatically determining the most effective behaviour in a context or environment (Kaelbling, Littman, & Moore, 1996). The goal of this type of learning is to take action to increase rewards or minimise risks based on the insights obtained from environmental activists (Mohammed et al., 2016).

The reinforcement learning offers a great tool for automating or optimising complex systems such as autonomous driving tasks, robotics, manufacturing, and supply chain logistics; however, it is not suitable for solving simple problems or straightforward tasks (Sarker, 2021). In human resource management all functions are dominated by decision trees and text-mining algorithms for classification, and ML applications are commonly used in HR recruitment and performance management (Garg, Sinha, Kar, & Mani, 2021).

As the processes of machine learning may need to be modified, it does not explicitly follow a method or framework. Some AI development process models have been proposed that illustrate how ML is conducted. In the following section these models are explained.

2.4.4 The development process of AI

There are two well-known process models in data related projects: the Cross-Industry Standard Process for Data Mining (CRISP-DM) (Schröer, Kruse, & Gómezb; 2021), and the Team Data Science Process (TDSP) (Microsoft, 2020). The CRISP-DM is one of the common process models used by industries and organisations for data mining projects that has been proved to be beneficial for data science and analytics projects. The CRISP-DM methodology includes six phases (Table 2-5): business understanding, data understanding, data preparation, modelling, evaluation, and deployment (Schröer et al., 2021).

Table 2-5- CRISP-DM process model phases (Schröer et al., 2021)

Phase	Description
Business understanding	Business understanding refers to assessing the business situation to determine which resources are available and required as well as determine the data mining goal, data mining type (e.g. classification), and data mining success criteria.
Data understanding	Data understanding is the process of collecting data from data sources, analysing, describing and checking for quality.
Data preparation	Data preparation explains selecting data based on the definition of inclusion and exclusion criteria, and poor data quality can be addressed by cleaning the data.
Modelling	Modelling includes selecting a modelling technique, creating test cases, and building the model. The choice of the model is mainly dictated by the data and the business problem. Parameters need to be set before the model can be built and after building the models it should be evaluated based on evaluation criteria to select the most appropriate ones.
Evaluation	Evaluation is the process of checking the results against the business objectives. Thus, it is essential to interpret the results and define the subsequent actions.
Deployment	Deployment consists of planning, monitoring, and maintenance and an overview of this phase is provided in the user guide as a final report or a software component.

TDSP is “an agile, iterative data science methodology” (p. 6) developed by Microsoft that helps improve team collaboration and learning by suggesting how team roles work best together. The TDSP methodology is similar to CRISP-DM and includes five iterative phases (Table 2-6): business understanding, data acquisition and understanding, modelling, deployment, and customer acceptance (Microsoft, 2020).

Table 2-6 TDSP process model phases

Phase	**Description**
Business understanding	Business understanding outlines the objectives by understanding and identifying the business problem and finding data sources that are relevant to the objectives of the project.
Data acquisition and understanding	Data acquisition and understanding includes collecting data from sources such as on-premises and cloud databases, creating the pipeline to extract data, and data cleansing and visualisation.
Modelling	Modelling includes the process of incorporating, aggregating, and transforming raw data to create the features needed for the analysis, model training, and model evaluation.
Deployment	Deployment refers to scoring and monitoring the performance of the algorithm and deploying the predictive models in the production environment.
Customer acceptance	Customer acceptance is the stage of generating a technical report on the project's completion for the customer. This technical report also incorporates all the information on the project that will be useful in learning how to operate the system.

Moreover, Long and Kelly (2015) developed a Data Analytics Lifecycle that describes the analytics process and includes six iterative phases (Table 2-7): discovery, data preparation, model planning, model building, communicating results, and operationalising. This lifecycle is based on established methods from the field of data analytics and decision science that provided input into various steps of the Data Analytics Lifecycle.

Table 2-7 Data analytics lifecycle (Long & Kelly, 2015)

Phase	Description
Discovery	Discovery refers to the process of learning and investigating the business problem, developing context and understanding, identifying relevant data sources, and formulating the initial hypothesis to test with data later.
Data preparation	Data preparation is iteratively exploring, pre-processing and conditioning data before modelling and analysis.
Model planning	Model planning includes exploring data and learning about the relationships between variables and the most suitable model.
Model building	Model building refers to developing datasets for training, testing, and executing models.
Communicating results	Communication results include comparing outcomes of modelling based on the established criteria for success and failure and developing a narrative to summarise and convey findings to stakeholders.
Operationalising	Operationalising is developing a pilot project to deploy the work in a controlled manner before expanding the work to the entire enterprise. This phase assists the development team in learning about the performance of the model in a production environment on a small scale to adjust the model before full deployment. It also includes delivering the final reports and codes to stakeholders.

The methods that have been used in this data analytics lifecycle are scientific method, CRISP-DM (Wirth & Hipp, 2000), Data, Enterprise, Leadership, Targets, and Analysts (DELTA) (Davenport, Harris, & Morison, 2010), Applied Information Economics (AIE) (Hubbard, 2007), and Magnetic, Agile, and Deep (MAD) skills (Cohen, Dolan, Dunlap, Hellerstein, & Welton, 2009).

CRISP-DM as a popular data mining approach provides useful input on framing analytical problems. DELTA offers a framework for developing data analytics projects that incorporates the skills, data, and leadership engagement of the organisation. AIE offers guidelines for developing decision models, calibrating expert estimates, and determining the expected value of information as well as measuring intangibles (Long & Kelly, 2015). The MAD skills provide a number of techniques that focus on model planning, execution, and key findings. .

Long and Kelly (2015) explained that scientists have long used the scientific method to think about and solve problems by providing a solid framework for thinking through and deconstructing problems into their constituent parts. The scientific method provides a basis for developing a process model for new analytics techniques such as artificial intelligence (AI) that assists in prescribing the steps in detail and gives practical advice for each step.

Some researchers have developed AI lifecycles in different contexts such as healthcare (Hwang, Kesselheim, & Vokinger, 2019) and fintech (Haakman, Cruz, Huijgens, & Deursen, 2020). In healthcare, machine learning promises the prevention of bias in diagnosis and treatment, since a computer algorithm can objectively synthesise and interpret data in the medical records (Gianfrancesco, Tamang, Yazdany, & Schmajuk, 2018). However, ML in healthcare may be prone to biases that are related to missing data and patients not identified by algorithms, sample size and underestimation, and misclassification and measurement error (Gianfrancesco, Tamang, Yazdany, & Schmajuk, 2018).

In fintech AI also shows discrimination rather than protecting the minority groups it is meant to protect. For example, the Apple Card was accused of unfair discrimination against women in its algorithmic lending decisions. It was also found that several insurance companies such as Allstate, Geico, and Liberty Mutual were discriminating against minority groups in their use of ML algorithms for the pricing of car insurance (Kelley & Ovchinnikov, 2020). The reason for this discrimination is using historical data to develop and test the ML model can sometimes reflect historical or latent bias (Nasiripour & Farrel, 2021),

In the context of the R&S process, Cowgill (2020) suggested that algorithms could eliminate bias and improve decision-making by removing human subjectivity from the process of judging and comparing individuals. The use of algorithms will increase consistency by reducing or even eliminating decision makers' biases, since they rely on a mathematical logic that converts quantitative and qualitative data into numerical factors (Newman, Fast & Harmon, 2020). However, there is concern about reproducing biases in algorithmic outcomes and the inability to detect biases (Houser, 2019).

Although discovery of cognitive biases in AI is new, cognitive bias in software engineering is well documented. In the following section, cognitive biases in software engineering and AI are explained.

2.4.5 Cognitive biases in software engineering and Artificial Intelligence

There are different categories of cognitive biases in computer systems (Friedman & Nissenbaum, 1996; Edwards & Rodriguez, 2019). Friedman and Nissenbaum (1996) proposed three categories of biases in computer systems based on analysing actual cases. These three categories are: pre-existing bias, technical bias, and emergent bias. Pre-existing bias stems from institutions, practices, and attitudes within society. Biases that result from technical constraints or considerations are known as technical biases and emergent biases occur as a result of the use of the system.

Edwards and Rodriguez (2019) offered a framework for potential types of bias in analytic-based systems based on Hammond, Keeney, and Raiffa's (2006) work on identifying human sources of bias. They focus on the phases that are most closely related to the business – acquisition and interpretation – and explain how biases such as anchoring, status-quo, sunk-cost, confirmation bias, and framing bias happen in these phases.

In general, software engineering has challenges in terms of cognitive biases in all phases of the development process that need to be addressed (Mohanani et al., 2018). Software engineering is a systematic approach used to develop a system and the development process includes analysis, design, assessment, implementation, testing, maintenance, and reengineering of software (Laplante, 2007). Most articles in information systems research choose one or two phases and identify biases (Table 2-8). For example, Tang (2011) specified biases in designing while Salman (2016) investigated biases in software quality and testing.

Table 2-8 Cognitive biases in information systems based on the literature

Software development processes	**Biases**
Requirements	Anchoring and adjustment (system analysts and project manager), availability bias, representativeness (Browne & Ramesh, 2002), overconfidence, miserly information processing, bandwagon effect and status quo bias (Mohanani, Salman, Turhan, Rodriguez, & Ralph, 2018), Framing desiderata as requirements (Mohanani et al., 2014)
Design	Anchoring and adjustment (Tang, 2011), availability (over-representation of specific code) (Salman, 2016), mere exposure effect, Parkinson’s law effect, representativeness, confirmation (Mohanani et al., 2018), fixation requirements and attentional bias (Mohanani et al., 2014)

Testing	Confirmation (test cases, software documentation), representativeness, availability, positive test bias (Salman, 2016)
Estimating time and cost	Availability bias, overconfidence, sunk-cost fallacy (Mohanani et al., 2018)

Since AI does not follow the traditional software engineering model and is basically extracting important relationships and correlations among correlated data (Nilsson, 1998), the sources of biases in AI are different. Researchers have explained that there are two main drivers of biases in AI: training datasets and algorithms (Kaplan & Haenlein, 2019; Shrestha, Ben-Menahem, & von Krogh, 2019). Fazelpour and Danks (2021) proposed another category of algorithmic bias that includes biases in problem specification, data, modelling and validation, and deployment. Table 2-9 outlines the definitions of these sources of algorithmic bias.

Table 2-9- Sources of algorithmic bias

Sources of algorithmic bias	**Definition**
Biases in problem specification	Bias in problem specification is often seen in the translation of unclear goals into precise measurements.
Biases in data	Algorithmic biases are often caused by existing biases in the real-world systems that are evident in the data. Additionally, measurement methods have limitations and biases that may result in biased data. A simple example is non-representative input data, which can result in the algorithms being less effective on groups that are under-represented.
Biases in modelling and validation	At this stage multiple value judgements are made, and the choice of objective function introduces values into the algorithms.
Biases in deployment	Algorithms implement values based on what they were trained to optimise; important biases can arise if user values diverge from algorithms' values.

As mentioned above, biases in historical datasets are one source that leads to developing biased AI. Research shows that recruiters, including external agents, line managers, and HR staff, can be biased when they evaluate candidates (Linos & Reinhard, 2015). Thus, using historical datasets in recruitment and selection might develop biased AI-Recruitment Systems. The following sections discuss the recruitment and selection process and cognitive biases in this process. Following that, the use of AI in recruitment and selection is explained.

2.5 Recruitment and selection process

The recruitment and selection process (R&S) is considered an important part of human resource management as it has significant strategic value, especially for skilled/scarce workers (Ekwoaba, Ikeije, & Ufoma, 2015). The goal of effective R&S is to match the right candidate with the right job (Newell, 2005). Business success increasingly depends on the ability to attract high-quality employees who can cope with competition, innovation and increasing consumer expectations (O'Meara & Petzall, 2013). Additionally, individuals bring perspectives, values, and attributes to the organisation so acquiring and retaining them is important (Bas, 2012).

Recruitment and selection are categorised into two different processes. Recruitment in general means attracting job applicants who meet key person specifications that must be met for a job to be successfully performed. Selection refers to the process by which the differences between candidates are measured to find the individual who best matches the specifications outlined in the job description (Graham & Benett, 1995). In essence, selection involves differentiating between candidates based on objective criteria and measures that are mainly based upon the personnel performance evaluation (Boran & Yavuz, 2008).

To choose the right type of employees there are some predictors such as person-organisation fit and person-job fit. Person-organisation fit is typically measured by how well a person's perceptions of company values match up with the values the person holds personally (Cable & Judge, 1996). Person-job fit describes how well a person's abilities and personality match the demands and requirements of a particular job (Edwards, 1991). Thus, person-job fit and person-organisation fit play relatively important roles as selection criteria for hiring various types of employees (Sekiguchi, 2007). However, research has highlighted the various ways that this may not be the case in practice as employers cannot directly observe the productivity of external applicants: instead, they rely on "signals" that they believe are related to the underlying capacity to produce (Rivera, 2015).

Assessment centres aim to assess candidates' productivity directly by simulating work-related situations; however, there are a few factors that affect the effectiveness of their testing. According to Kleinmann and Ingold (2019) assessors must be able to work under pressure, elaborate various information in a social situation, and determine the best way to rate different behaviours. These conditions can be conducive to the influence of System 1 processes based on dual process theory, leading to dominant initial general impressions formed quickly due to limited information that result in biased dimension ratings (Ingold, Dönni, & Lievens, 2018).

The biased ratings may be a reflection of implicit or explicit stereotypes such as an assessment of average group abilities, or personal experience (Spence, 2002). For example, Gorman (2005) proposed that candidates' gender affects the selection criteria and decision maker preferences for same-gender applicants exacerbate gender inequality in hiring. Pager (2003) examined the candidates' race by investigating how incarceration affects the employment outcomes of black and white job seekers. These examples represent some common cognitive biases that happen in the R&S process. In the following section, cognitive biases that have been identified in the R&S process are discussed.

2.5.1 Cognitive biases in the recruitment and selection process

There is a complication concerning how employers define or assume who will fit the role (Linos & Reinhard, 2015). Whysall (2017) explained that biases affect the main purpose of selection and assessment, i.e., to identify the extent to which a candidate has or is able to demonstrate certain key characteristics relevant to the role being recruited. There are some biases that have been studied specifically in the R&S process such as race/ethnicity and gender/sexual orientation. To identify existing relevant work on cognitive biases in R&S the researcher conducted a systematic literature review.

The researcher followed methodological guidelines developed by Linnenluecke, Marrone, and Singh (2020) in conducting a systematic literature review in management sciences. This guideline includes identification of literature for inclusion, data cleaning, analysis and synthesis, and presentation of results.

Identification of literature for inclusion: To study cognitive biases in recruitment and selection a search strategy was applied, including the selection or combination of keyword(s) and database(s). Literature for inclusion is often identified by Boolean searches within existing search engines, such as Scopus or Web of Science (Linnenluecke et al., 2020). In this study, relevant articles were identified using Scopus and Web of Science. Additionally, the Business Resource Complete database was used to focus on business/management papers. These databases enable a search for publications based on keywords predefined in the title, abstracts, or keywords. The search keywords were bias or prejudice or discrimination or inequality, and hiring decision or recruitment and selection process.

Data cleaning: As soon as the range of suitable studies was identified, duplicates and studies that were not relevant to the R&S process were removed from the analysis. In addition, in this step cited references were examined to make sure no contributions had been missed.

Analysis and synthesis: It is vital for any systematic review to analyse and synthesise the available evidence, which is determined by factors such as the number of studies that will be included in the review, the type of research method(s) used in individual studies (if applicable), and the quality of the evidence. In this study there was no limitation to the number of studies to be reviewed; all types of research – qualitative and quantitative – and articles published in scholarly journals, conferences, and proceedings during the twelve-year period of 2010-2022 were retrieved. Within this period Human Resource Information Systems (HRIS) started to be used extensively.

Presentation of results: For studies that use qualitative data, the researcher can conduct a qualitative analysis, but not necessarily offer a statistical analysis. Some descriptive statistics (e.g. frequency tables) to summarise basic information, such as the number of publications pertaining to the topic over time, can be used. Cognitive biases in the recruitment and selection process are presented in Table 2-10 and the number of publications for each category of biases per database is shown.

Table 2-10- Cognitive biases in the recruitment and selection process

Biases	Database	Frequency / database
Race/ethnicity	Web of science	8
	Business-Resource-Complete	35
	Scopus	1
Gender bias/sexual orientation discrimination	Web of science	6
	Business-Resource-Complete	20
	Scopus	14
Stereotype bias/discrimination in general (age, race, religion, national origin, colour, sex, pregnancy or disability, social categories, no	Web of science	-
	Business-Resource-Complete	20
	Scopus	-

tattoos, ethnicity)		
Age discrimination	Web of science	1
	Business-Resource-Complete	14
	Scopus	4
Attractiveness bias/appearance/obese candidates	Web of science	-
	Business-Resource-Complete	- 8
	Scopus	5
Bias against disabled candidates (physical or psychological)	Web of science	1
	Business-Resource-Complete	4
	Scopus	1
Similar-to-me (culture, personal values, religious practice, veteran)	Web of science	-
	Business-Resource-Complete	6
	Scopus	-
Bias due to credit history	Web of science	-
	Business-Resource-Complete	5
	Scopus	-
Bias against criminal records	Web of science	-
	Business-Resource-Complete	5
	Scopus	-
Bias towards or against qualifications	Web of science	-
	Business-Resource-Complete	3
	Scopus	1
Caregiving discrimination	Web of science	-
	Business-Resource-Complete	3
	Scopus	-
Decoy effect	Web of science	-
	Business-Resource-Complete	1
	Scopus	1
Working class background/social category	Web of science	1
	Business-Resource-Complete	-

	Scopus	1
Discrimination against smokers	Web of science	-
	Business-Resource-Complete	2
	Scopus	-
Pessimism bias	Web of science	1
	Business-Resource-Complete	-
	Scopus	-
Attentional bias	Web of science	1
	Business-Resource-Complete	-
	Scopus	-
Discrimination based on genetic information	Web of science	-
	Business-Resource-Complete	1
	Scopus	-
Bias towards insider/outsider candidates	Web of science	-
	Business-Resource-Complete	1
	Scopus	-
Covert discrimination	Web of science	-
	Business-Resource-Complete	1
	Scopus	-
First impression	Web of science	-
	Business-Resource-Complete	1
	Scopus	-

As shown in Table 2-10 two groups of biases – race/ethnicity and gender/sexual orientation discrimination – are more frequent in the recruitment and selection studies. Some studies consider these two biases subgroups of stereotypes (Shtulman & Schulz, 2008; Ndobo, Faure, Boisselier, & Giannaki, 2018). Stereotypes initially affect people, who often filter and interpret information through stereotypes and disregard information that contradicts the stereotype (Jonas, Schulz-Hardt, Frey, & Thelen; 2001). Consequently, it leads to selective attention and confirmation bias, reinforcing stereotypes (Whysall, 2018).

Some scholars believe that AI can help HR professionals to identify and remove bias in the R&S process, thereby improving their recruitment decisions and welcoming diverse candidates (Ahmed, 2018; Park, Ahn, Hosanagar, & Lee, 2021). There are many AI applications in the R&S process such as screening chatbots, automated social media scraping tools, and gamifications following selection, in the use of AI in R&S.

2.5.2 AI applications used in recruitment and selection

AI-enabled recruiting tools have been employed across the four main stages of the recruitment and selection process: outreach, screening, assessment and coordination of candidates across stages (Black & van Esch, 2020). In the outreach stage, organisations identify candidates and contact prospective candidates about the job position in ways that will prompt them to apply. Once candidates apply for the positions, employers screen their applications. Then, employers assess those candidates who pass the initial screening to find the most suitable candidate for the job position.

In the outreach stage, organisations need to find the right people by looking for the right candidates broadly and purposefully. The ideal candidate pool consists of both active job and passive job candidates. Identifying both active and passive candidates is critical to create the best possible candidate pool (Guinan, Parise, & Rollag, 2014). AI applications are being used to scrape data from social media such as LinkedIn to match candidates to the job (Campbell, Sands, Ferraro, Tsao, & Mavrommatis, 2020). Moreover, AI applications can form the exact wording and description of the job position and figure out the right presentation methods (e.g., banner ads, email, text) to place job opportunities for the optimal uptake and response by candidate profile (Black & van Esch, 2020).

In the screening stage, AI-enabled screening tools can help achieve significant reductions in lead time (Black & van Esch, 2020). AI-enabled screening tools are not just looking for keywords but instead inferring capabilities that have not been specified in specific words. For example, for a particular job position, persistence might be a required characteristic. AI can infer persistence from sentences instead of scanning for the term or common synonyms (Black & van Esch, 2020).

In the assessment stage, AI assists in a variety of forms such as gamification tests that provide insight into skills, capability, and even personality (Black & van Esch, 2020). For the assessment process, AI can be used to interview candidates and ask candidates various questions and candidates submit their recorded responses (Hamilton & Davison, 2018). Being interviewed by AI allows candidates to have more control over the process and participate in the virtual interview on any day or at any time convenient to them within a several-day window (Hamilton & Davison, 2018).

In the coordination of candidates across stages, AI can enhance the applicants 'experience even for those who get rejected. When a candidate has positive recruiting experience, there is a high chance of saying yes to the offer at the end (Jarrahi, 2018). AI systems can create a positive experience for candidates by asking them to submit their LinkedIn profiles and search through the candidates' profiles and fill in the application for them. Throughout the recruitment process AI can communicate with candidates by answering candidates' common questions about the organisation or the job position such as salary range or education reimbursement, asking candidates questions to fill in any missing or unclear bits of information (Black & van Esch, 2020).

Albert (2019) categorises AI applications that can be employed in various areas and for different proposes in R&S. Table 2-11 outlines the areas where AI applications can be employed to support R&S.

Table 2-11 Areas AI applications can be employed to support R&S (Albert, 2019, pp. 217-218)

AI applications	**Solutions**
Vacancy prediction software	Analysing employees' behavioural data, predicting the likelihood of employees' leaving and reducing costs
Job description optimisation software	Providing recommendations for optimising job descriptions and tailoring language to suit different types of candidates
Targeted job advertising optimisation	Targeting accurate recommendations to the right candidates
Multi-database candidate sourcing	Scanning multiple databases such as LinkedIn and other social media profiles much more quickly and accurately than human recruiters
CV screening software	Filtering out and ranking the best CVs from large amounts of applicants

AI-powered background checking	Enhancing candidate experience while assessing candidates simultaneously
Employer branding monitoring	Scanning across multiple databases to verify candidates’ details including criminal records, credit ratings, and references
Candidate engagement chatbot/CRM	Scanning public data to measure overall sentiment and determine weak points in the R&S process
Automated scheduling	Detecting scheduling expressions and automatically performs these administrative duties.

Ahmed (2018) classifies the AI applications that improve human resources staff’s ability to predict a candidate's future success with the organisation. These AI applications range from basic tools to more advanced AI solutions.

Basic AI applications: Artificial intelligence programs such as screening chatbots and automated social media scraping tools can be useful when sourcing and screening candidates. A chatbot interacts with applicants to confirm they meet job requirements, answer questions and update them about their application's status. Chatbots also provide continuous support through chat, text message, Skype or email, and will contact a human if it cannot complete a task. Social media scraping tools can collect and use vast amounts of data from an applicant's social media profile in order to predict future behaviour, such as engagement levels (Pillala, 2021; Ahmed, 2018).

Intermediate AI applications: There are intermediate AI applications that are used in hiring, such as tests, gamifications, and simulations to collect data directly from the applicant. For example, applicants are required to play some neuroscience-based games for around twenty minutes. Most of these intermediate AI applications give hiring managers an indication of whether a candidate will be a good fit for the job; however, these predictors usually do not focus on specific job metrics (Georgiou, Gouras, & Nikolaou, 2019; Ahmed, 2018).

Advanced AI applications: These AI applications involve algorithms that link unique job performance measures with candidates who possess the most of the required traits (Ahmed, 2018). For example, the video interview develops specific questions to elicit responses that predict job success and identify the right behaviours by analysing each applicant's answers, body language, tone, emotional intelligence, and honesty (Sołek-Borowska & Wilczewska, 2018).

2.6 Summary

This chapter provided a review of decision-making and cognitive biases in decision-making. Next, decision support methods and tools as well as AI, its components and the development process of AI, and cognitive biases in software engineering and AI were explained. Then, the R&S process, cognitive biases in R&S and AI applications used in R&S were described. To illustrate the research questions and the key points of the literature review a conceptual model is depicted below (Figure 2-3).

This model demonstrates the combination of the two cognitive systems: reasoning and intuition (Dual Process Theory). The model shows that AI is considered an adjunct to the reasoning system. Intuitive skills are useful to “monitor the bigger picture in a more holistic fashion” (Hodgkinson & Sadler-Smith, 2018, p. 475) to deal with the details without cognitive effort and to perform tasks efficiently (Hodgkinson & Sadler-Smith, 2018). However, less cognitive effort through using heuristics may result in cognitive biases (Bazerman & Moore, 2013). Cognitive biases influence both HR managers and AI developers which results in developing biased AI-Recruitment Systems.

Researchers have studied cognitive biases and how to mitigate them in different decision-making contexts such as clinical decision-making (Wang et al., 2019). Despite the profound impact AI is likely to have on the HR field, in particular the recruitment and selection process, there is a lack of empirical studies on mitigating cognitive biases in developing AI in the R&S context. This study aimed to identify the common biases that might occur in the development process of AIRS and investigate the approaches to mitigate AIRS cognitive biases. In the following chapter, the research methodology and grounded theory will be discussed.

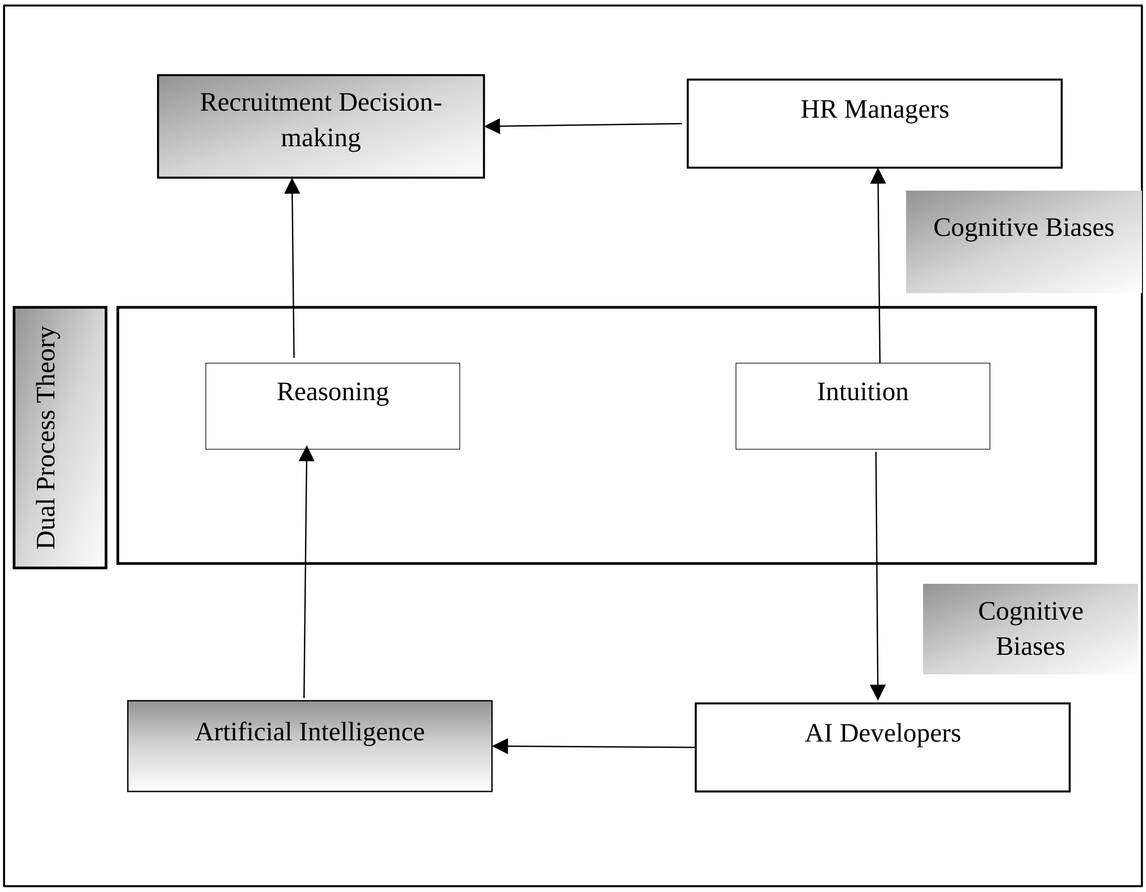

Figure 2-3 Integration of AI and decision-making in relation to cognitive biases in the R&S process

Chapter 3 Research Methodology

As explained in Chapter 1, this study aims to develop a theoretical model that explains how to mitigate cognitive biases in developing AI-Recruitment Systems (AIRS). To achieve this, this study identifies common cognitive biases in recruitment and selection and investigates approaches to mitigate cognitive biases in the development process of AI for recruitment and selection. This chapter explains the research design of this study. The chapter is organised as follows. First the researcher's epistemological stance is explained. Then the logic of choosing the research methodology is discussed. This is followed by a description of the data collection and coding process.

3.1 Research Philosophy

There are different philosophies concerning the nature of social reality and how to examine it. A researcher's philosophical assumptions should correspond to his/her chosen research. Thus, clarifying the underlying philosophical assumptions of a study is necessary (Creswell, 2014). Ontological views of existence and the epistemic relationships between knower and known are key factors determining how social researchers approach a phenomenon (Lincoln, Lynham, & Guba, 2011). This section discusses the ontology and epistemology underlying the study, which guides the methodological framework and data gathering and analysis.

Ontology refers to the interpretation of reality by the researcher and making assumptions about the nature of the understudied phenomenon (Hudson & Ozanne, 1988). Researchers are subjectivist or objectivist according to how they interpret reality. Objectivists consider social and natural reality as an independent entity existing prior to humans' cognitive abilities, whereas subjectivists believe that reality is the result of human cognition (Johnson & Duberley, 2003).

Subjectivists hold that reality is constructed based on human experiences and that there is nothing fundamentally real about it. In contrast, objectivists contend that there is only one reality as social reality is external to the researcher (Collis & Hussey, 2013). On the spectrum between subjectivism and objectivism, social constructivism lies in between and closer to subjectivism (Bell, Bryman, & Harley, 2018). In the social constructivism perspective, social reality can be viewed as subjective and influenced by daily interactions and interpretations of social actors. Thus, social actors play an important role to shape social phenomena and their meanings (Bell et al., 2018).

I believe that reality is not external to individuals and is constructed by an individual. Consequently, this study is based on social constructivism as an ontological paradigm. There are two reasons why social constructivism is suitable for this study. First, this study aims to develop an in-depth understanding of the HR managers and AI developers' perceptions on mitigating cognitive biases in developing AI for the recruitment and selection process. Second, the researcher can engage in in-depth conversations with participants and construct her knowledge of the studied phenomenon by interacting with the participants. Hence, the purpose of this study and its nature are in accordance with social constructivist ontology, which informs the epistemological perspective of the research.

Epistemological views pertain to the nature of knowledge (Collis & Hussey, 2013). It is the process by which we know what we know (Crotty, 1998). The researcher conducted this study by taking an interpretivist stance as an epistemological approach. Interpretivism assumes that knowledge is created or constructed by a person through their lived experience and is based on the assumption that social reality is not objective (Collis & Hussey, 2013). Interpretivism adheres to meaningful social action through the subjective perspectives of the people involved in a specific context and time (Hudson & Ozanne, 1988).

A researcher's ontology and epistemology can inform the methodology that the researcher uses in a study. The following section explains the reasons for choosing the research methodology and design of this study.

3.2 Methodology and Design

Methodology refers to the process of conducting research and may include multiple methods, which should align with the ontological and epistemological approaches (Collis & Hussey, 2013). The constructivist/interpretivist researchers most often use qualitative data collection and analysis methods or mixed methods (Seaman, 2008). Klein and Myers (1999) pointed out that the interpretive research in the field of information systems assists in "producing deep insights into information systems phenomena including the management of information systems and information systems development" (p. 67). Thus, conducting an exploratory study will provide insight into the development process of AI-Recruitment Systems (AIRS) and mitigating biases.

3.3 Qualitative and exploratory research

According to Morse (1991), the research problem can best be addressed by qualitative methods when i) the area under study is "immature" because of a lack of theory and previous research, ii) the notion of the available theory is not accurate, iii) the phenomenon needs to be explored and then a theory can be developed, and/or iv) quantitative measures cannot measure the phenomenon. In this study, the notion of cognitive bias in AI for recruitment and selection is novel and there is a lack of theories in this field. Approaches to mitigate cognitive biases in each phase of the development process of AIRS need to be explored that can lead to developing the process model of unbiased AIRS. Hunter (2004) believes that qualitative researchers try to make sense of the phenomena they observe and conduct their investigations "in the field". In this study, "in the field" investigation was deemed suitable to understand the process of developing AIRS and identify the source and/or cause of cognitive biases for developing AI-Recruitment systems.

