The International Journal of Human Resource Management

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rijh20

Reducing AI bias in recruitment and selection: an integrative grounded approach

Melika Soleimani, Ali Intezari, James Arrowsmith, David J. Pauleen & Nazim Taskin

To cite this article: Melika Soleimani, Ali Intezari, James Arrowsmith, David J. Pauleen & Nazim Taskin (20 Mar 2025): Reducing AI bias in recruitment and selection: an integrative grounded approach, The International Journal of Human Resource Management, DOI: 10.1080/09585192.2025.2480617

To link to this article: https://doi.org/10.1080/09585192.2025.2480617

THE INTERNATIONAL JOURNAL OF HUMAN RESOURCE MANAGEMENT
https://doi.org/10.1080/09585192.2025.2480617

Reducing AI bias in recruitment and selection: an integrative grounded approach

Melika Soleimani[a], Ali Intezari[b], James Arrowsmith[c], David J. Pauleen[d] and Nazim Taskin[e]

[a]Southern Cross Healthcare, Auckland, New Zealand; [b]University of Queensland Business School, Brisbane, Australia; [c]School of Management, Massey University, Auckland, New Zealand; [d]Graduate Institute and Department of Business Administration, National Chung Cheng University, Taiwan, Republic of China; [e]Department of Management Information Systems, Boğaziçi University, Istanbul, Turkey

ABSTRACT

Artificial Intelligence (AI) is transforming business domains such as operations, marketing, risk, and financial management. However, its integration into Human Resource Management (HRM) poses challenges, particularly in recruitment, where AI influences work dynamics and decision-making. This study, using a grounded theory approach, interviewed 39 HR professionals and AI developers to explore potential biases in AI-Recruitment Systems (AIRS) and identify mitigation techniques. Findings highlight a critical gap: the HR profession's need to embrace both technical skills and nuanced people-focused competencies to collaborate effectively with AI developers and drive informed discussions on the scope of AI's role in recruitment and selection. This research integrates Gibson's direct perception theory and Gregory's indirect perception theory, combining psychological, information systems, and HRM perspectives to offer insights into decision-making biases in AI. A framework is proposed to clarify decision-making biases and guide the development of robust protocols for AI in HR, with a focus on ethical oversight and regulatory needs. This research contributes to AI-based HR decision-making literature by exploring the intersection of cognitive bias and AI-augmented decisions in recruitment and selection. It offers practical insights for HR professionals and AI developers on how collaboration and perception can improve the fairness and effectiveness of AIRS-aided decisions.

KEYWORDS

Decision-making; Artificial Intelligence (AI); bias mitigation; Human Resource Management (HRM); recruitment and selection; grounded theory

CONTACT Melika Soleimani Soleimanimelika@gmail.com Senior Data Analyst, Southern Cross Healthcare, Auckland, New Zealand.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/09585192.2025.2480617.

Introduction

Illustrative case #1: Consulting firm's AI recruitment challenge

A consulting firm employed an AI recruitment tool to fill a senior leadership role, prioritizing traits like assertiveness and clarity as indicators of good communication. Consequently, the tool overlooked candidates with more reserved but equally effective communication styles. The firm recognized this limitation after observing feedback from recent hires and noticing a pattern of missed opportunities for candidates with alternative communication strengths. This experience has prompted the firm to refine its AI system to better recognize diverse leadership qualities, moving beyond narrow interpretations of communication skills.

Artificial Intelligence (AI) is a collection of programs, algorithms, systems, and machines that mimic human intelligence and is increasingly used in organizational decision-making, including recruitment and selection (R&S) (Upadhyay & Khandelwal, 2018). The adoption of AI technologies in R&S presents opportunities for objective evaluation and recruitment, background checks, compensation development, and potential job-fit prediction (Vrontis et al., 2021). However, challenges such as algorithmic biases may limit the effectiveness of AI in delivering on its promises (Rana et al., 2022).

Algorithmic biases refer to process biases and/or discriminatory outcomes that can occur due to the design, development, or implementation of algorithms in computational systems, including AI systems (Barocas & Selbst, 2016). The embedded source of bias in AI originates from the human behavior it replicates. Bias can stem from multiple sources, including the datasets used to train the algorithms, the design of the algorithms themselves, and human decision-making throughout the development process. If the AI's behavior seems problematic, it is important to remember that it is simply reflecting human actions, as it learns directly from them (Polli, 2019).

Artificial Intelligence biases can manifest when algorithms systematically favor or disadvantage certain individuals or groups. This has been reflected in AI applications for R&S, including candidate ranking and facial recognition, where biases have been identified (Mujtaba & Mahapatra, 2019). These biases often revolve around legally protected characteristics like race, gender, and age (Barocas & Selbst, 2016) and can extend to more subtle characteristics, such as personality types, which although not legally protected, can equally be subject to unfair treatment, thus creating a broader scope for potential bias (Mann & Matzner, 2019).

Cognitive biases have been extensively studied in the psychology and behavioral economics domains (Simon, 1997), with researchers investigating their impact on manager decision-making (Kahneman et al., 2011). Although ethical, legal, and design implications of algorithmic bias have been discussed conceptually (Shams et al., 2025), there is a lack of empirical study exploring these areas in AI (Kordzadeh & Ghasemaghaei, 2022) and this has circumscribed the ability of AI developers and HR professionals to identify and mitigate biases in AIRS.

In R&S, biases pose a formidable challenge in developing AI-based decision-making systems. Addressing these challenges necessitates a cross-functional approach involving the deployment of technologies and human knowledge aimed at reducing biases from algorithms' design and application (Bhattacharya, 2021). By integrating expertise from fields such as information technology (i.e. AI developers) and management, HR professionals and AI developers can better detect and mitigate biases, ensuring AI systems are both effective and ethically sound. This cross-functional collaboration enhances trust in AI-based decision-making (Gressel et al., 2020), as it combines diverse insights to create more robust frameworks for understanding and managing biases inherent in AI applications, including those specific to R&S (de Bruijn et al., 2022).

This study empirically explores HR and developer perceptions of AI biases and their mitigation specifically within the context of R&S. It focuses on potential cognitive biases that may emerge when integrating AI into R&S decisions. In addition to furthering an understanding of AI applications in general, the findings offer specific insight into the application of AI in R&S, addressing the context-sensitive nature of the relationship between technology and HR, as encouraged in recent HR research (Kim et al., 2021). The study emphasizes the importance of cross-disciplinary training between AI and HR professionals to enhance mutual understanding and communication, thereby addressing the often complex and ambiguous issues encountered throughout the AIRS development process. The identification and mitigation of such biases is important as they influence the quality of hiring decisions and thus affect organizational effectiveness and fairness as well as the equity of outcomes for individual candidates (Mann & O'Neil, 2016).

This study addresses the question: How can AI developers and HR professionals collaborate to jointly apply technical skills and people-focused competencies to reduce cognitive biases in AI-based recruitment decision systems? To this end, we first sought to identify the cognitive biases that are more likely to occur in R&S decisions and then explore how these biases may be embedded in AIRS. Based on the findings, we propose a model of AIRS development that integrates these essential competencies. Applying the theory of perception, we further explain the complex mechanisms through which biases can become entrenched in AIRS.

Literature review

As the pioneering computer scientist Joseph Weizenbaum (1976) pointed out, humans exercise judgment while machines make calculations. While AI systems may reflect human cognitive patterns, their biases do not always stem directly from human thought processes but can arise in unexpected ways due to the structure of data and the design and training of ML models within the R&S ecosystem. This literature review focuses on AI as a decision support system in the R&S ecosystem. The R&S ecosystem includes HR professionals, who interpret and evaluate candidate information, and AI developers, who design and implement AI systems like AIRS. Both groups influence how data is perceived and processed, with HR professionals often relying on intuition and experience (indirect perceptions), while AI developers focus on quantifiable data (direct perceptions). Drawing on theories of perception, the review explores how organizing, identifying, and interpreting information shapes cognitive biases that affect recruitment decisions. Figure 1 provides an overview of these interactions, illustrating the flow of perceptions, biases, and decision-making within the R&S ecosystem.

Artificial Intelligence in recruitment and selection

Applications of AI in R&S include automation of tasks, such as candidate screening, shortlisting, and ranking, as well as evaluating their compatibility with a team and predicting their likelihood of retention, leading to a more data-driven approach to recruitment (Ore & Sposato, 2022). The contribution of AI stems from its ability to learn from experience and to use knowledge representation for problem-solving. The 'thinking' capability of AI, which separates it from other decision tools, encompasses problem-solving, reasoning, and learning (Russell & Norvig, 2010). However, challenges exist in accurately capturing and representing real world complexity (Abebe et al., 2020).

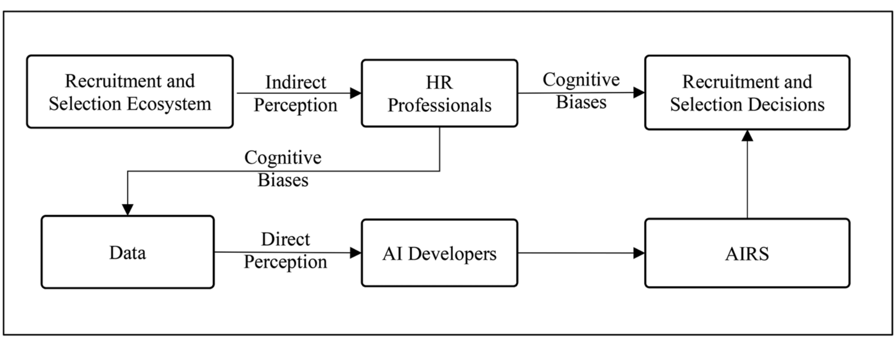

Figure 1. AI-assisted recruitment and selection ecosystem.

