The Ethical Risks of Overlooking Human Potential in AI Recruitment: Current Unethical Outcomes and Future Mitigation Strategies

Johanna Marie Prill Gonzalez

Dissertation written under the supervision of professor
Rosa Fioravante

Dissertation submitted in partial fulfillment of requirements for the MSc in Management with Specialization in Strategy, Entrepreneurship, and Impact at the Universidade Católica Portuguesa, 06.01.2025.

Abstract

The rapid integration of Artificial Intelligence (AI) into recruitment processes has revolutionized hiring practices, offering significant opportunities while introducing profound ethical risks. This thesis aims to provide an overview of current debates on the ethical risks and opportunities posed by AI recruitment and their mitigation strategies, with a particular focus on the underexplored risk of AI-Defined Candidate Evaluation Filters Overlooking Human Potential.
The central research question is: *How can organizations deploy AI recruitment systems without risking overlooking human potential, especially among candidates with more individualistic or non-conforming backgrounds?*

Utilizing a qualitative systematic literature review, the study synthesizes peer-reviewed articles as well as established books in the field to assess AI's ethical challenges and identify gaps in the literature. Findings reveal that AI-defined evaluation filters can exclude capable candidates from diverse backgrounds, amplify biases, undermine inclusivity, and devalue human judgment. These various outcomes result in workforce homogeneity, stifling innovation and adaptability. To better understand the interconnected risks, the framework categorizes these into three dimensions: Inclusivity & Accessibility, Depth & Accuracy of Candidate Evaluation, and Organizational Adaptability & Long-Term Impact.

Finally, the thesis addresses these challenges by proposing a mitigation framework based on two dimensions: resource intensity and implementation horizon. Strategies include expanding outreach channels, conducting algorithmic bias audits, refining data practices to exclude sensitive attributes, and fostering human-AI collaborative evaluation processes.

Title: The Ethical Risks of Overlooking Human Potential in AI Recruitment: Current Unethical Outcomes and Future Mitigation Strategies

Author: Johanna Marie Prill Gonzalez

Keywords: Artificial Intelligence, Recruitment, Ethical Risks, Human Potential, Bias Mitigation

Sumário

A rápida integração da Inteligência Artificial (IA) nos processos de recrutamento transformou as práticas de contratação, oferecendo oportunidades significativas, mas também introduzindo riscos éticos profundos. Esta tese avalia os debates atuais sobre os riscos éticos do recrutamento com IA e suas estratégias de mitigação, com foco no risco pouco explorado de Filtros de Avaliação de Candidatos Definidos por IA que Ignoram o Potencial Humano. A questão central é: Como as organizações podem implementar sistemas de recrutamento com IA sem ignorar o potencial humano, especialmente de candidatos com perfis individualistas ou não conformistas?

Baseando-se em uma revisão sistemática qualitativa da literatura, o estudo identifica que filtros definidos por IA podem excluir candidatos qualificados de origens diversas, amplificar preconceitos, minar a inclusão e desvalorizar o julgamento humano, levando a uma força de trabalho homogênea e reduzindo a inovação. Os riscos são categorizados em três dimensões: Inclusividade e Acessibilidade, Profundidade e Precisão na Avaliação de Candidatos e Adaptabilidade Organizacional e Impacto de Longo Prazo.

Para mitigar esses riscos, a tese propõe um quadro baseado em dois dimensões: intensidade de recursos e horizonte de implementação, incluindo estratégias como expansão de canais de alcance, auditorias de viés algorítmico, exclusão de atributos sensíveis e processos colaborativos entre humanos e IA.

Este estudo contribui para a governança ética da IA, promovendo práticas de recrutamento que equilibrem eficiência tecnológica com inclusão e responsabilidade.