Even though qualitative methods allow the researcher to examine the complexity of the issue deeply rather than abstract it away (Seaman, 2008), quantitative data may enhance or support qualitative data, effectively deepening the description (Mackenzie & Knipe, 2006). For example, one way to quantify qualitative data is the survey method which has been extensively applied in IS research (Gable, 1994). However, to design survey research, a prior understanding of the context and history of a certain computing phenomenon is required to develop a detailed model of the expected relationship based on independent and dependent variables (Pinsonneault & Kraemer, 2016). Since there is a lack of research on AI cognitive biases, particularly in the recruitment and selection process, developing a model of variables and their relationships and using a survey method does not suit this study. Therefore, these approaches need to be explored prior to measuring the efficiency of approaches to mitigate cognitive biases in developing AIRS.

A commonly used qualitative method is the case study, a well-established way to advance knowledge and discovery in IS (Benbasat, Goldstein, & Goldstein, 1987; Cavaye, 1996). Benbasat, Goldstein, and Goldstein (1987) review case study research in IS and conclude that the "case study strategy is well suited to capturing the knowledge of practitioners and developing theories from it" (p. 370). As opposed to surveys and experiments, case studies provide a more holistic view of a problem (Blumberg, Cooper, & Schindler., 2011). Case studies are more appropriate for researchers who lack a priori knowledge of the variables to be evaluated (Benbasat et al., 1987).

Due to the lack of a priori knowledge of cognitive biases in developing AIRS, both the case study method and grounded theory seem suitable for this research. It is appropriate to conduct case research when theoretical knowledge of a phenomenon is limited, or context must be captured (Cavaye, 1996). Cavaye (1996) further explains that IS research strategies are always suited to capturing the context, which is important when discussing people-related and organisational phenomena. In this study, the researcher employs classic grounded theory (Glaser & Strauss, 1967) as a guide to data collection and analysis. Cavaye (1996) points out that typical qualitative data analysis focuses on achieving rich descriptions and addressing issues of voice and contexts, whereas grounded theory is concerned with conceptualisations that are abstracts of time and place.

The case study differs from grounded theory in that it involves intensive descriptions and analyses of a singular unit or bounded system such as a single organisation, program, event, or group (Smith 1978). However, the assumption of grounded theory is that each group shares an unarticulated social problem (Laws, 2006). Grounded theory has the advantage of providing a detailed study of a micro issue of a larger reality within the context of a particular setting (Glaser & Strauss, 1967). Thus, the grounded theory allows the researcher to provide detailed information about a specific phenomenon and to be influenced by the context within which it is conducted (Laws, 2006). In this study, grounded theory helps the researcher understand cognitive biases in R&S and approaches to mitigate them in developing AIRS from various perspectives rather than targeting a singular unit.

Moreover, Fernandez and Lehmann (2011) compare the grounded theory method with case studies in regard to memo-writing and using literature reviews. They explain that the slices of data researchers collect using the case study method extend beyond the traditional data collection in observation and interview data in grounded theory (Fernandez & Lehmann, 2011). Data in grounded theory includes literature (which is used after the data analysis) and memo-writing produced by the researcher during the process of analysing data. The memos then form a second-order text that assists the researcher in achieving further conceptualisation

(Fernandez & Lehmann, 2011). In this study, the memos assisted the researcher in identifying cognitive biases in R&S and concepts and processes of developing unbiased AIRS.

To conduct interpretive research, other methodologies such as descriptive phenomenology, content analysis, and thematic analysis can be used (Vaismoradi, Turunen, & Bondas, 2013). Sandelowski and Barroso (2003) explained a scale for the research findings that can be transformed from description into interpretation. Based on this scale, for the qualitative descriptive approaches such as descriptive phenomenology, content analysis, and thematic analysis, less interpretation is required. Interpretive approaches such as grounded theory and hermeneutic phenomenology need more interpretation (Vaismoradi, Turunen, & Bondas, 2013). This is consistent with the researcher's interpretivist instance.

In addition, the researcher's interpretation was inevitable in this study since biases in R&S and the concepts of mitigation techniques may not be obvious to the participants. Therefore, the researcher needed to delve into and interpret what the participants mentioned to identify biases. Furthermore, the researcher interprets the concepts of the data concerning methods for mitigating cognitive biases to categorise meaningful and uniform data in each phase of the development process of AIRS.

At first, the researcher was going to use thematic analysis to derive knowledge and develop classifications that would lead to establishing meaning and solid findings (Sandelowski, 2010; Holloway & Todres, 2003). Thematic analysis provides a systematic element to data analysis and demonstrates the data in detail and manages interpretation of diverse subjects (Boyatzis, 1998). According to King (2004), thematic analysis is a useful method that helps the researcher to examine perspectives of different participants, and find out similarities and differences along with generating unanticipated insights. Namey, Guest, Thairu, and Johnson (2008) asserted that "thematic [analysis] moves beyond counting explicit words or phrases and focuses on identifying and describing both implicit and explicit ideas. Codes developed for ideas or themes are then applied or linked to raw data as summary markers for later analysis, which may include comparing the relative frequencies of themes or topics within a data set, looking for code co-occurrence, or graphically displaying code relationships." (p. 138).

However, for this study, grounded theory was deemed more suitable than thematic analysis as grounded theory could assist the researcher in gathering and interpreting data through a systematic iterative process of data gathering and analysis to discover the notion of cognitive biases and techniques to mitigate them in the development process of AIRS. The grounded theory relies on theoretical sampling that is determined in the process of data collection (Glaser & Strauss, 1967), and further data collection is grounded on what has been analysed previously (Strauss & Corbin, 1990). Grounded theory is explained in more detail in the following section. The variants of grounded theory and reasons why the Glaserian variant was a suitable methodology for this study are also discussed.

3.4 Grounded Theory

Grounded Theory as an explorative and interpretative qualitative research method was developed by Barney Glaser and Anselm Strauss in the late 1960s (Glaser & Strauss, 1967). Grounded Theory aimed for construction of new theories grounded in data (Fassinger, 2005; Charmaz, 2014). This type of research is characterised by the following attitude towards empirical work and data: the theoretical structuring of an issue under study is postponed until the structuring of the issue under study has been determined by the person being studied. When there is a delay in structuring, it implies that ex-ante formulation of hypotheses is abandoned. Accordingly, the research question can be outlined under the heading of theoretical aspects (Flick, 2019). Grounded Theory is a systematic multi-phased analytical process that could be complicated to conduct for novice researchers. For this reason, it is important to explain the main aspects of the method before describing the data collection procedure and coding process.

3.4.1 The Main Aspects of Grounded Theory

The main aspects of Grounded Theory are theoretical sensitivity, theoretical sampling, theoretical saturation, and constant comparison. These four aspects of Grounded Theory are explained below.

3.4.1.1 Theoretical Sensitivity

The principle of theoretical sensitivity is the ability to generate concepts from data, combine the concepts that are relevant, and fit them into the emerging concepts and hypotheses (Glaser & Strauss, 1967). Theoretical sensitivity entails remaining open to the data and the emerging theory. To be theoretically sensitive, it is necessary to perceive what is emerging from the data rather than using preconceived ideas and hypotheses (Glaser, 1978).

Theoretical sensitivity assists researchers in seeing data in new ways through the theory development process (Hallberg, 2006) and explains data in a way that best manifests reality (Hall & Callery, 2001). Thus, the theoretical sensitivity is completed when the research participants' interaction and perspective influence the analysis and results (Hall & Callery, 2001).

3.4.1.2 Theoretical Sampling

Glaser and Strauss state that "theoretical sampling is the process of data collection" (1967, p. 45) which means the researcher collects related data to "elaborate and refine categories in the emerging theory" (Charmaz, 2014, p. 192). The actual theoretical sampling starts after the first data collection and first data analysis so that the researcher can apply theoretical sampling for extending and refining theoretical categories (Charmaz, 2014). After elaborating and checking the theoretical categories, the researcher decides what data needs to be collected and from where (Glaser & Strauss, 1967). The process of data collection and data analysis proceed until theoretical saturation is achieved.

3.4.1.3 Theoretical Saturation

The definition of theoretical saturation is closely linked to theoretical sensitivity and theoretical sampling in that "no additional data are being found whereby the sociologist can develop properties of the category" (Glaser & Strauss, 1967, p. 61). Theoretical saturation occurs when similar instances appear repetitively, and it tells the researcher to stop sampling the different groups of a category as the category is saturated (Glaser & Strauss, 1967). Theoretical saturation is the endpoint of the coding process as fitting in new data would not develop new theoretical insights (Flick, 2019).

3.4.1.4 Constant Comparative Method

The constant comparative method is the analytical process to "generate theory more systematically … by using explicit coding and analytic procedures" (Glaser & Strauss, 1967, p. 102). The constant comparative method involves exploring variations, similarities, and differences in data by constantly comparing every part of the data with all other parts of the data. The constant comparative method of grounded theory assists researchers in investigating the content and meaning in the data (Hallberg, 2006). Moreover, constant comparison allows the researcher to identify the themes that are concerns for most participants. Grounded theory must fit and be relevant to the data, meaning that emerging categories should explain the collected data instead of imposing preconceived concepts and ideas on the collected data (Glaser, 1992). Thus, constant comparison advances the focus, relevance and accuracy of coding, categorisation, and conceptualisation.

The four main aspects of grounded theory help to generate theory derived from the data through a systematic and transparent process. Theory can be developed by logico-deductive methods; however, grounded theory is inductively derived from data (Goulding, 2002). There are various approaches in grounded theory. Even though there are commonalities in all approaches, some factors such as the philosophical position of the researcher, the use of the literature, and the coding method differentiate different variants of grounded theory (Chun Tie, Birks, & Francis, 2019). In the following section, grounded theory approaches are explained.

3.4.2 Grounded theory approaches

Barney Glaser and Anselm Strauss introduced grounded theory in their 1967 book *The Discovery of Grounded Theory: Strategies for Qualitative Research*, which defined grounded theory as a method of generating new theories from data through systematic qualitative analysis. They explained that to develop new theories, researchers should conduct data collection, coding, and analysis simultaneously (Glaser & Strauss, 1967). After publishing the book in 1967, the two authors have taken different approaches to grounded theory.

Glaser published a book that explained the grounded theory methodology further, titled *Theoretical Sensitivity* (1978), and Strauss published *Qualitative Analysis for Social Scientists* (1987). It seems that the reason Glaser and Strauss went different ways is Strauss believed that the researcher finds an objective external 'real reality' in an objective and neutral way (Hallberg, 2006) since Glaser mentions that all is data and data 'emerges' without being influenced by the researcher (Glaser, 2002).

In 1990, Strauss and Corbin published their book *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. They explained grounded theory and its procedures and called it "the reformulated grounded theory". Strauss and Corbin (1990) directly stated that reality can be interpreted; however, it cannot be fully known. They explain that "doing analysis is, in fact, making interpretations" (p. 59). Strauss and Corbin also proposed a coding paradigm that assists researchers in explaining the conceptual relationships between concepts/categories and their properties in the theory development (Strauss & Corbin, 1998). According to Strauss and Corbin (1998), the findings of a grounded theory study are validated through the process, and the result of a grounded theory study, which is a developed theory, should be useful in practice.

However, Glaser claimed that using the coding paradigm forces preconceived concepts and categories upon data. Glaser also argued that the grounded theory procedure of Strauss and Corbin does not support theoretical sensitivity and inductive openness towards the data (Glaser, 1992).

In 1991, Schatzman, who was Strauss's student and colleague, developed 'dimensional analysis'. According to Schatzman (1991), although dimensional analysis and grounded theory are both informed by symbolic interaction and both were designed to generate theory directly from data, dimensional analysis overcomes the limitations of early grounded theory. Schatzman believed that dimensional analysis is broader than grounded theory and considers the complexity of analytic processes by focusing on 'what all is involved' in the data that can lead to a more grounded understanding of a phenomenon (Morse et al., 2016).

To develop rich and comprehensive dimensions, Schatzman (1991) emphasised that the researcher should encourage informants to describe concepts, dimensions, and properties. In contrast to Strauss' (1987) approach where the analyst was engaged in comparison right from the beginning, Schatzman (1991) suggested that a larger dimension of data should be identified before the comparative analysis.

In 1995, Charmaz introduced 'constructivist grounded theory'. Constructivist grounded theory is a revision of Glaser and Strauss' classic grounded theory. She explained that constructivist grounded theory aims at getting close to the empirical realities as opposed to the classic grounded theory (Charmaz, 2006) and constructivism believes that there might not be only one social reality and thus there might be multiple social realities simultaneously. Social constructivists are against the idea of studying without prior knowledge and theories about the phenomena under study (Charmaz, 2008).

Charmaz (2008) explained that "objectivity is a questionable goal, and what researchers define as objective still reflects partial knowledge and particular perspectives, priorities, and positions" (p. 402). Conversely, Glaser (2003) discussed that the comparison method built into grounded theory helps researchers to analyse and conceptualise data without rendering the data objective.

Moreover, from a constructivist view, a real world cannot be separated from the viewer who observes it from multiple viewpoints that may conflict with participants' viewpoints and realities. The multiplicity of perspectives and multiple realities lead to having different ways that both participants and the researcher construct meaning (Morse et al., 2016). However, instead of giving priority to the researchers' views, constructivists consider participants' views as integral to the analysis (Charmaz, 2008).

In 2005, Clarke developed 'situational analysis' by considering grounded theory as a "theory/methods package" (p. 81), originating from symbolic interactionism. Social interactionism is considered as a theoretical framework used to understand people's behaviour and perspectives, and the researcher describes the processes of human interaction (Zeegers & Barron, 2015). The key part of symbolic interactionism is people who make their own social realities and viewpoints of their world using responses from the environment and different sociocultural relationships within which they interact. Clarke (2005) further explained that in addition to humans, nonhuman objects such as cultural objects, technologies, animals, and media should be taken into consideration. To analyse the situations adequately, the nonhuman objects must be included explicitly and in significant detail (Morse et al., 2016).

To conduct grounded theory analysis, Strauss specified structural conditions and developed the conditional matrix to specify the salient structural conditions for the phenomenon under study in the analysis. The matrix consists of different levels such as international, national, and governmental, and depending on where the research is undertaken, community, organisational, institutional, or local group and individual/(inter)actional settings. The international level consists of economic, cultural, religious, scientific, and environmental issues and the national level includes political, governmental, cultural, economic, gender, age, ethnicity, race, and particular national issues (Strauss & Corbin, 1990). For Strauss the core of the matrix is action and he applied his sociology of work into theorising routine and nonroutine action (Clarke, 2008).

Clarke (2008) further argued that the conditional matrix only contains the structural elements of situations which are not explained in detail and clearly as a necessary part of grounded theory analysis. Therefore, Clarke believed that conditional matrices do not do the conceptual analytic work that Strauss expected (Morse et al., 2016). In the situational matrix, the situation needs to be specified in the analysis. “Regardless of whether some actors might construe them as local or global, internal or external, close-in or far away, or whatever, the fundamental question is: ‘How do these conditions appear — make themselves felt as consequential — as integral parts of the empirical situation under examination?’ At least some answers to that question can be found through doing situational analyses” (Morse et al., 2016, p. 208).

Table 3-1 summarises the grounded theory approaches.

Table 3-1 Grounded theory approaches

Theorists	**Description**
Glaser and Strauss (1967)	Glaser and Strauss defined grounded theory as a systematic qualitative analysis that develops new theories derived from data and researchers should engage in the process of data collection, coding, and analysis simultaneously (Glaser & Strauss, 1967).
Strauss and Corbin (1990)	Strauss and Corbin developed a reformulated grounded theory by offering a coding paradigm that helps researchers to analyse relationships between categories and concepts (Strauss & Corbin, 1990).
Schatzman (1991)	Schatzman emphasised a more grounded understanding of a phenomenon and developing inclusive dimensions through explaining concepts, dimensions, and properties. However, he believed it is necessary to identify larger dimensions of data before comparative analysis (Schatzman, 1991).
Charmaz (1995)	Charmaz developed the constructivist grounded theory as a method to seek and understand a social process through building inductive analysis from the data. However, the constructivist grounded theory assumes that researchers are part of the research and that knowledge is co-developed (Charmaz, 1995).
Clarke (2005)	Clarke presented situational analysis as an extension of grounded theory that considers the situation as the unit of analysis and explains human and nonhuman objects in detail (Clarke, 2005).

In this study, the researcher used the Glaserian variant. The reasons why grounded theory, and more specifically the Glaserian variant, suits this study are explained in the following two sections.

3.4.3 Why grounded theory?

Grounded theory was applied in this study due to four main reasons: a paucity of pre-developed theories, engaging with data, incorporating the context, and providing a documented record of the progress of the analysis. In the following sections, each reason is discussed in detail.

3.4.3.1 A paucity of pre-developed theories

Von Krogh (2018) observed that studying AI in the field of decision-making and problem-solving in organisations is a new research area. He pointed out that academic work needs to be done to assist practitioners to be informed and have a realistic approach to AI. One of the expectations of HR managers in using AI in the recruitment and selection process is mitigating cognitive biases in recruitment decisions. However, cognitive biases are considered as one of the new issues and challenges in AI-assisted decision-making (Kaplan & Haenlein, 2019) and there is a lack of theory about mitigating cognitive biases in the development process of AI, specifically in the HR sector.

In this study, grounded theory was used as it allowed the researcher to develop a theoretical account of the general features of a topic while simultaneously grounding the account in empirical data (Martin & Turner, 1986; Glaser & Strauss, 1967). Moreover, grounded theory is suitable for both Information Systems (IS) and Human Resources (HR) fields. There are numerous empirical studies in the IS field that applied grounded theory (Pauleen, 2001; Intezari, 2014; Nguyen, 2021) as well as HR (Mergenthaler et al., 2021).

3.4.3.2 Engaging with data

According to Glaser (2005), grounded theory is a “research paradigm for discovery” (p. 145). As Walsh et al. (2015) argue, grounded theory is more than a methodology and it can be considered as an approach, being a “meta-theory of inductive research design” (p. 584). Grounded theory helps demonstrate the outlines of social phenomena through engaging with existing data and “discovering theories in rupture with existing literature” (Walsh et al., 2015, p. 584).

Grounded theory shapes the data collection while doing data analysis simultaneously. The iterative data analysis and data gathering throughout the data collection process visualises the emerging patterns, categories, and dimensions (Strauss & Corbin, 1998). Such an iterative process can certify that the developed theory is relevant to the phenomenon under study (Howard-Payne, 2016). Grounded theory generates enough data so that the illuminated patterns, concepts, categories, properties, and dimensions of the given phenomena can emerge (Glaser & Strauss, 1967; Strauss & Corbin, 1998).

The under-researched phenomenon in this study is ‘cognitive biases’ and how to mitigate them in AI in the context of the R&S process. To follow the grounded theory approach, this study began with selecting through the ‘initial sampling’ to get into the field and in touch with the first participants and their insights. Since the context of the decision-making is recruitment decisions, the first sample was chosen from HR managers who had experiences in recruiting. The data gathered from the first sample helped the researcher to identify common cognitive biases in R&S which was not possible if AI developers were the first sample.

Then, the researcher carried on with more purposeful strategies of sampling (theoretical sampling) based on the categories and dimensions emerging from the data. After interviewing HR managers, to gain a deeper understanding of the development process of AIRS and techniques to mitigate cognitive biases in developing AI, the researcher sought informants who had experience in developing AI for the R&S processes. In grounded theory researchers engage a phenomenon through the lens of those experiencing it and gain new theoretical insights (Corley, 2015).

In this study grounded theory allowed the researcher to interpret and analyse the informants’ perceptions on cognitive biases in the R&S process and the development process of unbiased AIRS. To gain theoretical insight, developing a specific plan for sampling cannot be clearly determined at the beginning because sampling decisions should be governed by theory. The researcher stopped collecting data and elaborated the development process of unbiased AI i) “When no new or relevant data regarding categories seemed to emerge”, ii) “the categories were well developed in terms of their properties and dimensions demonstrating variation”, and iii) “the relationships among categories were well established and validated” (theoretical saturation) (Strauss & Corbin, 1998, p. 212).

3.4.3.3 Incorporating the context of the 'under-researched phenomena'

The grounded theory process directs researchers' attention to the context in which individual behaviour takes place. Context is an important part of an individual's experience at work (Murphy, Klotz, & Kreiner 2017). For example, to produce accurate and useful results, the complexities of the organisational context have to be incorporated into an understanding of the phenomena (Oriikowski, 1993; Martin & Turner, 1986; Pettigrew, 1990).

In this research, grounded theory helped to understand how cognitive biases find their way into AIRS from the perceptions of experts involved in the development process of AIRS. The possible cognitive biases happen in the recruitment and selection process as one of the sources of training datasets are identified by the experts involved (i.e., HR managers). Additionally, grounded theory allowed the researcher to take into account the nuances of different contexts in which HR managers and AI developers work. For example, HR managers perceive cognitive biases in developing AI for R&S more subjectively, whereas AI developers' mindset about this phenomenon is more algorithmic and rational.

3.4.3.4 Providing a documented record of the progress of the analysis

Generating a documented record of the progress of the analysis is one of the strengths of using grounded theory. This detailed documentation, e.g. through memo-writing, assists the researcher to be able to check back through the data and memos to derive any new concepts or models (Pidgeon, Turner, & Blockley, 1991). Memo-writing is the essential process of researcher and data engagement that assists the researcher in coding the 'raw' data, analytically interpreting data, and transforming data into 'grounded' theory (Lempert, 2016).

As Charmaz (2006) explains, memo-writing is a methodological practice where the researcher analyses the data and simultaneously increases the level of abstraction of his/her analytical ideas. By writing memos continuously throughout the research process, the researcher explores, explicates, and theorises the emerging patterns. The researcher used memos "to create social reality" (Richardson, 1998, p. 349) by discursively organising and interpreting the participants' perceptions to conceptualise the data in narrative form which results in creating the development process of unbiased AIRS.

Table 3-2 summarises the reasons for using grounded theory in this research.

Table 3-2 Reasons for using grounded theory in this study

Reasons	Description
A paucity of pre-developed theories	Studying AI in the field of decision-making and problem-solving in organisations is a new research area (Von Krogh, 2018). Therefore, there is a lack of developed theory regarding using AI-assisted decision-making in organisations to mitigate cognitive biases.
Engaging with data	Grounded theory shapes the data collection and data analysis simultaneously. In this study, the iterative process of grounded theory helped the researcher to collect data and analyse it in a more focused way that lead to generating a theory of the development process of unbiased AIRS.
Incorporating the context of the 'under-researched phenomena'	To produce accurate and useful results the complexities of the context of the phenomena under study should be incorporated. This study used grounded theory to understand how cognitive biases happen in the R&S process based on the perceptions of experts contributing to developing AIRS such as HR managers and AI developers.
Providing a documented record of the progress of the analysis	Documenting the progress of the analysis through memo-writing helped the researcher to check back through the data and memos and derive new concepts to provide a nuanced understanding of how cognitive biases might appear in designing AI for the recruitment and selection process as well as the mitigation techniques.

So far, the reason grounded theory was deemed suitable for this study has been explained. Given that a specific variant of grounded theory, the Glaserian approach, has been used in this study, the reasons the variant was chosen are explained below.

3.4.4 Why the Glaserian Approach was Chosen

Rieger (2019) explained that a researcher can choose the methodology for a grounded theory study based on the philosophical point of view, the purpose of the study, and pragmatics. The four reasons comprise philosophical points of view, literature review, purpose of the study, and pragmatics.

3.4.4.1 Philosophical points of view

Numerous scholars have argued that the chosen grounded theory approaches should be compatible with researchers' personal beliefs (Corbin & Strauss, 2008; Evans, 2013). However, scholars such as Charmaz (2006), Holton (2008), and Alammar, Intezari, Cardow, and Pauleen (2019) stated that grounded theory is not related to the researcher's philosophical position and a researcher with a positivist or interpretivist approach can undertake each of the grounded theory approaches.

Although grounded theory can be neutral, the way grounded theory is used is not neutral and it depends on the researcher's philosophical views (Charmaz, 2006). According to Glaser (1992), Strauss and Corbin's structured design forces data and analysis into preconceived categories. In contrast, Strauss and Corbin (1990) believed that the structured design assists in making sense of the data and developing a theory.

According to Urquhart (2013), the philosophical positions of the Glaserian and the Straussian approaches are still uncertain and different authors have reached different conclusions. Due to the different philosophical viewpoints, there is a difference in the Glaserian and Straussian approaches to using literature.

3.4.4.2 Literature review

Doing a literature review is now acceptable in various levels to some grounded theorists of both approaches such as Charmaz (2006), Martin (2006), and Urquhart and Fernández (2006). As PhD students have to follow university procedures, doing a literature review is usually necessary as it helps students to find an area of research and prove that they have the ability to do research (i.e. the confirmation process) (Alammar et al., 2019).

In this study, the researcher, as a PhD student, conducted an initial literature review before data collection and data analysis to discover the area of interest, familiarised herself with the research area and developed an open and broad research question. After data collection, the data revealed itself to the researcher through conceptualisation and constant comparison to understand common cognitive biases in recruitment decisions. Then, the researcher narrowed down the research question by asking about challenges of developing AI that lead to biased AI and ways to mitigate cognitive biases in developing AI-Recruitment Systems (AIRS). Throughout the data collection and analysis process, the researcher revisited the literature to comprehend and explain the findings with relevant literature.

3.4.4.3 Purpose of the study

The grounded theory approach must fit the study's purpose (Corbin & Strauss, 2008; Evans, 2013). This study was qualitative and exploratory. Although the notion of cognitive bias in decision-making is not new, the use of AI in management decision-making particularly in recruitment decisions is rather new (Von Krogh, 2018). The purpose of this study was to explore and study cognitive biases in the R&S process and how to mitigate cognitive biases in the development process of AIRS. According to Rieger (2019), when a researcher aims at doing explanatory research and identifying variables, classic Glaserian theory can be more suitable.

3.4.4.4 Pragmatics

Although the Straussian approach provides more clearly defined procedural guidance (Strauss & Corbin, 1990) and has been widely applied in IS research (Urquhart, Lehmann, & Myers, 2010), the Glaserian approach has flexibility in process guidelines (Glaser, 1978).

The conditional matrix developed by Strauss and Corbin (1990) helps to generate multiple viewpoints on a phenomenon under study (Charmaz, 2006). However, Glaser (1992) suggests that "using constant comparison method gets the analyst to the desired conceptual power quickly, with ease and joy. Categories emerge upon comparison and properties emerge upon more comparison and that is all there is to it" (p. 43). Likewise, Cooney (2010) explained that the Glaserian approach may be more user-friendly for a novice researcher.

To follow the Straussian approach, the researcher hypothesises and relates categories and their properties together to create a theory or model. Conversely, Glaser (1992) believed that the concept and categories should emerge without forcing data into preconceived concepts and efforts to make relations between them.

This study explored how cognitive biases can be mitigated in the development process of AI-Recruitment Systems and the researcher did not consider any preconceived categories and hypotheses. The researcher let "the hypotheses and concepts systematically work out in relation to the data" (p. 56) while doing the research (Glaser & Holton, 2007).

Table 3-3 summarises the reasons for using the Glaserian approach in this study.

Table 3-3 Reasons for using the Glaserian approach

Reasons	Description
Philosophical points of view	Grounded theory can be neutral; however, using grounded theory depends on the researcher's philosophical points of view. In this study, the researcher did not consider any preconceived categories to analyse data. Therefore, the Glaserian approach is more aligned with interpretative points of view.

Literature review	Although the literature review was conducted in this study before data collection and analysis, the researcher was trying to derive concepts out of data.
Purpose of the study	This study was exploratory in nature. It explored cognitive biases in the recruitment and selection process and how to mitigate cognitive biases in the development process of AIRS. To accomplish exploratory research the Glaserian approach can be more suitable (Rieger, 2019).
Pragmatics	In this study the researcher approached hypotheses and concepts in relation to data without considering any preconceived categories and hypotheses.

As mentioned before, data collection and analysis are conducted in a recursive process. In the following sections the data collection procedure is explained. Then, the processes of open coding, selective coding, and theoretical coding are illustrated.

3.5 Data collection procedure

Interviews were conducted in English with twenty-two Human Resource (HR) managers based in New Zealand and seventeen AI developers globally (New Zealand, Australia, the USA, Germany, Israel, and India). The interviews started with HR managers in New Zealand; however, there are not many organisations in New Zealand using AI in their R&S process. A few organisations that are using AIRS purchased their system from overseas. Thus, in the third and fourth rounds of data collection, the researcher interviewed the AI companies working with HR departments in New Zealand and other AI companies around the globe that are working with HR departments in New Zealand.

The researcher recruited these informants through social networking sites such as LinkedIn, and personal connections. A range of criteria were used to identify candidates in different phases, including: 1) HR managers with work experience in recruiting and selecting candidates (regardless of the organisation size, and the industry); 2) HR managers who possess conceptual and or practical knowledge on AI and the AI development process; and, 3) AI developers with expertise in AI development for recruitment functions.

All interviews were audio-recorded and transcribed by the researcher. The researcher interviewed the HR managers over two phases of data collection – ten interviews in phase one and twelve interviews in phase two – to first explore potential biases from HR respondents. In the third phase of data collection, the researcher conducted ten interviews with AI developers and in the last phase, seven AI developers were interviewed.

Out of twenty-two HR managers, seventeen were female, and five were male, with an average of fourteen years of work experience in HR in New Zealand. HR managers' familiarity with AI varied, ranging from having limited knowledge of AI applications to being conceptually and technically knowledgeable about AI. The demographic information on HR managers is shown in Appendix 2- HR managers' demographics The average length for the interviews was 45 minutes.

Among seventeen AI developers, two were female, and fifteen were male, from AI companies in various countries. All AI developers had more than 2 years of experience in developing AI-Recruitment Systems (AIRS). The number of AI developers interviewed in this study was lower than the number of HR participants because the available pool of participants who are experts in developing AI for the recruitment and selection process was very limited. For the demographic information of AI developers see Appendix 3- AI developers' demographics

In the first phase of data collection from HR managers, all interviews were done face-to-face. Other phases of data collection were conducted via Zoom due to the Covid-19 pandemic as most people worked from home and preferred not to be interviewed face-to-face. Data collection using Voice Over Internet Protocol (VoIP) technologies has some common technical challenges such as loss of Internet connection (Fox, Morris, & Rumsey, 2007) and poor sound/video quality (Sullivan, 2012; Williams, Sheffield, & Knibb, 2015). In addition, VoIP technologies have some ethical, practical, and interactional issues (Seitz, 2016; Weller, 2015). However, Archibald, Ambagtsheer, Casey, and Lawless (2019) showed that Zoom is a better platform for qualitative interviewing because of ease of use, cost-effectiveness, data management features, security options, and building rapport (Archibald et al., 2019)(Archibald et al., 2019).

In grounded theory, data can be collected through multiple data sources such as interviews and observations of behaviour (Goulding, 1998), and in-depth interviews and observation are the most frequently used methods of data collection in grounded theory. Scholars mention that field data can be collected through remote interviews such as phone and video conferencing. According to the literature on interviewing, one of the major characteristics of high-quality interviews is their depth of detail (Hermanowicz, 2002). The richness of data is achieved through strategies such as probing, carefully crafted and sequenced interview guides, and listening to what is said and unsaid (Hermanowicz, 2002; Rubin & Rubin, 2011).

Additionally, Cachia and Millward (2011) state that nonverbal cues observed in person can be followed up with verbal cues or replaced with specific probing questions. Hanna (2012) and Janghorban, Latifnejad, and Taghipour (2014) also note that the interaction when using the web-camera is comparable to an onsite experience based on presence of nonverbal and social cues.

In this research, semi-structured interviews were conducted by asking open-ended questions. Open-ended questions were asked to allow interviewees to talk freely and to allow the researcher to gain as much information as possible (Glaser, 1998). The researcher conducted three pilot interviews to generate open-ended questions. Participants provided feedback to the researcher on their experiences with the questions, the interview process, and any significant absences. The participants' feedback assists the researcher in analysing the feasibility of the interview format and gauging the impact of the questions formulated to explore the issues (Burck, 2005).

3.5.1 Interview questions

The researcher asked both substantive questions and a set of theoretically sensitive questions to collect data (see Appendix 4). In the first round of data collection, the researcher only asked substantive questions about cognitive biases in the R&S process and using decision support systems such as AI in R&S. Then, from round 2 to 4, to follow the theoretical sensitivity, as more interviews were conducted and more data were analysed, the interview guideline was modified and developed further by adding new and more theoretically sensitive and purposeful questions. The questions were about cognitive biases that are likely to occur in the recruitment and selection process and might happen in AIRS as training datasets, the development process of AIRS, and/or strategies to mitigate cognitive biases in the development process of AIRS. Appendix 4- Interview questions outlines the interview questions in each phase.

3.6 Coding Process

To identify categories of the collected data, three coding strategies – open coding, selective coding, and theoretical coding – were used (Glaser & Strauss, 1967). Data collection and data analysis took place simultaneously in that the data collection was influenced by the development of the sub-core categories and core categories (theoretical sensitivity) (Glaser, 1978). Figure 3-1 outlines the data collection and data analysis process.