Technologies underpinning AI systems, such as machine learning (ML), neural networks, and deep learning, automate and enhance the accuracy of recruitment and selection (R&S) decisions (Davenport, 2018). ML refers to a set of algorithms that allow systems to learn from data, while neural networks are a specific type of ML model inspired by the structure of the human brain (Goodfellow et al., 2016). These networks consist of interconnected layers of nodes, or 'neurons,' that process data and adjust their parameters based on the patterns they detect, enabling more complex forms of learning and decision-making (Goodfellow et al., 2016). Deep learning, a subset of neural networks, utilizes multiple layers to analyze complex patterns in data. Together, these techniques enable AI to adapt and optimize recruitment decisions (Kaplan & Haenlein, 2019). However, the subjective components of R&S, such as job descriptions or character assessments, make the use of ML-based systems problematic (Bogen et al., 2018), a problem likely to be compounded by the emergence of Artificial General Intelligence (Salmon et al., 2023).

Three factors play important roles in determining the reliability of the ML-based systems: data, algorithms, and human perception. Firstly, the utility and reliability of ML-based decision support systems are highly dependent on the quality and volume of data. This challenge distinguishes ML-based systems from conventional decision support systems, where data quality and volume also matter but are not as central to their reliability after development (Davenport & Kalakota, 2019). While large quantities of data empower ML algorithms to discern patterns and enhance their predictive accuracy, poor data quality can undermine the ability of ML models to deliver accurate predictions, particularly when balancing accuracy and fairness, which can lead to biased decision-making (Barocas & Selbst, 2016).

Second, algorithms are the foundations of AI systems, shaping their ability to perceive patterns, make inferences, and execute decisions (Bishop, 2020). The design and decision-making processes inherent in developing these algorithms could introduce biases, significantly affecting the system's outputs. According to Kusner et al. (2017), biases can be introduced in the course of algorithmic design at several key junctures, such as the choices made in model selection, the decision-making process on how to measure the performance or effectiveness of that model, or the methods used to encode fairness into the algorithm.

For instance, if the selection or weighting of certain variables during algorithm development unintentionally favors certain outcomes, or if an algorithm otherwise emphasizes certain features while neglecting others, the resulting decision-making could become biased (Danks & London, 2017). This bias can manifest in the design of the decision-making process within the algorithm itself, where decision boundaries or thresholds set to favor certain outcomes can lead to skewed results (Grgic-Hlaca et al., 2018). As such, even seemingly innocuous design choices can significantly affect the fairness of algorithmic decisions. Notably, defining the target variable in R&S is particularly challenging, as definite performance metrics are often unavailable (Albaroudi et al., 2024). While some objective factors, such as sales numbers, can be used as a metric, they may not be appropriate in all cases. For example, sales volumes may be biased and influenced by external factors like organizational climate or opportunities rather than a candidate's skills and qualifications (Albaroudi et al., 2024).

The third important factor, to which we next turn, is human perception—our understanding and interpretation of the hiring context and AI outputs—which can significantly affect the integration of AI into R&S processes. While the quality of data and the design of algorithms play pivotal roles in developing unbiased AI systems, human perception remains a key influence.

The role of perception in integrating less biased AI for R&S

There are two main schools of thought in the psychology of perception which reflect different perspectives on the role of cognitive processes in shaping our perceptual experiences: the direct (Gibson, 1966) and the indirect accounts of perception (Gregory, 1970, 1980). The former, as exemplified by Gibson's Ecological Approach to Perception, emphasizes the direct and immediate nature of perception and suggests that perception is primarily driven by external information and not subject to hypothesis testing (Gibson, 1979). The indirect account of perception, as exemplified by Gregory's Indirect Theory of Perception, underscores the constructive nature of perception by highlighting the significant role of internal cognitive processes—such as memory, attention, and expectations—in shaping our understanding of the world (Gregory, 1970, 1980). This perspective emphasizes the critical interaction between the perceiver and the environment, illustrating how the brain actively processes sensory information by leveraging past experiences and anticipations to form our perceptual experiences (Chalk et al., 2010).

This dynamic process reflects the interplay of both direct and indirect perceptions, where sensory input and cognitive factors interact to shape our perceptual experiences. While the direct account emphasizes the role of external information in driving perception, the indirect account highlights the influence of internal cognitive processes such as pre-existing knowledge and hypothesis formation, which are heavily swayed by context, task, and individual differences (Feldman, 2015). Central to these cognitive factors are our individual mental models, which are developed through our upbringing and educational experiences (Gressel et al., 2020). These models deeply influence how we interpret sensory inputs and integrate them with our cognitive processes, thereby forming our interpretation of the world (Pearson et al., 2008). Understanding this intricate balance between sensory data and cognitive processes is integral to comprehending how we perceive and interpret our environment (Spivey, 2003).

As discussed, the development of AI for R&S carries the potential to enhance the profiling of candidates, projecting their fit within a team and likelihood of good performance and retention, but it requires a comprehensive set of competencies, such as gathering and amalgamating quality data, devising algorithms for training, and overseeing the learning process of the algorithm. Thus, the involvement of individuals with varied and specialized skills becomes an integral part of the AI development process.

Gregory's Theory of Perceptions as Hypothesis underscores the need for varied expertise in the realm of AI development. As mentioned, the theory posits that an individual's 'knowledge base'—a collection of their prior knowledge, beliefs, experiences, and expectations—directs their responses and interpretations of external stimuli, including environmental information (Gregory, 1980). Effective AI predictions hinge significantly on domain-specific data (Chowdhury et al., 2023), necessitating a substantial degree of cognitive engagement and leading to the inevitable formation of hypothesized perceptions.

In the AIRS development process, AI developers and HR professionals are not simply passive absorbers of R&S-related data points and information. On the contrary, they actively interpret this data, leveraging their unique knowledge bases to navigate multifaceted HR phenomena, ethical conundrums, and employee responses to AI, all of which could contribute to potential algorithmic biases (Giermindl et al., 2022).

The type of input AI developers and HR professionals engage with shapes their reliance on different types of perception. AI developers, trained to work primarily with objective, quantifiable data, may tend to rely more on direct perception, focusing on observable metrics and patterns. In contrast, HR professionals, who deal with more complex, contextualized and subjective human-related phenomena, often rely on indirect perception, drawing on intuition and experience (Highhouse, 2008). The following example highlights how these differing inputs might work in R&S.

For example, AIRS is used in preliminary video interviews to leverage direct perception by analyzing structured and observable data during the interview process. For instance, the AI can evaluate clear communication using speech-to-text transcription to assess sentence coherence, logical flow, and keyword relevance. It can analyze active listening by checking whether responses align with the questions asked and by measuring thoughtful response timing. For adaptability, AIRS might assess behavioral question responses for patterns indicating problem-solving or learning agility, alongside sentiment analysis to gauge attitudes toward challenges. Similarly, immersive technologies like Virtual Reality (VR) and Augmented Reality (AR) could complement AIRS by allowing candidates to engage in simulated job scenarios (Ferreira et al., 2021), such as a hospital simulating operating room conditions to assess a nurse's performance under pressure.

However, AIRS can only complement human final decision-making, as HR professionals might observe that a candidate with moderate technical scores demonstrates exceptional interpersonal skills, such as active listening, clear communication, and empathy when describing past teamwork experiences (Highhouse, 2008). They might also recognize adaptability through examples where the candidate successfully handled sudden changes in project scope or quickly learned a new tool to meet deadlines. These observations, rooted in indirect perception, highlight nuanced qualities that are not immediately quantifiable but are critical for workplace success (Ryan & Ployhart, 2014).

To be clear, this is not a zero-sum relationship. Both roles are essential to the AIRS development process, and their differing perceptions—objective and subjective—complement each other. This balance between direct and indirect perception, as reflected by their respective responsibilities and data focus, is illustrated in Figure 2.

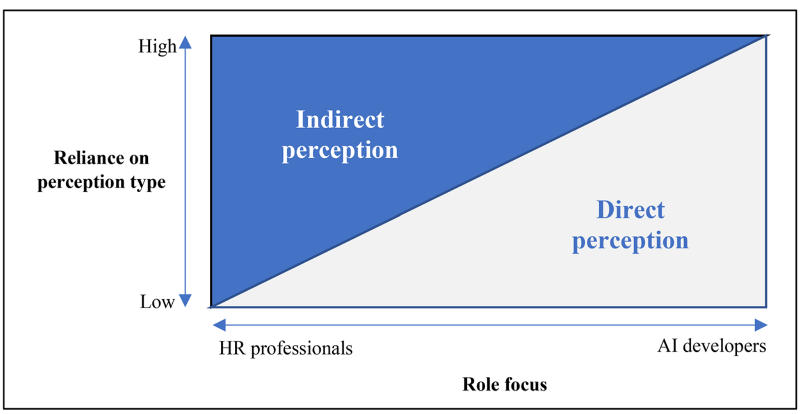

Figure 2. Balance of perception types across professional roles in AI recruitment systems development.

Decision-making and cognitive biases

Illustrative case #2: AI screening bias in a changing talent landscape

> A creative agency's AI tool for candidate screening, initially trained to prioritize traditional career trajectories and industry-specific qualifications, began overlooking a new wave of applicants with non-linear paths—such as career switchers and freelancers—as the market evolved. These candidates, who offered adaptability and innovation suited to the agency's needs, were filtered out due to embedded biases favoring conventional markers of success. Realizing the tool's misalignment with the changing talent landscape, the agency revised its AI criteria to better recognize diverse experiences, enhancing its capacity to attract talent attuned to current demands.

In organizational contexts, recruitment and selection decisions are influenced by individuals' decision-making perceptions, shaped by their unique expectations and past experiences (Gregory, 1980). These perceptions can introduce cognitive biases that may skew the outcome of recruitment decisions (Derous & Ryan, 2019). Understanding the relationship between perception and decision-making can help explain how cognitive biases can arise and exert influence throughout the recruitment decision-making process.

Decision-making is a mental process of 'option generation and comparison' (Schraagen et al., 2008, p. 4) to choose among alternatives (Galotti et al., 2006). It is a dynamic and non-linear process blending rational thought and intuition (Simon, 1993). Decisions required for many sensory-motor tasks can be thought of as a form of statistical inference, underlining the complexity of decision-making processes (Gold & Shadlen, 2007). Personal experience and perceptions substantially influence decision-making processes, notably in the face of incomplete information or amidst environmental changes, which may lead to cognitive biases (Kahneman & Tversky, 1979).