Título: Os Riscos Éticos de Negligenciar o Potencial Humano no Recrutamento por IA: Resultados Antiéticos Actuais e Estratégias Futuras de Mitigação

Autor: Johanna Marie Prill Gonzalez

Palavras-chave: Inteligência Artificial, Recrutamento, Riscos Éticos, Potencial Humano, Mitigação de Vieses

Table of Content

Table of Figures

“By far, the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.” - Eliezer Yudkowsky, 2008

1. Introduction

While AI has the potential to improve business operations, identify new markets, and promote innovative business models (Brynjolfsson & McAfeer, 2014), its creation and implementation pose significant ethical risks (Florida et al. 2018; Haenlein et al. 2022). The ethical risks associated with AI systems, as defined by Reid Blackman in 2022, include the potential for harm caused that may result in unjust, biased, discriminatory, or opaque outcomes. Such risks violate privacy trust, cause moral damage, and have adverse legal, reputational, and regulatory effects. Organizations that fail to tackle ethical risks face "ethical nightmares" (Blackman, 2022), where they lose control over AI and its impact on internal and/or external stakeholders (Grote et al. 2024; Kellogg et al. 2020). Losing control of AI represents a broader risk of losing control over the organization itself, as it becomes subjected to mechanized processes where technological advancement becomes the sole organizational rationale, replacing human values and human-centered goals (Lindebaum et al., 2023; Tasioulas, 2022).

Recruitment is one of the most automated organizational processes and has long been regarded as one of the most sensitive areas of application, experiencing an exponential rise in AI deployment. These practices are also crucial regarding the foundation of an organization’s culture and long-term performance by balancing financial goals with moral standards (Hatcher, 2002). Moreover, while being an intrinsically complicated matter, it directly impacts the organization's success by shaping a skilled and driven workforce (Cote, 2002; Pfeffer, 1998). Therefore, researchers stress that efficient hiring promotes fairness, accountability, and transparency while cultivating flexibility, competitiveness, and ethical integrity (Conaboy & Richard, 2005; Brand, 2008). Since it must respect values like justice, accountability, transparency, privacy, security, and general moral responsibility, hiring is also an ethically significant process (Dion et al., 2022; Ely and Thomas, 2020; Islam, 2012; Quintelier et al., 2021). Thus, the goal of recruitment is to evaluate the company's human potential while appreciating and comprehending each individual as a whole.

Automation and the advantages of AI have become alluring alternatives to the high expenses of hiring new employees, which demand a substantial investment of time, energy, and resources. The automation of repetitive activities such as reviewing resumes allows recruiters to concentrate on strategic tasks like making decisions, solving problems, and finding creative solutions for HR issues (Hackman et al., 1975; Davenport & Kirby, 2016; Wilson & Daugherty, 2018). Also, the precision of candidate evaluations is enhanced by analyzing large datasets to predict job performance and cultural fit. It also enables the identification of root causes of various issues such as employee engagement, while additionaly providing personalized responses (Beckman & Barry, 2007; Roozenburg, 1993; Raisch & Krakowski, 2021). Furthermore, AI improves the hiring process by detecting less obvious issues through pattern recognition and decreasing human cognitive biases that can cause distortion of decision-making (Garbuio & Lin, 2021). Therefore, fostering collaboration, developing central knowledge repositories, and supporting candidate evaluation (Card & DiNardo, 2002; Barney, 1991; Raisch et al., 2021) enhances overall organizational capabilities and sustains competitive advantages for recruiters. Lastly, new roles, such as algorithmic curators and brokers, are opening the door for skill-building and workforce adaptation (Kellogg, Valentine & Christin, 2020). This ultimately results in AI improving recruitment practices by enhancing efficiency, creativity, and reliability.

However, despite its benefits, the use of AI in recruitment has raised significant ethical risks, which have been examined from diverse perspectives. A significant drawback is that AI models cannot accurately predict the complexities of human behavior, which can lead to flawed hiring decisions (Van den Broek, Sergeeva. & Huysman, 2021). In addition, when employing AI in recruitment practices, there are ethical implications such as bias in HR decisions, perceived randomness, biased results based on incomplete data, reinforcing stereotypes, distorting human judgment, undervaluing conscientiousness, and job displacement. Underestimating these risks can result in replicating existing biases, as AI perpetuates inequalities, ultimately reducing diversity (Meyer, 2018; Tambe et al., 2019). Additional risks include biased decision-making due to oversimplified human factors (Barends & Rousseau, 2018) and eroded trust in management caused by the lack of transparency in AI-driven decisions (Liu & Denrell, 2018).