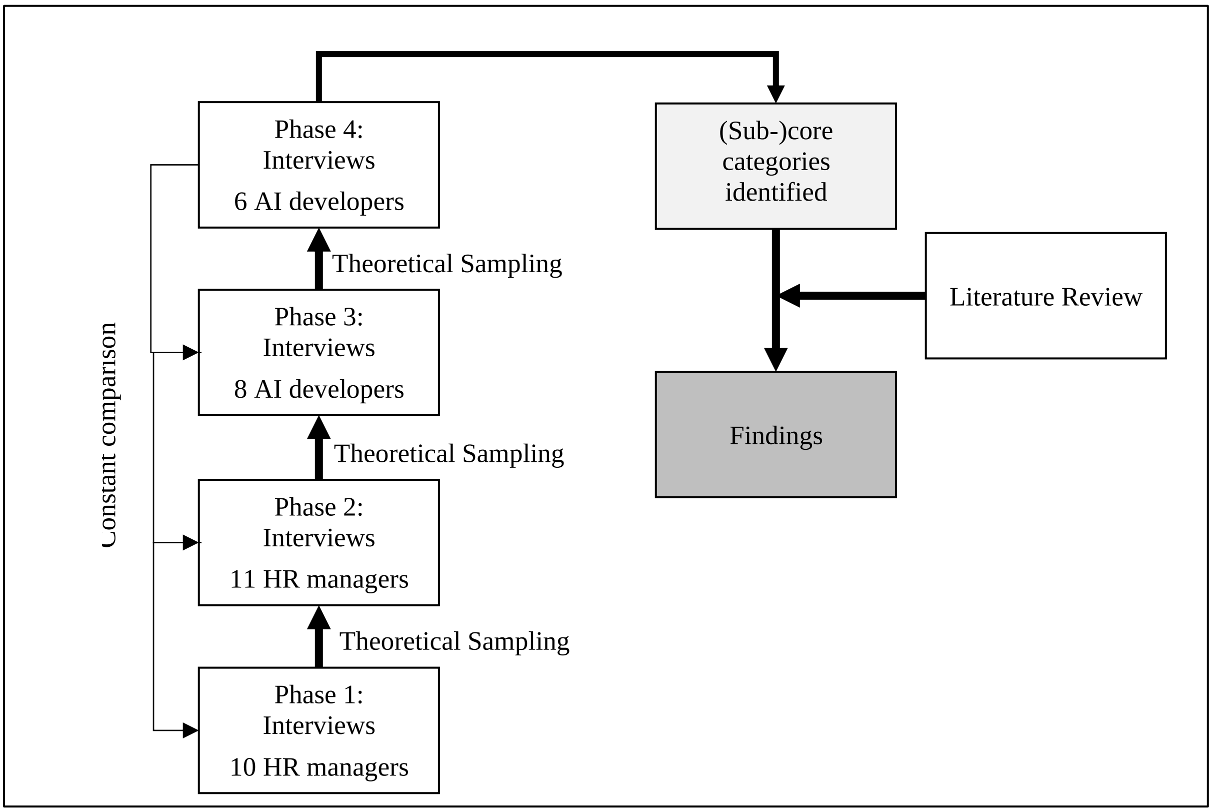

Figure 3-1 Data collection and data analysis process

Data analysis began with reading the transcriptions and memo-writing to gain an understanding of the context and main points of the data in relation to the research questions. During the coding process, the researcher assigned open codes to each paragraph or sentence, whichever conveyed one specific message or topic (Glaser, 1978). Open codes were assigned either using informants’ words (in vivo codes) or the researcher's knowledge and interpretation of the data in its context. Table 3-4 presents examples of open coding.

Table 3-4- Examples of open coding

Informant quotes	**Open coding**
Having enough data reflects the bias problem we're trying to solve	Enough datasets, mitigating biases
and the ability to communicate to the user of the AI	Communication

product is also very important. *As having a good feedback loop from users helps developers to mitigate biases in AI (AI developer, 7).*	Feedback
I know that AI can only learn as more data [is]put into it. *And the software that we use are trained with large datasets as this AI company works with companies all around the world (HR manager, 22).*	Enough datasets Large datasets, sources of datasets

In a similar manner, all transcripts were coded for each interview. Then, the researcher was able to decide which data to collect next (theoretical sampling). Throughout the data collection and data analysis process, the researcher ensured theoretical sensitivity and identified patterns among incidents by asking three questions that Glaser (1978, p. 57) suggested: “What is this data a study of?”, “What category does this incident indicate?”, and “What is actually happening in the data?”.

Not all open codes could be placed in the emerging conceptual categories. Codes that did not appear to be a concern for most informants were excluded from the analysis. For example, because one informant mentioned “Educating AI users” only once and it did not appear in the subsequent data analysis, the researcher did not include this code in the core category.

When the conceptual codes were derived, the incidents were grouped under as many categories as possible (Glaser, 1978). That is what Strauss and Corbin (1998) introduced as more abstract explanations, i.e. conceptual categories (p. 114). The data collected in the new interviews were compared with the data collected in the previous interviews (constant comparative method) (Glaser & Strauss, 1967). Thus, the researcher refined categories and decided whether a new category should be developed, or the existing ones were sufficient (Wastell, 2001).

To analyse the relationships between structure and process, as well as the links between categories and sub-core categories, the Glaser theoretical coding families known as the six C’s were used. The six C’s comprise six coding questions to find the ‘causes’, ‘consequences’, ‘contingency,’ ‘condition’, ‘covariance’, and the ‘context’. Causes are the reasons behind the observed phenomenon. Consequences cover the effects of the observed phenomenon. Contingency explains the moderating role of the phenomenon and/or other phenomena in causing the observed phenomenon to occur. Condition is related to time, place, and duration. Covariance identifies the correlation between categories and context is the circumstance where the phenomenon was observed (Glaser, 1978).

There were 925 codes in total, 453 of which were used for the analysis, and six conceptual categories were identified. Table 3-5 represents the frequency of informants mentioning the code.

Table 3-5 Frequency of informants mentioning the code

Conceptual Categories	**Conceptual codes**	**Frequency (Total 453)**	**Percentage**
Cognitive bias	Similar-to-me bias	95	20.97%
	Stereotype bias	75	16.55%
HR managers' assumptions and job positions requirements	Understanding job functions	56	12.36%
	Engaging with HR managers and employees	34	7.5%
	Asking good questions	8	1.76%
	Articulating job position requirements	27	5.96%
Data collection	Collecting and preparing enough, diverse, and accurate datasets	95	20.97%
Data preparation	Labelling datasets	5	1.10%
	Augmenting datasets	8	1.76%
	Filling out missing data points	4	0.88%
Training ML models	Choosing predictor variables	32	7.06%
Validating ML models	Using validation techniques, tools and frameworks	34	7.50%
	Removing biased data points	16	3.53%
	Tuning ML algorithms	7	1.54%
Monitoring and	Giving feedback by testing ML models	39	8.60%

retraining ML models	Applying feedback into the model	19	4.19%

3.7 Summary

In this chapter, the research methodology and design were explained. The exploratory and qualitative nature of this study was discussed. Grounded theory was introduced, and it was explained that grounded theory was considered appropriate due to four reasons: a paucity of pre-developed theories, engaging with data, incorporating the context of under-researched phenomena, and providing a documented record of the progress of the analysis. Different grounded theory approaches were presented. The classic grounded theory was justified as the adopted approach in this study and the analytical process was discussed. In the next chapter, how grounded theory was implemented in this research will be explained.

Chapter 4 The Research Findings

This chapter presents the findings by answering two research questions: 'Which cognitive biases are more likely to be observed in recruitment and selection decisions?' and 'How can cognitive biases be mitigated in developing AI-Recruitment Systems?' The main output of this study is a 'process model' that illustrates the process of developing unbiased AI in the Recruitment and Selection process (AIRS). To understand the development process of unbiased AIRS, the researcher needed to explore what cognitive biases can happen in recruitment decisions before understanding what design process could be developed to mitigate the biases.

This chapter illustrates the categories and sub-core categories that emerged from the data analysis (Table 4-1). For each category, some samples of the informants' quotes are provided, which are interpretively explained. The findings are presented in two sections. First, cognitive biases that might happen in developing AIRS are explained. Then, the process of mitigating biases during the design process of AIRS is described.

Table 4-1- Conceptual codes and categories

Theoretical Coding		**Selective coding**	**Open Coding**
Core Category	Sub-core Categories	Conceptual Categories	Conceptual Codes
Developing process of unbiased AIRS	Understanding the ML model requirements	HR managers' assumptions and job position requirements	Understanding job functions
			Engaging with HR managers and employees
			Asking good questions
			Articulating job position requirements
	Managing datasets	Data collection	Collecting and preparing sufficient, diverse, and accurate datasets
		Data preparation	Labelling datasets
			Augmenting datasets

			Filling out missing data points
	Developing and retraining machine learning model	Training ML models	Choosing predictor variables
		Validating ML models	Using validation techniques, tools and frameworks
			Removing biased data points
			Tuning ML algorithms
		Monitoring and retraining ML models	Giving feedback by testing ML models
			Applying feedback into the model

4.1 Cognitive biases in the recruitment and selection process

The findings indicate that participants are consistent about two major cognitive biases in the recruitment and selection: the similar-to-me and the stereotype biases. Some informants point to other cognitive biases such as conformation bias, first impression, bandwagon bias, and fatigue bias that have occurred in R&S. These biases, however, did not seem to be concerns for most participants. In grounded theory, the researcher focuses on the issues that are concerns for most participants (theoretical sensitivity) (Glaser, 1978).

Stereotype bias involves being unduly favourable towards or against different ethnic groups, candidates having a specific number of years of experience, working for famous companies, and working for companies in the same country (i.e., having local experience). Similar-to-me bias encompasses favourable tendencies towards candidates who graduated from a similar school as the recruiter, have shared habits and interests, and those who the recruiters perceive that they can ‘have more fun’ with.

4.1.1 Similar-to-me bias

Frank and Hackman (1975) explained that similarity increases the likelihood of receiving consensual approval of one’s views. Interviewers have a strong desire to obtain social validation of their opinions and views which leads to finding similarity-favourableness relationships with interviewees (Frank & Hackman, 1975). Participants gave examples of how having different values in terms of religion might result in dismissing the candidate:

“When I'm a Christian Catholic who doesn't believe in certain things and I look at someone's profile [and] see their values are not similar to my personal values, then I might rule them out as a potential candidate” (HR manager, 17).

One HR manager pointed out that similar-to-me bias happens in hiring decisions because HR managers can easily communicate with some candidates when they find similarities between themselves and candidates:

“Just a connection that the person is just like me, I could work with them, they're going to be easy, they've got a nice smile and they're going to be great with the customers” (HR manager, 2).

Another HR manager indicated that having similar values and points of view as interviewees can cover some missing requirements in their CV:

“Sometimes hiring managers ignored some big issues like the candidate was only three months at that last job. That's because the hiring manager thinks they are similar and have similar values” (HR manager, 16).

Similarly, HR manager 11 mentioned that HR managers judge candidates as a good team fit based on criteria such as having “the same hobbies” that might not be relevant to viewing a candidate as a team player:

"So, I remember a candidate who went through a few interviews and then went to the final round. The final round was meeting the team, and they met at a bar on a Friday at lunchtime. The candidate is Muslim and he needed a job to get a visa to stay in the country. He has spent his life savings to get there, he went to the bar, and they asked, do you want to drink and he said no, then it seemed like he is not a team player because he did not have the same hobbies as other team members and he got rejected" (HR manager, 11).

Likewise,

Candidates put on their hobbies and interests, and sometimes their hobbies can go for them. Someone might say, I love playing rugby, and we’ll have a rugby person to watch rugby with (HR manager, 7).

However, another HR manager believed that there might not be a clear distinction between similar-to-me bias and getting a good fit with the team. HR manager 12 mentioned that seeing a difference between similar-to-me bias and finding a good team fit is difficult:

> *"Everybody talks about it as a cultural fit or a team fit because it's like, ah, they can both do the job. But who would I have more fun working with? Who would fit in with our team better? I think that's a big one. You hear about it all the time, for example, you can have fun, or you can feel more comfortable with candidate A but not with candidate B because candidate A is more like you. So how can you judge? I think that's where it gets a little bit grey" (HR manager, 12).*

In addition to the similar-to-me bias, the data showed another category of biases that revolve around how HR managers rely on factors such as candidates' experience and ethnicity when assessing candidates and their applications. These findings were categorised under the title of stereotype bias.

4.1.2 Stereotype bias

The literature defines stereotype bias as a rigid or discriminatory view of a person because of their social category membership such as their ethnicity, age, or gender which undercuts their potential and ignores the diversity of the group (Hinton, 2019). HR manager 13, who has had sixteen years of experience in the hiring process, explained some items such as only looking at the candidates' place of birth and English skills that indicate bias towards an ethnicity:

> *"They could be biased on somebody's ethnicity, and making an assumption, maybe all that person might not have as great of English skills, or they might look at somebody who has a date of birth on there" (HR manager, 13).*

Similarly, another HR manager who has had the experience of working in two different countries explained the issue of judging candidates based on their name as an example of being biased towards an ethnicity:

> *"People get judged on their names: if it's a John Smith, managers are interested. But if it's a name that doesn't look like it's a familiar one, they are not interested" (HR manager, 3).*

HR managers sometimes have conscious biases towards candidates, although they would not necessarily identify such judgements as biased; rather, this kind of judgement could be useful tacit knowledge. One of the HR managers, who had fourteen years of experience and is now working as a principal consultant, explained his bias when he judged a candidate based on having five years of experience in a specific organisation because he knew that organisation:

> *"So I know the engineers at that company are really good. So, she gets a little bit more credibility when I read her CV because I know from experience that she's a really good engineer if she's been working for five years in that company" (HR manager, 15).*

Likewise, HR manager 5 pointed out that having work experience ensures a better chance of success in hiring a candidate who is able to conform and adapt to the organisational culture:

> *"First of all looking at experience and duration of time spent in employment that it's going to be easier to have a good cultural fit if you've had X number [years] of working experience. When you've been in positions for a long time will show that you're capable of culturally fitting into an organisation solar system" (HR manager, 5).*

The findings show that the stereotype and similar-to-me biases are very common in the R&S process. However, when biases can be mitigated in developing AIRS, it can help identify and reduce the impact of HR managers' biases in the R&S process. HR managers believed that AI assesses all applications and measuring capabilities consistently, based on measuring criteria “statistically”. Two HR managers explained the benefits of assessing all candidates similarly with the help of AI:

> *"AI can review every single application and proposal with the same eye consistently, and consistently absorbed information" (HR manager, 10).*

Similarly,

> *"You can build a really robust tool that matches the criteria with great performance confidently. Then, it's incredibly effective because it's more statistically based than saying that I met Anna, and she was a really nice person. She had a good sense of humour. Did that manager measure capabilities and then interview or were they just basing it on the fact that she is a nice person, so AI obviously is more robust if done*

well. The other thing is, it provides a lot more productivity, and productivity is key when you think about costs" (HR manager, 13).

Likewise, one of AI developers explained that all candidates were assessed by AI in the same way, as AI cognitive capacity is not limited. In addition, he mentioned that human assessment cannot be as fair as an unbiased AI due to human cognitive limitations:

"As human beings, when we got sent 500 or 1000, video introductions, we don't have the time and bandwidth to process all of those videos. What AI can do is AI can basically have a limitless look at all of these candidates. So, everybody has an equal level of visibility and consideration, and bring everybody into that process" (AI developer, 5).

HR manager 22 believed that although AI could initially process a large number of candidates, HR managers could gain information on a much deeper level by interviewing a candidate:

"I think allowing the use of AI means getting access to a larger amount of people and having HR people means that you can kind of get a different level of information from candidates as opposed to surface level. I believe a conversation with a human can pick up voice tones and body language that may allow us to probe further and uncover a deeper level of information" (HR manager, 22).

The HR managers also expected that with AIRS assistance, they could gain some level of understanding and interpretation of what has happened in the R&S process based on objective features. One of the HR managers who was experienced in HR and had a Master's degree in technological futures for HR stated that AIRS can give HR managers feedback in the interview sessions and act as one of their colleagues whose decisions are objective and unbiased:

"Sometimes, what happens is that a lot of hiring managers become really indecisive, and they can't make a call. But if you had somebody giving them really objective, factual data after each interview, it would be great. For example, for a strategy manager role, and if I am a very lenient person and give everybody a five out of five and the other colleague is more objective and give a two out of five then if we have AI

to work with us, it might make me think about things that I might have overlooked" (HR manager, 12).

Likewise, another participant pointed out one of the benefits of using AI is providing the opportunity for reviewing the decision-making through the R&S process and finding the decisions' errors:

"So sometimes it is hard to kind of go back and look at what/where the error was, could have been on bad referencing? Or it's just got to be that the manager didn't ask the right questions in the interview" (HR manager, 13).

In this section, two cognitive biases – 'similar-to-me' and 'stereotype' – have been identified that are very likely to be built into AI-Recruitment Systems (AIRS). According to the findings, developing unbiased AIRS can assist HR managers in mitigating biases in the recruitment and selection process. The following section explains the development process of unbiased AIRS as the core category along with the sub-core categories.

4.2 The Core Category: The Development Process of Unbiased AI-Recruitment Systems (AIRS)

The core category represents the process model of designing unbiased AIRS. This model explains how biases can be mitigated during the AIRS design process. The core category consists of three sub-core categories: *understanding the ML model requirements*, *managing datasets*, and *developing and retraining ML models*. Six conceptual categories emerged during axial coding (Table 4-1).

4.2.1 Understanding the ML model requirements

The development process of ML-based software systems starts with generating rules based on training datasets. To find relevant and accurate training datasets, understanding the domain context is required. According to the findings, understanding the R&S process requirements depends on HR managers' assumptions and requirements for each job position.

HR managers need to understand job functions, and AI developers should engage with HR managers and employees who know about job functions. Moreover, understanding requirements helps AI developers to measure the algorithm's performance based on understanding the problem and Key Performance Indicators (KPIs). According to the AI developers, AI cannot assist HR managers in the R&S process if the requirements are not specified at the beginning of the process:

> *"I think the recruiters should understand that AI is not a magic thing, it is developed by developers who are prone to errors. So, they really need to set the expectations right in the first place" (AI developer, 10).*

Another AI developer, who was experienced in software development and ML, highlighted the key role of business analysts in understanding the requirements. However, he believed that it was becoming less common in the IT industry nowadays:

> *"The role of having a dedicated business analyst is probably something that is becoming less and less common in the IT industry. But having said that, like the type of work that traditional business analysts might have performed in the past is still really important to the success of a project. I mean, it's simple, it's common sense, you really have to understand what the problems are that you're trying to solve before you can actually do anything useful. And it's no different for an AI project, or an AI team" (AI developer, 15).*

Similarly,

> *"I think some of the most valuable things are just talking to people on the business, maybe shadowing their work, reviewing problems that they've had in the past, It would be the discipline where you're trying to sit with someone and really understand how they work, and understand what's slowing them down, or where they see opportunities" (AI developer, 13).*

AI developers mentioned that understanding the requirements for each business context prior to the development process of AI leads to properly managing datasets. Identifying suitable training datasets is a function of understanding the context and the requirements:

> *"What would happen is understanding what you're trying to do at first and then you find data that is suitable for your use case" (AI developer, 9).*

> *"In the next step, you'll be trying to source relevant datasets and depending on the context, it could be from a variety of different sources that you need to look at them" (AI developer, 15).*

4.2.2 Managing datasets

Managing datasets involves collecting and preparing datasets and the role of HR managers and AI developers in data collection and preparation. Collecting datasets refers to the strategies that AI developers would follow to compensate for a lack of datasets from the R&S process. Preparing datasets includes data augmentation, data labelling, and ‘curated features’ that help to train ML models with accurate, sufficient, and diverse datasets.

4.2.2.1 Data collection

To develop AIRS, two datasets are required for training and testing ML algorithms. For developing AIRS, both HR managers and AI developers believed that developers could not simply rely on datasets from HR managers, and there was a lack of datasets in the R&S process. One of the AI developers believed that the R&S process did not have a ‘data-driven’ approach and as a result did not generate sufficient datasets for training AIRS:

> *"The Human Resource area and especially in the recruitment stage, it hasn't been an area where our data is used a lot to make decisions and it's typically less data-driven. We need more data to train algorithms" (AI developer, 2).*

Similarly, an HR manager, who had twenty years of experience in HR, asserted a lack of datasets in the R&S process compared to other parts of HR such as pay and leave, i.e. the logistical side:

> *"HR hasn't been very data driven. A lot of the transactional data that they look at is more around leave and pay and some of those kinds of logistical sides of HR" (HR manager, 16).*

Nonetheless, the participants mentioned that there were some strategies to collect datasets, such as collecting data from candidates who have applied for job positions through job advertising platforms. One of the AI developers mentioned that they collected data from advertisers when candidates applied for job positions online:

> *“We collect a lot of data when people interact with jobs on site. And we also collect data from advertisers when they interact with candidates on our candidate portal. So for us, it would be trying to wrangle together all of those different datasets and get them in a place that we can access them easily” (AI developer, 15).*

He further highlighted the importance role of HR managers in speeding up and increasing the success rate of ML models by collecting as much data as they could:

> *"I guess, what AI developers would be hoping for is that recruitment managers have access to existing data about their hiring. I would encourage HR managers to collect as much data as possible about the processes. Because that would speed up and increase the likelihood that AI projects would be successful in the space" (AI developer, 15).*

Another AI developer mentioned two other ways of finding datasets: "crowdsourcing and asking people" and "finding similar datasets". He gave an example of collecting similar datasets for training a language model when there is a lack of datasets:

> *"In scenarios where there was limited data, either crowdsourcing and asking people was an option, or basically being able to go and find similar data and see how well it performs. So, for example, if I'm trying to train a language model for Turkish, I might go and collect data from Turkish as well as Azerbaijan as Azerbaijani is similar to Turkish. So, try to collect multiple sources of data that might be similar to improve the model" (AI developer, 9).*

4.2.2.2 Data preparation

According to the findings, datasets that are used to train ML models in R&S often cannot be readily used, and as a result huge efforts must be made to prepare data before training the ML models. Datasets should be prepared in a way that is sufficient, accurate, and diverse for training. The challenge of having enough datasets is that the R&S process is not rich in data, and as mentioned above, there is a lack of datasets in R&S. Diverse datasets focus on having inclusive and representative datasets to train ML models. Accurate datasets refer to valid and relevant datasets without missing data points.

4.2.2.2.1 Enough datasets

Although it is not easy to obtain enough datasets, there are some techniques such as "data augmentation", "synthetic data", and "aggregating datasets" that developers can use to compensate for the lack of datasets. These techniques are used to produce more datasets to develop and retrain ML models. For example, one AI developer pointed to a technique called "augmentation" that developers apply to increase small amounts of training datasets.

> *"One suggestion would be to try to generate your own data in creative ways, which is called augmentation, so you have more than one example of training AI. Let's say we're trying to learn how a person does transitions in terms of his career. For example, if I look at myself, I was an analyst in the past, and now I'm a data scientist, so if I have my CV, I know that transition and developers can understand from my curriculum [vitae]" (AI developer, 7).*

Another technique that the AI developers referred to in relation to increasing datasets was "synthetic data". As AI developer 1 explained, the technique, however, is risky, and developers need to make sure that the original datasets used in this technique are free of errors:

> *"The ability to generate synthetic data is super good these days. But that doesn't mean that it is good data. The problem with synthetic data is that if there's one little problem in one data point and you amplify that data to make more of it, you get to make more data points that have problems, so we need to be careful when we get data" (AI developer, 1).*

Another AI developer discussed the technique of "aggregating datasets", by which different sources are aggregated to build a larger dataset. He further explained that aggregation should be correct for the ML models to be able to find correlations within large datasets:

> *"You need to understand the challenges of how to first aggregate your dataset because if you get bad data in, you won't get good output from that data. So, you need to know first how to aggregate your data, how to find correlation within those very large datasets to draw meaningful conclusions that make the process innovative, efficient and better" (AI developer, 5).*

4.2.2.2.2 Diverse datasets

AI developers believed that having diverse datasets is required for training AIRS. Diverse datasets help AIRS to "find patterns" and "correlations"; however, existing datasets of the R&S process are not diverse enough. One of the AI developers defined diverse datasets as "representative datasets" so that ML models can "find correlations and conclusions" from them.

"Good data is a very broad cross-section of participants feeding data into that set that represent either the community that company recruits from, or the diversity represented just in the candidate pool in a certain geographic region or country. So, it should be a representative dataset that you can draw meaningful correlations and conclusions from that are valid and reliable" (AI developer, 5).

Another AI developer explained the reason that the R&S process does not have representative and diverse datasets. He believes that the work opportunities are not equal for all candidates. For example, datasets for the leadership positions do not contain data from minorities such as people of colour and women.

"You can see how many of these AI engines have biases right now because the dataset is not accurate. So, with inequality in the market, you don't have as many people of colour or as many women or as many people from minorities in leadership positions. So that biased dataset skews the AI engine and AI is going to learn from a dataset that is essentially corrupt" (AI developer, 6).

Another AI developer, however, believed that it was possible to generate diverse datasets for recruitment decisions. She explained that they interview people working in the job position and/or similar positions to understand which features are required. Then, they find more datasets that cover the necessary features:

"Having a diverse dataset to train AI is really important. That's not just one type of person you're looking for, but that you get different ways people spoke of something that we've all perceived as good or ended up being good for the job. We interview many people in a role, and who worked with people in that role to understand the competencies, like is teamwork very important? Is customer oriented very important? The algorithms are a prediction, they're not perfectly accurate and humans as well. So, you know, they're looking for patterns in the topic and what the person talked about and how they express themselves" (AI developer, 4).

4.2.2.2.3 Dataset accuracy checks

Accurate datasets refer to not having missing data points in datasets which is important for training AIRS. Having accurate datasets leads to training AIRS with more “objective” datasets as they are classified properly. AI developer 4 believed that the R&S process in some organisations has missing data points. This missing data is typically the final result of the hiring process, which is a valuable data point for AI.

> "*What is kind of difficult is that we don't always get the whole picture when it comes to data, like the applicant tracking system has a lot of data for this whole entire journey that a candidate takes. Depending on the customer, we don't always know if this person got hired; many of our customers use our interface to rank candidates and talk about them. But some don't use it at all. Some of them do that all through the applicant tracking system and that information just flows back into the applicant tracking system" (AI developer, 4).*

Likewise, the statement below shows that when HR managers assist AI developers in filling out the missing data points in datasets, it helps AI developers manage datasets properly for training.

> *"AI is nothing except some baseline screening and recommendation filtering. HR managers should not think that AI can do everything for them. They should be involved in the process of developing AI by filling out data points in datasets completely that help developers easily classify and categorise data and help us train AI with large scale data and objective datasets" (AI developer, 13).*

One of the HR managers also pointed out the probability of missing data points in the R&S datasets that may cause some issues when developers fill them out.

> *"We don't feed through, like total information into data, so they're not getting everybody's information" (HR manager, 18).*

4.2.3 Developing and retraining ML models

After collecting data, the development process starts with determining the true success criteria of job positions for ML algorithms. Based on the defined success factors (i.e., class target), ML algorithms find a pattern to map between a set of input features in training datasets and an output target. Developing ML algorithms is a continuous process, and different modelling approaches should be tested on test datasets to find the "optimised" ML model. In addition, the knowledge of experts in the field can provide an opportunity to improve the outcomes of algorithms.

One of the AI developers mentioned that the development phase of ML models is not a one-off design process. AI developers will try various modelling approaches to measure success and find the optimised model based on the key success metrics:

> *“The next phase would be typically quite iterative, and here, you might be trying out a number of different modelling approaches. So you might be trying a variety of algorithms that are kind of framed towards your success measures. It would be kind of an experimentation process where you'd have an evaluation dataset that you'll be using, and you'll be trying to optimise your key metrics against the evaluation dataset (AI developer, 15).*

The participants also pointed out that HR managers providing feedback to AI developers and communicating with them about the results assist in continuously improving the ML models:

> *“The second piece is that after six months down the road, you do an analysis and a review of the AI model. Suppose HR people communicate so that the AI developers are also provided with the result of what came out of the hiring decisions they make. In that case, I think it helps to make the AI ecosystem advanced” (AI developer, 14).*

As AI developer 5 explained, the developers can improve the ML model by integrating HR managers' knowledge in the initial stages of developing AI. They ask a group of HR managers to do a test on a sample set of data and shortlist the top ten candidates based on the required soft skills. Then, they put the exact datasets into AI and compare the results of HR decisions and AI decisions, and make changes in the ML model based on HR managers' knowledge:

> *"One of our very early stage phase one research projects will involve us taking a sample set of candidate videos, and feeding that to a test group of HR executives, who would rank these candidate videos in certain soft skill attributes up to 10 to see what they came up with. We had our AI go through the same process to make those same rankings based off how we had built it to assess the inputs it was receiving from the video feed and then we compare these two. Then we brought all this knowledge over from what the Human Resources said, and put it into our machine learning or modelling and maybe I'm looking at this batch of videos with human beings to see the missing parts in our outcomes" (AI developer, 5).*

Two AI developers explained that having a diverse group of people working on ML models is necessary. Having various perspectives helps a group to think about as many features as possible in ML models, leading to generating unbiased outcomes.

> *"It's important to have diverse groups of people working on building those algorithms. The more perspectives you have, the more diversity you have in building algorithms, the more representative it might be" (AI developer, 9).*

Likewise, other AI developers explained that having a diverse team working on ML models can decrease the risk of developing biased ML models. For example, AI developer 15 specifically highlighted having multi-cultural team members as a good strategy for mitigating biases.

> *"I mean, it can get quite risky if you've got like, essentially modelling decisions or algorithmic decisions, resting on one person. That's high risk because if that one person has some biases, then there's really no protection against them. So demystifying the work, good governance, good peer review was probably quite useful. So in that case, making sure that our teams are fairly diverse and not too monocultural would probably be a good move as well" (AI developer, 15).*

After developing ML models, monitoring them and providing feedback on ML modelling is essential. AI developers should retrain and improve ML models based on new requirements and provide feedback on ML model errors. The statements below emphasise the importance of continuous monitoring and users' feedback that leads to detecting errors, retraining, and improving the ML models.

> "*I think any model or any AI model is not perfect out when it comes to production. They always need to be tested like they test it and keep on training it. It's called retraining the model. Like we do our production models, we test them every two weeks, for any problems or any errors, and then we train it again, and that's how it improved by the time" (AI developer, 10).*

> *"We get user feedback about some kind of preferences say that they give us some signal that again, this one does not fit that one particular job and this can be like, no, missing the skills or the title does not match, or the location does not match or*

experience not enough or are qualified or experiences not enough, or some kind of signals the user can give to us" (AI developer, 13).

AI developers pointed to reinforcement learning as a technique in which a ML model analyses the feedback and tries to "learn from its mistake". The ML models that do not just rely on learning from a fixed dataset and interacting with an environment can benefit from the reinforcement learning technique. The statement below explains reinforcement learning as a technique to improve ML models by learning from their mistakes.

"I think there's one more thing that we can always use is reinforcement learning. Like, instead of just doing training, learn from what we missed the AI had, like learn from their mistakes like a child ... when we were growing up, we learn from our mistakes, we learn from our books, and then we learn from our mistake as well. So, there is not always the data that you can learn from and we can learn from our mistakes" (AI developer, 10).

Likewise,

So, you can do something like reinforcement learning, where you keep adding more and more data to help improve your model (AI developer, 9).

Moreover, AI developers brought up some techniques that mitigate cognitive biases and lead to improving ML models. For instance, AI developer 9 asserted that managing the 'parameters' to generate unbiased outcomes is important. For example, if the model is biased against a 'parameter' (e.g., gender), more data could be collected to leverage the ML output:

"All models have some parameters. I think it's just about like, what do you do with those parameters that might be causing the skew? I guess you could hide some of them, or try to find more data for each one" (AI developer, 9).

Different techniques such as the A/B test or F-score to test ML models and manage ML features can assist developers in mitigating biases. AI developers can apply such techniques to evaluate ML models for biases. The A/B test measures ML performance using new scenarios and new datasets. The F-score is a method to check models' biases by testing the effects of certain categories on the outcome of ML. AI developer 14 mentioned the F-score as one of the most straightforward techniques to check the ML models regarding biases. He explained the F test is done to check how certain categories make a difference to the ML outcomes.

> *"You can do hypothesis testing like F-score statistics, that's something that I sometimes use for when I have a category, like gender and we have some continuous variables. And let's say I put up my system discourse, and I tried to see whether certain attributes certain categories make a difference. I think F test is one of the good ones and easiest one to use and there are a bunch of others, too" (AI developer, 14).*

AI developer 15 explained the "A/B test". In this technique, AI developers measure the ML performance with new scenarios and new datasets:

> *"We run A/B tests, essentially, to measure the effectiveness of whatever we're doing in a real-world scenario with real people that the model hasn't seen. Depending on what we're trying to achieve, we might be looking for a commercial outcome, or an increase in some sort of engagement metric, or something like that. This is basically a final step, which is kind of try a model out in the real world" (AI developer, 15).*

The development process of AI that consists of three main parts – understanding the ML model requirements, managing datasets, and developing and retraining ML models – along with techniques to improve ML models and mitigate cognitive biases are explained in the above sections. As one of the objectives of this study is understanding *"How HR managers and AI developers' collaboration can mitigate cognitive biases in developing AI-Recruitment Systems (AIRS)"*, the following section explains some points based on informants' comments to address this research question.

4.2.4 How can HR managers and AI developers collaborate to mitigate cognitive biases in developing AIRS?

According to the findings, HR managers and AI developers can communicate and collaborate in different phases of developing AIRS to mitigate cognitive biases (Table 4-2). One of the AI developers stated that biases should not be considered only as a technical issue. Without communication and collaboration between AI developers and domain experts, AIRS biases cannot be mitigated:

"Bias is a really complex concept, and it cannot be solved technically. It has to be solved through communication and collaboration, it needs different stakeholders coming together and having just general conversations so that they can pick up on these things. It shouldn't be solved by a bunch of machine learning AI researchers sitting in an office somewhere, most of whom are male" (AI developer, 1).