Cognitive biases, viewed as irrational perceptions that affect decision-making (Simon et al., 2000), are crucial elements to consider in the decision-making process, as they can sway or even distort this process. These biases stem from heuristic thinking, or the mental shortcuts that individuals use to simplify complex problems or decisions (Tversky & Kahneman, 1974). Cognitive biases can influence managerial and organizational decision-making, creating potential pitfalls but also areas for improvement. By understanding and acknowledging these biases, decision-makers can employ strategies to mitigate their effects, promoting more accurate and balanced decision-making (Larrick, 2004).

In a study of the decision-making process, Acciarini et al. (2021) outline critical stages including objective identification, information gathering, strategy selection, action execution, and result evaluation. These stages inherently intertwine rigorous data analysis and human judgment, therefore introducing potential cognitive biases into the process. It is within this context that a delicate balance emerges between reliance on data-driven decisions and the influence of cognitive biases.

Despite the potential of high-quality data to mitigate decision-making risk (Merendino et al., 2018), a comprehensive understanding of the context and the application of analytical techniques remain vital for effective decision-making. This underlines the necessity of managing data analysis and cognitive biases simultaneously for better decision-making amid environmental transformations. The interplay between cognitive biases and data-centric methodologies is evident in numerous organizational decision-making processes, including R&S (Davenport et al., 2010; Klimoski et al., 2016; Rasmussen & Ulrich, 2015).

As noted above, AI systems, increasingly prevalent as a data-driven solution for mitigating biases in decision-making processes, are also susceptible to human-induced cognitive biases (Ntoutsi et al., 2020). These biases stem from two predominant sources: the regularities of training datasets and the algorithms themselves (Barocas et al., 2023). The dataset may not be representative of the target population, labeling may be inaccurate or coarse, and cognitive biases prevailing in the environment may be reflected in the data (Johnson, 2021). Algorithmic biases can also arise through design choices and assumptions made during development (Mitchell et al., 2021). These combined biases can lead to socially skewed outcomes, potentially exacerbating inequalities in the workplace and society (Kordzadeh & Ghasemaghaei, 2022). This is particularly concerning in recruitment decisions where biases can obstruct the development of a diverse and inclusive workplace (Whysall, 2018). Recognizing and addressing these algorithmic biases in AI systems is crucial (Friedler et al., 2019). This study therefore explores the potential scope and mitigation of these biases through the perceptions and experiences of the two key stakeholders: developers and HR professionals.

Research methodology

This study employs an exploratory inductive research design using a grounded theory (GT) approach (Glaser & Strauss, 1967). This is appropriate for examining complex, multi-dimensional phenomena such as group and individual decision-making (Intezari & Pauleen, 2018) and was deemed suitable for this study for four reasons. First, it aims to meet the immediate need to investigate methods to regulate potential adverse behaviors associated with AI usage in business decision-making, as highlighted in previous studies (Duan et al., 2019; Kordzadeh & Ghasemaghaei, 2022). Second, the recent emergence of biases in AI (Von Krogh, 2018) and the rapidly increasing use of AI in R&S (Alsaif & Aksoy, 2023) has resulted in a lack of theoretical frameworks and practical solutions for addressing biases in R&S decisions (Hunkenschroer & Luetge, 2022). Third, the study aims to explain, based on field research, how cognitive biases may be introduced into AIRS. Finally, the study seeks to identify effective techniques for mitigating cognitive biases through the varied experiences of practitioners.

The classic GT approach facilitates the inductive identification of variables without preconceived categories and hypothesis (Glaser, 1992). In the later phases of this study, particularly after the emergence of the core category, references were made back to the literature to gain a deeper understanding of the evolving concepts and categories. This comparison with the literature and conceptual mapping facilitated the emergence of the core category (Intezari & Pauleen, 2018).

Participants and data collection

There are growing concerns about a disconnection between AI developers and those who implement and utilize their tools in applied settings, as well as between the technical and social sciences that seek to understand them (Graziani et al., 2023). To address this issue, this study engages both AI developers for their technical insights and HR professionals for their practical perspectives. Despite not interacting directly in an organizational setting, the shared and integrated understanding of these two groups is critical to encompass the full lifecycle of AI in recruitment decisions—from its creation to deployment—for addressing algorithmic bias issues. In all, 39 informants were recruited, comprising 22 HR professionals (56%) and 17 AI developers (44%), who participated in a four-phase interview process.

HR professionals were interviewed in phases 1 and 2, and AI developers were interviewed in phases 3 and 4. This strategy ensured the selection of informants who met specific criteria: HR professionals were chosen based on their experience in R&S and their understanding of the key concepts, principles, and potential applications of AI in the recruitment process, though not necessarily having hands-on experience in developing AI systems. AI developers were selected for their expertise in developing AI for recruitment purposes. Interviewing HR professionals before AI developers was decided upon to first explore the real-world implications and practical issues surrounding potential biases in AIRS. This foundation then guided the exploration of technical solutions with AI developers, ensuring a more complete and grounded insight into the management of algorithmic bias.

Reflecting typical occupational gender distributions, 17 of the 22 HR professionals were women, and five were men, with an overall average of 14 years of work experience in HR; in contrast, of 17 AI developers, two were female and 15 were male. The average work experience of the AI developers was four years, and all had at least one university qualification.

Participants were recruited through social networking platforms, such as LinkedIn, personal connections, and snowballing techniques. The initial ten participants were interviewed in person, while subsequent interviews were conducted remotely *via* an online communication tool. Although in-person interviews were feasible, several participants, including international participants, preferred the convenience and comfort of remote interviews. This flexibility was maintained throughout the data collection process, which spanned a period during and after the COVID-19 pandemic (2020–2022). On average, the interviews lasted approximately one hour, with a minimum duration of 30 min and a maximum of one and a half hours (Appendix 1 and Appendix 2). All interviews were audio-recorded and transcribed by the principal researcher. The transcriptions were cross-checked with participants.

Following Glaser's (2003) approach, we conducted semi-structured interviews with open-ended questions (Appendix 3). In Phase 1, we interviewed 10 HR professionals who were selected based on their experience in recruitment and selection (R&S) processes, particularly in larger organizations where these practices are more formalized. These HR professionals were also chosen for their potential exposure to and familiarity with technological applications commonly used in HR, such as applicant tracking systems (ATS), automated resumé screening, and HR analytics tools. For Phase 2, we further focused on enhancing our theoretical sensitivity by interviewing 12 additional HR professionals. These participants were selected based on their experiences collaborating with AI developers, unlike in Phase 1, where the focus was broader and included HR professionals with varying levels of familiarity with AI systems.

Building on the findings from Phase 1 and Phase 2, it was revealed that HR professionals lacked sufficient understanding of the AI development process and the potential biases inherent in AI systems. In accordance with theoretical sampling (Glaser & Strauss, 1967), new informants (AI developers) were recruited in Phase 3 based on emerging themes such as 'missing data points in R&S datasets,' 'collecting R&S datasets,' and 'providing feedback to AI models.' To explore these issues further, we conducted interviews with eight AI developers. During these interviews, the analysis raised concerns about the techniques used to validate

ML models and mitigate cognitive biases. This led to the recruitment of nine additional AI developers in Phase 4, to explore and address the issues identified by the Phase 3 cohort.

Data analysis

The data collection and analysis process occurred concurrently and iteratively over the four phases, as illustrated in Figure 3. Three coding strategies were employed to identify categories: open coding, selective coding, and theoretical coding (Glaser & Strauss, 1967). Data analysis began with reading the transcriptions and memo-writing to comprehend the context and main points of the data in relation to the research questions.

To systematically analyze the interview data, the principal researcher used open coding, assigning codes to each sentence or paragraph in the transcripts. The goal was to code the data in 'every way possible' (Glaser, 1978, p. 56). Another researcher cross-checked the codes to ensure their accuracy, and any discrepancies were discussed and resolved to ensure consistency in the coding. After each interview analysis, the researchers reviewed the results to ensure complete consistency in their coding.

Similar conceptual codes were identified and grouped 'under more abstract explanatory terms known as conceptual categories' (Strauss & Corbin, 1998, p. 114). The analysis of data gathered in each phase was performed using the constant comparison method, where data was constantly compared with previously collected data to identify

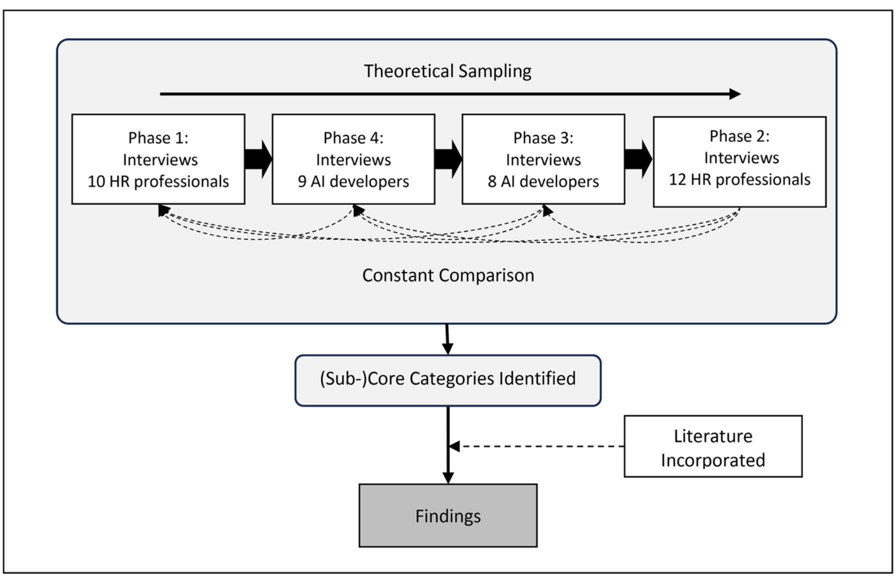

Figure 3. Data collection and analysis process.[3]

similarities, differences, and emerging patterns. This process was detailed and iterative, necessitating manual coding to accommodate the nuanced interpretation and flexibility needed for this analysis, which software like NVivo could not fully support (Davidson & Jacobs, 2008). Through constant comparison, each code and category were continually examined to determine whether a new category was required or the existing categories were sufficient. This process was conducted across interviews and all the interview phases. As more interviews were conducted, the conceptual codes and categories were refined.