In the presence of such risks, traditional risk management methods often fail to incorporate ethical consideration assessment (Guntzburger, Pauchant, & Tanguy, 2017; Liu, 2014). This highlights the need for further guidance on how to deal with such topics (Blackman, 2022).

To enhance our knowledge of ethical risks tied to AI recruitment and its possible mitigation strategies, this work seeks to answer the following research question:

How can organizations deploy AI recruitment systems without risking overlooking human potential, especially among candidates with more individualistic or non-conforming backgrounds?

To address the proposed research question, this work aims to utilize the well-established approach of theorizing from existing literature. The here applied method of a systematic qualitative literature review (Raisch and Krakowski, 2021) is a well-known technique that is frequently utilized in the areas of business, ethics, and organizational studies (Mori, Sassetti, Cavaliere, and Bonti, 2020). A literature analysis on ethical issues in AI recruitment, as demonstrated by Fraij and László (2021), is a valuable approach for analyzing organizational challenges supported by a rapidly expanding scientific discourse. This is particularly relevant because AI's organizational and ethical dimensions in recruitment are areas where quantitative methods often fall short in capturing the complexity and nuance involved (Reinecke & Palazzo, 2016). This thesis makes use of insights gathered from peer-reviewed articles and well-established books in the field, providing a comprehensive understanding of AI's role in recruitment, including its ethical implications, benefits, and associated risks (Figueroa-Armijos et al., 2023). All aspects are examined across the entire recruitment process and its four stages (Hunkenschroer and Luetge, 2022). A purposive sampling approach was employed to capture academic literature relevant to the thematic approach, including keywords and criteria for the data analysis (Steps 1 and 2 in Chapter 2). Furthermore, a systematic classification of the literature was developed based on journal type, with explicit inclusion and exclusion criteria established to enhance reliability (see Step 3, Chapter 2). By clustering the risks based on their theoretical underpinnings (Step 4, Chapter 2), the study provides a structured presentation for analyzing the ethical and organizational implications of AI-driven hiring. Lastly, this thesis identifies an under-discussed risk: AI-defined candidate evaluation filters, which cause organizations to overlook human potential.

Specifically, the study defines the risk as AI reinforcing biases, not only through the application of predefined filters but by autonomously choosing the criteria for candidate selection. By permitting AI to determine these criteria, unconscious biases can be reinforced, and new ones may be created by merely recognizing patterns present in historical data due to historical imbalances (e.g., male-dominated leadership roles). This focus on certain evaluation filters leads to an over-reliance on rigid metrics, such as university brand names, company prestige, or narrowly defined skills, resulting in biased hiring decisions and reduced inclusivity. It also disproportionately disregards qualitative attributes that may be essential, such as human potential and motivation. Moreover, this thesis offers a clear framework to categorize the impacts of the newly identified risk of AI-defined candidate evaluation filters. This risk results in many unethical implications, manifesting as the exclusion of diverse, qualified candidates from diverse backgrounds, undermining accessibility and inclusivity, bias reinforcement, human judgment devaluation, and the development of a homogeneous workforce that lacks diversity in thought processes and capabilities. To address these concerns, the thesis presents a framework categorizing the risks and outcomes of AI-defined evaluation filters into three interrelated dimensions— these are based on previous literature on risks and benefits: Inclusivity & Accessibility, Depth & Accuracy of candidate evaluation, and Organizational Adaptability & Long-term Impact (Pearl, 2018; Acemoglu & Restrepo, 2020; Kahn et al., 1964; Schad et al., 2016; Smith & Lewis, 2011). This methodical approach sheds light on how the emerging risks unfold throughout the four stages of the recruitment process and their diverse impacts. In the outreach stage, AI tools tend to show a preference for candidates from schools and influential social networks while leaving out those from underprivileged backgrounds. When it comes to the screening phase of the process, unconventional career paths, versatile skills, and promising candidates are often overlooked due to rigid evaluation criteria. In the assessment stage, AI misinterprets traits such as nervousness or creativity, which human evaluators might recognize as situational or indicative of underlying capabilities. Lastly, the selection stage emphasizing achievements rather than future skills can hinder the recognition of individuals, thereby curbing their potential for development and innovative thinking. Overall, these implications hinder inclusiveness, dampen innovation, and undermine long-term flexibility.