4.2.4.1 Understanding requirements of each job position

AI developers discussed that understanding the essential criteria of each job position leads to developing less biased AIRS. AI developers need to engage with HR managers and ask good questions to understand job position requirements. In addition, HR managers are responsible for articulating each job position's requirements properly. An AI developer, who has a Ph.D. in AI and ML and has experience working with AIRS development companies, proposed that understanding the criteria for choosing the right candidates is the first and most important phase of developing ML models:

"For the recruitment, it's all about first being educated about what it really means. So when you talk about job requirements and knowing what characteristics make that person the right person, it eases the pain on the AI developers on how they need to model those" (AI developer, 14).

The AI developer participants pointed out that AI developers need to develop their skills to ask the right questions of HR managers to find job position requirements:

"The developer needs to understand what questions to ask, and AI developers need to work on this skill" (AI developer, 1).

The participant further explained that the knowledge of employees who are working in the relevant job positions in the company helps developers understand the positions better to set the relevant parameters into their ML models:

"When we think about developing AI systems for HR within a particular business culture, we would expect that you would get input from stakeholders who are actually in that role within that company, and that is the only way we can make this work" (AI developer, 1).

Similarly, another AI developer discussed that asking the right questions helps AI developers figure out the "attributes of each job position" as the model variables:

"AI developers should ask the right questions from HR, like, which are going to be the attributes or parameters in each job position" (AI developer, 10).

According to AI developers, HR managers are responsible for helping AI developers understand the job position requirements. As quoted previously, the statement below points to the importance of HR managers' role to define the requirements at the beginning of the AIRS development process.

"I think HR managers should understand that AI is not a magic thing, it is something developed by someone who is prone to errors. So, they really need to set the expectations right in the first place" (AI developer, 10).

Likewise, another AI developer explained that HR managers can point out some details that AI developers might need to know about the job positions.

"Also, HR people should help developers by thinking of important points, for example, what little gotchas are there that maybe the machine learning person isn't aware of? For example, AI developers need to actually channel the functions and the specifics of what employers are doing in the role" (AI developer, 1).

One AI developer believed that to find out the "objective features" for choosing the best candidates, it was necessary to ask questions of many AIRS users and make conclusions based on common answers:

"We have to see how we can ask more questions to have more objective features and we need to get enough data from users. If I get ten users' opinions, maybe everyone has some different opinions and preferences ready to go. But if I can sell thousands of user input, maybe I can learn common knowledge common entailed, and then this is a good reason, this is a good, not a good thing. If you have a large enough database, then this problem will be eliminated to some extent" (AI developer, 13).

However, according to HR managers, articulating job requirements can be challenging, and HR managers claimed that one of the reasons that the hiring decisions might be biased is their inappropriate assumptions about job positions. According to HR managers, inappropriate assumptions happen due to a weak understanding of the industry and functioning of a job, inadequate understanding of required soft skills for a job position, and only looking for perfect candidates in terms of technical skills.

One of the HR managers pointed to the fact that HR managers sometimes do not know each job position's required skills and characteristics. Without conducting a detailed job analysis, the essential characteristics are vague, and the right applicants cannot be found quickly.

> *"People do not necessarily understand what they exactly want; people retrofitting a role around rather than considering what the role is and what they need first. To get your applicants quicker, you really need to understand what it is you're looking for? What skills? What age? What stage is going to be relevant for this" (HR manager, 7).*

In addition, a senior manager with twenty years of HR experience explained that HR managers often identify the skills based on the job description, which might be different from the actual requirements.

> *"Very often, teams are not even aware of the work they do. So, they hire for a defined role and not for a gap. They say 'we need another tester' instead of 'we need someone who's going to do a lot of the communications and upward reporting or spreading the information about the quality of the product" (HR manager, 1).*

Similarly,

> *"Sometimes, we are trying to recruit for a skill that we don't have the understanding of what we actually need and the person that is hired have a very little awareness of what the actual job needs" (HR manager, 5).*

As well as understanding the requirements of each job position, HR managers and AI developers can collaborate in managing datasets to make datasets unbiased for training AIRS.

4.2.4.2 Managing datasets

According to AI developers, HR managers can contribute to managing datasets before training algorithms. Managing datasets refers to labelling datasets and organising representative and accurate datasets to train algorithms that are not biased towards/against a particular group. The statement below reveals HR managers' role in helping ML models learn from meaningful datasets by adding unique and informative "labels" to datasets.

> *"HR managers can collaborate and do the annotation of data, making datasets to be supervised. So how you actually clean that dataset and make it something that is good enough, that you don't have two things that represent the same thing or don't have*

something completely wrong. So human experts can at least curate data labelling" (AI developer, 7).

Another AI developer discussed that data labelling is challenging as creating accurately labelled training data is difficult. Data labelling needs to be accurate enough to classify personalities that can be used for training AI.

> *"At the end of the day, AI is based on a supervised model. So, we have some labels that either a human has labelled manually or an individual has self-rated by taking tests such as the IBM test. In the case of personality, the way it's done is we had 40-45,000 candidates take our tests in the past, where they answered both the free text questions intuitions and a standard personality test. Then, we have the ground truth that offers automatic data labelling, and then we can build a machine learning model that is able to infer the self-rated personality from the text" (AI developer, 2).*

One AI developer, who was the founder of an AI recruitment company, explained that preparing diverse and representative datasets is not always easy. However, he believed that experts in the HR field, both practitioners and academics, could help AI developers to create representative datasets by targeting various populations:

> *"It's hard sometimes for us as individual builders of technology to effectively curate all of the datasets ourselves. We need research partners; we need academic partners who can help us do that. Having academics as research partners help us curate datasets that cover different ages and genders and demographics and education levels" (AI developer, 5).*

Moreover, the statement below highlights that HR managers can help AI developers to have accurate datasets to train AI by generating their real datasets under controlled conditions. This way, they reduce the need to clean data and use customers' data directly to train AI.

> *"As the data provided by customers is just a mess, we don't know exactly what it means or where it's coming from, it's kind of a hodgepodge of different things. So, working closely with customers and communicating on that is important. What we ended up building as a newer product are these that we use our own data and really controlled the data gathering at a level that we never could before when customers had to give it to us. Thousands of people answering the same question or a very*

similar question about teamwork. Then, we hired a panel of evaluators and trained them on kind of what 1, 2, 3, 4 or 5 rating on this question would mean and that data actually was amazing" (AI developer, 4).

HR managers' and AI developers' communication and collaboration can have an influence on the model training and evaluation by applying AIRS to the recruitment decisions and assessing the ML models with regard to biases.

4.2.4.3 Model training and evaluation

According to AI developers, HR managers can help AI developers to train and assess the ML models and find out biases based on the prediction outcomes. HR managers need to be committed to using AIRS for a long time and generating a complete record for each hiring decision. In addition, AI developers need to validate ML models with some tools and frameworks such as "Watson open scale", "IBM 360", and "Aequitas". If there are any biases in datasets, those biased instances should be removed, and parameters should be weighted differently.

One of the AI developers pointed out a way that HR managers could help validate algorithms. He explained his company's strategy of having HR managers evaluate the AIRS performance within a three to six months test to assess the prediction outcomes:

"Whenever we start a project with a customer, we do a pilot that usually runs within three to six months, depending on the number of roles they're hiring and how large the organisation is. So, in the pilot, candidates take our test, and they are assessed in the regular way that the organisation assesses the system. Then, we can see the correlation between our scores and the outcomes. Then, there's a discussion around that: does the outcomes make sense? If all parties agree the predictions are well aligned, then we go live with the product" (AI developer, 2).

However, another AI developer discussed that it is important that HR managers are committed to using AIRS for a long time and have complete records of each hiring decision to help AI developers improve their ML models.

"Every company is going to begin with a small dataset as a baseline and then gets better over time. That's why we want outcome data and a longer-term commitment from HR managers. The outcome data includes every piece of every hiring decision

and every outcome that occurs gives more data to learn from the back into AI" (AI developer, 5).

Moreover, AI developers explained their strategies for evaluating ML models, such as validation tests, using tools or frameworks, and tuning ML models. One AI developer asserted that in order to validate ML models and check biases, datasets should be checked first. Then, biased data points should be removed. He also believed that using tools such as IBM Watson Open Scale can assist AI developers in making changes in biased datasets by providing some advice:

> *"I feel like if you have a dataset and it is biased, and your AI model is showing bias, it's your responsibility to get it fixed. If you build an AI model, and that AI model ends up having some bias, you probably should go and look at your data and see why does it have it, where's this bias coming from: maybe it's the data that you've chosen. Either you should remove it, or try to add filters weighted differently. We also have some tools at IBM, like one tool is Watson open scale to figure out biases and get my model fixed. Then, it also provides you some tips on like, trying to modify this data" (AI developer, 9).*

He further explained that it is necessary to test the ML model with the same data points and a new set of data when removing the biased data points.

> *"I think if you remove the biased data points, you have to go and measure it and test it before you can make a claim it is removed. So if you remove some of the data that might be causing the skew, or you add additional data to neutralise that skew run your model, again, test it on that same dataset that initially gave you the skew as well as some new datasets. That's probably the best metric for to judge if your model actually removed the bias" (AI developer, 9).*

However, AI developer 7 believed that it is difficult to hide some data from algorithms to mitigate biases. He gives an example of how algorithms can find hidden data from other correlated data.

> *"It's not easy to hide information from an algorithm because it's very sneaky and sophisticated about the way it finds that information even when you hide it. For example, we have a CV in Israel, and in the Israeli army, the military service is four*

> *years for men and two years for women. So even if you hide the name or gender, it might learn that the duration of your role in the army just reflects your gender, or it's correlated to your gender. It suggests what's your gender, even if we are not showing the gender to the algorithms" (AI developer, 7).*

Another AI developer agreed about using frameworks such as “Aequitas” and “IBM 360” which define fairness well and allow developers to check ML models' errors.

> *"Fairness is still like an evolving field. So, one of the frameworks we have found is the Aequitas framework from the University of Chicago. Also, IBM has a framework called IBM 360. We studied both of them; they are very similar in what they propose. So, we've implemented that approach. With regard to fairness, we follow the University of Chicago's framework. They have very good definitions of fairness" (AI developer, 2).*

Furthermore, another technique is somewhat a trial-and-error technique in that developers change some parameters and run the algorithm on the same datasets again to analyse ML models' performance.

> *"There will be an aspect of the kind of tuning the AI engine that is being able to not consider certain parameters. For example, if I am looking for CIO from an insurance company, I want the AI engine to not consider or not even process the gender of that CIO, and it is only analysing the background and experience. So, any person, men or women that can meet certain criteria would be fit for that world" (AI developer, 8).*

Likewise,

> *You can review all the assumptions and all the candidates if you'd like to and even before letting it search for candidates, there's always this process of tuning the model (AI developer, 12).*

Table 4-2 summarises how AI developers and HR managers can collaborate to mitigate cognitive biases in the development process of AIRS.

Table 4-2 HR managers and AI developers' collaboration to mitigate cognitive biases in the development process of AIRS

Mitigating cognitive biases in the development process of AIRS	HR managers' roles	AI developers' roles	Reasons for the collaboration
Understanding the requirements of each job positions	• Articulating job position requirements in detail • Having a good understanding of the industry and functioning of the job position • Having a good understanding of required soft skills	• Engaging with HR managers and employees • Asking good questions • Choosing as many objective features as possible for ML by asking AIRS users	• Choosing the right candidate depends on having a good understanding of the essential criteria of each job position
Managing datasets	• Labelling datasets • Organising representative and accurate datasets • Generating datasets by using AI under a controlled situation	• Asking domain experts such as HR practitioners and academics for data collection	• Training ML algorithms is the process of learning patterns from training datasets and making datasets unbiased is an important step.
Model training and evaluation	• Being committed to using AIRS and having a complete record of each hiring decision • Giving feedback	• Validating the ML models with tools and frameworks • Removing biased data points and weighting parameters differently • Tuning algorithms	• Finding biases on the prediction outcomes, retraining the ML models and making the models unbiased

4.3 Summary

This chapter presented the findings that addressed research question: "which cognitive biases are more likely to be observed in recruitment and selection decisions" and "how can cognitive biases be mitigated in developing AI-Recruitment Systems?". The emerging conceptual categories and sub-core categories along with representative informants' comments were presented.

The findings show that there are two major cognitive biases in R&S – 'similar-to-me' and 'stereotype' bias – that were very likely to be embedded in AIRS. This chapter illustrated the core category of the development process of unbiased AIRS. The AIRS development process includes three phases: understanding the ML model requirements, managing datasets, and developing and retraining ML models. Further findings were reported on HR managers and AI developers' collaboration to mitigate cognitive biases in developing AIRS. The findings will be discussed in more detail in the following chapter. The next chapter discusses the development process of AIRS (by explaining each stage in detail) and mitigating cognitive biases in developing AIRS.

Chapter 5 Discussion

This chapter discusses the findings from Chapter 4. Chapter 4 described common cognitive biases in the recruitment and selection process along with the development process of unbiased AI-Recruitment Systems (AIRS) by introducing the conceptual and core categories. The purpose of this chapter is to position the findings within the extant literature. The grounded theory that has been developed in this study provides an explanation and interpretation of the conceptual model (Figure 2-3).

Figure 2-3 shows that cognitive biases influence HR managers and AI developers, resulting in biased AI-Recruitment Systems. Also, this study aims at explaining approaches to mitigate cognitive biases in the development process of AI-Recruitment Systems (AIRS). Thus, analysing empirical data using grounded theory, common cognitive biases that might happen in the R&S process and consequently exist in datasets for training ML algorithms, as well as techniques to mitigate AIRS cognitive biases, are explored. The chapter is structured as follows. The first part provides a brief introduction. The second part of the chapter presents the process model of developing unbiased AIRS. The last section summarises the chapter.

5.1 Mitigating cognitive biases in developing AIRS

Known cognitive biases in the recruitment and selection process are stereotypes related to different categories such as gender, age, and ethnicity (Table 2-10). Molenberghs and Louis (2018) explained that people from a wide range of groups, such as those who come from different races, nations, cultures, or ethnicities, as well as supporters of different sports teams are biased towards intergroup people.

Likewise, Whysall (2017) stated that in-group bias refers to similarity bias and affinity bias in that it is an inherent tendency to feel a greater sense of affinity with and trust in people who are similar to ourselves. For example, studies show that similar-to-me bias is demonstrated when interviewers select candidates based on their own perceptions of themselves that include values, habits, beliefs, demographics, and cultural characteristics (Prewett-Livingston, Veres, Field, & Lewis, 1996; Bagues & Perez-Villadoniga, 2012).

The findings of this study demonstrated two common cognitive biases – ‘Similar-to-me’ and ‘Stereotype’ bias – in recruitment decisions that can make AIRS biased since historical recruitment decisions are used as datasets to train ML models for the R&S process. ‘Similar-to-me’ bias is defined in the literature as hiring managers’ tendency to favourably evaluate candidates who have similar biographical backgrounds, attitudes, and perceived personalities to themselves (Anderson & Shackleton, 1990). ‘Stereotype’ bias is “a fixed, over-generalized belief about a particular group or class of people” (Cardwell, 2013, p. 227).

According to the literature there are two sources of biases in developing AI: biased datasets and algorithmic bias. The two biases that were identified in the findings were related to biased datasets, and not algorithmic biases. Datta, Tschantz, and Datta (2015) explained that any predictive algorithmic decision-making tool developed based on historical data may have historical biases. Likewise, Köchling and Wehner (2020) argued that biases might pre-exist and not be noticeable before developing AI for HR recruitment and development. However, algorithms replicate these pre-existing biases from the past decision (Köchling & Wehner, 2020).

The other source of bias, algorithmic bias, may happen due to poorly designed AI systems (Lattimore et al., 2020), such as failure to formulate users' assumptions objectively or using inaccurate selection criteria (Tambe, Cappelli, & Yakubovich, 2019). For example, suppose specific age ranges, genders, and ethnicities are used in algorithms. The algorithms will find the relationship between the chosen attributes and the target, leading to generating biased algorithmic outcomes (Saifee, 2020).

The biases in the relationship between the attributes and the outcome are not always obvious like the above example, and biases might happen due to the proxy attributes. Johnson (2020) believed that relying on proxy attributes is one of the reasons for algorithmic bias. Proxy attributes are “seemingly innocuous attributes that correlate with socially-sensitive attributes, serving as proxies for the socially-sensitive attributes themselves” (Johnson, 2020, p. 1).

Understanding the development process of unbiased AIRS based on the findings of this study enables controlling the sources of cognitive biases and developing AIRS that do not have the ‘Similar-to-me’ and ‘Stereotype’ biases. Thus, unbiased AIRS can assist HR managers in mitigating the two known biases in recruitment decisions. In the following sections, the development process of unbiased AIRS and strategies to mitigate cognitive biases are discussed.

5.2 The development process of unbiased AI-Recruitment Systems (AIRS)

The development process of unbiased AIRS has emerged from the field data and the researcher continuously investigated the literature and consulted with her supervisors regarding the categorisation of the codes until the sub-core categories and core category emerged. The process model of developing unbiased AIRS is a multi-phased and iterative process. The model (Figure 5-1) represents the three phases of developing AIRS: understanding the ML requirements, managing datasets, and developing and retraining ML models.

The model represents how human (i.e., both HR managers and AI developers) cognitive biases can arise in each phase of developing AIRS which may result in biased AI. Moreover, the model determines the collaboration of AI developers and HR managers in each phase to mitigate cognitive biases in developing AIRS. The findings also offered three strategies: HR managers' and AI developers' collaboration, diversity in the AI development team, and techniques to manage ML features such as the F-score technique and A/B test that need to be considered in all phases of the AIRS development process to mitigate cognitive biases.

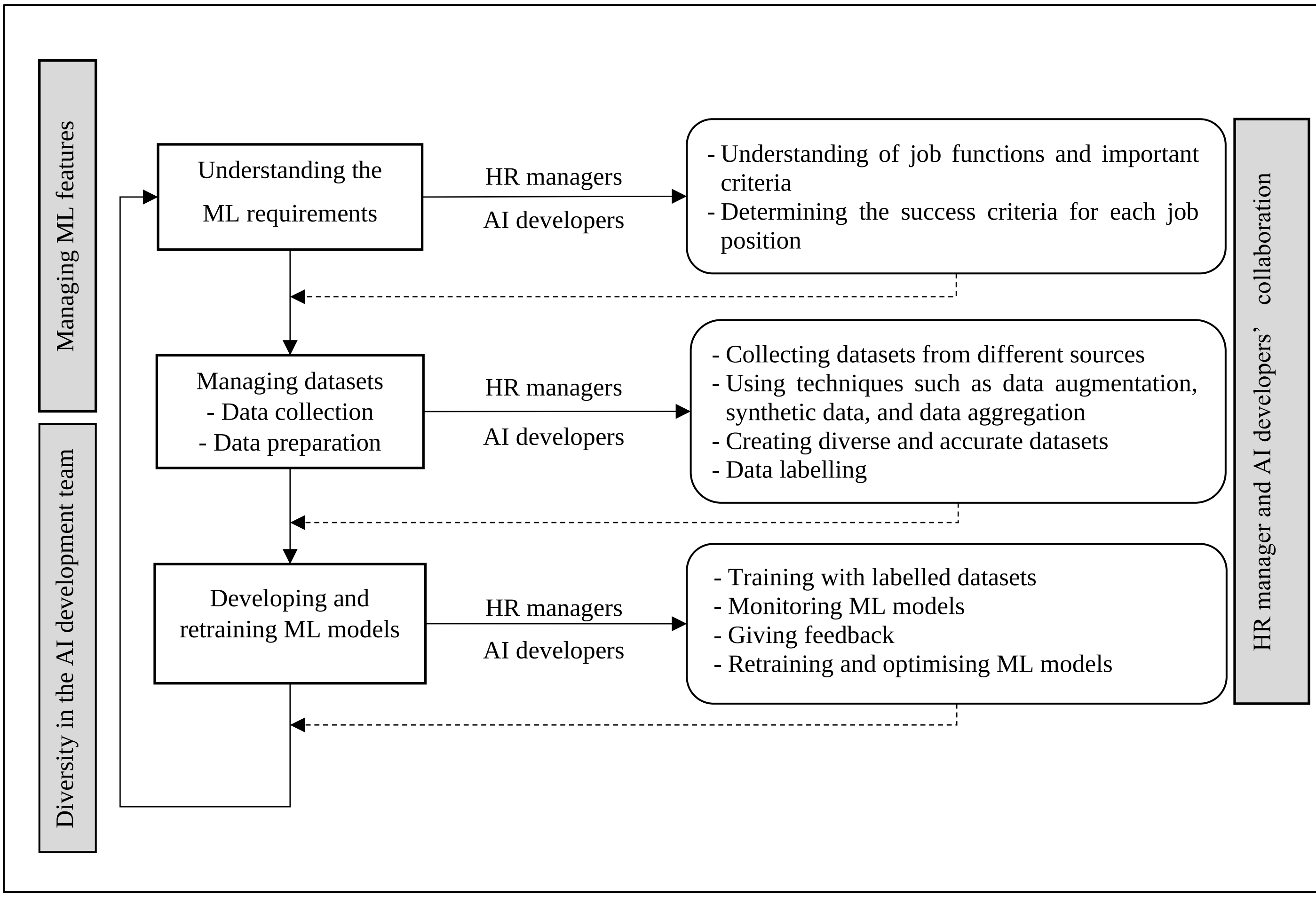

Figure 5-1 The development process of unbiased AIRS and cognitive bias mitigation techniques

The model illustrates that the development process of AIRS starts with understanding the ML model's requirements, which is the main phase of developing software systems in general. However, the actual development and training process of ML models begins with managing datasets and training ML models with them. The development process of ML-based software systems is different from the traditional software systems as the logic in AI development is not explicitly programmed (Lwakatare, Crnkovic, & Bosch, 2020).

Explicit programming is defined as a "human-executable procedure" (p. 2417) that entails the steps of retrieving and interpreting information from the real world to do a programming task (LaToza, Arab, Loksa, & Ko, 2019). However, AI automatically creates logic by learning from data (Rhem, 2020) and data has a key role in the development process of ML-based software systems (Baer & Kamalnath, 2017).

Different process models that describe the lifecycle of AI, such as the Cross-Industry Standard Process for Data Mining (CRISP-DM) (Schröer, Kruse, & Gómezb, 2021), and the Team Data Science Process (TDSP) (Microsoft, 2020) have been explained in section 2.4.4

(The development process of AI). Moreover, it has been discussed that some researchers have worked on adjusting the Machine Learning (ML) lifecycle to particular contexts, such as developing AI for healthcare (Hwang, Kesselheim, & Vokinger; 2019) and fintech (Haakman, Cruz, Huijgens, & Deursen, 2020). The findings of this study described the agile development process of unbiased AI in recruitment and selection that encompasses three main stages: understanding the ML model requirements, managing datasets, and developing and retraining ML models.

The findings of this study are more focused on selecting candidates in the R&S process through analysing CVs, résumés, and references, algorithmic videos, and chatbot analysis. Moreover, the researcher inferred that the ML models in the R&S process are based on supervised learning and/or semi-supervised learning. Supervised learning requires labelled data to train ML models, whereas when using semi-supervised learning, it is not necessary to train ML models with labelling every single training example. The findings show that AI developers mentioned training ML models for the R&S process with labelled data and unlabelled data (i.e. the ML models find patterns from unlabelled datasets). Table 5-1 outlines the activities that are performed in each phase of developing AIRS.

Table 5-1: Activities in each stage of developing AIRS

Phases	Activities
Understanding the ML model requirements	This phase involves understanding managers' assumptions and requirements for each job position to learn the prerequisites before beginning to train ML models and determining the success criteria and weighting them for each job position.
Managing datasets	Managing datasets encompasses two main activities: data collection and data preparation. Data collection is the process of finding relevant and accurate training datasets, and data preparation refers to having sufficient, diverse, and accurate datasets.
Developing and retraining ML models	Developing the ML models in the R&S context entails determining the success factors for choosing the best candidates and training the ML models with datasets. The training process is not a single and finite phase, and models need to be retrained to keep the model accurate as the variables in the R&S process evolve, leading to degrading the ML model's accuracy.

5.2.1 Understanding the ML model requirements

According to the findings, understanding the ML model requirements depends on understanding the important criteria for each job position. The job position criteria are features of ML models that are weighted to predict the best candidates. Job requirements are based on the job position success criteria that need to be determined clearly by HR managers, and ML algorithms can compare the defined criteria with candidates' information such as information from their CVs, interviews, and tests (Allal-Chérif, Aránega, & Sánchez, 2021).

The findings suggested that for defining the job position criteria as features of ML models and weighting them, communication between HR managers and AI developers is essential. AI developers' engagement helps HR managers to understand what ML is capable and not capable of doing in the R&S process. With the HR managers' assistance, AI developers can define job position criteria objectively by transforming the required criteria into measurable features.

Pessach et al. (2020) emphasise that developing ML models by asking HR managers to imitate their decision-making might not be the best approach, as HR managers' decisions are highly subjective and inaccurate due to hiring biases. Instead, objective measurements of the actual success of employee recruitment, such as post-hire prediction of turnover (Ribes, Touahri, & Perthame, 2017) and past employee performance should be considered (Kirimi & Moturi, 2016).

However, employees' requirements in terms of having a diverse team or properly balancing the workforce among different departments should be taken into account by considering successful recruitment from an organisational regulatory viewpoint, candidate viewpoint, and job allocation viewpoint (Pessach et al., 2020).

The importance of AI developers and HR managers' communication was also supported by Nalchigar, Yu, and Keshavjee (2021). They noted that field experts and stakeholders might not be clear about AI limitations and capabilities, and AI developers should provide an understanding of AI's potential in the R&S process. Wan, Xia, Lo, and Murphy (2020) also believe that AI developers need to recognise the areas of the business that can be modelled and benefit from ML.

Aizenberg and van den Hoven (2020) pointed out that design requirements are both social and technical (hence the term socio-technical), reflecting the interaction between human and technical artefacts and processes. This interaction includes not only the interactions between humans and technical artefacts, but also, and more importantly, the human-to-human interactions. Thus, based on business priorities and available training datasets, HR managers can contribute to developing AIRS by providing an insight into business priorities and requirements.

5.2.2 Managing datasets

The results of this study demonstrate that one of the main steps in developing AIRS is managing datasets, including data collection and data preparation. Data collection is gathering the right datasets, and data preparation is pre-processing the datasets to have sufficient, diverse and accurate datasets for training ML models.

There are various descriptions of dataset management in developing AI. Wan, Xia, Lo, and Murphy (2020) described managing datasets as ‘data curation’, which encompasses data collection from different sources, data pre-processing, training, validating, and testing datasets. Roh, Heo, and Whang (2021) defined this phase as the data collection process in general, which consists of data acquisition, data labelling, and improvement of existing data or models. However, according to the findings, this study considered data collection and data preparation as sub-processes of the dataset management phase.

5.2.2.1 Data collection

ML models, in general, function properly and can be more effective when they are trained by large datasets (Wang & Perez, 2017). However, the findings show that there are insufficient datasets in the R&S process. Pessach et al. (2020) believed that the challenge of the deficiency of data for training ML models not only exists in the R&S process but also in other areas in HR which have the same problem.

According to the findings, AI developers should find different sources of data, such as collecting data from candidates who have applied for job positions through job advertising platforms, and cannot only rely on datasets from HR managers. Likewise, Raghavan and Barocas (2019) pointed out that datasets from HR managers are inadequate for training ML algorithms. In addition to the need for large training datasets, Edwards and Rodriguez (2019) mentioned that when AI developers use only the data that organisations already have instead of looking for more data, algorithms can be affected by anchoring bias.

The findings also revealed that datasets from HR managers are often incomplete and biased, and as a result the datasets do not provide support for ML features. Strohmeier and Piazza (2013) mentioned that the available datasets are missing important data points as the data of candidates who are not recruited is not recorded by HR managers. Likewise, Raghavan and Barocas (2019) discussed that candidates' data included in the datasets is based on employers' choices and the criteria that represent a good employee in their past hires.

In addition to collecting data from different sources, there are some techniques that help AI developers to get enough datasets. The data preparation process that leads to having sufficient, diverse and accurate datasets is explained in the next section.

5.2.2.2 Data preparation

According to the findings, data preparation refers to techniques for having enough, diverse, and accurate datasets. Below, these techniques are explained in detail.

5.2.2.2.1 Enough datasets

To develop AIRS, AI developers can use some techniques such as data augmentation, synthetic data, and aggregating datasets to obtain sufficient data and improve the quality of ML models. Polyzotis, Roy, Whang, and Zinkevich (2017) define data augmentation to enlarge datasets as either enriching the existing features with new datasets or using the same datasets but conducting various transformations. For example, embedding a new word for text data or duplicating images through shifting, zooming in/out, rotating, flipping, distorting, and shading are samples of data augmentation (Wang & Perez, 2017).

Data augmentation is one way to generate synthetic data (Nowruzi, Kapoor, Kolhatkar, Hassanat, Laganiere, & Rebut, 2019). Two different methods can be used to create synthetic data: process-driven methods and data-driven methods. Process-driven methods are computational/mathematical models such as numerical simulations, Monte Carlo simulations, agent-based modelling, and discrete-event simulations. Data-driven methods are methods that "derive synthetic data from generative models that have been trained on observed data" (Nowruzi et al., 2019, p. 2).

The findings suggest that AI developers can also use another technique, data aggregation, to gather large amounts of data. The data aggregation technique collects data from multiple and various data sources to create relevant results. According to the literature, after data aggregation, data needs to be turned into meaningful and valuable information through a group of technical and business processes called data integration (IBM Corporation, 2016).

5.2.2.2.2 Diverse datasets

According to the findings, having diverse and representative datasets leads to ML algorithms finding more patterns and correlations. Sug (2018) stated that training ML algorithms with more diverse values in datasets results in better ML model performance. It is assumed that the training sets have the same proportion of representative instances; however, there are domains that have imbalanced datasets. Imbalanced datasets is the problem where fewer instances represent some classes while other classes have a large number of representative instances (Borovicka, Jirina, Kordik, & Jiri, 2012).

The findings show that one of the reasons that the R&S process is among the domains where datasets are not representative and diverse enough is inequality in working opportunities for different groups such as women and people of colour for leadership positions. Similarly, an example of imbalanced datasets for leadership positions was shown in research done by Lattimore et al. (2020). They discuss the inequality in leadership positions as there are fewer employment opportunities for women in leadership roles because women have family and caring responsibilities.

5.2.2.2.3 Accurate datasets

Accurate datasets refer to the datasets that have no missing values. However, the findings uncovered that in the R&S process, the final result of the hiring decision, which seems important for AI, is often not available. Köchling and Wehner (2020) noted that the absence of particular data points in datasets leads to representation bias. An example of representative bias is when males are overrepresented in the training datasets, and the algorithms' outcome is biased towards male candidates. According to the findings, one method of generating accurate datasets is when HR managers use AI in recruitment decisions and create datasets under a controlled situation.

5.2.3 Developing and retraining ML models

The findings show that ML algorithms for the R&S process are trained with data from labelled datasets that consist of a set of features, and/or the ML model learns to predict the label for new and unseen datasets. For example, Teodorescu, Morse, Awwad, and Kane

(2021) explained that developers must completely retrain the model instead of just adjusting the part of the code that lead to unfair results.

ML algorithms in the selection process should be able to predict and rank the best candidates for each job position based on the defined success criteria. Based on the above explanations, it is inferred that the findings of this study show the method of developing ML models for the R&S process on the basis of supervised learning algorithms and semi-supervised learning.

Wan, Xia, Lo, and Murphy (2020) defined ML model training as the step in the development process of AI where ML practitioners choose, train, and tune ML models based on the chosen features. Features are extracted and selected through the process of feature engineering. Feature engineering comprises activities for converting the input data to an easier and interpretable form. Next, the current model is tuned to identify potential issues and adjust parameters. Finally, the AI developers team evaluates the model's output with the test datasets (Wan et al., 2020).

The findings point to the importance of monitoring ML models continuously and testing different modelling approaches to find the "optimised" ML model. In addition, the field experts' feedback can improve the quality of ML algorithms. Likewise, researchers have asserted that the model performance and input data need to be monitored continuously to detect errors and whether the ML model needs to be changed over time (Wan et al., 2020). Thus, organisations need to monitor the model's decisions, react when unfairness is discovered, and, most importantly, provide on-going feedback to the system to help it achieve globally optimal fairness solutions (Teodorescu, Morse, Awwad, & Kane, 2021).

The ML models in the R&S process might change due to the rapid changes in employment patterns that have led to significant changes in the careers landscape in the last few decades (Xin, Zhou, Li, & Tang, 2020). To find the best possible prediction model, ML algorithms require retraining, and the central part of the ML process is training iterations (Lwakatare, Crnkovic, & Bosch, 2020).

Bogen (2019) emphasised monitoring ML algorithms to detect the adverse impact of the ML algorithms on the selection phase in case any latent bias exists or a new bias emerges. However, before checking the ML models to detect biases in the monitoring phase, checking the quality of input data is necessary. The input data are usually historical datasets, and algorithms learn from their examples (Barocas & Selbst, 2016; Danks & London, 2017). The main reason for biases in ML algorithms is the quality of input data (Köchling & Wehner, 2020).

5.2.4 Techniques to mitigate cognitive biases in all three phases of developing AI

The model (Figure 5-1) shows that there are other techniques such as collaboration between HR managers and AI developers, diversity in the AI development team, and managing ML features that need to be considered in all three phases of developing AI. In the following sections, each strategy is discussed in detail.