For instance, the initial conceptual categories of 'asking [the] right questions' and 'articulating job position requirements' were aggregated into the conceptual category of 'HR professionals' assumptions and job position requirements.' Codes or categories that were identified by only one or two participants and did not come up in other interviews were excluded from the analysis. This increasing theoretical sensitivity (Glaser & Strauss, 1967) led the data analysis to become increasingly focused on conceptual terms, such as the AI development process and the collaboration between HR professionals and AI developers in developing AI for R&S decisions. This constant comparison allowed for the refinement and elaboration of the properties of the initial concepts and categories (Glaser & Strauss, 1967).

In Phase 4, data analysis started becoming theoretically saturated (i.e. no new data or concerns emerged in the new interviews). To ensure theoretical saturation, we continued data collection until no new theoretical concepts emerged. By the end of Phase 4, 16 conceptual codes were identified, leading to the emergence of five conceptual categories and two sub-core categories.

Two sub-core categories were identified: cognitive biases, and how the biases are embedded in AIRS. The sub-categories represent potential cognitive biases and three aspects of how these biases are embedded within AIRS, which are discussed in more detail in the following sections.

Findings

The aim of this study was to explore practitioner perceptions of bias and specifically how to mitigate biases that could significantly influence AIRS-based recruitment decisions. To this end, we conducted semi-structured interviews with HR professionals and AI developers around the world. We were particularly interested in understanding (1) what cognitive biases are likely to slip into AIRS, and (2) how the biases are embedded within AIRS. This would help us understand how cognitive biases could be mitigated in AIRS. Following the presentation of the findings of which cognitive biases are embedded in AIRS and how, we will explain, in the discussion section, how the biases can be mitigated.

The sub-core categories and conceptual categories, along with their definitions, are presented in Appendices 4 and 5.

Cognitive biases in AIRS

The findings revealed that two common cognitive biases that happen in R&S decisions are likely to end up in AIRS: the *similar-to-me* bias and the *stereotype* bias. These two biases predominated in our findings, though other biases, such as the halo effect (Zebrowitz & Montepare, 2008) and the horns effect (Radeke & Stahelski, 2020), were also identified and are widely recognized in the literature. However, these effects can be viewed as forms of stereotypical inferences (Radeke & Stahelski, 2020; Shao, 2023).[1]

The similar-to-me bias involves preferences for candidates who, for example, attended the same school as the recruiter or share common habits, interests, or are seen as more enjoyable companions by the recruiters. Stereotype bias refers to a superficial evaluation of a candidate's background, which may result in undue favoritism or prejudice against certain ethnic groups, candidates with specific amounts of experience, or those employed by either renowned or local companies. While evaluating a candidate's background can be an effective way to assess their competencies, it can lead to biased decisions if done algorithmically (Barocas et al., 2023).

Similar-to-me bias

Similarity tends to enhance the chances of gaining agreement with one's opinions. Interviewers often have a pronounced inclination to seek social affirmation of their thoughts and beliefs, leading them to find favorable similarities with those they are interviewing (Fontana, 2023). Participants in this study provided instances where discrepancies in values, such as religious beliefs, could lead to a candidate being overlooked or rejected.

The findings revealed that one reason that the similar-to-me bias occurs in hiring decisions is that HR professionals often find it easier to communicate with candidates when they perceive similarities between themselves and the applicants (for evidence see Appendix 5). Moreover, they justify the bias as a criterion for good fit. For example, our participants noted that HR professionals might assess candidates for being a good fit within the team based on factors like sharing 'the same hobbies.' The following statement exemplifies this perspective among HR professionals:

> I remember a candidate who had a background in running community outreach programs—something the manager was personally passionate about—was seen as a better cultural fit. The manager mentioned that they found it easier to engage in conversation and felt the candidate would integrate more seamlessly with the existing team. Conversely, a technically stronger candidate, who did not share these interests, was seen as less of a match, simply because the rapport during the interview didn't feel as natural. (HR_2)

Such criteria might not actually be pertinent when considering whether a candidate would be a genuine team player. However, the normalization of the bias makes it unnoticeable during the AIRS design and development process. The concern was raised by the participants, stressing that there is no clear distinction between the 'similar-to-me' bias and ensuring a candidate is a good fit for the team. We did not find any information (either from the HR professionals or AI developers) as to whether and how the bias is counterbalanced or moderated in AIRS so that AIRS uses the 'similar-to-me' criterion to genuinely assess the good fit of the candidate.

While the 'similar-to-me' bias appeared to be the major cognitive bias that could imperceptibly become embedded in AIRS, we also found another category of biases, which could be similarly difficult to pick up on during the AIRS design and development: the stereotype biases.

Stereotype bias

Stereotype bias is defined in the literature as a fixed or prejudiced perspective on an individual based on their belonging to a particular social category, such as ethnicity, age, or gender. This view diminishes their potential and overlooks the variety within the group (Hinton, 2019). Our participants identified specific indicators of this bias, such as evaluating candidates solely based on their place of birth and proficiency in English, pointing to a bias towards certain ethnicities:

> 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_13).

What makes the stereotype bias undetectable during the design and development of AIRS is that it is rooted in the common practice of heuristically assessing a candidate's professional and personal background. While an HR professional may be able to moderate or augment their own experience-based preconceived view about a candidate's professional and personal background to assess their overall competences, AIRS do not have this ability. Once embedded into AIRS, the stereotype bias may continue unchecked. For example, the following sentiment shows how HR professionals sometimes have conscious biases towards candidates, although they would not necessarily identify such judgments as biased, perhaps considering it to be useful tacit knowledge. The participant explains his bias when he judged a candidate based on having five years of experience in a specific organization that he was familiar with:

> 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_15).

This kind of human perception and judgment is difficult, if not impossible, to include in AIRS. For this reason, AI developers rely on data-driven analysis to make ML models accurate. While this can be seen as a benefit of AIRS, the downside is that it might reduce the importance of human judgment and expertise. The focus of AIRS would likely shift from a deep, thoughtful evaluation of a candidate's suitability to an automated, algorithm-based assessment of many applications (See more quotes about 'Stereotype biases' in Appendix 5).

How the biases are embedded in AIRS

Our findings revealed that the biases are likely to be embedded in AIRS due to (1) flawed interpretation-algorithm transition, (2) flawed foundations-bias through data and design, and (3) flawed cycles-feedback stagnation and bias propagation.

Flawed interpretation-algorithm transition

Understanding the assumptions and requirements that HR professionals have for each job position is a pivotal aspect for AI developers who are seeking to select precise and relevant training datasets. For example, HR professionals may perceive that a good leader must have strong communication skills, and they rely on information and cues during interviews, such as observing how a candidate speaks, engages, and responds (Whysall, 2018).

This indirect perception (as it engages HR professionals' high-level cognitive processes such as memory and expectations) shapes their evaluation of candidates face-to-face (Hunkenschroer & Lütge, 2021). They often equate strong communication with traits like assertiveness and confidence, which are easily detected through direct interaction during the interview process.

In contrast, when AI developers are defining 'strong communication skills' for an algorithm, they rely on direct perceptions, shaped by available data and their technical knowledge (Marinucci et al., 2023). This translation process can introduce errors, as AI developers may define communication strength through quantifiable indicators like speaking clearly, forcefully, and without hesitation, leading AIRS to prioritize these measurable traits.

As a result, the system may end up overlooking those who possess strong leadership qualities but communicate in more collaborative or quieter ways (Albassam, 2023). Interviewees who engage in more thoughtful, reserved communication styles may be disadvantaged by the AI's narrow interpretation of what constitutes strong communication, which was shaped by indirect perceptions rather than the direct sensory cues available to HR professionals.

Understanding the contextual job requirements, however, can be highly challenging as the requirements may be interpreted differently by HR professionals and again when the understanding is transferred to the AI developers. This dilemma is captured by this participant:

> 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_7).

Both AI developers and HR professionals highlight the importance of domain expertise. Identifying appropriate training datasets relies heavily on an understanding of contextual knowledge. Without a comprehensive and precise understanding of domain-specific requirements, AI developers may encounter challenges with datasets that do not align with the R&S objectives. As this HR professional explains:

> What we need to do before AI can really help us [is] fix the start of the process in terms of understanding what it is that we're looking for first, so that everyone knows, everyone at the start of the process is aware of what we're looking for and drawing the right information out (HR_8). (For more supporting data, see Appendix 5).

For this reason, when HR professionals' perception and interpretation of the job position requirements are not transferred to the AI developers correctly, the algorithms that AI developers write fail to reflect the actual job position assumptions and requirements. Such a flawed interpretation-algorithm transition can lead to unintentional and undetectable biases in AIRS.

Flawed foundations-bias through data and design

When AI developers embark on creating AI systems, they often rely on historical data as the foundation for model development. However, addressing potential biases within this data is frequently an afterthought, rather than a priority, during the initial stages of development (Davies, 2023).

Historical data often contains biases due to several factors, such as the inclusion of irrelevant data points, imbalanced datasets, missing entries, flawed data labeling, and inconsistencies in data preprocessing methods. This collection of biases can significantly affect the fairness of AI systems; even when AI developers curate these biased datasets, their indirect perception come into play (Dwork, 2023; Gatzemeier, 2021).

The disparate impact of data and algorithms design is likely to happen in AIRS, because the R&S decisions are traditionally subjective (heavily engaging indirect perception), and not necessarily data driven. AI developers have access to limited datasets from past decisions related to specific job positions. As explained here:

> 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_2).