To address these issues, the thesis proposes managerial implications that utilize a mitigation framework that assesses each suggestion according to resource intensity and the time frame for implementation. Regarding the resource intensity, the strategies range from low-cost to high-cost, referring to human and financial resources needed to implement such techniques. The implementation horizon regards the time it takes for results to materialize and whether some approaches provide immediate benefits while others require continuous effort to achieve long-term benefits (Tambes et al., 2019; McKinsey, 2023). The framework provides a structured method for presenting feasible suggestions, such as expanding to a multi-channel outreach to diversify talent

pipelines, addressing systemic biases through algorithmic bias audits and adjustments, and excluding sensitive attributes from training data to encourage equal decision-making. Moreover, human-AI collaborative review panels with fixed questionnaires are suggested to integrate AI with human judgment for fair evaluations. Simultaneously, contextual candidate profiling and sentiment analysis of soft skills will capture both AI merged with human judgment for a balanced evaluation. These strategies offer a roadmap and managerial implications to address the identified risk: AI-Defined Candidate Evaluation Filters Overlooking Human Potential.
In this way, this study expands current knowledge on AI recruitment ethical risks, particularly those tied to the overlooking of human potential. Therefore, this also offers new insights into how AI-driven processes can perpetuate biases and exclude diverse talent (Fritts & Cabrera, 2021; Tilmes, 2022). Resulting in the study contributing to the literature on business ethics and organizational studies by first defining and addressing the underexplored risk of AI-defined candidate evaluation filters. Secondly, the thesis highlights that AI cannot assess qualitative characteristics essential for promoting diversity and organizational adaptation, such as human potential and cultural fit. Thirdly, it offers two organized frameworks to address these challenges. The first, a risk-clustering framework, provides a thorough understanding of these interrelated challenges by grouping the ethical and organizational ramifications into three dimensions: Inclusivity & Accessibility, Depth & Accuracy of Candidate Evaluation, and Organizational Adaptability & Long-Term Impact. The second, a mitigation framework, offers viable measures to promote workforce diversity, increase inclusivity, and improve evaluation accuracy while considering implementation horizon and resource intensity.

The rest of the thesis is organized as follows: Section two describes the methodology that was used in the systematic literature review and the analytic steps needed to identify gaps and risks in AI recruitment. Sections three to seven explain the role of AI in the recruitment processes, introduce the new risk-AI-defined candidate Evaluation Filters Overlooking Human Potential with a new framework, and present actionable mitigation strategies. Finally, Section 8 concludes the thesis by summarizing its key contributions, managerial implications, limitations, and suggesting directions for future research.

2. Literature review

With the growing body of research on AI recruitment (Kaplan & Haenlein, 2018; Albassam, 2023; Hunkenschroer & Kriebitz, 2022) and its significant ethical implications (Blackman, 2022; Florida et al., 2018; Haenlein et al., 2022), a systematic literature review (Raisch & Krakowski, 2021) is a valuable way to explore the ethical risks as highlighted in existing studies (Liu & Denrell, 2018; Tambe et al., 2019; Tang et al., 2021). Despite the information on ethical risks, there is limited focus on how AI-defined evaluation filters exclude candidates who diverge from conventional success markers (Acemoglu & Restrepo, 2020). So far, research has prioritized measurable metrics, often overlooking critical qualitative traits such as cultural fit and human potential (Tang et al., 2021; Barends & Rousseau, 2018). On the other hand, the advantages of AI in enhancing efficiency and competitiveness through enhanced decision-making have been widely acknowledged (Wilson & Daugherty in 2018 and Hackman et al. in 1975). The importance of AI in achieving success has also been underscored by McKinsey in 2023. This results in the need for strategies that retain these advantages while addressing associated risks. Building on the following literature review regarding AI recruitment, its benefits, risks, ethical considerations, and the gaps in current research, this study formulates its research question.