5.2.4.1 HR managers' and AI developers' collaboration

The findings of this study demonstrate the need for collaboration of HR managers and AI developers to mitigate biases in the development process of AI-Recruitment Systems (AIRS). HR managers' expertise that is accumulated through learning and experience within the recruitment and selection domain is complementary to AI developers' expertise in mitigating cognitive biases. HR managers and AI developers can collaborate on understanding the important criteria for each job position, preparing datasets for training ML models, testing ML models, and giving feedback, retraining, and improving ML models based on managers' feedback.

Understanding important criteria for each job position refers to requirements gathering in software development before starting the development process. Using inappropriate inputs based on managers' assumptions about required criteria for managerial decision-making (Tambe et al., 2019) and incorrect weighted criteria (Shrestha et al., 2019) might lead to developing biased AIRS.

In software development, identifying business requirements and understanding end-users is important (Courage & Baxter, 2004). In developing AIRS, HR managers' knowledge and expertise can assist AI developers in understanding the requirements. The findings show that as well as collaboration between AI developers and HR managers to find out the success criteria of each job position, AI developers need to engage with employees who are working in the same or similar roles and are familiar with the job functions and required criteria.

Moreover, HR managers' and AI developers' collaboration on datasets management leads to collecting sufficient, diverse and accurate datasets. HR managers can assist AI developers in labelling data for training AI based on their technical knowledge. Likewise, Manyika, Bughin, Silberg, and Gumbel (2018) pointed out that field experts' knowledge and effort are required to label datasets for training AI. Wick, Panda, and Tristan (2019) mentioned that ML models might mimic unfair patterns if they lack human intuition and understanding of how the "ground truth" labels were biased.

In addition, HR managers' feedback can assist AI developers in detecting biases, retraining ML models, and making their ML models better. In general, to develop software systems, applying users' feedback is vital and considered as part of knowledge sharing in software development (Williams, 2011). There are obstacles and facilitators such as contacting users, motivating users, facilitating and mediating meetings and offering points of focus for user contributions (Wilson, Bekker, Johnson, & Johnson, 1997). A limited amount of time for the first phase of the software development (i.e. understanding requirements) resulted in various decisions that hindered the involvement of users (Wilson et al., 1997). Gulliksen et al. (2003) also mention that software developers' resistance to iteratively evaluating and developing solutions in the software development process is an obstacle to user involvement in developing software. Users might not be motivated to get involved in the development process of software as they do not know how they can contribute and are not confident enough due to a lack of technical background. Facilitating meetings is another essential component of fostering user involvement in developing software. Lack of meetings often forces software users to agree with the software developers too quickly (Wilson et al., 1997).

In addition to collaboration between HR managers and AI developers, diversity in the AI development team can mitigate cognitive biases in developing AIRS, as explained in the following section.

5.2.4.2 Diversity in the AI development team

According to the findings, diversity in the AI development team is a non-technical strategy that helps to mitigate cognitive biases in developing AIRS. Mitigating cognitive biases in software development results in developing high-quality software. Diversity is considered a key element in developing a high-quality software system as diverse team members have broad perspectives and can be extremely innovative (Long, 2018). Hewlett, Marshall, and Sherbin (2013) categorised diversity into two groups: inherent and acquired. Inherent diversity refers to the attributes such as gender, ethnicity, and sexual orientation. Acquired diversity is defined as the team members' technical background and experience such as working in another country.

The findings show that having a diverse development team can mitigate cognitive biases since each team member has their own technical background, personal experience, perspectives, and cultural background to share. However, various studies have shown that there is a lack of diversity in the tech industry, for example, the under-representation of women and minority groups (Alegria, 2016; Parasurama, Ghose, & Ipeirotis, 2021).

Other than the diversity in the AI development team, managing ML features is considered as a technical strategy to mitigate cognitive biases in datasets and to develop and retrain ML models. In the below section, managing ML features is explained in detail.

5.2.4.3 Managing ML features

Cai, Luo, Wang, and Yang (2018) explained feature selection as an effective method that removes irrelevant and redundant features. Feature selection increases the ML model accuracy, provides a better understanding of the ML model and data, and operates faster (Hall, 1999). Cai, Luo, Wang, and Yang (2018) presented a framework that categorises the feature selection method in the development process of AI (Figure 5-2). This figure shows that the feature selection method can be used in all three kinds of ML models (supervised, unsupervised, and semi-supervised) and helps to detect and remove redundant and irrelevant features from training datasets and ML models.

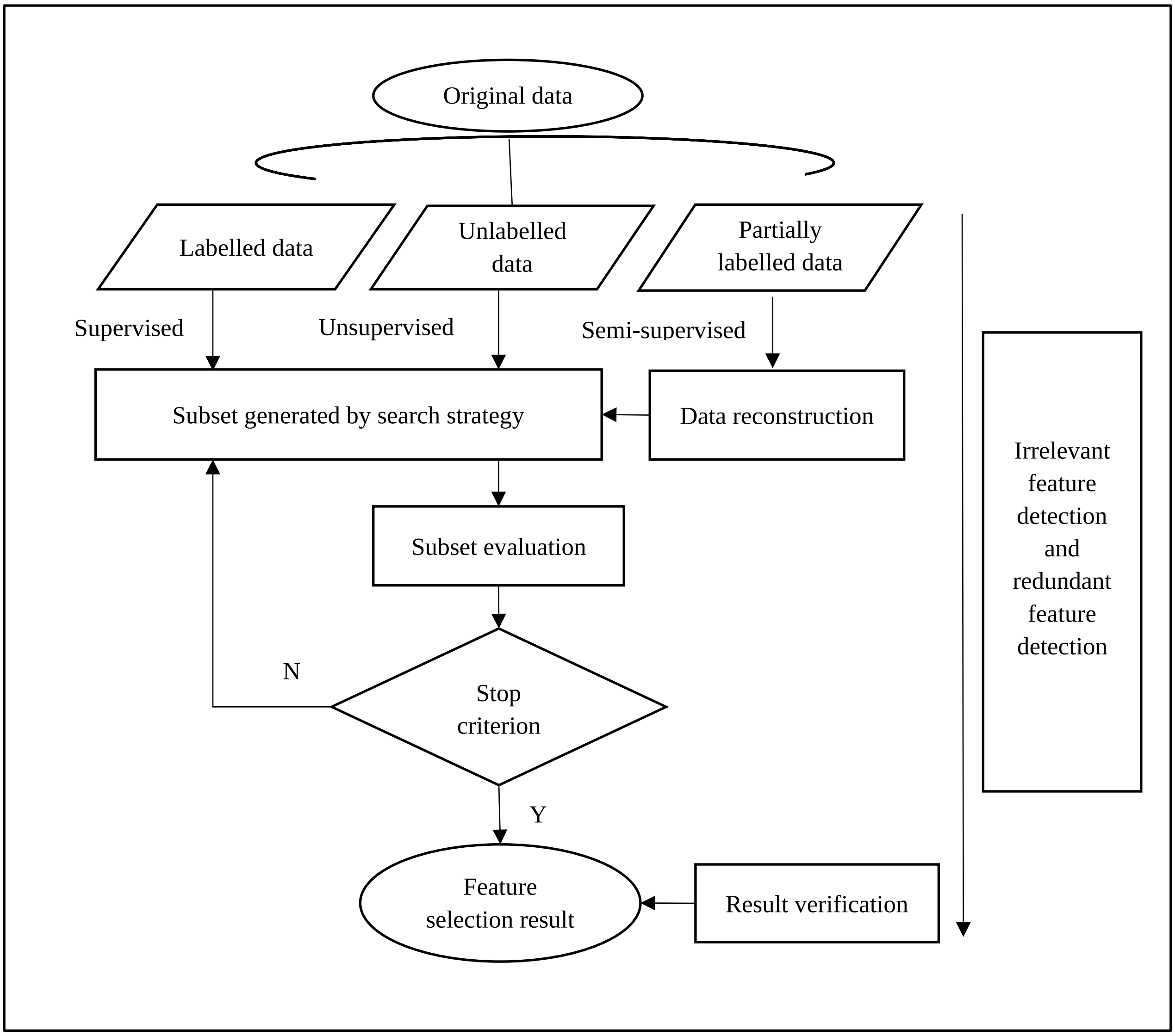

Figure 5-2 A framework for feature selection (Cai et al., 2018, p. 71)

There are three general classes of feature selection algorithms: filter methods, wrapper methods, and embedded methods. Filter methods pre-process and provide a ranking order of features by using a relevant index such as correlation coefficients or classical statistical tests (T-test, F-test, Chi-squared) to keep only the best features (Amaral, Lopes, Jansen, Faria, & Melo, 2012; Hall, 1999). Wrapper methods are efficient ways to find the best predictor, and the predictor is wrapped in a search algorithm that finds a subset to give the best predictor performance. Embedded methods select features in the training process without splitting the datasets into training and testing sets, and the feature selection result provides outputs automatically while the training process is finished (Guyon & Elisseeff, 2003).

Although managing ML features has been mentioned as a strategy to mitigate cognitive biases based on the findings of this study, there is some doubt that biases can be mitigated using feature selection techniques. Johnson (2020) mentioned that biases, which happen due to the proxy problem, cannot be mitigated by applying filtering techniques. For example, using a zip code is considered as a proxy that leads to discriminatory effects.

5.2.5 Summary

This chapter introduced the development process model of unbiased AI-Recruitment Systems (AIRS) and discussed the development process and techniques to mitigate cognitive biases in AIRS based on the findings of the study and the literature. It explained how the collaboration of AI developers and HR managers in each phase of developing AIRS could mitigate cognitive biases. In addition, diversity in the AI development team and managing ML features were discussed as two strategies that should be taken into account in all three phases of the development process of AIRS.

The following chapter first provides an overview of the findings. Next, the implications and contributions of the findings are discussed. Then, the limitations of this study and recommendations for future studies are explained where applicable.

Chapter 6 Implications and Contributions

This dissertation reported a study of cognitive biases in the development process of AI. More specifically, it considered the recruitment and selection process as a decision-making process using AI. This study developed a model which represents the process of developing unbiased AI-Recruitment Systems (AIRS). The research methodology used in this study was grounded theory which guided the collection and analysis of the research data.

The previous chapter introduced the model of developing unbiased AI-Recruitment Systems (AIRS). This chapter begins with a review of the research problem, objectives, and findings. Then it discusses the following implications for researchers: contributions to the theory of information system design, contribution to biases in developing AI and mitigation techniques, contribution to knowledge sharing, and AI transparency.

Additionally, the following implications for practitioners are provided: in the HR field, education system, insurance services, and educating human decision makers to make unbiased and fair decisions when using AI. This is followed by a discussion of the limitations of the study and suggesting directions for future research on biased AI. The chapter ends with the concluding statements of the study.

6.1 Research problem and objectives

In the wake of the rapid development of communication and information technologies, businesses have found themselves in new and diverse competitive situations (Akerkar, 2019). AI has the potential to drastically alter business decision-making processes (Walport & Sedwill, 2016). One of the areas in which AI has a significant impact on the decision-making process is recruitment and selection (Upadhyay & Khandelwal, 2018). Even though AI can improve R&S, evidence from industry proves that the way algorithms are developed and the datasets used to train them can lead to bias in AI (Manyika, Bughin, Silberg, & Gumbel, 2018).

Cognitive biases have been broadly studied in psychology and behavioural economics (Barnes, 1984, Simon, 1997), the business context (Roberto, 2002; Bohmer, Edmondson, & Roberto, 2004), and in the recruitment and selection process (Blair-Loy, Rogers, Glaser, Wong, Abraham, & Cosman; 2017; Woolf & Dixon, 2017). Moreover, studies have shown that cognitive biases are one of the biggest challenges in developing AI-software systems, and that identifying the biases and understanding how to mitigate them in the design and use of AI in business processes is crucial (Barocas & Selbst, 2016; Martin, 2018; Kaplan & Haenlein, 2019; Shrestha, Ben-Menahem, & Von Krogh, 2019; Tambe, Cappelli, & Yakubovich, 2019).

However, there is a lack of empirical studies that investigate cognitive biases in developing AI in a business context. Accordingly, this exploratory study aimed at identifying which biases are more likely to be observed in the recruitment decisions and how cognitive biases can be mitigated in developing AI-Recruitment Systems. To address the research questions, the classic grounded theory was adopted to investigate experts' (i.e. HR managers and AI developers) perceptions of cognitive biases incorporated in AIRS and how to mitigate them.

6.2 An overview of the research findings

The findings show that HR managers are consistent about the two common cognitive biases – stereotype bias and similar-to-me bias – that might happen in R&S. As ML models are trained based on historical recruitment decisions, these biases result in bias in AIRS. The HR managers in New Zealand that were interviewed had experiences in AIRS applications for screening CVs and chatbots in customer service positions in contact centres and retail stores. Based on the findings, the development process of unbiased AI-Recruitment Systems (AIRS) was provided which presents an iterative-lifecycle framework adapted to AI systems along with techniques to mitigate cognitive biases. The model includes three phases: understanding the ML model requirements, managing datasets, and developing and retraining ML models.

Understanding the ML model requirements: ML models are predicted on an understanding of the business case which is choosing the best candidate in the R&S process. Thus, data and ML models must be applied to the business case, resulting in a ML model that addresses choosing the best candidate. The features of ML models that are weighted to predict the best candidate are based on the job position success criteria that should be defined by HR managers. Defining the job position success criteria leads to identifying the ML model requirements.

Then, the ML algorithms can be used to compare the defined criteria with a candidate's information regarding their CV, interview, and test scores (Allal-Chérif, Aránega, & Sánchez 2021).

Managing datasets: Data is needed to develop the ML model and this phase includes data collection and preparation. The focus is on data-centric activities such as labelling datasets and augmentation necessary to make datasets to be used for training ML models. However, it is difficult to come up with fully fleshed-out datasets that are sufficient, diverse, and accurate the first time and datasets need to be improved over time.

Developing and retraining ML models: The actual development of ML models starts with determining the true success criteria of job positions for machine learning algorithms after collecting data. Then, a set of input features in a training dataset is mapped to a set of input features in the output dataset based on the defined success factors (i.e., class target) so that ML algorithms find a pattern within a dataset and start learning.

The findings also point to both technical and non-technical strategies to mitigate cognitive biases in developing AIRS. The non-technical strategies include collaboration between AI developers and HR managers and having a diverse AI development team. HR managers' collaboration assists in choosing the right candidate which depends on having a good understanding of the essential criteria of each job position. Training ML algorithms is the process of learning patterns from training datasets so making datasets unbiased is an important step, by finding biases in the prediction outcomes, retraining the ML models and making the models unbiased.

Moreover, the same as in a traditional software development team, having diverse team members in the AI team can result in contributing to developing a high-quality system (Long, 2018). As a technical strategy to mitigate biases in developing AIRS, the findings suggest ML model feature selection. Feature selection increases the model accuracy by modifying the feature set used to describe all instances (Cardie, 2009).

This study has significant theoretical and practical contributions. In the following section, the theoretical contributions and practical implications are discussed.

6.3 Theoretical contributions

This study contributes to the body of knowledge by extending theoretical contributions in information systems design theory, biases in developing AI and mitigation techniques, advancing AI transparency, and knowledge sharing concepts.

6.3.1 Contribution to the theory of information system design

System design theory as outlined by Walls, Widmeyer, and Sawy (1992) specifies how the process of design can be carried out so as to achieve both effectiveness and feasibility. In the same way that scientific theories are hypothesised, design processes create the artefacts they describe and are therefore only able to be verified by implementing the hypotheses. To the extent that the design incorporates the principles of the theory, a design can be examined in light of scientific theory (Walls et al., 1992).

Information system theory is based on Dubin's (1978) concept of theory building as well as Simon's (1976) idea of a science of the artificial. According to Dubin (1978), a theory generally explains and predicts a phenomenon which includes seven components: (1) units whose interactions are the topic of study, (2) laws determining how those units interact, (3) boundaries within which the hypothesis is supposed to hold, (4) system states where the units interact differently, (5) propositions or truth statements concerning the theory (laws can be considered as propositions), (6) empirical indicators of the terms used in the propositions, and (7) testable hypotheses with empirical indicators.

As both a noun and a verb, "design" refers to a process as well as a product (Walls et al., 1992), so it is used in a dual sense. The design theory, therefore, consists of two components: one dealing with the product and one with the process. Walls et al. (2015) also noted that since the design process leads to a final product, these aspects cannot be independent. As a product, design is considered a plan of what is going to be done or produced (Walls et al., 1992). In terms of a process, design is understood as planning and proportioning the components of a machine or structure in such a way that all requirements are satisfied (Walls et al., 1992).

Walls, Widmeyer, and Sawy (1992) proposed that design theories that focus on the output of design begin with a set of meta-requirements that define the objective classes that the theory applies to. In the case of design theories, meta-requirements are used instead of merely requirements since they address a class of problems instead of just one particular instance.

Second, a meta-design outlines a class of artefacts that are deemed to satisfy the meta-requirements. For example, a design theory is not concerned with the design of the payroll system for any corporation but it considers all transaction processing systems as a class of artefacts. The third component is a set of theories from the natural sciences and social sciences that govern design requirements. The final component is a group of testable hypotheses, which can be used to determine whether the meta-design fulfils the meta-requirements (Walls et al., 1992).

The design theory also focuses on the process of design as well as the design theory that relates to the product. The design process constitutes three elements: a design method, a set of kernel theories, and a set of testable design processes. The steps for making an artefact are defined as a design method. During the design process, kernel theories – which are fundamental to natural or social sciences – guide the process. Kernel theories in the design process might be different from theories related to the design product. Finally, a set of testable hypotheses for the design process gives insight into whether the method produces a design meeting the meta-design (Walls et al., 1992; Walls, Widmeyer, & Sawy, 2015).

The design theory framework proposed by Walls et al. (1992) was extended by Gregor and Jones (2007) as they believed that Walls and his colleagues failed to capture the range of ideas presented by Dubin (1978) and Simon (1981) or reported in other significant related works. First, the Walls et al. (1992) specification does not contain two of Dubin's mandatory theory components: "units" and "system states". Second, it was not explicitly addressed that it is necessary to specify a theory for methodologies rather than a theory for products. Third, their formulation was complex due to the fact that kernel theories for the designing product and designing process had to be separated (Gregor & Jones, 2007). Furthermore, Walls, Widmeyer, and Elsawy (2004) expressed concerns about their approach to displaying theory components being too complex for practical application.

As a result, the new design theory framework was improved by incorporating constructs, artefact mutability, and expository instantiation. Gregor and Jones (2007) stated that a comprehensive description of a design theory for designing an information system should include eight elements: (1) the purpose and scope, (2) the constructs, (3) the principles of form and function, (4) the artefact mutability, (5) testable propositions, (6) justificatory knowledge, (7) principles of implementation, and (8) expository instantiation. The components of an information system design theory (ISDT) are summarised in Table 6-1.

Table 6-1 Comparison of design theory approaches (Gregor & Jones, 2007, p. 28)

Proposed anatomical skeleton	Dubin (1978)	Walls et al. (1992)
Purpose and scope	Boundaries	Meta-requirements
Constructs	Units	
Principles of form and function	Laws of interaction	Meta-description
Artefact Mutability	System states	
Testable propositions	Propositions	Product hypotheses Process hypotheses
Justificatory knowledge		Product kernel theories Process kernel theories

Gregor and Jones (2007) argue that theorising about IS artefacts should include a notion of "mutability” as a fundamental component of IS design theory. However, mutability can be viewed from different perspectives (Gregor & Jones, 2007). The key difference between different perspectives of artefact mutability is the purposeful design of adaptable artefacts (in-design) and how artefacts evolve (in-use) over time (Pöppelbuß & Goeken, 2015).

Mutability-in-design refers to the concept of changing and adapting artefacts in order to fit different organisational contexts. Mutability-in-use reflects the notion that IT artefacts are not immutable outcomes of design processes, but are inherently dynamic (Pöppelbuß & Goeken, 2015). AI is developed by retraining the ML models as part of the development process, which illustrates the mutability of the system. In order to make the algorithm more accurate, AI developers consider the learning process that happens over many training cycles.

Teodorescu, Morse, Awwad, and Kane (2021) pointed out that the differences between ML and traditional IS fundamentally undermine many assumptions of previous generations of IS theory. For example, when unfairness is identified in AI, ML models need to be completely retrained, rather than simply tweaked as developers could do with previous generations of traditional IS systems. The need to completely retrain ML models has implications for different theories and models for software development (Kane et al., 2014).

The development process of unbiased AIRS shows retraining as part of the process that refers to mutability in design theory. Mutability in developing AIRS is the process where AI developers conduct a test and then retrain the ML models based on business use cases. The use of ML models varies across different business use cases, and it is essential to understand how the ML model will be used in production before building an automated pipeline for model retraining (Muthusamy, Slominski, & Ishakian, 2018).

In machine learning, continuous training is an approach that constantly retrains the models to adjust to changes before they are deployed. The retraining of ML models can be based on the feedback data, new training data, or a change to the model definition. The findings of this study (section 4.2.3) also point out that in developing AIRS, HR managers' feedback is vital and considered as an essential part of knowledge sharing in that it can help retrain ML models. Table 6-2 shows how Gregor and Jones' (2007) framework can be expanded by adding mutability as a component that refers to the continuous changing of ML models.

Table 6-2- The contribution of this study to design theory

Gregor and Jones (2007)	The contribution of this study
Purpose and scope	-
Constructs	-
Principles of form and function	-
Artefact mutability	Retraining the pre-developed model based on business use cases
Testable propositions	-
Justificatory knowledge	-
Artefact mutability	Continuously retraining ML models due to feedback data, new training data, or a change to the model definition

6.3.2 Contribution to biases in developing AI and mitigation techniques

The development process of unbiased AIRS can contribute to the literature on biases in the development process of AI and approaches to mitigate AI biases. While research on AI biases and mitigation techniques has been studied in decision-making contexts (Akter et al., 2022; Mehrabi et al., 2021), little is known about biases in AI and mitigation techniques in the context of R&S.

Research categorises biases and suggests techniques to mitigate biases in AI systems (Srinivasan & Chander, 2021; Mehrabi et al., 2021). Srinivasan and Chander (2021) developed a taxonomy of biases in developing AI and determined biases in data creation, problem formulation, data analysis and validation and testing. Mehrabi et al. (2021) proposed technical techniques for unbiasing datasets and algorithms in natural language processing. For example, having datasheets as a supporting document to report dataset creation method, its characteristics, motivations, and its skews assists in mitigating biases in datasets. For algorithms, technical techniques such as making gender-neutral words, using word2vec debiasing techniques, and using the AI fairness 360 toolkit are explained.

The AIRS development model categorises biases based on three main phases of developing AIRS. The model shows that biases are rooted in an inappropriate understanding of important criteria for each job position, inadequate, incomplete and unrepresentative datasets, and infrequently monitoring ML models that can results in biased AI decision-making.

The findings of this study offer techniques such as HR managers and AI developers' collaboration, diversity in the AI development team, and managing ML features to mitigate cognitive biases in developing AIRS. These techniques can mitigate biases related to formulating algorithms by having a good understanding of job position requirements to find relevant ML features, collecting enough and representative datasets, labelling and annotating datasets precisely, and monitoring AIRS constantly to detect errors and retrain the model (The techniques have been discussed in detail in section 5.2.4).

6.3.3 Contribution to knowledge sharing

The development process of unbiased AIRS can contribute to knowledge sharing and consider the AI development process as a knowledge-intensive process in which AI developers and experts (i.e. HR managers) with different backgrounds and cognitive styles learn from each other. While there is a comprehensive literature on knowledge sharing in software development (Ghobadi, 2015), relatively little is known about the cross-functional knowledge sharing between HR managers and AI developers leading to mitigating cognitive biases in developing AIRS (Soleimani, Intezari, & Pauleen, 2021).

The AIRS development model might assist both HR managers and AI developers in sharing their knowledge in all stages of the development process of AIRS and make it unbiased. To build ML models based on important criteria for each job position, AI developers need to engage with HR managers and employees who are working in the same or similar roles to be familiar with job functions and the required criteria and weight them.

Research has shown that ML systems fail to take into account the social system surrounding the application (Van den Broek, Sergeeva, & Huysman, 2021). By defining social concepts mathematically, ML systems often fail to recognise the full meaning behind social concepts such as fairness and justice that may be procedural, contextual, or even contestable (Selbst, Boyd, Friedler, Venkatasubramanian, & Vertesi, 2019; Van Den Broek, Sergeeva, & Huysman, 2019). This may lead to ML applications that are ineffective, inaccurate, or even misguided (Amershi, Cakmak, Knox, & Kulesza, 2014). Thus, HR managers and employees need the knowledge to consider the full meaning of social concepts in regard to the R&S process.

During the process of developing AIRS, HR managers assist AI developers in labelling training datasets based on their technical knowledge and expertise in the field. Manyika, Bughin, Silberg, and Gumbel (2018) also argue that human involvement is required to label datasets for ML training. Moreover, it is important for developing any software systems to incorporate users' feedback into the development process (Williams, 2011). Thus, incorporating HR managers' feedback into the AIRS development process as part of knowledge sharing leads to retraining ML models and makes them better.

6.3.4 Contribution to AI transparency

The findings of this study offer a contribution to the AI transparency literature. Roovers (2019) explained that there are technical steps for creating transparent AI, such as checking the technical correctness of the ML model, explaining how developers approached the problem and assessing the outcomes of the model in terms of biases.

AI models cannot identify biases in datasets, and only humans can discover biased results in the outcome of the ML model as humans understand the context in which the data has been collected (Roovers, 2019). This study identifies two known biases – similar-to-me bias and stereotype bias – that can exist in R&S datasets. Thus, the findings of this study can assist in assessing the outcomes of the model considering these two biases in R&S.

Moreover, Van Nuenen, Ferrer, Such, and Cote (2020) mention that the goal of AI transparency is human understanding of how an AI behaves. Additionally, an AI system's transparency refers to its openness and communication of both the data processed by the system as well as the mechanisms used to build it (Van Nuenen, Ferrer, Such, & Cote, 2020).

The development process of unbiased AIRS assists the users of the system (i.e. HR managers) and those who are affected by the AIRS decisions (i.e., job candidates) in having a good understanding of the data collection and preparation process as well as training and retraining the ML models.

In addition, transparency can be considered to be a sub-dimension of accountability and explainability (Strohmeier, 2020). The explainability of the algorithm and its procedure refers to whether the affected individuals can understand and explain the algorithmic decision. Responsibility is an important aspect of accountability, as it pertains to whether the responsibility is diffused, unclear, or can be assigned to known individuals or institutions (Strohmeier, 2020).

Teodorescu, Morse, Awwad, and Kane (2021) argued that both AI developers and managers should be responsible in regard to AI fairness. They believed that AI developers should be provided with clear fairness objectives and incentives to achieve them, and definitions of fairness and priority groups should be specified in advance by managers. They further mentioned that as it is unlikely that perfect fairness can be achieved, managers may prefer to perform well on more modest fairness goals than to fail to achieve more comprehensive ones.

The findings of this study clearly show that both HR managers and AI developers are responsible for developing unbiased AIRS and indicate their responsibilities in each phase of the process (Table 4-2). When HR managers are committed to using AIRS and having a complete record of each recruitment decision, they will have a good understanding of how AI reaches a conclusion. Thus, AI developers can manage the ML models by specifying the reasons for the issues and tuning the ML model to better fit the needs.

6.4 Practical implications

The findings of this study have implications for practitioners (i.e. developers and users) who are involved in the development process of AI to mitigate cognitive biases. First, it provides insights for HR managers by giving them sufficient information about how AI is developed as well as guiding HR managers to realise how they can contribute to mitigating biases in developing AIRS.

Second, practitioners from other fields/industries where AI is increasingly applied, for instance the education system and insurance services, can use the AIRS process to mitigate cognitive biases at each stage of the development process. The findings highlight the importance of close collaboration between AI developers and AI users at each stage of developing AI to prevent biases.

6.4.1 Implications for the HR field

The findings of this study contribute to knowledge around HR and AI developers' activities and collaboration in the development process of unbiased AIRS. As a predictive algorithmic decision-making tool developed with historical datasets, AI may have historical cognitive biases that might not be visible before development (Datta, Tschantz, & Datta, 2015; Köchling & Wehner, 2020). The findings of this study help both HR managers and AI developers to be aware of two known cognitive biases – the 'similar-to-me' and 'stereotype' bias – in the R&S process. Using domain experts' knowledge (i.e. HR managers and AI developers) and understanding how algorithms work lead to changing undesired AI outcomes.

> *"AI is basically learning about the past and the results might be undesired in terms of the biased solutions and are not aligned with what society now wants. If we're trying to change the past, we need to see beyond those steps. If we have the combination of HR knowledge, AI experts, and an understanding of how algorithms work, it can be very useful to provide those insights" (AI developer, 7).*

Moreover, the findings show how both AI developers and HR managers are responsible for developing less biased AI, and their collaboration can mitigate cognitive biases in the development process of AIRS in practice. Tambe, Cappelli, and Yakubovich (2019) argued that ML models might be biased as they are based on managers' assumptions about required criteria for managerial decision-making. Examining criteria separately and finding out how the criteria are weighted is difficult for humans (Shrestha, Ben-Menahem, Von Krogh, 2019). When HR managers understand the requirements of each job position and how criteria should be weighted, they can contribute to developing less biased AIRS by helping AI developers define job position criteria objectively.

The findings imply that there might not be sufficient data pertaining to a given job in a given organisation and for developing AIRS more generic datasets have to be used. AI developers can collaborate with HR managers from various organisations to use AIRS and prepare sufficient training datasets that assist AI developers in retraining AIRS and developing objective and context specific AIRS.

In addition to improvements to R&S, HR managers can also integrate AIRS with other HR functions such as performance management (Buck & Morrow, 2018) and employee retention (Johnson, Stone, & Lukaszewski, 2020) to test the AIRS efficacy and evaluate key indicators of employee success. This can further help them to reduce the workload, and secure HR resources to work on more strategic business goals.

6.4.2 Education system

Biases in education have been studied as a prevalent and pernicious phenomenon (Hodson, Dovidio, & Gaertner, 2002; Régner, Thinus-Blanc, Netter, Schmader, & Huguet, 2019). For example, black applicants have been judged as unsuitable for admission compared to identical white applicants (Hodson et al., 2002), scientific fields are unconsciously associated with men (Régner et al., 2019), and college students believe that male students are smarter than female students (Cooper, Krieg, & Brownell, 2018).

Even though the focus of this study is on the HR field, the findings can be extended to other fields, including education (Kim, Lee, & Cho, 2022; Kokku et al., 2018). The findings of this study can contribute to improving education conditions for disadvantaged groups by applying the development process of unbiased AI to education systems to provide an equal opportunity for all well-qualified students. Using unbiased AI assists admission officers in augmenting their decision-making process based on data-driven assessment. When education systems embrace diversity, there will be more diverse education leaders, which leads to improving diversity and inclusion in the workplace.

6.4.3 Insurance services

According to the McKinsey consultancy company, AI applications are being used to automate various tasks such as underwriting services, and it is expected that these tasks will be fully automated by 2030 (Balasubramanian, Libarikian, & McElhaney, 2021). IBM suggests that taking initial steps such as understanding the current decision management, recognising data depreciates and determining new sources of data is required to augment the capabilities of decision-making in insurance companies by using AI (Ramchandani & Anderson, 2018). However, there has not yet been any research to identify the barriers to developing AI for insurance decision-making.

This empirical study sheds light on understanding some challenges of developing AI in HR that can be applied to the insurance industry. The challenges include data labelling and lack of accurate training datasets in addition to solutions to these challenges. In addition, this study provides an insight into how field experts (experts from insurance companies) can collaborate with AI developers in developing AI. Understanding the challenges of developing AI and techniques to solve them can accelerate the development process of AI in other contexts which will save time for AI developers to increase the accuracy of ML models.

6.4.4 Educating human decision makers to make unbiased and fair decisions when using AI

The development process of unbiased AIRS provides a learning opportunity for human decision makers so that they understand ML tools' general functionality and their potential limitations (Kane et al., 2019). The decision maker can only critically analyse recommendations for potential biases and unfairness if they understand how the ML model may be systematically unfair.

If managers have a thorough understanding of these issues, they can begin to support the kind of oversight that is necessary for achieving or improving augmented approaches to fairness. This understanding will assist human decision makers in balancing performance outcomes with fairness outcomes in any given application of machine learning (Teodorescu et al., 2021). The findings of this study provide an understanding of existing biases in R&S datasets that might assist HR managers in talking about biases in their datasets without hesitation.

> *"It's a little bit tricky when you're in kind of a business relationship to say, we want to look at how biased you guys are before and after using our product, because a lot of companies are a little hesitant to have that be public or shared" (AI developer, 4).*

6.5 Limitations of the study

This study is subject to some limitations that can provide opportunities for future research. The first limitation is the limited generalisability of the findings due to the qualitative nature of the study (Walsham, 1995). This study identifies two biases that are very likely to be embedded in AIRS and the development process of unbiased AIRS might not account for mitigating all biases in developing AIRS. However, the insights from this study can provide useful guidelines for future research design and objectives for mitigating cognitive biases in developing AI. Moreover, AI developers comprised a smaller percentage of the study participants than HR participants because those with expertise in developing AI for recruitment and selection purposes are few in number.