Our findings reveal that the limited availability of datasets in R&S can increase the risk of biases within AIRS. Data scarcity underscores a prevailing challenge within HR, where decisions often lack a robust data-driven basis, potentially leading to indirect bias:

> What is kind of difficult is that we don't always get the whole picture when it comes to data, we don't always know if this person got hired (AI_4). (For more supporting data, see Appendix 5).

The findings also spotlight inconsistencies and lapses in the data harnessed for R&S decisions, which is a recurrent issue within many organizations. These shortcomings could undermine the completeness of the AI dataset, leading to a decrease in the overall accuracy of the AIRS. as explained by this participant:

> We don't feed through, like total information into data, so they're not getting everybody's information (HR_18).

These foundational biases in data and design pose significant challenges to the fairness and accuracy of AIRS. Without proactive efforts to address data limitations and inconsistencies, AIRS is likely to replicate historical biases, potentially exacerbating inequities in R&S.

Flawed cycles-feedback stagnation and bias propagation

Developing and retraining ML models emerges as a potential avenue through which biases may infiltrate AIRS. The ML model used in AIRS degrades over time due to the lack of continuous updating and retraining with new data. This stagnation can occur if there is no feedback from domain experts. AI developers may assume that once the model is operational, it continues to perform effectively without requiring significant updates or refinements, which can allow biases to persist over time.

This study's findings support the idea that HR professionals should be providing continuous feedback based on interpreting the outcomes of AIRS (indirect perception). By doing so, they can identify biases that developers might overlook due to their technical focus on models. HR professionals' involvement ensures the model remains dynamic and responsive, bringing in real-world contextual understanding that can make the system fairer and more adaptable to the complex social dynamics influencing hiring decisions.

> We often receive user feedback regarding their preferences. They might indicate that a candidate doesn't fit a specific job. This misfit could be due to various reasons: the candidate might lack certain skills, the title might not match, the location might be unsuitable, or the candidate might not have sufficient experience. Essentially, users provide us with various signals that guide our understanding (AI_13).

This research underscores the need for an adaptive approach to ML development in AIRS, highlighting that building a model is just the start; ongoing monitoring and refinement based on user feedback and evolving requirements are crucial. The findings indicate that AI developers must be vigilant in retraining and updating ML models in response to these ongoing changes. Failure to do so, according to the insights gathered, could result in issues, such as incorrect predictions (i.e. predictions that might be accurate but not fair), which can lead to undetected algorithmic biases within AIRS.

> I think 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, for any errors, and then we train it again, and that's how it improved by the time (AI_10). (For more supporting data, see Appendix 5).

Essentially, the persistence of cognitive biases within AIRS calls for an adaptive approach where continuous feedback and targeted retraining become integral. The following section presents a structured process for refining AIRS to remain responsive and minimize embedded biases over time.

Discussion

This study aimed to identify the scope and nature of AI bias in R&S, and how these might be mitigated, based on the perceptions and experiences of the two key practitioner stakeholders. Our findings indicated that at least two biases, 'similar-to-me' and 'stereotype,' are at risk of becoming integrated into AIRS, due to three factors: flawed interpretation-algorithm transition, flawed foundations-bias through data and design, and flawed cycles-feedback stagnation and bias propagation. Drawing upon the findings, we propose a model of the development process for a less biased AIRS. We employ the theory of perception to explain the intricacies of the HR-AI multidomain expertise inherent in AIRS development.

The AIRS development model

This model presents a multi-phase process comprising three phases—understanding the ML model requirements, managing datasets, and developing and retraining ML models—in an iterative manner as depicted in Figure 4. Each phase engages multiple techniques that need to be implemented by the HR professionals and AI developers to reduce cognitive biases.

Phase 1: Understanding the requirements of the ML models

In the initial stage of the development process for AIRS, understanding the requirements of ML models takes center stage. HR professionals need

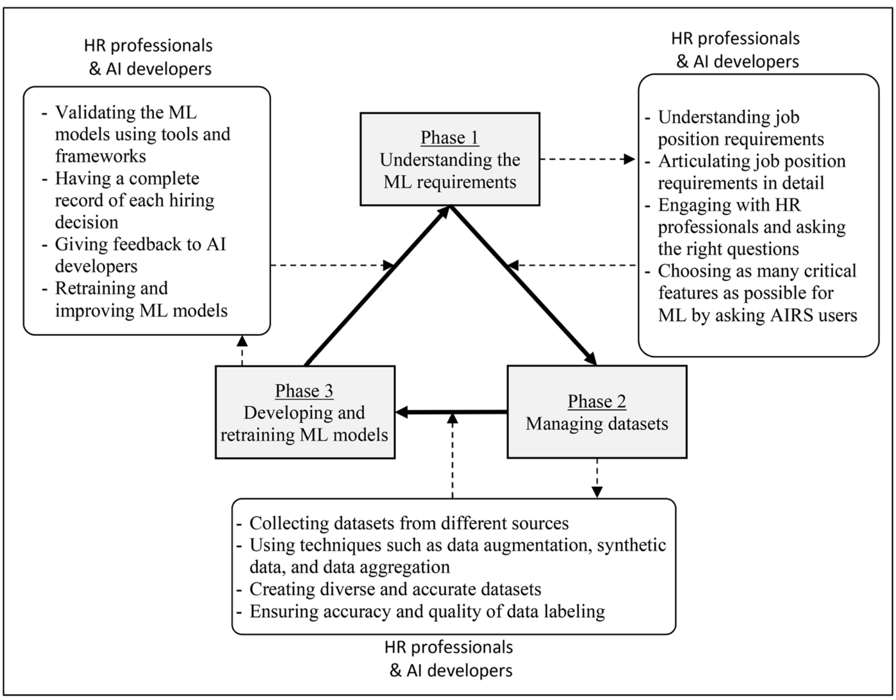

Figure 4. The development process of unbiased airs and cognitive bias mitigation techniques.

to communicate their expectations and assumptions to AI developers.[2] The findings suggest that these expectations often stem from high-level cognitive interpretations, which are difficult to fully translate into the structured and measurable formats AI developers require. This complexity is consistent with Gregory's theory of perception, where perception is not a simple reflection of reality but rather an interpretation or hypothesis formed from prior knowledge (Gregory, 1980).

The challenge lies not just in different perception styles between HR professionals and AI developers but also in how these perceptions are communicated and understood. Specifically, HR professionals' understanding of R&S requirements must be clearly conveyed so that AI developers can interpret and implement them effectively. Miscommunication or reliance on inappropriate selection criteria can increase the risk of biases being introduced into the system (Tambe et al., 2019). Therefore, both HR professionals and AI developers need to be mindful of these details and skilled in ensuring a clear and accurate exchange of information.

In light of these communication challenges, a deeper understanding of the nuances in R&S is essential. This understanding extends beyond surface-level knowledge, encompassing the intricate domain knowledge, contextual relevance of each role, and the distinct needs of the organization. Without capturing this depth, as highlighted in the findings, AI developers may select datasets that fail to reflect nuanced job requirements, leading to misaligned outputs. Addressing these biases early, particularly with historical data, is critical to prevent systemic issues from embedding within the AI system.

To minimize potential biases, AI developers must gain a solid grasp of the domain context to select appropriate and reliable training datasets (Barocas et al., 2023). Since ML-based systems derive their rules from these datasets, it is essential that AI developers ask the right questions and fully understand the job requirements communicated by HR. While training on the fundamentals of R&S can enhance communication between HR professionals and AI developers, it is essential for AI developers to develop a deeper understanding that extends beyond basic knowledge. Ideally, they should strive to grasp the high-level cognition involved in R&S. This can be achieved, for example, by having AI developers study the R&S scenarios previously handled by HR professionals. The cases should closely resemble the job(s) for which AI developers develop AIRS.

Similarly, HR professionals should aim to comprehend the complex technical processes of AI development by examining cases that mirror the challenges developers face (Tambe et al., 2019). This mutual learning will foster more effective communication between HR professionals and

AI developers which ensures that the system aligns with business needs and that key performance indicators (KPIs) are met. The ability to identify critical features for candidate selection and mitigate bias depends heavily on this collaborative effort (Bodie et al., 2017), as the findings emphasize the importance of aligning HR professionals' domain knowledge with AI developers' technical expertise.

Whether an AI model is customized for a specific role or designed to serve multiple positions, the risk of bias through misinterpretation persists. For example, some AI hiring platforms, like HireVue, rely on pre-built models adaptable to a range of job types rather than bespoke models tailored for every position (HrTechcube, 2018). These models adapt to competencies such as leadership or technical skills, though custom solutions are sometimes developed, as seen with the tailored models for 26 North American employers (Kassir et al., 2023).

Phase 2: Managing datasets

Once the requirements for the ML model are established, the next crucial step is managing datasets, which involves careful data collection and preparation (Goodfellow et al., 2016). Our study underscores that relying solely on datasets provided by HR professionals can be insufficient for training ML algorithms (Raghavan & Barocas, 2019). For instance, critical data points—such as the final outcomes of recruitment decisions—may be missing, introducing potential biases due to incomplete information (Köchling & Wehner, 2020).

Imbalanced datasets present another challenge, as overrepresentation of certain groups can skew the model's predictions (Mehrabi et al., 2021). Our findings corroborate this issue, revealing that limited datasets in R&S often fail to capture the diversity of candidate backgrounds. For example, if candidates with corporate backgrounds dominate the dataset, the model might prioritize corporate skills over equally valuable traits found in non-traditional backgrounds, potentially disadvantaging capable candidates with entrepreneurial experiences (Bohnet, 2016). Additionally, irrelevant datapoints may be included based on the mistaken belief that certain characteristics (overemphasis on formal qualifications and specific skills relevance) are more predictive of success without supporting evidence (Raghavan et al., 2019).