To do so, this study follows a five-step methodology, here presented in detail. (see Figure 1)

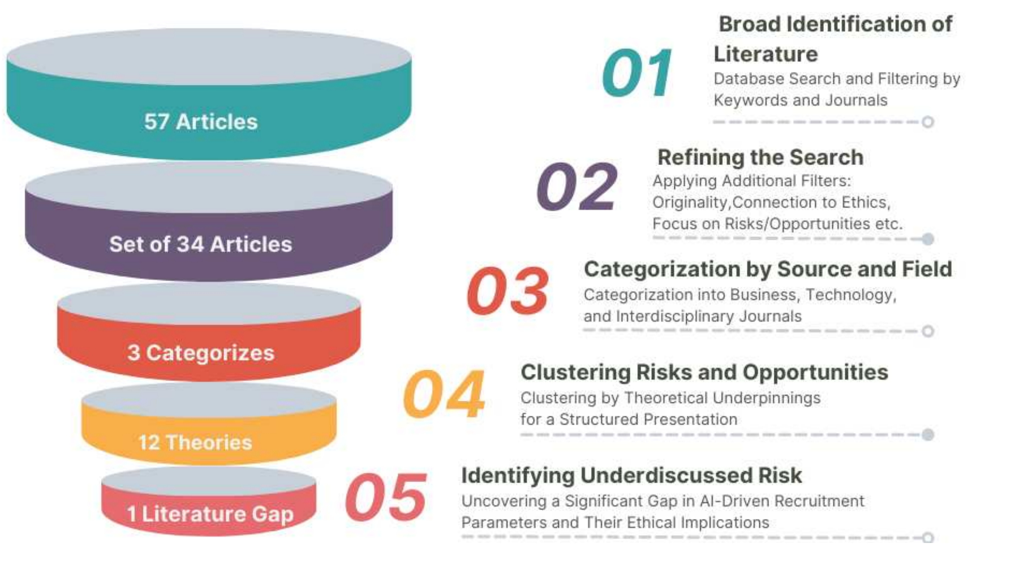

Figure 1: Visualized Systematic Literature Review

Step 1: Identification of Risks and Opportunities

The initial phase of the systematic literature review involves an extensive search of secondary sources, including books such as Ethical Machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI by Blackman (2022). The investigation preserves the main focus on peer-reviewed journal articles in the following domains: AI, design, ethics, business, and management. Although there were no restrictions regarding the publication dates, the search has taken place in a limited time frame between late August and the end of October 2024.

In order to comprehensively explore the subject, a range of keywords tailored to various aspects of AI in recruitment were implemented, and the first three result pages reviewed. The foundational investigation keywords were "AI in recruitment" and searched using ScienceDirect and EBSCO for a broad understanding of the implementation and implications of AI technologies in hiring processes.

To explore the opportunities of AI in recruitment, a focused search was conducted using the keyword "AI recruitment opportunities" on Google Scholar. Specific filters were applied to include articles from renowned journals such as the Academy of Management Journal and the Strategic Management Journal, ensuring access to high-quality and impactful research.

For insights into the company and business perspective on the risks of AI in recruitment, the keyword "AI recruitment risks" was used on EBSCO. This helped identify studies addressing organizational challenges. To refine the search further, the filters "Automation--Economic aspects," "Industrial robots," and "Personnel management" were applied. Additionally, searches for "Artificial Intelligence" and "Artificial Intelligence in Business" were filtered to include articles from the Harvard Business Review, known for its authoritative perspectives on AI. Moreover, this search looked into two categories within EBSCO: "Business & Economics / Human Resources & Personnel Management" and "Business & Economics / Management". Given the abundance of results, the thesis prioritized top-tier journals, including Oxford University Press, the Academy of Management Annals, and Harvard Business Review. EBSCO proved particularly valuable, as it provided the majority of the high-quality and diverse literature referenced

For the ethical considerations of AI in recruitment, the terms “ethics in AI recruitment,” “AI recruitment ethical risks,” and “AI recruitment ethics” were used across Google Scholar, EBSCO, and ScienceDirect. This uncovered literature discussing bias, fairness, and broader ethical implications associated with automating recruitment processes.