More specifically, the limitation related to the grounded theory methodology is a research bias since the coding and category development have been done by the researcher as a PhD student. For example, two major cognitive biases emerged from the findings while someone else may come up with more cognitive biases if s/he codes the same data or, indeed, interviewed a different sample. However, analysing qualitative data can be improved when the analysis is done by more than one researcher. Thus, the researcher can be reasonably confident that his or her coding is reproducible (J. L. Campbell et al., 2013).

The information provided by AI developers is considered to be limited as AI developers could not give information about some of the approaches that they were applying for developing their product and mitigating biases due to their companies' internal confidentiality policies as well as their unwillingness to talk about the negative aspects of their products. This limitation could not have been avoided; however, the researcher asked the interviewees to check with the product owner of the development team to find out to what extent s/he could give information on their product and the mitigation approaches. Additionally, information from HR managers about AI was not sufficient due to their limited exposure to AIRS. Thus, the researcher interviewed AI developers to better understand the concept of AI in R&S. This might not be a limitation in future studies as the implementation of AI for human resource management, specifically the R&S function, is increasing.

The imbalanced number of male and female participants among both AI developers and HR managers is considered as a limitation. In the HR field, the number of women is larger than that of men. In contrast, in software engineering and AI, men outnumber women. It is important to note that this study is not the first in the field of information systems to have low female participation rates (Thomas & Bostrom, 2010).

This study examines AI applications in general, such as scanning resumes and chatbots. Since using AI in the R&S process is rather new, the number of AI developers who are experts in developing AI for different areas in the HR field, specifically for the R&S process, is quite low around the globe. Thus, the unbiased development process in this study does not specify an AI application in the R&S process. In addition, future research can focus on each of the AI applications in the R&S process separately and expand the model for every single AI application.

6.6 Future research recommendations

The findings of this study identify important areas for future research. The development process of unbiased AI is embedded in the context of the R&S process. This model can be developed and applied to other business contexts such as logistics or marketing that are prone to algorithmic biases. Furthermore, future research projects can test the findings as propositions and hypotheses in different HR contexts or other industrial fields.

Moreover, there could be a co-design exercise (i.e. longitudinal/case-study research) with HR managers and AI developers to examine the relationship in depth and test or further develop the model. This co-design approach would help to validate solutions to mitigate cognitive biases in each phase of the development process and make decisions through collaboration in an iterative process.

Even though men are over-represented in the AI field all around the world, balancing the number of female and male AI developers could be considered in future studies by targeting countries with more female graduates in the field.

Moreover, Kordzadeh and Ghasemaghaei (2021) mentioned that it is still unknown how interactions between individuals and algorithmic systems might activate or inhibit data-driven biases in organisational decision-making. The findings of this study point to the HR managers’ interaction with AI systems that can lead to mitigating biases in developing AI for recruitment decisions by giving feedback to AI developers. Future studies can test the effectiveness of HR managers’ feedback to mitigate cognitive biases in AIRS.

6.7 The PhD Journey Reflection

Throughout the past few years during my PhD study, some of the most memorable experiences came from side trips that took me out of my comfort zone and changed me for good. I learnt to be more resilient in finding the answer to my questions and the challenges I faced both in academia and my own personal life. Studying cognitive biases in AI is a rather new field and one would expect to have difficulties at first in adopting and applying a suitable research methodology for its research questions. The most challenging part of my study was the data collection. Due to the limited usage of AIRS in New Zealand there is a lack of understanding about developing AI in R&S. To overcome this issue, I collected data from AI developers globally. Finding AI developers was difficult as there are not many AI developers who are experts in developing AI for R&S. Moreover, AI developers were more reticent about sharing information about their AI applications and their techniques to mitigate biases in developing AIRS because of the Intellectual Property (IP). Thus, I had to conduct more interviews to gain more information about AI and mitigation techniques.

The Covid pandemic also affected my PhD study as it happened at the beginning of my data collection stage. All my appointments with interviewees were cancelled at once, and the coordination of interviews became more challenging. As a novice researcher, this was my first experience of conducting semi-structured interviews and collecting data virtually was difficult at first compared to in-person interviews. As I gained more experience in interviewing through Zoom, I became more comfortable when listening to and engaging participants. Nevertheless, these challenges in my PhD journey had valuable lessons that are applicable to my future professional career and personal life.

Although doing a PhD is largely an independent undertaking, it is a valuable opportunity to work as part of a research group, learn to listen to others, and make your own suggestions to the group without being afraid of making inappropriate decisions. I learnt there is not a single right answer to each challenge, and I need to push through negative emotions and keep myself motivated to find my way to the best answer and outcome possible. This long and unique journey has been a transforming experience that affected me.

6.8 Summary

This study adopted an exploratory research design to address two research questions: (1) Which biases are more likely to be observed in recruitment decisions? and (2) How can the cognitive biases be mitigated in developing AI-Recruitment Systems? A total of twenty-one HR managers and fourteen AI developers across the globe, including New Zealand, Australia, the United States, Germany, Israel, and India, were interviewed in English. The data from semi-structured interviews were analysed using the classic grounded theory (Glaser & Strauss, 1967). Based on the findings a development process model of unbiased AI for the recruitment and selection process was created.

The model shows the three phases of developing AIRS: understanding the machine learning requirements, managing datasets, and developing and retraining the machine learning models. The model determines how the collaboration of AI developers and HR managers should be structured in each phase as well as offering some technical and non-technical approaches to mitigate cognitive biases during the development of AIRS. The theoretical contributions and practical implications of the study were discussed. The limitations of the study as well as directions for future research were explained.

References

Ahmed, O. (2018). Artificial Intelligence in HR. *International Journal of Research and Analytical Reviews*, *5*(4), 971–978.

Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data and Society*, *7*(2). https://doi.org/10.1177/2053951720949566

Akerkar, R. (2013). *Big Data Computing* (R. Akerkar (ed.)). CRC, Taylor & Francis Group.

Akerkar, R. (2019). *Artificial Intelligence for Business*. Springer Nature Switzerland AG.

Akter, S., Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. *Journal of Business Research*, *144*(March 2021), 201–216. https://doi.org/10.1016/j.jbusres.2022.01.083

Alammar, F., Intezari, A., Cardow, A., & Pauleen, D. (2019). Grounded Theory in Practice: Novice Researchers' Choice Between Straussian and Glaserian. *Journal of Management Inquiry*, *28*(2), 228–245. https://doi.org/10.1177/1056492618770743

Albert, E. T. (2019). AI in talent acquisition: a review of AI-applications used in recruitment and selection. *Strategic HR Review*, *18*(5), 215–221. https://doi.org/10.1108/shr-04-2019-0024

Alegria, S. N. (2016). *A Mixed Methods Analysis of the Intersections of Gender, Race, and Migration in the High-Tech Workforce* (Issue July) [University of Massachusetts Amherst]. http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1681&context=dissertations_2

Allal-Chérif, O., Yela Aránega, A., & Castaño Sánchez, R. (2021). Intelligent recruitment: How to identify, select, and retain talents from around the world using artificial intelligence. *Technological Forecasting and Social Change*, *169*(January). https://doi.org/10.1016/j.techfore.2021.120822

Amaral, J. L. M., Lopes, A. J., Jansen, J. M., Faria, A. C. D., & Melo, P. L. (2012). Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease. *Computer Methods and Programs in Biomedicine*, *105*(3), 183–193. https://doi.org/10.1016/j.cmpb.2011.09.009

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35*(4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513

Anderson, N., & Shackleton, V. (1990). Decision making in the graduate selection interview: A field study. *Journal of Occupational Psychology*, *63*(1), 63–76.

Archibald, M., Ambagtsheer, R., Casey, M., & Lawless, M. (2019). Using Zoom

Videoconferencing for Qualitative Data Collection: Perceptions and Experiences of Researchers and Participants. *International Journal of Qualitative Methods*, *18*, 1–8. https://doi.org/10.1177/1609406919874596

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*, 124–140. https://doi.org/10.1016/0749-5978(85)90049-4

Baddeley, A. D. (1992). Working Memory. *Science*, *255*(ii), 556–559. https://doi.org/10.4249/scholarpedia.3015

Baer, T., & Kamalnath, V. (2017). *Controlling machine-learning algorithms and their biases*. McKinsey & Company. https://www.mckinsey.com/business-functions/risk/our-insights/controlling-machine-learning-algorithms-and-their-biases

Bagues, M., & Perez-Villadoniga, M. (2012). Do recruiters prefer applicants with similar skills? Evidence from a randomized natural experiment. *Journal of Economic Behavior and Organization*, *82*(1), 12–20. https://doi.org/10.1016/j.jebo.2011.12.004

Balasubramanian, R., Libarikian, A., & McElhaney, D. (2021, March). Insurance 2030—The impact of AI on the future of insurance. *McKinsey & Company*. https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance

Banerjee, S., Kumar Singh, P., & Bajpai, J. (2015). Nature Inspired Computing. *Advances in Intelligent Systems and Computing*, 203–210.

Barnard, C. (1968). *The functions of the executive*. Harvard University Press.

Barnes, J. (1984). Cognitive Biases and Their Impact on Strategic Planning. *Strategic Management Journal*, *5*(2), 129–137.

Barocas, S., & Selbst, A. (2016). Big Data ’ S Disparate Impact. *Law Rev*, 671–732.

Bas, A. (2012). Strategic HR Management: Strategy Facilitation Process by HR. *Procedia - Social and Behavioral Sciences*, *58*(2006), 313–321. https://doi.org/10.1016/j.sbspro.2012.09.1006

Bazerman, M. H., & Moore, D. A. (2009). *Judgment in managerial decision making* (7th ed.). John Wiley & Sons.

Bazerman, M. H., & Moore, D. A. (2013). *Judgement in Managerial Decision making*. Wiley.

Bazerman, M. H., & Sezer, O. (2016). Bounded awareness: Implications for ethical decision making. *Organizational Behavior and Human Decision Processes*, *136*, 95–105. https://doi.org/10.1016/j.obhdp.2015.11.004

Beach, L. R., & Connolly, T. (2005). *The psychology of decision making: People in organizations* (2nd ed.). Sage Publications.

Bell, D., Raiffa, H., & Tversky, A. (1988). *Decision Making: Descriptive, Normative, and Prescriptive Interactions*. Cambridge University Press.

Bell, E., Bryman, A., & Harley, B. (2018). *Business Research Methods*. Oxford University Press.

Bellman, R. (1978). *An Introduction to Artificial Intelligence: Can Computers Think?* (6th ed.). Boyd & Fraser Pub. Co.

Benbasat, I., Goldstein, D., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369–386. https://www.jstor.org/stable/248684?seq=1#page_scan_tab_contents

Benjamins, R., Barbado, A., & Sierra, D. (2019). Responsible AI by Design in Practice. *ArXiv Preprint ArXiv:1909.12838, Ec*. http://arxiv.org/abs/1909.12838

Betsch, T., & Held, C. (2012). Rational decision making: Balancing RUN and JUMP modes of analysis. *Mind and Society*, *11*(1), 69–80. https://doi.org/10.1007/s11299-011-0097-2

Bhatt, G., & Zaveri, J. (2002). The enabling role of decision support systems in organizational learning. *Decision Support Systems*, *32*(3), 297–309. https://doi.org/10.1016/S0167-9236(01)00120-8

Bishop, J. M. (2021). Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It. *Frontiers in Psychology*, *11*(January), 1–18. https://doi.org/10.3389/fpsyg.2020.513474

Black, J. S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it? *Business Horizons*, *63*(2), 215–226. https://doi.org/10.1016/j.bushor.2019.12.001

Blair-Loy, M., Rogers, L. E., Glaser, D., Wong, Y. L. A., Abraham, D., & Cosman, P. C. (2017). Gender in engineering departments: Are there gender differences in interruptions of academic job talks? *Social Sciences*, *6*(1), 1–19. https://doi.org/10.3390/socsci6010029

Bloom, P., & Weisberg, D. (2007). Younger Children Often Distort the Scientific Not Linked To the Brain. 06520,. *Science, 316*(May), 996–997.

Blumberg, B., Cooper, D., & Schindler, P. (2011). *Business Research Methods* (3 ed). McGraw-Hill Education.

Blumenthal-Barby, J. S. (2016). Biases and Heuristics in Decision Making and Their Impact on Autonomy. *American Journal of Bioethics*, *16*(5), 5–15. https://doi.org/10.1080/15265161.2016.1159750

Boden, M. (2016). *AI: Its Nature and Future* (1st ed.). Oxford University Press.

Bogen, M. (2019). All the Ways Hiring Algorithms Can Introduce Bias. *Harvard Business Review*, 1–5. https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias

Bohmer, M. ., Edmondson, A. ., & Roberto, M. A. (2004). Columbia’s final mission. *Harvard Business School*.

Bonabeau, E. (2003). Trust Your Gut? Don’t. *Harvard Business Review*, 116–123.

Bonde Thylstrup, N., Flyverbom, M., & Helles, R. (2019). Datafied knowledge production: Introduction to the special theme. *Big Data and Society*, *6*(2), 1–5.

https://doi.org/10.1177/2053951719875985

Booth, C. (1998). Beyond Incommensurability in Strategic Management: A Commentary and an Application. *Organization*, *5*(2).

Boran, S., & Yavuz, E. (2008). A Study On Election Of Personnel Based On Performance Measurement By Using Analytic Network Process (ANP). *Journal of Computer Science*, *8*(4), 333–338.

Borovicka, T., Jirina, M., Kordik, P., & Jiri, M. (2012). Selecting Representative Data Sets. *Advances in Data Mining Knowledge Discovery and Applications*. https://doi.org/10.5772/50787

Bouteska, A., & Regaieg, B. (2018). Loss aversion, overconfidence of investors and their impact on market performance evidence from the US stock markets. *Journal of Economics, Finance and Administrative Science*. https://doi.org/10.1108/jefas-07-2017-0081

Brescoll, V., Uhlmann, E., & Newman, G. (2013). The effects of system-justifying motives on endorsement of essentialist explanations for gender differences. *Journal of Personality and Social Psychology*, *105*(6), 891–908. https://doi.org/10.1037/a0034701

Browne, G. J., & Ramesh, V. (2002). Improving information requirements determination: A cognitive perspective. *Information and Management*, *39*(8), 625–645. https://doi.org/10.1016/S0378-7206(02)00014-9

Buchanan, L., & O’Connell, A. (2006). Leigh Buchanan (2006), A brief History of Decision Making, HBR, January 2006. *Harward Business Review*, *January*. http://www.samuellearning.org/decisionmaking/handout1.pdf%0Ahttps://hbr.org/2006/01/a-brief-history-of-decision-making

Buck, B., & Morrow, J. (2018). AI, performance management and engagement: keeping your best their best. *Strategic HR Review*, *17*(5), 261–262. https://doi.org/10.1108/shr-10-2018-145

Bumblauskas, D., Nold, H., Bumblauskas, P., & Igou, A. (2017). Big data analytics: transforming data to action. *Business Process Management Journal*, *23*(3), 703–720. https://doi.org/10.1108/BPMJ-03-2016-0056

Burck, C. (2005). Comparing qualitative research methodologies for systemic research: The use of grounded theory, discourse analysis and narrative analysis. *Journal of Family Therapy*, *27*(3), 237–262. https://doi.org/10.1111/j.1467-6427.2005.00314.x

Burgelman, R. (1991). Intraorganizational Ecology of Strategy Making and Organizational Adaptation : Theory and Field Research Author (s): Robert A . Burgelman Published by : INFORMS Stable URL : http://www.jstor.org/stable/2634929 REFERENCES Linked references are available. *Organization Science*, *2*(3), 239–262.

Cable, D. M., & Judge, T. A. (1996). Person - Organization fit, job choice decisions, and organizational entry. *Organizational Behavior and Human Decision Processes*, *67*(3), 294–311. https://doi.org/10.1006/obhd.1996.0081

Cachia, M., & Millward, L. (2011). The telephone medium and semi-structured interviews: A complementary fit. *Qualitative Research in Organizations and Management: An International Journal*, *6*(3), 265–277. https://doi.org/10.1108/17465641111188420

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, 70–79. https://doi.org/10.1016/j.neucom.2017.11.077

Callegaro, M. (2011). Social desirability. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods*. SAGE Publications, Inc. https://doi.org/https://dx.doi.org/10.4135/9781412963947

Campbell, C., Sands, S., Ferraro, C., Tsao, H. Y. (Jody), & Mavrommatis, A. (2020). From data to action: How marketers can leverage AI. *Business Horizons*, *63*(2), 227–243. https://doi.org/10.1016/j.bushor.2019.12.002

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods and Research*, *42*(3), 294–320. https://doi.org/10.1177/0049124113500475

Cardie, C. (2009). Feature Selection for Case-Based Learning : A Cognitive Bias Approach. In *Language Learning*.

Cardwell, M. (2013). *Dictionary of Psychology*. Fitzroy Dearborn.

Cavaye, A. (1996). Case study research: a multifaceted research approach for IS. *Information Systems Journal*, *6*(3), 119–130.

Chaffee, E. (1985). Three Models of Strategy. *Academy of Management Review*, *10*(1), 89–98.

Chaiken, S. (1980). Heuristic Versus Systematic Information Processing and the Use of Source Verusu Message Cues in Persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766.

Chalmers, D. (2011). A Computational Foundation for the Study of Cognition. *Journal of Cognitive Science*, *12*(4), 325–359. https://doi.org/10.17791/jcs.2011.12.4.325

Charmaz, K. (1995). Grounded theory. In *Rethinking methods in psychology* (Smith, J., pp. 27–49). Sage.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

Charmaz, K. (2008). Constructionism and the Grounded Theory Method. In *Handbook of Constructionist Research* (pp. 397–416). The Guilford Press.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Sage.

Charniak, D., & McDermott, E. (1985). *Introduction to Artificial Intelligence*. Addison-wesley.

Chen, H., Chiang, R., & Storey, V. (2012). Business Intelligence and Analytics: From Big Data To Big Impact. *MIS Quarterly*, *36*(4), 1165–1188.

https://doi.org/10.2307/41703503

Chen, T. Y., & Huang, J. H. (2013). Application of data mining in a global optimization algorithm. *Advances in Engineering Software*, *66*, 24–33. https://doi.org/10.1016/j.advengsoft.2012.11.019

Chugh, D., Bazerman, M. H., & Banaji, M. R. (2005). Bounded ethicality as a psychological barrier to recognizing conflicts of interest. *Conflicts of Interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy*, 74–95. https://doi.org/10.1017/CBO9780511610332.006

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, *7*, 205031211882292. https://doi.org/10.1177/2050312118822927

Clark, A. (2018). The machine learning audit CRISP-DM Framework. *ISACA Journal*, *1*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85046297653&partnerID=40&md5=3d0f666512933f42377610101717e9a2

Clarke, A. (2005). Pushing and Being Pulled Around the Postmodern Turn. In *Situational Analysis*. SAGE Publications, Inc. https://doi.org/10.4135/9781412985833.n1

Clarke, A. (2008). Sex/gender and race/ethnicity in the legacy of Anselm Strauss. *Studies in Symbolic Interaction*, *32*(08), 161–176. https://doi.org/10.1016/S0163-2396(08)32012-2

Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J. M., & Welton, C. (2009). MAD skills. *Proceedings of the VLDB Endowment*, *2*(2), 1481–1492. https://doi.org/10.14778/1687553.1687576

Collis, J., & Hussey, R. (2013). *Business Research: A Practical Guide for Undergraduate and Postgraduate Students*. Macmillan International Higher Education.

Cooney, A. (2010). Choosing between Glaser and Strauss: an example. *Nurse Researcher*, *17*(4), 18–28.

Cooper, A., & Abrams, E. (2021). Emergent Unfairness in Algorithmic Fairness-Accuracy Trade-Off Research. *AIES 2021 - Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 46–54. https://doi.org/10.1145/3461702.3462519

Cooper, K. M., Krieg, A., & Brownell, S. E. (2018). Who perceives they are smarter? Exploring the influence of student characteristics on student academic self-concept in physiology. *Advances in Physiology Education*, *42*(2), 200–208. https://doi.org/10.1152/advan.00085.2017

Corbin, J., & Strauss, A. (2008). *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory*. SAGE Publications.

Corley, K. G. (2015). A Commentary on “What Grounded Theory Is…”: Engaging a Phenomenon from the Perspective of Those Living it. *Organizational Research Methods*, *18*(4), 600–605. https://doi.org/10.1177/1094428115574747

Cormen, T. (2009). *Introduction to algorithms*. MIT Press.

Courage, C., & Baxter, K. (2004). *Understanding your users : a practical guide to user* *requirements methods, tools, and techniques*. Morgan Kaufmann.

Cowgill, B. (2020). Bias and Productivity in Humans and Machines: Theory and Evidence from Resume Screening. In *Columbia Business School*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3584916

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Creswell, J. (2014). *Research Design Qualitative Quantitative and Mixed Methods Approaches* (4th ed.). SAGE Publications Ltd. https://doi.org/0.1177/2050312117740990

Crotty, M. (1998). *The foundations of social research : meaning and perspective in the research process*. Sage.

Dane, E., Rockmann, K. W., & Pratt, M. G. (2012). When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. *Organizational Behavior and Human Decision Processes*, *119*(2), 187–194. https://doi.org/10.1016/j.obhdp.2012.07.009

Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. *IJCAI International Joint Conference on Artificial Intelligence*, *0*(January), 4691–4697. https://doi.org/10.24963/ijcai.2017/654

Das, T. k., & Teng B.S. (1999). Cognitive biases and strategic decision making processes: An integrative perspective. *Journal of Management Studies*, *36*(6), 757–778.

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies*, *2015*(1), 92–112. https://doi.org/10.1515/popets-2015-0007

Davenport, T. (2006). Competing on Analytics. *Harvard Business Review.*, *84*(1), 98.

Davenport, T., Harris, J., & Morison, R. (2010). Analytics at Work: Smarter Decisions, Better Results. In *Harward Business Review*. Brighton. https://doi.org/10.1002/9781118983836

Davenport, T., & Michelman, P. (2018). *The AI Advantage : How to Put the Artificial Intelligence Revolution to Work*. MIT Press.

Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation? *AI Magazine*, *14*(1), 17–33. https://doi.org/10.1609/AIMAG.V14I1.1029

Dillion, S. M. (2014). Descriptive Decision Making: Comparing Theory with Practice. *33rd Annual Operational Research Society of New Zealand Conference*, *May*, 99–108. http://orsnz.org.nz/conf33/papers/p61.pdf

Dubin, R. (1978). *Theory Building*. Free Press.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D.

(2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *August*, 0–1. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Edwards, J. (1991). Person-job fit: a conceptual integration, literature review, and methodological critique. In C. COOPER & I. ROBERTSON (Eds.), *International review of industrial and organisational psychology* (pp. 283–357). Chichester: John Wiley.

Edwards, J., & Rodriguez, E. (2019). Remedies against bias in analytics systems. *Journal of Business Analytics*, 1–14. https://doi.org/10.1080/2573234x.2019.1633890

Eisenhardt, K. M. ., & Zbaracki, M. J. . (1992). Strategic Decision Making. *Management Journal*, *13*(Special Issue : Fundamental Themes in Strategy Process Research), 17–37.

Ekwoaba, J., Ikeije, U., & Ufoma, N. (2015). The impact of recruitment and selection criteria on organizational performance. *Global Journal of Human Resource Management*, *3*(2), 22–33. https://doi.org/10.1017/CBO9781107415324.004

Evans, E., & Rosengren, K. S. (2018). Cognitive Biases or Cognitive Bridges?: Intuitive Reasoning in Biology. In *In Teaching biology. in schools* (pp. 9–21). Routledge.

Evans, G. (2013). A Novice Researcher 's First Walk Through the Maze of Grounded Theory: Rat ionalizat ion for Classical Grounded Theory. *The Grounded Theory Review*, *12*(1), 37–56.

Evans, J. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*, 451–468.

Evans, M. (2001). Cognitive and Contextual Factors in the Emergence of Diverse Belief Systems: Creation versus Evolution. *Cognitive Psychology*, *42*(3), 217–266. https://doi.org/10.1006/cogp.2001.0749

Ezzamel, M., & Willmott, H. (2004). Rethinking strategy: contemporary perspectives and debates. *European Management Review*, *1*(1), 43–48. https://doi.org/10.1057/palgrave.emr.1500009

Fassinger, R. E. (2005). Paradigms, praxis, problems, and promise: Grounded theory in counseling psychology research. *Journal of Counseling Psychology*, *52*(2), 156–166. https://doi.org/10.1037/0022-0167.52.2.156

Fazelpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, *16*(8).

Feldman, G., Kutscher, L., & Yay, T. (2018). *Omission and commission in judgment and decision making : A review of biases and cognitive asymmetries related to action and inaction* (Issue October). https://doi.org/10.13140/RG.2.2.15958.09289

Felfernig, A. (2014). Biases in decision making. *CEUR Workshop Proceedings*.

Fernandez, W. D., & Lehmann, H. (2011). Case Studies And Grounded Theory Method In Information Systems Research: Issues And Use. *Journal of Information Technology Case and Application Research, 13*(1), 4–15.

https://doi.org/10.1080/15228053.2011.10856199

Flick, U. (2019). Doing grounded theory: Key components, process and elements. *SAGE Publications Ltd*, 17–30. https://doi.org/10.4135/9781529716658.n2

Fox, F., Morris, M., & Rumsey, N. (2007). With Young People : Methodological Reflections. *Qualitative Health Research*, 539–547.

Frank, L., & Hackman, R. (1975). Effects of interviewer-interviewee similarity on interviewer objectivity in college admissions interviews. *Journal of Applied Psychology*, *60*(3), 356–360. https://doi.org/10.1037/h0076610

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347. https://doi.org/10.4324/9781315259697-23

Gable, G. (1994). Integrating Case Study and Survey Research Methods: An Example in Information Systems. *European Journal of Information Systems*, *3*(2). https://doi.org/10.5603/KP.2015.0050

Galotti, K. M. (2002). *Making decisions that matter: How people face important life choices*. Lawrence Erlbaum Associates.

Gardner, W. L., & Avolio, B. J. (2003). Waking up! Mindfulness in the Face of Bandwagons. *Academy of Management Review*, *28*(1), 54–70. https://doi.org/10.2307/259098

Garg, S., Sinha, S., Kar, A. K., & Mani, M. (2021). A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management*. https://doi.org/10.1108/IJPPM-08-2020-0427

Georgiou, K., Gouras, A., & Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, *27*(2), 91–103. https://doi.org/10.1111/ijsa.12240

Ghobadi, S. (2015). What drives knowledge sharing in software development teams: A literature review and classification framework. *Information and Management*, *52*(1), 82–97. https://doi.org/10.1016/j.im.2014.10.008

Gianfrancesco, M., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential Biases in ML Algorithms Using EHR Data. *JAMA Internal Medicine*, *178*(11), 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763.Potential

Gino, F. (2016). What we miss when we judge a decision by the outcome. *Harvard Business Review*, 1–4.

Ginsberg, M. (1993). *Essentials of Artificial Intelligence*. Morgan Kaufmann.

Glaser, B. (1978). *Theoretical Sensitivity: Advances in the methodology of grounded theory*. Sociology Press.

Glaser, B. (1992). *Basics of Grounded Theory Analysis*. Sociology Press.

Glaser, B. (1998). *Doing Grounded Theory: Issues and discussions*. Sociology Press.

Glaser, B. (2002). *Constructivist grounded theory?* Forum: Qualitative Social Research

(Online Journal). https://doi.org/https://doi.org/10.17169/fqs-3.3.825

Glaser, B. (2003). *The grounded theory perspective II: Description's remodeling of grounded theory*. Sociology Press.

Glaser, B. (2005). *The grounded theory perspective III: Theoretical coding*. Sociology Press.

Glaser, B., & Holton, J. (2007). Remodeling Grounded Theory. *Historical Social Research/Historische Sozialforschung. Supplement*, *19*, 47–68.

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

Gorman, E. H. (2005). Gender Stereotypes, Same-Gender Preferences, and Organizational Variation in the Hiring of Women: Evidence from Law Firms. *American Sociological Review*, *70*, 702–728.

Goulding, C. (1998). Grounded theory: the missing methodology on the interpretivist agenda. *Qualitative Market Research: An International Journal*, *1*(1), 50–57.

Goulding, C. (2002). *Grounded theory : a practical guide for management, business and market researchers*. Sage.

Graham, H., & Benett, R. (1995). *Human resources management* (Eighth edi). Pitman publishing.

Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, *8*(5), 312–335. https://doi.org/10.1057/9781137399625

Guinan, P. J., Parise, S., & Rollag, K. (2014). Jumpstarting the use of social technologies in your organization. *Business Horizons*, *57*(3), 337–347. https://doi.org/10.1016/j.bushor.2013.12.005

Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). Key principles for user-centred systems design. *Behaviour and Information Technology*, *22*(6), 397–409. https://doi.org/10.1080/01449290310001624329

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*, 1157–1182.

Haakman, M., Cruz, L., Huijgens, H., & van Deursen, A. (2020). AI lifecycle models need to be revised an exploratory study in fintech. *ArXiv*.

Haley, U. C. V., & Stumpf, S. A. (1989a). Cognitive Trails in Strategic Decision-Making: Linking Theories of Personalities and Cognitions. *Journal of Management Studies*, *26*(5), 477–497. https://doi.org/10.1111/j.1467-6486.1989.tb00740.x

Haley, U. C. V., & Stumpf, S. A. (1989b). *Cognitive trials in strategic deciison-making: linking theories of personalities and cognitions*. *26*(5), 477–497.

Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning* (Issue April). Hamilton, New Zealand.

Hall, W., & Callery, P. (2001). Enhancing the rigor of grounded theory: Incorporating reflexivity and relationality. *Qualitative Health Research*, *11*(2), 257–272. https://doi.org/10.1177/104973201129119082

Hallberg, L. (2006). The "core category" of grounded theory: Making constant comparisons. *International Journal of Qualitative Studies on Health and Well-Being*, *1*(3), 141–148. https://doi.org/10.1080/17482620600858399

Hamilton, R. H., & Davison, H. K. (2018). The search for skills: Knowledge stars and innovation in the hiring process. *Business Horizons*, *61*(3), 409–419. https://doi.org/10.1016/j.bushor.2018.01.006

Hammond, J., Keeney, R., & Raiffa, H. (2006). The hidden traps in decision making. *Harvard Business Review*, *84*(1).

Han, J., & Kamber, M. (2011). Data mining: concepts and techniques. In *Elsevier* (Second). https://doi.org/10.5860/choice.49-3305

Hanna, P. (2012). Using internet technologies (such as Skype) as a research medium: A research note. *Qualitative Research*, *12*(2), 239–242. https://doi.org/10.1177/1468794111426607

Harrington, R. J., & Ottenbacher, M. C. (2009). Decision-making tactics and contextual features: Strategic, tactical and operational implications. *International Journal of Hospitality and Tourism Administration*, *10*(1), 25–43. https://doi.org/10.1080/15256480802557259

Harrison, E. F., & Pelletier, M. A. (1998). Foundations of strategic decision effectiveness. *Management Decision*, *36*(3), 147.

Haugeland, J. (1981). *Mind Design: Philosophy, Psychology, Artificial Intelligence*. The MIT Press.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. MIT Press.

Hermanowicz, J. (2002). The great interview: 25 strategies for studying people in bed. *Qualitative Sociology*, *25*(4), 479–499. https://doi.org/10.1023/A:1021062932081

Hertwig, R., Fanselow, C., & Hoffrage, U. (2003). Hindsight bias: How knowledge and heuristics affect our reconstruction of the past. *Memory*, *11*(4–5), 357–377. https://doi.org/10.1080/09658210244000595

Hewlett, S. A., Marshall, M., & Sherbin, L. (2013). How Diversity Can Drive Innovation. *Harward Business Review*, *December*. https://doi.org/10.1177/1745790414546037

Hinton, P. (2019). *Stereotypes and the Construction of the Social World*. Routledge.

Hitt, M. A., Ireland, R. D., & Hoskisson, R. E. (1999). *Strategic Management* (3rd ed.). South-Western College Publishing.

Hodgkinson, C. G. P. (2001). Cognitive Processes in Strategic Management: Some Emerging Trends and Future Directions. In N. Anderson, D. Ones, K. Handan, & C. Viswesvaran (Eds.), *Handbook of Industrial , Work & Organizational Psychology - Volume 2 : Organizational Psychology Cognitive Processes in Strategic Management :* (Vol. 2, pp. 416–440). SAGE Publications Ltd.

https://doi.org/http://dx.doi.org/10.4135/9781848608368.n22

Hodgkinson, G., & Sadler-Smith, E. (2018). The Dynamics of Intuition and Analysis in Managerial and Organizational Decision Making. *Academy of Management Perspectives*, *32*(4), 473–492. https://doi.org/10.5465/amp.2016.0140

Hodson, G., Dovidio, J. F., & Gaertner, S. L. (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin*, *28*(4), 460–471. https://doi.org/10.1177/0146167202287004

Hogarth, R. (1980). Human judgement: An overview. In *Judgement and choice the psychology of decision*. John Wiley & Sons.

Holloway, I., & Todres, L. (2003). The status of method: flexibility, consistency and coherence. In *Qualitative Research* (Vol. 2). SAGE Publications.

Holton, J. (2008). Grounded Theory as a General Research Methodology. *The Grounded Theory Review*, *7*(June 2008), 67–94.

Houser, K. A. (2019). Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making. *Stanford Technology Law Review*, *22*(2), 291–353.