To address these issues, our findings suggest the necessity of robust data preparation techniques—such as data augmentation, synthetic data generation, and dataset aggregation—are essential to ensure a more comprehensive and less biased training dataset. Data augmentation can balance existing datasets by introducing new or counterpart data or transforming existing data to expand the dataset's scope, although this is complex in practice (Polyzotis et al., 2017). In addition, using synthetic data generation can help to fill in gaps where real-world data is scarce (Nowruzi et al., 2019). For example, synthetic candidate profiles can be generated to simulate candidates from underrepresented groups (e.g. applicants from entrepreneurial backgrounds or those with career gaps due to caregiving) and test the features that have been selected for ML models (Bolón-Canedo et al., 2013). These profiles can introduce diversity into the dataset and help the system learn patterns beyond traditional candidates.

However, the use of synthetic data requires careful consideration to avoid perpetuating biases to implement context-aware generation techniques and validate synthetic data against real-world scenarios to ensure it aligns with the intended diversity objectives without compromising data integrity (Lu et al., 2021). If synthetic data is generated based on biased patterns from historical data, it could end up replicating those biases. To mitigate this, synthetic data generation in R&S should involve fairness-aware algorithms, which explicitly correct for overrepresented features (e.g. corporate background, specific qualifications) and ensure balanced representation of diverse candidate profiles (Varshney, 2018). For instance, synthetic candidates might be created with different combinations of qualifications and experiences, focusing on skills rather than purely formal titles, to avoid overemphasis on a specific type of career path (e.g. corporate vs. entrepreneurial).

By diversifying data sources and expanding the dataset, AI developers can reduce the influence of biases, transforming them into more balanced, objective insights. Consistent with our findings, participants highlighted the importance of integrating diverse data sources to enhance the representativeness of the training dataset. This process enhances the fairness and robustness of the AI system, reducing the risk of bias replication (Lepri et al., 2018; Mehrabi et al., 2021). Data integration consolidates data from various sources (IBM Corporation, 2020), such as industry-wide databases, public employment statistics, or anonymized candidate pools from various regions so the AI system can better account for a broader range of job requirements and applicant profiles. This strategy helps in reducing the risk of perpetuating biases by increasing the representativeness of the training dataset (Gebru et al., 2021).

Flaws in data preparation can also emerge through improper handling of missing entries, inaccurate or coarse data labeling, and disregarding outliers that may represent marginalized groups. Our findings indicate that these issues are prevalent in many organizations, undermining the completeness and accuracy of AI datasets. Addressing missing entries often involves imputation or normalization, yet inconsistent application across subgroups can distort results (Little & Rubin, 2020). For instance, gaps in employment history can introduce bias if interpreted inconsistently. A woman's employment gap might be assumed to be due to maternity leave and viewed more favorably, while a similar gap for a man could be associated with career instability, potentially leading to unfair treatment.

Inaccurate or overly simplistic data labeling further compounds biases (Suresh & Guttag, 2017). When labels rely solely on observable attributes, such as job titles or years of experience (e.g. 'Junior Developer,' 'Senior Developer'), they may ignore the full scope of a candidate's skills and achievements. This reliance on surface-level metrics can produce misleading results, as deeper qualifications and skills go unrecognized.

Additionally, downplaying outliers that do not align with dominant trends can introduce bias, particularly if those outliers reflect the unique experiences of minority candidates. For example, in NLP-based resumé parsers, international qualifications or experiences may be misinterpreted or undervalued, as assumptions about equivalence in educational terms go unexamined (Pagano et al., 2023). By failing to account for these differences, the system risks skewing assessments toward more familiar qualifications and undervaluing diverse backgrounds.

A structured, fairness-aware approach to data preparation that considers subgroup-specific contexts, applies consistent methods, and values diverse profiles can mitigate these risks. Through thoughtful data labeling and accurate handling of unique experiences, the AIRS system can better represent varied candidate profiles, reducing biases introduced through labeling and preprocessing inconsistencies.

Phase 3: Developing and retraining ML models

If a machine learning model is being developed in collaboration with domain experts, continuous feedback is essential to refine and improve the model. As highlighted in the findings, the lack of continuous updating and retraining can lead to bias propagation within AIRS, underscoring the necessity of ongoing feedback loops. Without this iterative feedback, the model might not evolve to meet the specific requirements, nuances, or complexities of the domain. This form of stagnation is less about the mathematical or algorithmic aspects and more about the model's alignment with domain-specific needs and expertise.

The development of ML algorithms in AIRS is an iterative process that commences with determining the genuine success criteria of job positions. The findings indicate that HR professionals play a crucial role in providing continuous feedback, which helps in identifying biases that might be overlooked by AI developers focused solely on model accuracy. During the AIRS development process, it is important to minimize feedback loop stagnation to ensure that the ML models are continually monitored, developed, and retrained to identify factors leading to successful candidate selection.

At this juncture, it is critical to acknowledge the role of human perception in this process. The ecological theories of perception, as proposed by Gibson (1950) and Gregory (1980), suggest that HR professionals play a crucial role in providing this feedback by interpreting AI outputs. Their indirect perceptions—shaped by their experience, expectations, and understanding of R&S—are essential to spotting biases that AI developers may overlook, given the AI developers' reliance on their direct perceptions, such as the models' accuracy.

According to our findings, HR professionals should maintain a comprehensive record of every hiring and non-hiring decision made using AIRS over time. Furthermore, ML models must be recalibrated as variables in R&S evolve, especially given the rapid employment changes affecting the career landscape in recent decades (Bessen, 2018). Thus, the ongoing development process of ML models in AIRS requires the expertise of HR professionals to enhance algorithm performance, as highlighted by the participants.

The complexity of developing ML models involves selection, adjustment, and training for effective performance (Goodfellow et al., 2016). To ascertain the optimal ML model, different modeling approaches are evaluated on test datasets. During this phase, AI developers experiment with various modeling approaches to assess performance and determine the optimal model based on success metrics, enabling AIRS to continually improve and adapt to evolving circumstances.

Cai et al. (2018) describe the process of feature selection, which involves mitigating irrelevant and redundant features, as a means of enhancing ML models. Careful selection and engineering can significantly improve the accuracy of ML models, streamline the understanding of both the model and the underlying data, and bolster overall performance (Zheng & Casari, 2018).

Moreover, research demonstrates that simplifying neural network models through moderate compression not only streamlines the training process but also enhances their fairness (Good et al., 2022). By strategically pruning less crucial connections within these networks, developers can achieve a more uniform performance across diverse groups, thereby reducing bias and improving the accuracy of the models. This underscores our findings, emphasizing the importance of adaptive ML development in addressing and minimizing embedded biases over time. This approach aligns with ongoing efforts to refine ML model training and feature selection, further boosting the robustness and adaptability of AI systems (Good et al., 2022).

While ML models excel at recognizing patterns, they struggle to grasp the deeper sociocultural and contextual subtleties that influence bias, particularly in hiring decisions. For example, while a model may detect protected categories and apply corrective measures, it still lacks a full understanding of the broader social dynamics and nuances (Albaroudi et al., 2024). Our findings support this limitation, highlighting that without continuous feedback from HR professionals, AIRS may perpetuate biases despite technical safeguards. In R&S, bias is shaped by various intricate contextual factors, and AI's inability to fully capture these ongoing nuances makes it vulnerable to perpetuating biases in different contexts.

Conclusions, implications, and future research

This study underscores the potential problem of biases in AIRS that can affect recruitment decision-making, guided by insights obtained from interviews with HR professionals and AI developers. The study identifies the 'similar-to-me' and 'stereotype' biases as the most common ones that frequently appear in historical R&S data and risk becoming embedded in AIRS. Based on these findings, this study proposes a three-phase iterative process for constructing less biased AIRS, encompassing understanding ML model requirements, managing diverse and substantial datasets, and continually developing and retraining ML models.

This study makes a theoretical contribution by bridging cognitive psychology, particularly the notion of perception, and the software design field. It integrates Gibson's direct perception theory (1950) and Gregory's indirect perception theory (1980) within the AI development process for recruitment decision-making (Figure 5).

The theoretical contribution of this study sheds light on the process leading to biased AIRS, which is challenging to identify, as the agent of biases is often hidden, and the cause is not easily recognizable in the post-development phase. Further, this study, built on a grounded theoretical framework, suggests that biases in AIRS can be reduced through understanding ML requirements, developing diverse datasets, and iterative model development and training. However, this process requires not only collaboration between HR professionals and AI developers, but also adequate training for both parties to bridge their different paradigms and ensure effective communication.

Our findings suggest that AI developers' technical approach to interacting with datasets aligns with Gibson's theory of direct perception (Gibson, 1950), where understanding is based on observable data with minimal subjective interpretation. In contrast, HR professionals, drawing from their experience and domain knowledge, reflect Gregory's theory of indirect perception (Gregory, 1980), where decision-making is shaped by prior knowledge, assumptions, and interpretations. This complementary

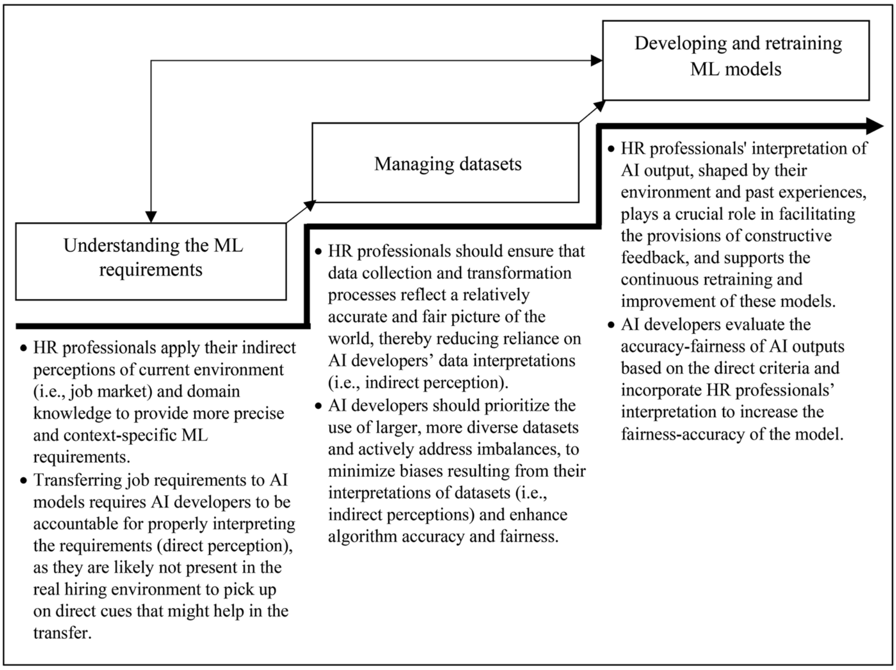

Figure 5. Theoretical contribution to Gregory's theory of perception in the airs development process.

interaction between AI developers and HR professionals highlights the need for both approaches in the development of AIRS.