Furthermore, to explore the domain of design and innovation, the keywords "Design thinking" and "Design theory" were used, while also filtering for the journal Design Studies, a leading academic platform for design research.

It is important to note that while emphasis was placed on certain journals due to their prominence in specific fields, other high-quality sources were not excluded, ensuring a balanced and inclusive selection of literature. Using the designated keywords and filters, this procedure ultimately produced a pool of 420 articles from the first three pages of results across three databases. Out of these, 57 articles were found to be significant (see Appendix A for a list of all 57 articles, including six that were counted twice since they appeared under several keywords).

Nevertheless, given that many of the articles didn't meet the following requirements, a number of them were left out of the final thesis:

- Mentions Risks or Opportunities: In the abstract or first few pages, risks or opportunities are discussed.

- Academic Standards: Sufficient citations or references to established studies that seem credible when looking through their list of references.

- Up-to-date Information: Offered insights aligned with advancements in AI or recruitment, not contradicted by newer articles, thereby applicable to the thesis.

Step 2: Refining the Search

Out of the 57 articles found to be significant, 34 articles were left to be reviewed in Chapters 4.1 and 4.2 after the application of additional filters. This selection includes articles that focus on broader, more general applications of AI and provide conceptual frameworks relevant to the research question.

Articles were excluded if they met any of the following criteria:

- Primarily addressed technical aspects without a connection to ethics or management.
- Focused on AI broadly without discussing its specific benefits, risks, or impacts on the recruitment process.
- Were repetitive or duplicative of other articles already included in the analysis.

Step 3: Categorization by Source and Field

The final set of 34 articles can be categorized by their disciplinary focus to understand the context of the discourse:

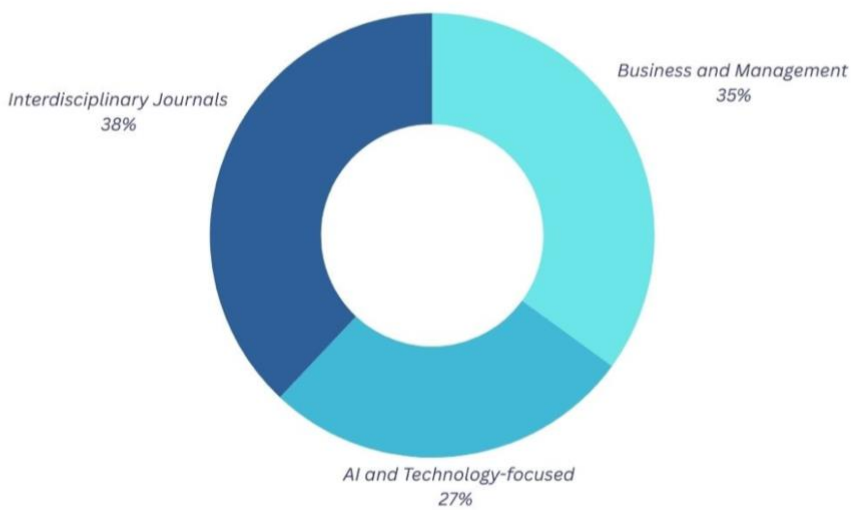

Figure 2: Journal Category Distribution (%)

- **12 articles** were from business and management journals, emphasizing practical and organizational lenses.
- **9 articles** were from AI and technology-focused publications, focusing on technical perspectives.
- **13 articles** were from interdisciplinary journals, integrating ethical implications, societal impacts, and human-centered design in AI recruitment.