Howard-Payne, L. (2016). Glaser or Strauss? Considerations for selecting a grounded theory study. *South African Journal of Psychology*, *46*(1), 50–62. https://doi.org/10.1177/0081246315593071

Hubbard, D. (2007). *How to Measure Anything : Finding the Value of "intangibles" in Business*. John Wiley & Sons.

Hudson, L. A., & Ozanne, J. L. (1988). Alternative Ways of Seeking Knowledge in Consumer Research. *Journal of Consumer Research*, *14*(4), 508–521.

Hunter, G. (2004). Qualitative Research in Information Systems: An exploration methods. In *The handbook of information systems research* (pp. 291–304). IGI Global. https://doi.org/10.4135/9781849209687

Hwang, T., Kesselheim, A., & Vokinger, K. (2019). Lifecycle Regulation of Artificial Intelligence- and Machine Learning-Based Software Devices in Medicine. *JAMA*, *322*(23), 2285–2286.

IBM. (2018). *Many AI systems are trained using biased data*. https://www.research.ibm.com/5-in-5/ai-and-bias/

IBM Corporation. (2016). *Analytics Solutions Unified Method*.

Ingold, P. V., Dönni, M., & Lievens, F. (2018). A dual-process theory perspective to better understand judgments in assessment centers: The role of initial impressions for dimension ratings and validity. *Journal of Applied Psychology*, *103*(12), 1367–1378. https://doi.org/10.1037/apl0000333

Intezari, A. (2014). Wisdom and decision making: Grounding theory in management practice. In *School of Management: Vol. PhD*. Massey University.

Intezari, A., & Gressel, S. (2017). Information and reformation in KM systems: big data and strategic decision-making. *Journal of Knowledge Management*, *21*(1), 71–91. https://doi.org/10.1108/JKM-07-2015-0293

Intezari, A., & Pauleen, D. (2013). Students of Wisdom: An Integral Meta-competencies Theory of Practical Wisdom. In W. Kupers & D. Pauleen (Eds.), *The Practical Wisdom in Leadership ,Organization and Integral Business Practice* (First). Routledge.

Intezari, A., & Pauleen, D. (2019). *Wisdom, Analytics and Wicked Problems: Integral Decision Making for the Data Age*. Routledge.

Intezari, A., & Pauleen, D. J. (2018). Conceptualizing wise management decision-making: A Grounded Theory approach. *Decision Sciences*, *49*(2), 335–400.

Janghorban, R., Latifnejad, R., & Taghipour, A. (2014). Skype interviewing: The new generation of online synchronous interview in qualitative research. *International Journal of Qualitative Studies on Health and Well-Being*, *9*(1). https://doi.org/10.3402/qhw.v9.24152

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Johnson, G. M. (2020). Algorithmic bias: on the implicit biases of social technology. *Synthese*, *198*(10), 9941–9961. https://doi.org/10.1007/s11229-020-02696-y

Johnson, P., & Duberley, J. (2003). Reflexivity in management research. *Journal of Management Studies*, *40*(5), 1279–1303. https://doi.org/10.1111/1467-6486.00380

Johnson, R., Stone, D., & Lukaszewski, K. (2020). The benefits of eHRM and AI for talent acquisition. *Journal of Tourism Futures*, *7*(1), 40–52. https://doi.org/10.1108/JTF-02-2020-0013

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*(4), 557–571. https://doi.org/10.1037/0022-3514.80.4.557

Kaelbling, L., Littman, M., & Moore, A. (1996). Deep reinforcement learning: a survey. *Journal of Arti?Cial Intelligence Research 4*, *4*, 237–285. https://doi.org/10.1631/FITEE.1900533

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*, *93*(5), 1449–1470. https://doi.org/10.1128/AAC.03728-14

Kahneman, D., Lovallo, D., & Sibony, O. (2011). Before you make the big decision. *Harvard Business Review*, *June*.

Kahneman, D., & Tversky, A. (1979). *Prospect Theory: An Analysis of Decision under Risk*. *47*(2), 263–292. https://www.jstor.org/stable/1914185

Kane, G., Alavi, M., Labianca, G., & Borgatti, S. (2014). What’s different about social media networks? A framework and research agenda. *MIS Quarterly: Management Information*

Systems, *38*(1), 275–304. https://doi.org/10.25300/misq/2014/38.1.13

Kaplan, A., & Haenlein, M. (2019). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons*, *63*(1), 37–50. https://doi.org/10.1016/j.bushor.2019.09.003

Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, *142*(4), 1074–1083. https://doi.org/10.1037/a0030399

Keller, J. (2005). In genes we trust: The biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology*, *88*(4), 686–702.

Kelley, S., & Ovchinnikov, A. (2020). Anti-discrimination Laws, AI, and Gender Bias in Non-mortgage Fintech Lending. In *SSRN Electronic Journal* (Issue November). https://doi.org/10.2139/ssrn.3719577

Khemlani, S. (2018). Reasoning. In S. Thompson-Schill (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–45). Wiley & Sons. https://doi.org/10.1002/9781119170174.epcn311

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: perspectives of leading teachers for AI in education. In *Education and Information Technologies* (Issue 0123456789). Springer US. https://doi.org/10.1007/s10639-021-10831-6

King, N. (2004). Using templates in the thematic analysis of text. In G. Cassell, C., Symon (Ed.), *Essential guide to qualitative methods in organizational research* (pp. 257–270). Sage.

Kirimi, J., & Moturi, C. (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

Klein, G. (2015). A naturalistic decision making perspective on studying intuitive decision making. *Journal of Applied Research in Memory and Cognition*, *4*(3), 164–168. https://doi.org/10.1016/j.jarmac.2015.07.001

Klein, H., & Myers, M. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, *32*(1), 67–94.

Kleinmann, M., & Ingold, P. V. (2019). Toward a Better Understanding of Assessment Centers: A Conceptual Review. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*, 349–372. https://doi.org/10.1146/annurev-orgpsych-012218-014955

Kline, T. J. B. (1994). Measurement of tactical and strategic decision making. *Educational and Psychological Measurement*, *54*(3), 745–756. https://doi.org/10.1177/0013164494054003021

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

Kokku, R., Sundararajan, S., Dey, P., Sindhgatta, R., Nitta, S., & Sengupta, B. (2018). Augmenting Classrooms with AI for Personalized Education. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2018-April*, 6976–6980. https://doi.org/10.1109/ICASSP.2018.8461812

Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, *00*(00), 1–22. https://doi.org/10.1080/0960085X.2021.1927212

Kumar, R. (2017). *Machine Learning and Cognition in Enterprises* (First). Apress.

Laplante, P. (2007). *What Every Engineer Should Know about Software Engineering* (1st ed.). CRC Press. Taylor & Francis

LaToza, T. D., Arab, M., Loksa, D., & Ko, A. J. (2019). Explicit programming strategies. *ArXiv*, 2416–2449.

Lattimore, F., O'Callaghan, S., Paleologos, Z., Reid, A., Santow, E., Sargeant, H., & Thomsen, A. (2020). *Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias*.

Laws, K. (2006). Case study and grounded theory : Sharing some alternative qualitative research methodologies with systems professionals. *Proceedings of the 22nd International Conference of the System Dynamics Society*, *Forrester 1961*, 1–25. http://www.systemdynamics.org/conferences/2004/SDS_2004/PAPERS/220MCLEO.pdf

Leighton, J. (2010). Cognitive Biases. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of Case Study Research* (pp. 430–434). SAGE Publications, Inc. https://doi.org/10.4135/9781412957397

Lele, A. (2018). *Disruptive Technologies for the Militaries and Security*. Springer.

Lempert, L. (2016). Asking Questions of the Data: Memo Writing in the Grounded Theory Tradition. *The SAGE Handbook of Grounded Theory*, 245–264. https://doi.org/10.4135/9781848607941.n12

Li, X., & Ling, W. (2015). How Framing Effect Impact on Decision Making on Internet Shopping. *Open Journal of Business and Management*, *03*(01), 96–108. https://doi.org/10.4236/ojbm.2015.31010

Lincoln, Y., Lynham, S., & Guba, E. (2011). *Paradigmatic controversies, contradictions, and emerging confluences, revisited* (N. Denzin & Y. Lincoln (eds.); 4th ed.). The Sage handbook of qualitative research.

Linnenluecke, M., Marrone, M., & Singh, A. (2020). Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*, *45*(2), 175–194. https://doi.org/10.1177/0312896219877678

Linos, E., & Reinhard, J. (2015). *A head for hiring: The behavioural science of recruitment* *and selection* (Issue August).

Lockton, D. (2012). Cognitive Biases, Heuristics and Decision-Making in Design for Behaviour Change. *Ssrn*, 1–19. https://doi.org/10.2139/ssrn.2124557

Long, C., & Kelly, T. (2015). *Data science & big data analytics. Discovering, analyzing, visualizing and presenting data*. John Wiley & Sons.

Long, R. (2018). *Diverse Teams Are Essential for Quality Software*.

Luger, G. F. (2004). *Artificial Intelligence Structures and Strategies for Complex Problem Solving*. Pearson Education.

Lwakatare, L. E., Crnkovic, I., & Bosch, J. (2020). DevOps for AI - Challenges in Development of AI-enabled Applications. *28th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2020*. https://doi.org/10.23919/SoftCOM50211.2020.9238323

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues In Educational Research*, *16*(2), 193–205.

Manyika, J., Bughin, J., Silberg, J., & Gumbel, P. (2018). The promise and challenge of the age of Artificial Intelligence. *McKinsey & Company*, *October*, 8.

Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2021). Algorithmic pollution: Making the invisible visible. *Journal of Information Technology*, *36*(4), 391–408. https://doi.org/10.1177/02683962211010356

Martin, K. (2018). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, 1–16. https://doi.org/10.1007/s10551-018-3921-3

Martin, P., & Turner, B. (1986). Grounded Theory and Organizational Research. *The Journal Of Applied Behavioral Science*, *22*(2).

Martin, V. (2006). The Relationship between an Emerging Grounded Theory and the Existing Literature: Four phases for consideration. In *The Grounded Theory Review: An International Journal* (Vol. 5, Issue 2, pp. 47–50). Sociology Press.

McCarthy, J. (2007). *What Is Artificial Intelligence*. http://www-formal.stanford.edu/jmc/

McFall, J. P. (2015). Rational, normative, descriptive, prescriptive, or choice behavior? The search for integrative metatheory of decision making. *Behavioral Development Bulletin*, *20*(1), 45–59. https://doi.org/10.1037/h0101039

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). https://doi.org/10.1145/3457607

Mergenthaler, J. V., Chiong, W., Dohan, D., Feler, J., Lechner, C. R., Starr, P. A., & Arias, J. J. (2021). A Qualitative Analysis of Ethical Perspectives on Recruitment and Consent for Human Intracranial Electrophysiology Studies. *AJOB Neuroscience*, *12*(1), 57–67. https://doi.org/10.1080/21507740.2020.1866098

Messick, D. M., & Bazerman, M. H. (1996). Ethical leadership and the psychology of decision making. *The New Phase of Business Ethics*, *37*(2), 9–22.

Microsoft. (2020). *What is the Team Data Science Process?* https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The Structure of " Un- structured ". *Administrative Science Quarterly*, *21*(June), 246–275.

Mohammed, M., Khan, M. B., & Bashie, E. B. M. (2016). Machine learning: Algorithms and applications. In *Machine Learning: Algorithms and Applications* (Issue July). https://doi.org/10.1201/9781315371658

Mohanani, R., Ralph, P., & Shreeve, B. (2014). Requirements fixation. *Proceedings of the 36th International Conference on Software Engineering - ICSE 2014*, 895–906. https://doi.org/10.1145/2568225.2568235

Mohanani, R., Salman, I., Turhan, B., Rodriguez, P., & Ralph, P. (2018). Cognitive Biases in Software Engineering: A Systematic Mapping Study and Quasi-Literature Review. *IEEE Transactions on Software Engineering*, *October*. https://doi.org/10.1109/TSE.2018.2877759

Molenberghs, P., & Louis, W. (2018). Insights from fMRI studies into ingroup bias. *Frontiers in Psychology*, *9*(OCT), 1–12. https://doi.org/10.3389/fpsyg.2018.01868

Morse, J. (1991). Approaches to qualitative quantitative methodological triangulation. *Nurssing Research*, *40*(1), 120–123.

Morse, J., Stern, P., Corbin, J., Bowers, B., Charmaz, K., & Clarke, A. (2016). *Developing Grounded Theory: The Second Generation* (J. Morse (ed.); Second). Routledge.

Mullainathan, S., & Thaler, R. (2000). Behavioral Economics. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, *3*, 437–442. https://doi.org/10.1016/B978-0-08-097086-8.71007-5

Murphy, C., Klotz, A. C., & Kreiner, G. E. (2017). Blue skies and black boxes: The promise (and practice) of grounded theory in human resource management research. *Human Resource Management Review*, *27*(2), 291–305. https://doi.org/10.1016/j.hrmr.2016.08.006

Muthusamy, V., Slominski, A., & Ishakian, V. (2018). Towards enterprise-ready AI deployments. *2018 First International Conference on Artificial Intelligence for Industries (AI4I)*, 108–109. https://doi.org/10.1109/ai4i.2018.00034

Nalchigar, S., Yu, E., & Keshavjee, K. (2021). Modeling machine learning requirements from three perspectives: a case report from the healthcare domain. *Requirements Engineering*, *26*(2), 237–254. https://doi.org/10.1007/s00766-020-00343-z

Namey, E., Guest, G., Thairu, L. and Johnson, L. (2008). Data Reduction Techniques for Large Qualitative Data Sets. In *Handbook for team-based qualitative research*. Rowman Altamira.

Nasiripour, S., & Farrel, G. (2021). *Goldman Cleared of Bias in New York Review of Apple Card*. https://www.bloomberg.com/news/articles/2021-03-23/goldman-didn-t-discriminate-with-apple-card-n-y-regulator-says

Ndobo, A., Faure, A., Boisselier, J., & Giannaki, S. (2018). The ethno-racial segmentation jobs: The impacts of the occupational stereotypes on hiring decisions. *Journal of Social Psychology*, *158*(6), 663–679. https://doi.org/10.1080/00224545.2017.1389685

Newell, A., & Simon, H. (1972). *Human Problem Solving*. Prentice-Hall.

Newell, S. (2005). Recruitment and selection. Managing. In *human resources: Personnel management in transition* (pp. 115–147).

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*(June), 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

Nguyen, H. (2021). *Supply Chain Information Visibility and its impact on Decision-Making: An Integrated Model in the Pharmaceutical Industry*. Massey University.

Nilsson, J., & Nilsson, J. (1998). *Artificial intelligence: a new synthesis*. Morgan Kaufmann.

Nilsson, N. J. (1998). Introduction to machine learning. *An Early Draft of A Proposed Textbook*. https://doi.org/10.1016/j.neuroimage.2010.11.004

Nilsson, N. J. (2010). *The Quest for Artificial Intelligence*. Cambridge University Press.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, *5*(1), 17–37.

Nowruzi, F. E., Kapoor, P., Kolhatkar, D., Hassanat, F. Al, Laganiere, R., & Rebut, J. (2019). How much real data do we actually need: Analyzing object detection performance using synthetic and real data. *ArXiv*.

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., … Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), 1–14. https://doi.org/10.1002/widm.1356

Nürnberger, M., Nerb, J., Schmitz, F., Keller, J., & Sütterlin, S. (2016). Implicit Gender Stereotypes and Essentialist Beliefs Predict Preservice Teachers Tracking Recommendations. *Journal of Experimental Education*, *84*(1), 152–174. https://doi.org/10.1080/00220973.2015.1027807

O'Meara, B., & Petzall, S. (2013). *Handbook of Strategic Recruitment and Selection: A Systems Approach*. Emerald Group Publishing.

Öllinger, M., Jones, G., & Knoblich, G. (2008). Investigating the effect of mental set on insight problem solving. *Experimental Psychology*, *55*(4), 269–282. https://doi.org/10.1027/1618-3169.55.4.269

Oriikowski, W. (1993). CASE Tools as Organizational Change Investigating Incremental and

Radical Changes in Systems Development. *MIS Quarterly*, *September*, 309–341.

Osmont, A., Cassotti, M., Agogué, M., Houdé, O., & Moutier, S. (2015). Does ambiguity aversion influence the framing effect during decision making? *Psychonomic Bulletin and Review*, *22*(2), 572–577. https://doi.org/10.3758/s13423-014-0688-0

Parasurama, P., Ghose, A., & Ipeirotis, P. (2021). *Determinants of Occupational Segregation across Race and Gender : Evidence from Sourcing , Screening , and Hiring in IT Firms*. https://ssrn.com/abstract=3672484

Park, H., Ahn, D., Hosanagar, K., & Lee, J. (2021). Human-AI interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411764.3445304

Pauleen, D. (2001). *A grounded theory of virtual facilitation: Building relationships with virtual team members* (Issue June) [Victoria University of Wellington]. http://researcharchive.vuw.ac.nz/handle/10063/363

Pauleen, D. J., Rooney, D., & Intezari, A. (2017). Big data, little wisdom: trouble brewing? Ethical implications for the information systems discipline. *Social Epistemology*, *31*(4), 9–33.

Pearl, J. (2018). Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution. In *WSDM'18* (Issue February 5-9). https://doi.org/10.1145/3159652.3176182

Pessach, D., Singer, G., Avrahami, D., Ben-Gal, H., Shmueli, E., & Ben-Gal, I. (2020). Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. *Decision Support Systems*, *134*(January). https://doi.org/10.1016/j.dss.2020.113290

Pettigrew, A. (1992). The Character and Significance of Strategy Process Research. *Strategic Management Journal*, *13*, 5–16.

Pettigrew, A. M. (1990). Longitudinal Field Research on Change: Theory and Practice. *Organization Science*, *1*(3), 267–292. https://doi.org/10.1287/orsc.1.3.267

Petty, R., & Cacioppo, J. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2

Pidgeon, N. F., Turner, B. A., & Blockley, D. I. (1991). The use of Grounded theory for conceptual analysis in knowledge elicitation. *International Journal of Man-Machine Studies*, *35*(2), 151–173. https://doi.org/10.1016/S0020-7373(05)80146-4

Pillala, A. (2021). Role of Artificial Intelligence in Human Resources. *NVEO-Natural Volatiles & Essential Oils*, *8*(4), 297–307. https://doi.org/10.4018/978-1-7998-7959-6.ch011

Pinsonneault, A., & Kraemer, K. (2016). Survey Research Methodology in Management Information Systems: An Assessment. *Journal of Management Information Systems*, *July*, 1–23.

Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. University of Chicago Press.

Polli, F. (2019). Using AI to eliminate bias from hiring. *Harvard Business Review*, 1–5. https://hbr.org/2019/10/using-ai-to-eliminate-bias-from-hiring

Polyzotis, N., Roy, S., Whang, S. E., & Zinkevich, M. (2017). Data management challenges in production machine learning. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, *Part F1277*, 1723–1726. https://doi.org/10.1145/3035918.3054782

Pomerol, J.-C. (1996). Human Decision Making and Artificial Intelligence. *ELSEVIER European Journal of Operational Research*, *99*, 3–25. https://doi.org/10.1145/2987491.2987493

Pompian, M. (2011). Conservatism bias. In *Behavioral Finance and Wealth Management* (Second). John Wiley & Sons. https://doi.org/10.1002/9781119202400

Poole, D., Mackworth, A., & Goebel, R. (1998). *Computational Intelligence.*

Pöppelbuß, J., & Goeken, M. (2015). Understanding the Elusive Black Box of Artifact Mutability. *Association for Information Systems*, *2015*, 1557–1571. http://www.wi2015.uni-osnabrueck.de/Files/WI2015-D-14-00212.pdf

Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. Free Press.

Prewett-Livingston, A., Veres, J., Feild, H., & Lewis, P. (1996). Effects of race on interview ratings in a situational panel interview. *Journal of Applied Psychology*, *81*(2), 178–186. https://doi.org/10.1037/0021-9010.81.2.178

PWC. (2016). *Data-driven : Big decisions in the intelligence age*. https://www.pwc.com/us/en/services/consulting/library/view-of-decisions.html

Raghavan, M., & Barocas, S. (2019). *Challenges for mitigating bias in algorithmic hiring*. https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/

Ragsdale, C. (2007). *Spreadsheet Modeling and Decision Analysis* (5th ed.). Thomson South Western.

Ramchandani, K., & Anderson, J. (2018). *Cognitive decision making in insurance Data , data everywhere - From art to science*. https://public.dhe.ibm.com/common/ssi/ecm/77/en/77019477usen/77019477usen-00_77019477USEN.pdf

Rangel, U., & Keller, J. (2011). Essentialism goes social: Belief in social determinism as a component of psychological essentialism. *Journal of Personality and Social Psychology*, *100*(6), 1056–1078.

Régner, I., Thinus-Blanc, C., Netter, A., Schmader, T., & Huguet, P. (2019). Committees with implicit biases promote fewer women when they do not believe gender bias exists. *Nature Human Behaviour*, *3*(11), 1171–1179. https://doi.org/10.1038/s41562-019-0686-3

Rhem, A. J. (2020). AI ethics and its impact on knowledge management. *AI and Ethics*, *1*(1), 33–37. https://doi.org/10.1007/s43681-020-00015-2

Ribes, E., Touahri, K., & Perthame, B. (2017). Employee turnover prediction and retention policies design: a case study. *ArXiv*, 1–12. http://arxiv.org/abs/1707.01377

Rich, E., & Knight, K. (1991). *Artificial Intelligence*. McGraw-Hill.

Richardson, L. (1998). Writing: A Method of Inquiry. In N. Denzin & Y. Lincoln (Eds.), *Collecting and Interpreting Qualitative Materials* (pp. 345–371). SAGE Publications.

Rieger, K. L. (2019). Discriminating among grounded theory approaches. *Nursing Inquiry*, *26*(1), 1–12. https://doi.org/10.1111/nin.12261

Rivera, L. A. (2015). Go with your gut: Emotion and evaluation in job interviews. *American Journal of Sociology*, *120*(5), 1339–1389. https://doi.org/10.1086/681214

Roberto, M. A. (2002). Lessons from Everest: The interaction of cognitive bias, psychological safety, and system complexity. *California Management Review*, *45*(1), 136–158. https://doi.org/10.2307/41166157

Robillard, P. N. (1999). The role of knowledge in software development. *Communications of the ACM*, *42*(1), 87–92. https://doi.org/10.1145/291469.291476

Roh, Y., Heo, G., & Whang, S. E. (2021). A Survey on Data Collection for Machine Learning: A Big Data-AI Integration Perspective. *IEEE Transactions on Knowledge and Data Engineering*, *33*(4), 1328–1347. https://doi.org/10.1109/TKDE.2019.2946162

Roovers, R. (2019). *Transparency and Responsibility in Artificial Intelligence: A call for explainable AI*.

Rosenfeld, A., & Kraus, S. (2018). *Predicting Human Decision-Making From Prediction to Action* (R. J. Brachman & P. Stone (eds.)). Morgan & Claypool. https://doi.org/10.2200/S00820ED1V01Y201712AIM036

Rubin, H. J., & Rubin, I. S. (2011). *Qualitative Interviewing : The Art Of Hearing Data Description*. Sage.

Russell, S., & Norvig, P. (2010). *Artificial Intelligence A modern Approach third edition*. Pearson Education. https://doi.org/10.1017/S0269888900007724

Sadler-Smith, E., & Shefy, E. (2004). The intuitive executive: Understanding and applying 'gut feel' in decision-making Eugene. *Academy of Management Executive*, *18*(4), 76–91. https://doi.org/10.1177/106591299104400315

Saifee, M. (2020). *Can AI Algorithms be Biased?* https://towardsdatascience.com/can-ai-algorithms-be-biased-6ab05f499ed6

Salin, E. ., & Winston, P. (1992). Machine Learning and Artificial Intelligence. *Analytical Chemistry*, *64*(1).

Salman, I. (2016). Cognitive biases in software quality and testing. *Proceedings of the 38th International Conference on Software Engineering Companion - ICSE '16*, 823–826. https://doi.org/10.1145/2889160.2889265

Sandelowski, M. (2010). What's in a name? Qualitative description revisited. *Research in Nursing and Health*, *33*(1), 77–84. https://doi.org/10.1002/nur.20362

Sandelowski, M., & Barroso, J. (2003). Classifying the findings in qualitative studies. *Qualitative Health Research*, *13*(7), 905–923. https://doi.org/10.1177/1049732303253488

Santos, R., & Bernardino, J. (2008). Real-time data warehouse loading methodology. *ACM International Conference Proceeding Series*, *299*(January), 49–58. https://doi.org/10.1145/1451940.1451949

Sarker, I. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 1–21. https://doi.org/10.1007/s42979-021-00592-x

Sarker, I., Kayes, A., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00318-5

Scanlan, C. (2011). Primacy Effect. In *Encyclopedia of Survey Research Methods*. SAGE Publications, Inc. https://doi.org/https://dx.doi.org/10.4135/9781412963947

Schank, R. C. (1987). What is AI anyway? *The Foundations of Artificial Intelligence*, *8*(4), 59–65. https://doi.org/10.1017/cbo9780511663116.003

Schatzman, L. (1991). Dimensional analysis: Notes on an alternative approach to the grounding of theory in qualitative research. In D. Maines (Ed.), *Social organization and social process: Essays in honor of Anselm Strauss* (pp. 303– 314). Aldine De Gruyter.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying on Applying CRISP-DM Process Model. *ScienceDirect*, *181*(2019), 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Schwenk, C. R. (1982). *Cognitive Bias in Strategic Decision-Making: Some Conjectures*. Faculty Working Paper NO. 863 College.

Scott, M. J. (2021). Reasons Things Happen for a Reason: An Integrative Theory of Teleology. *Perspectives on Psychological Science*, *1987*. https://doi.org/10.1177/1745691621995753

Seaman, C. (2008). Qualitative Methods. In F. Shull, J. Singer, & D. Sjøberg (Eds.), *Guide to Advanced Empirical Software Engineering* (pp. 1–388). Springer. https://doi.org/10.1007/978-1-84800-044-5

Secchi, D., & Bardone, E. (2009). A Model of Organizational Bandwagon. In *College of Business Administration* (Issue July).

Seitz, S. (2016). Pixilated partnerships, overcoming obstacles in qualitative interviews via Skype: a research note. *Qualitative Research*, *16*(2), 229–235. https://doi.org/10.1177/1468794115577011

Sekiguchi, T. (2007). A contingency perspective of the importance of PJ fit and PO fit in employee selection. *Journal of Managerial Psychology*, *22*(2), 118–131. https://doi.org/10.1108/02683940710726384

Selbst, A., Boyd, D., Friedler, S., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, *1*(1), 59–68. https://doi.org/10.1145/3287560.3287598

Shapiro, S. (1992). Artificial Intellignce. In *Encyclopedia of Artificial Intelligence* (Second). John Wiley & Sons. https://doi.org/10.1007/978-3-658-00456-9

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), 941–945. https://doi.org/10.1016/j.cub.2011.10.030

Shiffrin, M., & Schneider, W. (1984). Automatic and controlled processing revisited. *Psychological Review*, *91*, 269–276.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review*, 000812561986225. https://doi.org/10.1177/0008125619862257

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, *32*(6), 1049–1062. https://doi.org/10.1080/03640210801897864

Siegerink, B., & Rohmann, J. L. (2018). Impact of your results: Beyond the relative risk. *Research and Practice in Thrombosis and Haemostasis*, *2*(4), 653–657. https://doi.org/10.1002/rth2.12148

Simon, H. (1947). *Administrative Behavior: A Study of Decision Making Processes in Administrative Organization*. MacMillan Company.

Simon, H. (1976). *Administrative behavior. A study of decision making processes in administrative organization* (third). The Free Press, Collier Macmillan.

Simon, H. (1981). *The Sciences of the Artificial*. MIT Press.

Simon, H. (1997). *Models of Bounded Rationality : Empirically Grounded Economic Reason*. The MIT Press.

Simon, H. A. (1960). *The new science of management decision*. Harper & Row.

Smith, G. (2003). Using integrated spreadsheet modeling for supply chain analysis. *Supply Chain Management: An International Journal*, *8*(4), 285–290. https://doi.org/10.1108/13598540310490044

Soleimani, M., Intezari, A., & Pauleen, D. (2021). Mitigating Cognitive Biases in Developing AI-Assisted Recruitment Systems. *International Journal of Knowledge Management*, *18*(1), 1–18. https://doi.org/10.4018/ijkm.290022

Sołek-Borowska, C., & Wilczewska, M. (2018). New Technologies in the Recruitment Process. *Economics and Culture*, *15*(2), 25–33. https://doi.org/10.2478/jec-2018-0017

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American Economic Review*, *92*(3), 434–459. https://doi.org/10.1257/00028280260136200

Spiegelman, D., & Valanis, B. (1998). Correcting for bias in relative risk estimates due to exposure measurement error: A case study of occupational exposure to antineoplastics in pharmacists. *American Journal of Public Health*, *88*(3), 406–412. https://doi.org/10.2105/AJPH.88.3.406

Srinivasan, R., & Chander, A. (2021). Biases in AI systems. *Communications of the ACM*, *64*(8), 44–49. https://doi.org/10.1145/3464903

Stanovich, K., & West, R. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–726. https://doi.org/10.1017/S0140525X00003435

Strauss, A. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.

Strauss, A., & Corbin, J. (1990a). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage.

Strauss, A., & Corbin, J. (1990b). *The basics of qualitative analysis: Grounded theory procedures and techniques*. SAGE Publications.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research : techniques and procedures for developing grounded theory*. SAGE Publications.

Strohmeier, S. (2020). Algorithmic Decision Making in HRM. In T. Bondarouk & S. Fisher (Eds.), *Encyclopedia of Electronic HRM* (pp. 54–60). De Gruyter Oldenbourg.

Strohmeier, S., & Piazza, F. (2013). Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications*, *40*(7), 2410–2420. https://doi.org/10.1016/j.eswa.2012.10.059

Sudhamathy, G. (2016). Credit risk analysis and prediction modelling of bank loans using R. *International Journal of Engineering and Technology*, *8*(5), 1954–1966. https://doi.org/10.21817/ijet/2016/v8i5/160805414

Sug, H. (2018). Performance of Machine Learning Algorithms and Diversity in Data. *MATEC Web of Conferences*, *210*, 1–5. https://doi.org/10.1051/matecconf/201821004019

Sullivan, J. R. (2012). Skype: An Appropriate Method of Data Collection for Qualitative Interviews? *The Hilltop Review*, *6*(1), 54–60. internal-pdf://152.210.174.6/Skype_ An Appropriate Method of Data Collectio.pdf

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: C hallenges and a Path Forward . *California Management Review*, 000812561986791. https://doi.org/10.1177/0008125619867910

Tang, A. (2011). Software designers, are you biased? *ACM*, 1. https://doi.org/10.1145/1988676.1988678

Teodorescu, M. H. M., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of fairness in automation require a deeper understanding of human–ml augmentation. *MIS Quarterly: Management Information Systems*, *45*(3), 1483–1499. https://doi.org/10.25300/MISQ/2021/16535

Thaler, R. H. (2000). From Homo Economicus to Homo Sapiens. *Journal of Economic Perspectives*, *14*(1), 133–141. https://doi.org/10.1257/jep.14.1.133

Thomas, D. M., & Bostrom, R. P. (2010). Vital Signs for Virtual Teams: An Empirically Developed Trigger Model for Technology Adaptation Interventions. *MIS Quarterly*, *34*(1), 115–142.

Tripathi, K. P. (2011). Decision Support System Is a Tool for Making Better Decisions in the Organization. *Indian Journal of Computer Science and Engineering*, *2*(1), 112–117.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315.

Tversky, A., & Kahneman, D. (2013). Judgment under Uncertainty: Heuristics and Biases. *Handbook of the Fundamentals of Financial Decision Making. Part I*, *185*(4157), 261–268.

Tvrdíková, M. (2007). Support of decision making by business intelligence tools. *6th International Conference on Computer Information Systems and Industrial Management Applications, CISIM*, 364–368. https://doi.org/10.1109/CISIM.2007.64

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

Urquhart, C. (2013). *Grounded theory for qualitative research: A practical guide*. Sage.

Urquhart, C., & Fernández, W. (2006). Grounded theory method: The researcher as blank slate and other myths. *ICIS 2006 Proceedings - Twenty Seventh International Conference on Information Systems*, 457–464.

Urquhart, C., Lehmann, H., & Myers, M. (2010). Putting the “theory” back into grounded theory: Guidelines for grounded theory studies in information systems. *Information Systems Journal*, *20*(4), 357–381. https://doi.org/10.1111/j.1365-2575.2009.00328.x

Uzonwanne, F. (2018). Rational Model of Decision Making. In A. Farazmand (Ed.), *Global Encyclopedia of Public Administration, Public Policy, and Governance* (1st ed.). Springer International Publishing. https://doi.org/10.1007/978-3-319-31816-5

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing and Health Sciences*, *15*(3), 398–405. https://doi.org/10.1111/nhs.12048

Van de Ven, A. (1992). Suggestions for Studying Strategy Process: A Research Note. *Strategic Management Journal*, *13*, 169–191.