This interaction underscores the critical importance of HR professionals' experiential knowledge in identifying biases that AI developers, focused on direct data interpretation, might overlook. It also reflects the value of AI developers' data-driven insights and their direct perception in enhancing the accuracy of ML models. By combining these perspectives, the cross-functional collaboration becomes essential for mitigating biases in AI systems. Such a collaborative approach not only helps reduce algorithmic bias throughout each stage of AIRS development but also emphasizes the need for a comprehensive strategy that integrates technical, ethical, legal, and design considerations.

The study's findings have practical implications for individuals involved in the development and implementation of AIRS. First, it highlights the necessity of technical due diligence and data literacy for both AI developers and HR professionals. Organizations must ensure that developers understand the organizational context and how to create appropriate and unbiased algorithms, while HR professionals need to be equipped to communicate requirements and effectively use AI tools. Second, bias mitigation techniques are essential for developing ethical AI systems. AI developers and HR professionals must collaborate to integrate bias mitigation strategies during the tool's design phase. For example, some AI solutions proactively eliminate gendered language in CVs to reduce unconscious bias. Furthermore, by incorporating social category data, AI developers can identify and address patterns of discrimination that may otherwise remain hidden. Regular auditing of AI tools to detect bias or data inaccuracies is necessary to maintain fairness throughout the hiring process. Such auditing is a cross-functional task and therefore must be conducted by a cross-disciplinary team consisting of both HR professionals and AI developers.

Finally, the potential of AI in promoting more objective and relevant decision-making, and as a consequence facilitating diversity, is its key practical contribution. However, this requires investment in both technology and human expertise. AI developers and HR professionals must be trained to understand and mitigate biases in AI systems. By continually refining AI tools and maintaining a proactive approach to auditing, companies can use AI to foster greater diversity and equity in their hiring processes, ultimately leading to improved organizational performance and innovation.

The present study contains limitations which in turn offer opportunities for future research. While the qualitative approach provides context-sensitive insights, future studies could test their broader applicability through qualitative research in other domains or quantitative methods across varied organizational and occupational contexts. Interviews with both successful and unsuccessful candidates could offer deeper insights into AIRS's impact on outcomes and perceptions.

This study identifies two broad categories of biases, but its ultimate goal is to contribute to their mitigation. Future research could expand these categories into more detailed subcategories to refine strategies for addressing biases. Incorporating mixed methods, such as case studies of AI design and implementation or studies of HR-AI dyads, would provide more nuanced insights. Additionally, investigating alternative interview formats, such as panel or group assessments, and involving a wider range of participant roles could further validate and enhance our understanding of biases in recruitment.

In this study, the gender and age imbalance among AI developers and HR professionals is acknowledged as a limitation (Charmaz, 2014), though one that tends to reflect the current realities of the HR and AI disciplines. Grounded theory demands careful equilibrium between pursuing the emerging theory and encompassing a variety of experiences and perspectives. Statistical representativeness is not the primary objective, but the depth and richness induced by diversity in theoretical sampling are indispensable for crafting a robust and nuanced theory that elucidates the complexity and diversity of the studied phenomenon (Charmaz, 2014).

The shared responsibility of HR managers and AI developers in designing and using AIRS underscores the importance of their collaboration, as AI alone cannot be held fully accountable. Future research could explore which recruitment and selection processes are most suitable for full or partial AI delegation. Addressing the gender and age imbalance among HR professionals and AI developers could also enrich theoretical diversity and enhance representation.

Lastly, given the global scarcity of AI developers with R&S expertise, this study's proposed framework provides a starting point for reducing biases in AI applications. Future research could focus on individual applications to enhance the specificity and effectiveness of bias mitigation strategies.

Notes

1. While identifying biases in this study is important, the primary objective is to mitigate them.
2. Communicating expectations and assumptions between HR professionals and AI developers is critical but presents significant challenges. This complexity may warrant further exploration or future research to identify effective approaches.
3. The ordering of the phases in this figure is designed to emphasize the iterative process of constant comparison, highlighting how insights from earlier phases informed subsequent ones, rather than strictly reflecting the chronological sequence of data collection.

Acknowledgments

The authors sincerely appreciate the editor and reviewers for their valuable feedback and constructive suggestions, which have greatly contributed to improving this work. We also extend our gratitude to the study participants for generously sharing their time and insights. This research was supported by a Massey University scholarship awarded to the first author for her PhD. The dataset used in this study is openly available in the Massey Research Online repository: http://hdl.handle.net/10179/17686.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

James Arrowsmith http://orcid.org/0000-0002-9205-4190

References

Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. G. (2020). Roles for computing in social change. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 252–260.

Acciarini, C., Brunetta, F., & Boccardelli, P. (2021). Cognitive biases and decision-making strategies in times of change: A systematic literature review. *Management Decision*, *59*(3), 638–652. https://doi.org/10.1108/MD-07-2019-1006

Albaroudi, E., Mansouri, T., & Alameer, A. (2024). A comprehensive review of AI techniques for addressing algorithmic bias in job hiring. *AI (Switzerland)*, *5*(1), 383–404. https://doi.org/10.3390/ai5010019

Albassam, W. A. (2023). The power of Artificial Intelligence in recruitment: An analytical review of current AI-based recruitment strategies. *International Journal of Professional Business Review*, *8*(6), e02089. https://doi.org/10.26668/businessreview/2023.v8i6.2089

Alsaif, A., & Aksoy, S. (2023). AI-HRM: Artificial Intelligence in Human Resource Management: A literature review. *Journal of Computing and Communication*, *2*(2), 1–7. https://doi.org/10.21608/jocc.2023.307053

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press.

Barocas, S., & Selbst, A. (2016). Big data's disparate impact. *Law Review*, *104*(3), 671–732.

Bessen, J. (2018). AI and jobs: The role of demand. In *NBER WORKING PAPER SERIES*.

Bhattacharya, D. (2021). Competing in the age of AI: Strategy and leadership when algorithms and networks run the world. *Strategic Analysis*, *45*(3), 264–266. https://doi.org/10.1080/09700161.2021.1918951

Bishop, J. M. (2020). Artificial Intelligence is stupid and causal reasoning will not fix it. *Frontiers in Psychology*, *11*(January), 513474. https://doi.org/10.3389/fpsyg.2020.513474

Bodie, M. T., Cherry, M. A., McCormick, M. L., & Tang, J. (2017). The law and policy of people analytics. *University of Colorado Law Review*, *88*, 961–1027.

Bogen, M., Rieke, A., & Ahmed, A. (2018). Help wanted: An examination of hiring algorithms, equity, and bias. *Upturn* December.

Bohnet, I. (2016). *What works: Gender equality by design*. Harvard University Press.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems*, *34*(3), 483–519. https://doi.org/10.1007/s10115-012-0487-8

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

Chalk, M., Seitz, A. R., & Seriès, P. (2010). Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision*, *10*(8), 2–2. https://doi.org/10.1167/10.8.2

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

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

Danks, D., & London, A. J. (2017). *Algorithmic bias in autonomous systems* [Paper presentation]. IJCAI International Joint Conference on Artificial Intelligence, *0*(January), 4691–4697.

Davenport, T. (2018). *The AI advantage: How to put the artificial intelligence revolution to work*. MIT Press.

Davenport, T. H., Harris, J., & Shapiro, J. (2010). Competing on talent analytics. *Harvard Business Review*, *88*(10), 52–58, 150.

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Davidson, J., & Jacobs, C. (2008). The implications of qualitative-research software for doctoral work considering the individual and institutional context. *Qualitative Research Journal*, *8*(2), 73–80. https://doi.org/10.3316/QRJ0802072

Davies, J. (2023). *Discrimination and bias in AI recruitment: A case study*. Lewis Silkin.

de Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, *39*(2), 101666. https://doi.org/10.1016/j.giq.2021.101666

Derous, E., & Ryan, A. M. (2019). When your resume is (not) turning you down: Modelling ethnic bias in resume screening. *Human Resource Management Journal*, *29*(2), 113–130. https://doi.org/10.1111/1748-8583.12217

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data: Evolution, challenges and research agenda. *International Journal of Information Management*, *48*(February), 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Dwork, C. (2023). *How can bias be removed from artificial intelligence-powered hiring platforms?* Harvard-Led Institute to Pursue Fairness in Online Systems.

Feldman, J. (2015). Bayesian models of perception: A tutorial introduction. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization* (pp. 1008–1026). Oxford University Press.

Ferreira, P., Meirinhos, V., Rodrigues, A. C., & Marques, A. (2021). Virtual and augmented reality in human resource management and development: A systematic literature review. *IBIMA Business Review*, *2021*, 1–18. https://doi.org/10.5171/2021.926642

Fontana, R. (2023). *Similar-to-me effect vs halo effect: Impacts of political affiliation and physical attractiveness*. California State University, Stanislaus.

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

Galotti, K. M., Ciner, E., Altenbaumer, H. E., Geerts, H. J., Rupp, A., & Woulfe, J. (2006). Decision-making styles in a real-life decision: Choosing a college major. *Personality and Individual Differences*, *41*(4), 629–639. https://doi.org/10.1016/j.paid.2006.03.003

Gatzemeier, S. (2021). *AI bias: Where does it come from and what can we do about it?*. Data Science W231 | Behind the Data: Humans and Values Ethical Legal Data Science. what is this?