This distribution shows that the majority of the literature has its roots in management and business settings, emphasizing an applied and pragmatic approach to AI hiring. However, an underrepresentation of articles employing an interdisciplinary approach and articles outlining ethical viewpoints can be observed, suggesting an opportunity for additional contribution.

Step 4: Presentation of Currently Debated Risks and Opportunities

Two main clusters of research have been identified, and their corresponding theoretical foundations analyzed: On the one hand, risks posed by AI in recruitment, and on the other hand, AI's opportunities in recruitment. In the following chapters, these clusters are organized based on underlying theoretical frameworks, allowing a more detailed review of causes and impacts to expand the understanding of risks and opportunities (refer to Figures 3 and 4).

Additional details will be outlined in Chapters 4.2 and 4.3. By outlining these theories each risk and opportunity is presented with context, which enables the analysis of contributing factors as well as creates a theoretical basis for the discussion. This presentation of the current debate and its theoretical basis confirms a significant risk being overlooked: AI-defined candidate evaluation filters overlooking human potential (see Step 5).

Step 5: Identifying Underdiscussed Risks Through Analysis

The preceding analysis reveals a significant gap in the existing literature regarding AI-Defined Candidate Evaluation Filters Overlooking Human Potential and the corresponding ethical and organizational implications. While this thesis outlines many theories and perspectives on the topic, the focus should be on contributing to the following aspects within the broader discourse.

The following aspects remain underexplored:

- Over-reliance on rigid parameters like university brand names or company prestige leads to biased hiring outcomes.
- AI's inability to assess nuanced qualities, such as cultural fit, is critical for identifying diverse and fitting talent.
- Ethical considerations and long-term organizational outcomes include homogeneity in the workforce and its impact on innovation and a toxic culture.
- AI-driven outreach can inadvertently exclude under-networked communities, reducing the diversity of the initial candidate pool.
- Misalignments with organizational inclusivity efforts and messaging undermine the effectiveness of diversity initiatives.
- The inability to assess qualitative attributes like human potential and motivation limits its ability to identify candidates who may excel in dynamic roles.
- Stage-specific analysis of AI's impacts instead of generalizing risks across the entire recruitment process.

This gap exists due to:

- A predominant focus on measurable short-term outcomes, overlooking broader implications.
- Limited interdisciplinary research integrating technical, ethical, as well as organizational perspectives.
- Generalized discussions that fail to analyze risks and opportunities within specific stages of recruitment.
- The prioritization of AI's speed and cost-saving benefits often overshadows its broader ethical and organizational consequences.

Finally, in subsequent chapters, this thesis contributes to literature by exploring every underexamined aspect and offering additional insights. Key theoretical contributions address critical gaps in the current literature regarding necessary mitigation strategies. A central focus is the reliance on rigid filters and how these algorithmic frameworks embed bias, ultimately hindering diversity over time. To counter these biases, the thesis proposes various strategies, such as algorithmic audits, to foster more inclusive candidate selection.

The literature review also emphasizes AI's struggle to assess traits essential for identifying quality talents aligned with organizational culture. To tackle this issue head-on, the dissertation expands recruitment theories by introducing strategies such as contextual profiling and soft skill sentiment analysis. These methods aim to incorporate qualitative perspectives into the assessment, thereby enhancing the theoretical significance of that stage.

By discussing the underexplored aspect of workforce homogeneity, the thesis connects biased hiring to hindered innovation and fostering toxic work environments. Thus, highlighting the dangers of uniformity in organizations. It further addresses AI's neglect of marginalized communities by using algorithms to decide the targets of the outreach offers. This adds to existing knowledge by defining this new challenge as well as suggesting various approaches to expand the range of potential candidates.

Additionally, the thesis investigates the misalignment between AI-driven hiring practices and inclusivity efforts. Thereby looking into the diversity-related repercussions of these contradictions while stressing the significance of collaboration across departments. Ultimately, it enhances theoretical understanding by proposing blended assessment approaches that incorporate intuition and AI to evaluate qualitative characteristics.