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing ai for hiring. *MIS Quarterly: Management Information Systems*, *45*(3), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

Van Den Broek, E., Sergeeva, A., & Huysman, M. (2019). An Ethnography of Fairness in Practice. *Fortieth International Conference on Information Systems, Munich*. https://aisel.aisnet.org/icis2019/future_of_work/future_work/6

Van Nuenen, T., Ferrer, X., Such, J., & Cote, M. (2020). Transparency for Whom? Assessing Discriminatory Artificial Intelligence. *IEEE Computer Society*, *53*(11), 36–44. https://doi.org/10.1109/MC.2020.3002181

Varshney, K. (2018). *Introducing AI Fairness 360*. IBM. https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/

Vasconcelos, M., Cardonha, C., & Gonçalves, B. (2017). Modeling Epistemological Principles for Bias Mitigation in AI Systems: An Illustration in Hiring Decisions. In *IBM Research Brasil*.

Vizecky, K., & El-Gayar, O. (2011). Increasing research relevance in DSS: Looking forward by reflecting on 40 years of progress. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 1–9. https://doi.org/10.1109/HICSS.2011.239

Von Krogh, G. (2018). Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing. *Academy of Management Discoveries*, *4*(4), 404–409. https://doi.org/10.5465/amd.2018.0084

Walls, J. G., Widmeyer, G. R., & EL Sawy, O. A. (2015). Assessing Information System Design Theory In Perspective: How Useful Was Our 1992 Initial Rendition? *Journal of Information Technology and Application (JITTA)*, 26–36.

Walls, J. G., Widmeyer, G. R., & Elsawy, O. A. (2004). ASSESSING Information System Design Theory in Perspective: How useful was our 1992 initial rendition? *Journal of Information Technology Theory and Application (JITTA)*, *6*(2), 43–58.

Walls, J. G., Widmeyer, G. R., & Sawy, O. A. El. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research*, *3*(1), 36–59.

Walport, M., & Sedwill, M. (2016). *Artificial intelligence: an overview for policy-makers*. GOV.UK. https://www.gov.uk/government/publications/artificial-intelligence-an-overview-for-policy-makers

Walsh, I., Holton, J. A., Bailyn, L., Fernandez, W., Levina, N., & Glaser, B. (2015). What Grounded Theory Is…A Critically Reflective Conversation Among Scholars. *Organizational Research Methods*, *18*(4), 581–599. https://doi.org/10.1177/1094428114565028

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

Wan, Z., Xia, X., Lo, D., & Murphy, G. C. (2020). How does Machine Learning Change

Software Development Practices? *IEEE Transactions on Software Engineering*, *5589*(c), 1–15. https://doi.org/10.1109/TSE.2019.2937083

Wang, D., Yang, Q., Abdul, A., & Lim, B. (2019). Designing Theory-Driven User-Centric Explainable AI. *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15.

Wang, J., & Perez, L. (2017). The effectiveness of data augmentation in image classification using deep learning. *ArXiv*.

Wang, Y., Liu, D., & Wang, Y. (2003). Discovering the capacity of human brain. *Brain and Mind*, *4*, 189–198.

Wastell, D. G. (2001). Barriers to effective knowledge management: Action research meets grounded theory. *Journal of Systems and Information Technology*, *5*(2), 21–36. https://doi.org/10.1108/13287260180000764

Watson, H., & Marjanovic, O. (2012). Big data: The fourth data management generation. *Business Intelligence Journal*, *18*(3), 4–9.

Weller, S. (2015). The Potentials and Pitfalls of Using Skype for Qualitative (Longitudinal) Interviews. *National Centre for Research Methods Working Paper*, 50. http://eprints.ncrm.ac.uk/3757/1/Susie Weller.pdf

Whysall, Z. (2018). Cognitive Biases in Recruitment, Selection, and Promotion: The Risk of Subconscious Discrimination. In V. Caven & S. Nachmias (Eds.), *Hidden inequalities in the workplace : a guide to the current challenges, issues and business solutions* (pp. 215–243).

Wick, M., Panda, S., & Tristan, J. B. (2019). *Unlocking fairness: A trade-off revisited* (Vol. 32).

Williams, C. (2011). Client-vendor knowledge transfer in IS offshore outsourcing: Insights from a survey of Indian software engineers. *Information Systems Journal*, *21*(4), 335–356. https://doi.org/10.1111/j.1365-2575.2010.00354.x

Williams, S., Sheffield, D., & Knibb, R. C. (2015). 'Everything's from the inside out with PCOS': Exploring women's experiences of living with polycystic ovary syndrome and co-morbidities through Skype™ interviews. *Health Psychology Open*, *2*(2). https://doi.org/10.1177/2055102915603051

Wilson, R. A., & Keil, F. C. (1999). THE MIT Encyclopedia of the Cognitive Sciences. In *A Bradford Book*. The MIT Press. https://doi.org/10.1016/S0004-3702(01)00095-9

Wilson, S., Bekker, M., Johnson, P., & Johnson, H. (1997). Helping and hindering user involvement - a tale of everyday design. *Conference on Human Factors in Computing Systems - Proceedings*, 178–185. https://doi.org/10.1145/258549.258699

Winston, P. (1992). *Artificial Intelligence* (Vol. 110). Addison-wesley.

Wirth, R., & Hipp, J. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, *24959*, 29–39. https://doi.org/10.1.1.198.5133

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: a review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, *29*(1), 64–77. https://doi.org/10.1080/1359432X.2019.1681401

Woolf, J., & Dixon, J. C. (2017). You're hired! a hiring simulation for sport management students that incorporates the hidden profile phenomenon. *Sport Management Education Journal*, *11*(2), 106–119. https://doi.org/10.1123/smej.2016-0028

Wright, J., & Atkinson, D. (2019). The impact of artificial intelligence within the recruitment industry: Defining a new way of recruiting. *Charmicael Fisher*, 1–39.

Xin, L., Zhou, W., Li, M., & Tang, F. (2020). Career Success Criteria Clarity as a Predictor of Employment Outcomes. *Frontiers in Psychology*, *11*(April), 1–11. https://doi.org/10.3389/fpsyg.2020.00540

Zander, T., Öllinger, M., & Volz, K. (2016). *Intuition and Insight: Two Processes That Build on Each Other or Fundamentally Differ?* Frontiers in Psychology. https://doi.org/https://doi.org/10.3389/fpsyg.2016.01395

Zeegers, M., & Barron, D. (2015). *Milestone Moments in Getting your PhD in Qualitative Research*. Elsevier.

Appendices

Appendix 1- Extra codes

Conceptual codes	Quotes
Cognitive biases	*"I think recruiters tend to talk a lot about local experience. And they don't trust the experience that people bring from other countries which is a real problem because it means that highly skilled individuals coming in on special work visas, then are not recognized for those skills and find it really hard to get work and end up kind of going down the chain that I see as a massive problem" (HR manager, 11)* *"It's a little bit tricky when you're in kind of a business relationship to say, we want to look at how biased you guys are before and after using our product, because a lot of companies are a little hesitant to have that be public or shared" (AI developer, 4).* *"I think what's important for HR executives to do is to be aware of the potential biases and these risks associated with exploring AI. Most developers have no concept of what those risk factors are, their task is just to build something that completes a certain task and is functional. But HR professionals and executives understand the risks involved from an equal opportunity standpoint, from a biased standpoint. What it's important for developers to do is to build it from the end user perspective of what do they need to do? What are their concerns? What are the risks?" (AI developer, 5).* *"The first thing is you're being careful about the data that you actually use to train the models and thinking critically whether or not they're a victim for biases. I mean, it is possible like not easy but possible to potentially debias machine learning models as well. Like after they've been trained, you can potentially look to analyse their output to understand if they do have ingrained biases, and perhaps look to mitigate those. But I think it's a lot more difficult than just making sure that your input data is as nonbiased as you can" (AI developer, 15).* *"You can see how many of these AI engines have biases right now because the dataset is not accurate. So, with inequality in the market, you don't have as many as people of colour or as many women or as many people from minorities in leadership positions. So that biased dataset skews the AI engine and AI is*

	going to learn from a dataset that is essentially corrupt" (AI developer, 6). *"To improve our model, we are working on our decision tree. Also we are trying to influence managers to take different people on board to find out biases. Because, again, the managers who are hiring for the bias, I mean, they call it mainly unconscious bias, they don't even know what they are doing" (AI developer, 11).*
Understanding requirements	*"Merely for recruitment, the context is interviewing process. But there are other opportunities or areas where we can use this. The intention was like, normally when there is an any interview, there's always usually a training course and that usually is related to how a person can match you know, like how culturally match a person is to a company like every company has a culture, and can we match that person capabilities? Or are you know, character, personality with company culture? So those questions that we were supposed to answer mostly, like binary questions like yes or no question or one liner questions, mostly binary, so that and we can calculate the interviewee response based on those binary answers and categorize if this person is going to be culture fit with the organisation they're interviewing for. So, usually, those interviews are conducted by a human we thought of replacing with an AI (AI developer, 10).* *"Through talking to clients, because clients have different priorities. In the end once you have a product concept, you come up with a product roadmap, with different pin-points you're trying to solve, you prioritize them. And then there's also strategic planning in terms of a competitive landscape. What, who's out there? And what's a better solution? And once you do that you work with designers to come with the flows, that they have the whole user experience. If the user is a recruiter who's using this, or a candidate, what is the flow A to Z? So obviously, there's the focus groups and lots of feedback from different people on how intuitive it is, how useful this is; once all this is done, development starts" (AI developer, 12).* *"HR people want us like this job requires to have like this kind of characteristics that characteristics and they were working with professional psychometrics. Anyway, they were working with universities, they were working with really big professionals, and try to translate the job description and the job requirements into what is required for the big five ranking for*

	that person and then they would instantly like when the job was posted" (AI developer, 14). *"I think like, some of the most valuable things are just talking to people on the business, maybe shadowing their work, reviewing problems that they've had in the past, I guess you'd call it. Business Analysis would be the discipline where you're trying to sit with someone and really understand how they work, and understand what's slowing them down, or where they see opportunities. So I think it's a really important part of the process (AI developer, 15).* *You might have data scientists like myself, but you might also have developers, testers, delivery managers; different roles and things work really well when all of the team is invested in building that kind of understanding at the start of a project. Yeah, so there's kind of a shift away, I guess it's similar to, you know, the shift from waterfall development approaches to agile development approaches where it essentially becomes part of the whole team's remit to actually work on the business problem, rather than just a business analyst (AI developer, 15).* *I think collaborating would be a good one, like, I think, you know, just thinking of an ideal scenario would be actually taking subject matter experts from, you know, the business or the organisation and embedding them with a development team would be so valuable, where, you know, they're constantly dedicated to the success of the project in they can bring such a rich understanding of the subject matter (AI developer, 15).* *I mean, I, I would guess that HR managers, like most managers, they'll have a set of quantitative data that they use to judge the success of their processes such as time to hire a candidate, length of tenure for new people, there'll be a set of quantitative metrics that we can kind of explore, to help understand to help HR managers and help ourselves to understand how we can make things better. And some of those metrics may end up even being things that we try to optimize for and our machine learning kind of projects" (AI developer, 15).* *"At the beginning, we do not have any inputs from our customer and our model is based on job descriptions. So basically the job description can be a real description of the job requirements, description, all can be a search field and criteria that are from the user, but you can consider it as generalisation there is a job*

	description, there is a position that required what kind of skills what kind of backgrounds, what kind of expected experience, degrees, what are these kind of requirements, then, on the other side, that are every kind of data profile" (AI developer, 13). *"The main goal is when people are applying for a role and the recruiters are recruiting, the retention rate is low and staff turnover is high and was the ethos element in the whole process and we are trying to find the best fit for a job position" (AI developer, 11).*
Data collection	*"Our starting point was, we had a ranking model, which was very purely statistical mathematical model, mainly used questionnaire, to do the ranking. And as we advertised more as we got more into more people applying once we reached a threshold of about 10,000 people, you know, in our database, which means we had their CV, we had a questionnaire, we had some of those standard questions that they had answered the pertinent to most of the jobs. At this point, we developed our AI predictive model" (AI developer, 11).* *"In the worlds of language, for example, if this language is similar to another language, how can we use that data to expand our current dataset? Or how can we kind of go and try to reach out to people to generate that data? And so, you know,* *"Within our internal protocols, it can be tricky that sometimes we have all this data, but a lot of it has been deleted, or we do not have permission to use it to train. So that's something we've to work with customers on like, here's what we're proposing, we would like to use your data for pre-built assessments" (AI developer, 4).* *"Other examples will be that maybe there are articles in Google, that people will say that analysts can be scientists. So maybe this type of information exists, It's a well-known fact that the dynamics and tasks of these two roles are somehow related. And if I had millions of series, of course, I could learn from that. But if we don't have enough data, maybe we could use open-source data or try to adjust it" (AI developer, 7).*
Data preparation	*"You can see how many of these AI engines have biases right now because the dataset is not accurate. So, with inequality in the market, you don't have as many as people of colour or as many women or as many people from minorities in leadership*

	positions. So that biased dataset skews the AI engine and AI is going to learn from a dataset that is essentially corrupt" (AI developer, 6). *"Having a diverse dataset to train AI is really important. That's not just one type of person you're looking for, but that you get different ways people spoke of something that we're all perceived as good or ended up being good for the job. We interview many people in a role, and who worked with people in that role to understand the competencies, like is teamwork very important? Is customer oriented very important? The algorithms are a prediction, they're not perfectly accurate and humans as well. So, you know, they're looking for patterns in the topic and what the person talked about and how they express themselves" (AI developer, 4).* *"Platform using organisational questionnaire that helps understanding people wrote in about their personality was looking for a job, and we collect that piece of information, and then there's definitely your CV, that had a lot of information" (AI developer, 11).* *"When you have a very controlled situation like that, we saw very little bias in the data at all. Some of the best models that we've ever built in this way, it's a little more straightforward to predict how well this person answered that question. So that's been very powerful that we've created from our own data, we continue to add to it and add different competencies, like a willingness to learn" (AI developer, 4).* *"It's hard sometimes for us as individual builders of technology, to effectively curate all of the datasets ourselves. We need research partners, we need academic partners who can help us do that. And that way, we were really confident as we got data back in from those studies. It was good data that we can build our ML model. I guess recruiters and companies can trust AI that is developed with this way of approaching data. One of the things that we did when we were building our AI product is we forged a research partnership with Michigan State University to basically have them as a research project, go out and intentionally curate a very diverse set of research participants that cover different ages and genders and demographics and education levels. So, we can get very broad representative datasets" (AI developer, 5).*

	"You can see how many of these AI engines have biases right now because the dataset is not accurate. So, with inequality in the market, you don't have as many as people of colour or as many women or as many people from minorities in leadership positions. So that biased dataset skews the AI engine and AI is going to learn from a dataset that is essentially corrupt" (AI developer, 6). *"So I'm not sure if you know this, but every video on YouTube is scanned by AI for any wrong content on it, but actually, Google have people watching those videos to flag and you don't content as well. So, you can post bad content on YouTube video and to not be found out for some days or sometimes until like and after some time to flag. So, it is so they are learning in a way and neural nets and image conditions is very good in terms of advancement as compared to NLP. So like just, we have to think it in a way that it is not a perfect system. If someone is telling you, it can be perfect they're just lying to the face" (AI developer, 10).* *"If you are not mindful of equally, assembling and aggregating the data, then you can get certain pockets of data that aren't valid or relevant, because they're missing from other components of that dataset. So, you need to be mindful on the process for data collection, data analysis and building of the data, so that you can account for those gaps or holes that might make nothing to build flow" (AI developer, 5).* *"To give you an example of increasing the quality of datasets, let's say, in web search, so you want to label webpages. And if you want to create a real good data, datasets and data database of different label, like this page is related to cooking, that page related to electronics, that page related to toys, things of that nature, you would need to do millions and millions of these labels then, right. And it's infeasible, timely, and it's expensive, because each of these pieces will take quite a bit of money to have people label it for you, right? Even if that's a very simple task. So instead, one approach is to do a very small set of these randomly, and then use that as your training set and then go on whenever you have more money, you bring people label a little more you improve the accuracy (AI developer, 14).* *When you're looking at the Big Five scores, some of them are different for women versus men. The scale of that attribute, and that score is different. So when you want to analyse that, and*

	when you want to get the labels and you want to really use those label data as your training set, you need to know the skills are different. And if you don't do that your system will always be biased, biased toward one gender versus another. Especially when you talk about psychology, because of the differences in culture because of differences in general the scales are different and if you don't account for that, if you don't use augmentation, if you don't use a bunch of other things, then your training sample set will be skewed and that will show itself as an excuse or as a bias in your testing results from AI" (AI developer, 14). *"We're looking at videos and games, and we have varied input data. We have quite a bit of data that we can get rid of any data that does have differences in groups. For example, if two different groups pronounce a word differently, we can simply not pay attention to that word at all. And that's kind of the simple version of what you can do to remove bias in a model" (AI developer, 4).* *"Even though AI is perfectly capable of predicting my skin tone, and therefore potentially my ethnic background, and my gender, and maybe even my age, we don't want those data points going into the analysis process. So, we specifically stripped them out of all the data that AI would look at them and make calculations. All we want to look at is how do plot points change on the face for each individual candidate? Over time there are micro facial expressions and all kinds of meaningful, valuable data that's sitting there and technology could look at, but we strip out all the personal identifiers around race around the age around gender, so that we have a completely neutral process. It only focuses on relevant data, and not data that feeds biases into the process" (AI developer, 5).*

Developing the ML model	*"One of our very early stage phase one research projects will involve us taking a sample set of candidate videos, and feeding that to a test group of HR executives, who would rank these candidate videos in certain soft skill attributes up to 10 to see what they came up with. We had our AI go through the same process to make those same rankings based off how we had built it to assess the inputs it was receiving from the video feed. And then we compare these two. Then we brought all this knowledge over from what the Human Resources said, and put it into our machine learning or modelling and maybe I'm looking at this batch of videos with human beings to see the missing parts in our outcomes" (AI developer, 5).* *"For probably more than 80% of the jobs, the psychometric analysis of the person is important. So when the behaviour of the person is really important, but in order to do that you need that person to be invited or even like to a video call, and you need to have an expert on the other side that does good. Say whatever they call the Big Five, that's the minimum, right? So the big five attributes analysis on that person, and it's expensive to do. So it's time consuming to do. So what they actually put together was to have a very simple workflow. And it's very simple structure in place that would require so anybody that applies for a job description, these guys were working with a company and they would just send them in a link, right" (AI developer, 14).* *"All machine learning algorithms have some sort of parameters, where we can tune them for biases. Like for logistic regressions or for any type of method where we are doing any binary classification or we are doing any sort of grouping each algorithm has some sorts of parameters which can be used for biases. I don't think those parameters help to avoid biases. It would be validation with the test dataset, and those parameter tuning" (AI developer, 10).* *"I think I mean, probably the first step would be assembling up first of all, your bit about developing hypotheses about what biases exists. So, you know, might be gender, gender bias, and hiring, for example. So, you'd be looking at your input data, and kind of distilling all of the qualitative information that you have to try and understand, you know, form a view on what biases might be there" (AI developer, 15).* *"So and even before even letting it search for candidates, it there's always this process of tuning the model. And sometimes,*

	for example, the program gets read a job description, it should select the right candidate for. So at this point, if it if it does, like it has to first understand the job description before finding a candidate. And it's likely that it doesn't understand that very well. So but it will do a very strong first ask, let's say. So in this scenario, where we feel like it's understanding of the job description is off, that basically people in the operations team will basically just shown the output. So in that intermediary step before going on to search for the right candidate, because if the understanding of the job description is wrong, it will search for the wrong candidate. Sometimes it's not necessary that the algorithm didn't work well. It could be that the job description is poorly written, or it's very sparse. So that's why there is this concept of what do they call a human in the loop? So yeah, yeah. Yeah. So yeah, it's not completely hands off" (AI developer, 11). *"Every now and then it basically, when the AI when the program tunes the parameters, it will tell you which ones seem very important, right? And so for example, if it will seem something like or the time of the day they applied, you will see random ones like that, and it will use and it will show you it's one of the and you realize No, no, how did I forget this parameter, and you remove it, just because for whatever reason, when obviously, the program ingested the data, and try to fit its understanding to the data, it crunched this the time of the day and used it to decide maybe a selection process or whatnot, when in reality, that should not be a factor" (AI developer, 12).* *"We usually have a human picking up a few short, shortlisting people manually, and then we're going and going, all doing the same job. But the same role the results and going yeah, is the model doing better or worse? And we look at accuracy ratio of our model, as well" (AI developer, 11).* *"The next phase would be typically quite iterative and here, you might be trying out a number of different modelling approaches. So you might be trying a different variety of algorithms that are kind of framed towards your success measures. It would be kind of an experimentation process where you'd have an evaluation dataset that you'll be using, and you'll be trying to try to optimize your key metrics against the evaluation dataset" (AI developer, 15).* *"A proper testing methodology to be able to validate models probably from an independent kind of person on that process,*

	not the HR person, not the developer is important. Someone independent that can validate the result and being able to give impartial view of that process is critical because of all the ethical reasons. Probably to some extent, it is very important to have ethical regulators or ethical analysts that they analyse some ethical sides of AI such as Are they biased? Are they discriminating people based on gender, or colour. So, the ethical side of AI needs to be pretty solid for us" (AI developer, 6). *"For checking ML models in terms of biases I think this comes down to Validation Test. When we are doing the Validation Test, we can put that check in our Validation Test that when we are testing it" (AI developer, 10).* *"There's quite a lot of research, but there are ways you can take a model and essentially start to unpack the layers that it uses to actually predict. And numerically you can kind of understand how different features shift the model's predictions. So, you might take gender as a feature, for example, and you can kind of see how that feature shifts the predictions. And, you know, with other variables being held fixed, and you can take, you can summarize those shifts in behaviour numerically and potentially apply them post hoc to the model to try and correct for any biases that exist. Yeah, I mean, that's probably I would prefer, like, you know, research and understanding before you try models to try and understand if biases are there, and you may find that the data that you're working with is just too biased is that might be a lot of events that perhaps the hiring intentions at this company aren't particularly fair. In which case, that's probably a point where you'd stop and actually raise that issue in, you know, communicate back to the client say, look, we found, you know, this is what we're seeing in the data. We don't believe that we have a solid foundation for building models at this point" (AI developer, 15).* *"The bigger blanket solution would be adding noise in data and trying to change the actual data to hide more information from the algorithm, like the example of the duration of the army. In this example, developers randomly add or subtract one year, so it will not be sure if this guy is a man or woman, that's one solution. So, you miss information here because the data is changed. If we want to be accurate, adding noise is not a good solution. However, if we want to be less biased, adding noise can help us to achieve that solution" (AI developer, 7).*

	"We are not even using things like how much time you spend trying to answer, how many errors you make, how many times you deleted things. We have all that data, but we don't use it because we feel it is a bit unfair at the end of the day, we only care about your expression of the answer, but not these other things" (AI developer, 2). *"We can get a biased algorithm, but once we start removing data that's causing the bias, we kind of reach this balancing point where we still have predictive power, but without the bias. For example, we're ignoring anything that's predictive of gender" (AI developer, 4).*

Appendix 2- HR managers' demographics

ID	Gender	Age	Experience (Year)	Field	Position	Academic Qualification	AI familiarity Scale (1-10)	
							Conceptually	Technically
1	Female	42	22	Technology	HR manager	PhD of HR	7	0
2	Female	50	17	Sport	General manager	High school	0	0
3	Female	44	14	Engineering	People manager	Bachelor of education	6	0
4	Male	41	18	Consultancy	Associate director	Master of commerce	9	3
5	Male	43	15	Education institute	Middle manager	Master of business	8	0
6	Male	28	1	Recruiter agency	Recruiter	Bachelor of management	7	3
7	Female	59	25	Technology	HR manager	Master of psychology	5	0
8	Female	29	7	Technology	Partner	Master of	9	3

					development manager	technology		
9	Female	32	7	Finance	HR researcher	Master of psychology	7	0
10	Male	57	14	Human Resources New Zealand (HRNZ)	Chief executive officer	Post-graduate of HR	5	0
11	Female	49	12	Technology	HR manager	Master of technology	5	2
12	Female	34	17	Telecommunication	People manager	Master of HR	6	2
13	Female	45	16	Consultancy	Principal consultant	Qualified degree	6	2
14	Female	-	15	Telecommunication	Senior recruiter	Bachelor of social work	8	3
15	Male	40	14	Consultancy	Principal consultant	Tertiary hospital	5	2
16	Female	44	20	Bank	Banking and finance	Master of tech futures	9	5
17	Female	47	9	Consultancy	Consultant	Master of tech futures	7	5
18	Female	56	20	Telecommunication	HR lead	Master of tech futures	7	5

19	Female	50	17	Technology	Global HR manager	Qualified degree	5	5
20	Female	47	18	Technology	Co-founder of a start-up	Post-graduate of business	8	3
21	Female	32	8	Telecommunication	HR manager	Qualified degree	8	5
22	Female	40	20	Telecommunication	HR manager	Post-graduate of HR	5	5

Appendix 3- AI developers' demographics

ID	Gender	Age	Experience (Year)	Position	Academic Qualification	Country
1	Female	30	3	Data scientist	PhD	Australia
2	Male	32	5	Data scientist	Masters	Australia
3	Male	35	5	AI Engineer	Masters	Germany
4	Female	35	4	Data scientist	PhD	United States
5	Male	44	3	Project Manager	College degree	United States
6	Male	35	3	Solution Engineer	Bachelors	New Zealand
7	Male	31	5	Data scientist	Masters	Israel
8	Male	30	3	AI Engineer	Bachelors	New Zealand
9	Male	27	4	AI Engineer	Bachelors	United States
10	Male	27	4	AI Engineer	Masters	India
11	Male	40	4	AI Engineer	Masters	New Zealand
12	Male	30	3	AI Engineer	Masters	United States
13	Male	50	6	AI Engineer	PhD	United States
14	Male	34	4	AI Engineer	PhD	United States
15	Male	30	4	AI Engineer	Bachelors	New Zealand
16	Male	27	5	Data scientist	Masters	New Zealand
17	Male	35	12	Data scientist	PhD	New Zealand

Appendix 4- Interview questions

Phases	Interview questions
Phase 1 (HR managers)	Have you ever seen a biased hiring decision? • Were there any personal or organisational factors that influenced the decision-making process? Why do you think these factors mislead the decision maker? • If you were the decision maker how would you have changed the actions which were taken? In general, which factors do you think affect the recruitment and selection process? How do you define rational decision-making in the recruitment and selection process? Are there any programs, software, applications that you know about them or your organisation uses in the recruitment and selection process? • For what purposes are these systems and tools being used? And what are the main benefits of these systems? How do you think AI is different from other decision support systems in the recruitment and selection process? Did some criteria from the candidates' CV, interviews or other tests that are not task-related out-weight the task-related criteria? Can you tell me which criteria? Are there any differences in the steps to recruit someone for a new and novel job position? In which steps of the recruitment and selection process do you think experience and fixed solutions to similar situations can lead up to selecting the right candidate?

Phase 2 (HR managers)	To what extent can you trust AI to streamline some repetitive tasks? Do you think using AI to speed up some parts of recruiting through automation helps you not to lose the best talent and you can compete with your competitors? Do you think the screening and sourcing of the candidates can be done objectively? Do you think AI can improve the quality of hiring through standardized job matching? What are the value-add contributions of AI systems at each stage of the recruitment and selection process? How do you think experience can help more? Do you think a system like AI can help you match between candidates' experience, knowledge, and skills and the requirements of the job? Do you think AI can assess the candidates fit for a role and even the organisation and its culture? How do you think the combination of AI and human intelligence maximize value throughout the recruitment and selection process? How do you think the interaction between AI and human can make both parties smarter over time? How do you think the unique strengths of humans and AI can act synergistically? How do you think bigger picture thinking is required in the recruitment and selection process? AI developers claim that they remove biases by removing the gender, name, age, to what extent you think that removing these items makes the decisions less biased? AI developers claim that AI can standardize the process to more objectively assess a candidate's ability and skills by removing the inherent biases? How do you think that relying too much on social media profiles would make you biased?

Phase 3 (AI developers)	How is AI-enabled recruitment developed? What are the challenges of developing AI-Recruitment Systems? • How do you collect datasets to train AI? • What are the sources of data? • What kind of algorithms are being used? Have you ever seen any signs of biases in AI-Recruitment Systems? • How do you test the AI system to find biases? How do you think AI- Recruitment Systems can be biased? How do you think AI systems can support and augment HR managers decision-making? • How can AI assess interpersonal skills such as confidence, motivation, critical thinking? Do you think AI systems can reduce HR managers biases? If so, how?
Phase 4 (AI developers)	How do you think that biases can be mitigated for developing AI? How can AI models be validated?

Appendix 5- Ethics approval

Gmail

melika soleimani <soleimanimelika@gmail.com>

Human Ethics Notification - 4000021182

humanethics@massey.ac.nz <humanethics@massey.ac.nz> Thu, Jun 6, 2019 at 9:02 PM
To: Melika.Soleimani.1@uni.massey.ac.nz, D.Pauleen@massey.ac.nz, N.Taskin@massey.ac.nz, T.A.Bentley@massey.ac.nz
Cc: humanethics@massey.ac.nz

HoU Review Group
Prof Tim Bentley

Ethics Notification Number: 4000021182
Title: Decision-making through Artificial Intelligence and human insight

Thank you for your notification which you have assessed as Low Risk.

Your project has been recorded in our system which is reported in the Annual Report of the Massey University Human Ethics Committee.

The low risk notification for this project is valid for a maximum of three years.

Please note that travel undertaken by students must be approved by the supervisor and the relevant Pro Vice-Chancellor and be in accordance with the Policy and Procedures for Course-Related Student Travel Overseas. In addition, the supervisor must advise the University's Insurance Officer.

A reminder to include the following statement on all public documents:

"This project has been evaluated by peer review and judged to be low risk. Consequently it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named in this document are responsible for the ethical conduct of this research.
If you have any concerns about the conduct of this research that you want to raise with someone other than the researcher(s), please contact Professor Craig Johnson, Director (Research Ethics), email humanethics@massey.ac.nz. "

Please note that if a sponsoring organisation, funding authority or a journal in which you wish to publish require evidence of committee approval (with an approval number), you will have to complete the application form again answering yes to the publication question to provide more information to go before one of the University's Human Ethics Committees. You should also note that such an approval can only be provided prior to the commencement of the research.

You are reminded that staff researchers and supervisors are fully responsible for ensuring that the information in the low risk notification has met the requirements and guidelines for submission of a low risk notification.

If you wish to print an official copy of this letter, please login to the RIMS system, and under the Reporting section, View Reports you will find a link to run the LR Report.

Yours sincerely

Professor Craig Johnson
Chair, Human Ethics Chairs' Committee and
Director (Research Ethics)

Appendix 6- Participant consent form

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

MASSEY BUSINESS SCHOOL

Doctoral Research Project

Decision Making through Human Insight and Artificial Intelligence

Researcher: Melika Soleimani

PARTICIPANT CONSENT FORM - INDIVIDUAL

I have read the information sheet and have had the details of the study explained to me. Any questions I had have been answered to my satisfaction, and I understand that I may ask further questions at any time. I have been given sufficient time to consider whether to participate in this study and I understand participation is voluntary.

I agree/do not agree to the interview being sound recorded.

I wish/do not wish to have my recordings/transcripts returned to me.

I agree to participate in this study under the conditions set out in the Information Sheet.

Signature: ________________________ **Date:** ________________

Full Name-printed ________________________

Te Kunenga ki Pūrehuroa

School of Management
http://www.massey.ac.nz/school-management
Massey University, Private Bag 102904, Auckland 0745 **T** +64 9 414 0800
Massey University, Private Bag 11222, Palmerston North 4442 **T** +64 6 356 9099

Appendix 7- Information sheet

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

MASSEY BUSINESS SCHOOL

Information Sheet

Researcher Introduction

This interview is being conducted by Melika Soleimani as part of her PhD study at Massey University, Auckland, New Zealand. This project is supervised by Professor David Pauleen (Massey University), Dr Ali Intezari (University of Queensland), and Dr Nazim Taskin (Massey University).

Project Description and Invitation

The aim is to understand how Artificial Intelligence (AI) might assist managers to mitigate cognitive biases in management decision making. The focus is on the decisions in the recruitment and selection of personnel. Your involvement in this study will contribute to a better understanding of how AI may assist the process of recruitment and selection in practice.

I would like to invite you to participate in my research by agreeing to be interviewed. The interview will last one hour at a time suitable for you. This interview will be audio recorded and transcribed and you will be offered a transcript to read and comment on. All personal data and the transcription of the interview will be kept confidential. Participants will remain non-identifiable in the final report and publications resulting from this project.

I really appreciate your participation. However, your participation is voluntary. If you agree to participate, you have the right to:

- decline to answer any particular question;
- withdraw from the study prior to – or at any time during – the interview;
- ask for the recorder to be turned off at any time during the interview;
- ask any questions about the study at any time during participation;
- provide information on the understanding that your name will not be used unless you give permission to the researcher;
- be given access to a summary of the project findings when it is concluded.

Project Contacts

Thank you for your time. Please do not hesitate to contact me or any of my research supervisors if you have any questions.

The Researcher:	Melika Soleimani	m.soleimani@massey.ac.nz	+64 2041716063
Supervisors:	Prof. David Pauleen	d.pauleen@massey.ac.nz	
	Dr. Ali Intezari	a.intezari@uq.edu.au	
	Dr. Nazim Taskin	n.taskin@massey.ac.nz	

Te Kunenga ki Pūrehuroa

School of Management
http://www.massey.ac.nz/school-management
Massey University, Private Bag 102904, Auckland 0745 **T** +64 9 414 0800
Massey University, Private Bag 11222, Palmerston North 4442 **T** +64 6 356 9099