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J., Wallach, H., Daumé Iii, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86–92. https://doi.org/10.1145/3458723

Gibson, J. (1950). *Perception of the visual world*. Houghton-Mifflin.

Gibson, J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.

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

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

Glaser, B. (1992). *Basics of grounded theory analysis*. Sociology Press.
Glaser, B. (2003). *The grounded theory perspective II: Description's remodeling of grounded theory*. Sociology Press.
Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine De Gruyter.
Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038
Good, A., Lin, J., Yu, X., Sieg, H., Ferguson, M., Zhe, S., Wieczorek, J., & Serra, T. (2022). Recall distortion in neural network pruning and the undecayed pruning algorithm. *Advances in Neural Information Processing Systems*, *35*(NeurIPS 2022).
Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
Graziani, M., Dutkiewicz, L., Calvaresi, D., Amorim, J. P., Yordanova, K., Vered, M., Nair, R., Abreu, P. H., Blanke, T., Pulignano, V., Prior, J. O., Lauwaert, L., Reijers, W., Depeursinge, A., Andrearczyk, V., & Müller, H. (2023). A global taxonomy of interpretable AI: Unifying the terminology for the technical and social sciences. *Artificial Intelligence Review*, *56*(4), 3473–3504. https://doi.org/10.1007/s10462-022-10256-8
Gregory, R. (1970). *The intelligent eye*. Weidenfeld and Nicolson.
Gregory, R. (1980). Perceptions as hypotheses. *Biological Sciences*, *290*(1038), 181–197.
Gressel, S., Pauleen, D., & Taskin, N. (2020). *Management decision-making, big data and analytics*. SAGE Publications Ltd.
Grgic-Hlaca, N., Redmiles, E. M., Gummadi, K. P., & Weller, A. (2018). Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. *The Proceedings of the World Wide Web Conference*, 903–912.
Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, *1*(3), 333–342. https://doi.org/10.1111/j.1754-9434.2008.00058.x
Hinton, P. R. (2019). *Stereotypes and the construction of the social world*. Routledge.
HrTechcube. (2018). *HireVue launches AI-Driven pre-built assessments*. HrTechcube. https://hrtechcube.com/hirevue-launches-ai-driven-pre-built-assessments/
Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics*, *178*(4), 977–1007. https://doi.org/10.1007/s10551-022-05049-6
Hunkenschroer, A., & Lütge, C. (2021). How to improve fairness perceptions of AI in hiring: The crucial role of positioning and sensitization. *AI Ethics Journal*, *2*(2), 1–31. https://doi.org/10.47289/AIEJ20210716-3
IBM. (2020). *Multicloud data integration that fuels AI*. IBM Corporation. Available at https://tdwi.org
Intezari, A., & Pauleen, D. (2018). Conceptualizing wise management decision-making: A grounded theory approach. *Decision Sciences*, *49*(2), 335–400. https://doi.org/10.1111/deci.12267
Johnson, G. M. (2021). Algorithmic bias: On the implicit biases of social technology. *Synthese*, *198*(10), 9941–9961. https://doi.org/10.1007/s11229-020-02696-y
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*. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185
Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kassir, S., Baker, L., Dolphin, J., & Polli, F. (2023). AI for hiring in context: A perspective on overcoming the unique challenges of employment research to mitigate disparate impact. *AI and Ethics*, *3*(3), 845–868. https://doi.org/10.1007/s43681-022-00208-x

Kim, S., Wang, Y., & Boon, C. (2021). Sixty years of research on technology and human resource management: Looking back and looking forward. *Human Resource Management*, *60*(1), 229–247. https://doi.org/10.1002/hrm.22049

Klimoski, R., Paul, K., Rushing, C. M., Rynes, S., Schmit, M. J., Schultz, J. R., & Tomas, J. (2016). *Use of workforce analytics for competitive advantage*. SHRM Foundation.

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, *13*(3), 795–848. https://doi.org/10.1007/s40685-020-00134-w

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

Kusner, M., Loftus, J., Russell, C., & Silva, R. (2017). Counterfactual fairness. *Neural Information Processing Systems*, *30*, 4066–4076.

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–337). Blackwell Publishing.

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy & Technology*, *31*(4), 611–627. https://doi.org/10.1007/s13347-017-0279-x

Little, R., & Rubin, D. (2020). Missing data in experiments. In *Statistical Analysis with Missing Data* (3rd ed., pp. 29–46). Wiley Series in Probability and Statistics.

Lu, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., & Wei, W. (2021). Machine learning for synthetic data generation: A review. *arXiv*. https://arxiv.org/abs/2302.04062

Mann, M., & Matzner, T. (2019). Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination. *Big Data & Society*, *6*(2), 205395171989580. https://doi.org/10.1177/2053951719895805

Mann, G., & O'Neil, C. (2016). Hiring algorithms are not neutral. *Harvard Business Review*. https://hbr.org/2016/12/hiring-algorithms-are-not-neutral

Marinucci, L., Mazzuca, C., & Gangemi, A. (2023). Exposing implicit biases and stereotypes in human and artificial intelligence: State of the art and challenges with a focus on gender. *AI & Society*, *38*(2), 747–761. https://doi.org/10.1007/s00146-022-01474-3

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6).

Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L., & Canhoto, A. (2018). Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, *93*(vember 2017), 67–78. https://doi.org/10.1016/j.jbusres.2018.08.029

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Prediction-based decisions and fairness: A catalogue of choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, *8*(1), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

Mujtaba, D. F., & Mahapatra, N. R. (2019). Ethical considerations in AI-based recruitment. *IEEE International Symposium on Technology in Society (ISTAS) Proceedings*.

Nowruzi, F. E., Kapoor, P., Kolhatkar, D., Hassanat, F. A., Laganiere, R., & Rebut, J. (2019). How much real data do we actually need: Analyzing object detection performance using synthetic and real data. *ArXiv*. https://arxiv.org/abs/1907.07061

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.

Ore, O., & Sposato, M. (2022). Opportunities and risks of artificial intelligence in recruitment and selection. *International Journal of Organizational Analysis*, *30*(6), 1771–1782. https://doi.org/10.1108/IJOA-07-2020-2291

Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (2023). Bias and unfairness in machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big Data and Cognitive Computing*, *7*(1), 15. https://doi.org/10.3390/bdcc7010015

Pearson, J., Clifford, C. W., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology*, *18*(13), 982–986. https://doi.org/10.1016/j.cub.2008.05.048

Polli, F. (2019). Using AI to eliminate bias from hiring. *Harvard Business Review*. 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.

Radeke, M. K., & Stahelski, A. J. (2020). Altering age and gender stereotypes by creating the halo and horns effects with facial expressions. *Humanities and Social Sciences Communications*, *7*(1), 1–11. https://doi.org/10.1057/s41599-020-0504-6

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2019). Mitigating bias in algorithmic employment screening: Evaluating claims and practices. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3408010

Raghavan, M., & Barocas, S. (2019). *Challenges for mitigating bias in algorithmic hiring*. https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/

Rana, N., Chatterjee, S., Dwivedi, Y., & Akter, S. (2022). Understanding dark side of artificial intelligence (AI) integrated business analytics: Assessing firm's operational inefficiency and competitiveness. *European Journal of Information Systems*, *31*(3), 364–387. https://doi.org/10.1080/0960085X.2021.1955628

Rasmussen, T., & Ulrich, D. (2015). Learning from practice: How HR analytics avoids being a management fad. *Organizational Dynamics*, *44*(3), 236–242. https://doi.org/10.1016/j.orgdyn.2015.05.008

Russell, S., & Norvig, P. (2010). Artificial Intelligence a modern approach. erd ed. Pearson Education.

Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology*, *65*(1), 693–717. https://doi.org/10.1146/annurev-psych-010213-115134

Salmon, P. M., Baber, C., Burns, C., Carden, T., Cooke, N., Cummings, M., Hancock, P., McLean, S., Read, G. J. M., & Stanton, N. A. (2023). Managing the risks of artificial general intelligence: A human factors and ergonomics perspective. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *33*(5), 366–378. https://doi.org/10.1002/hfm.20996

Schraagen, J., Klein, G., & Hoffman, R. (2008). *The macrocognition framework of naturalistic decision making*. Ashgate Publishing Limited.

Shams, R., Zowghi, D., & Bano, M. (2025). AI and the quest for diversity and inclusion: A systematic literature review. *AI and Ethics*, *5*(1), 411–438. https://doi.org/10.1007/s43681-023-00362-w

Shao, Y. (2023). *How do the halo effect and horn effect influence the human resources managers recruitment decision in an occupation interview?* [Paper presentation]. Proceedings of the International Conference on Social Psychology and Humanity Studies, *6*(1), 187–195.

Simon, H. (1993). Strategy and organizational evolution. *Strategic Management Journal*, *14*(S2), 131–142. https://doi.org/10.1002/smj.4250141011

Simon, H. (1997). *Models of bounded rationality: Empirically grounded economic reason*. The MIT Press.

Simon, M., Houghton, S. M., & Aquino, K. (2000). Cognitive biases, risk perception, and venture formation: How individuals decide to start companies. *Journal of Business Venturing*, *15*(2), 113–134. https://doi.org/10.1016/S0883-9026(98)00003-2

Spivey, M. (2003). *The continuity of mind*. Oxford University Press.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. SAGE Publications.

Suresh, H., & Guttag, J. V. (2017). A framework for understanding unintended consequences of machine learning. *ArXiv*. https://arxiv.org/abs/1901.10002

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science (New York, N.Y.)*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

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

Varshney, K. (2018). *Introducing AI Fairness 360*. IBM. https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/

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

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2021). Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review. *International Journal of Human Resource Management*, 1–30.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman and Company.

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). Palgrave Macmillan.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: Principles and techniques for data scientists*. O'Reilly Media, Inc.