Lastly, this thesis addresses the underexplored aspect: the generalization risks across the process by offering a stage-specific analysis of the outcomes of the newly identified risk.

3. AI-Driven Recruitment: Revolutionizing the Process

To thoroughly understand the use of AI in recruitment, it's essential to understand its tools, their applications, and how they are used across various stages of the recruitment process.

Kaplan and Haenlein (2018) define AI as "a system's ability to interpret external data correctly, learn from it, and use those learnings to achieve specific goals and tasks through flexible adaptation". AI recruitment refers to the application of artificial intelligence in organizational processes related to recruiting and selecting candidates (Hunkenschroer & Kriebitz, 2022). Consequently, AI is increasingly applied across several stages of the recruiting process.

The foundation of AI systems lies in analyzing large datasets using intelligent, iterative algorithms to identify patterns and features (CSU Global, 2024). Commonly used AI tools in recruitment include Textkernel and SAP's Resume Matcher, which offer automated resume scanning, reducing manual labor and improving efficiency (Fraij & László, 2021). Enhanced analytical capabilities are also enabled by converting unstructured into structured data formats through Natural Language Processing (NLP) and Generation (NLG) technologies (Fraij & László, 2021).
In real-time, chatbots such as XOR and Talkpush are used to engage candidates and gather key information (Fraij & László, 2021). Predictive analytics further helps the hiring quality by identifying candidates with the highest likelihood of success based on prior data (Fraij & László, 2021). Based on these advancements, 38% of organizations had already adopted AI in Human resource management (HRM) by 2018, with the rest expected to follow soon (Erickson, 2018). By 2023, McKinsey reported that already 60% of organizations were using AI for talent management (The State of AI in 2023: Generative AI's Breakout Year, 2023). A recent survey found that 99% of Fortune 500 companies now use AI tools, such as applicant tracking systems, to screen candidates during recruitment (Das, 2023). Notable examples are Mastercard, which saw an increase in hiring from less than 200 in 2021 to almost 2,000 in two years, and Kuehne+Nagel, with a 74% employee satisfaction rate achieved through AI-enhanced recruitment processes (Blehar, 2020; Kaushik, 2020). Other companies utilizing AI in recruitment are Delta Air Lines and Siemens AG (Blehar, 2024; Kaushik, 2240).

A clearer understanding of how AI is used in recruitment necessitates examining its application across the four stages: outreach, screening, assessment, and selection. These four are the key stages of the recruitment process and where AI technologies are most commonly deployed (Albassam, 2023). Recruitment outreach is the process of identifying and communicating with candidates who exhibit exceptional qualities for the positions offered by an organization. Although it's often done through email, other methods like LinkedIn also effectively reach potential candidates (Chatelaine, 2024). In the outreach phase, AI can focus on communication across online platforms and social media to ensure that the job advertisement reaches the appropriate audience (Hunkenschroer & Kriebitz, 2022). For instance, by making job ads gender-neutral, AI can attract more diverse applicants and help de-bias job candidates (Rb-Kettler & Lehnervp, 2019). Screening entails assessing job applications, reviewing resumes, and selecting candidates who meet the role's requirements by considering their qualifications, experience, and skills (peopleHum, 2023). During this stage, AI algorithms can analyze large volumes of resumes and quickly identify the most qualified candidates based on predetermined criteria, improving application management's speed and accuracy (Hunkenschroer & Kriebitz; 2022). As part of candidate assessment and to determine their suitability for the role, an applicant's qualifications, skills, and experience are evaluated through interviews, tests, or work samples (Gasparyan, 2024). In this stage, face recognition software and other AI tools can be utilized to assess video interviews, providing insights into personality traits and competencies that may not be immediately evident to human evaluators (Albassam, 2023). During the final stage of selection, the employer narrows the applicant pool to identify the most qualified candidate for the role (Recruitment and Selection Process: Best Practices & Overview, 2022). It further evaluates candidates' qualifications and experience through additional interviews and tests. AI facilitates this process by automating administrative tasks like scheduling interviews and assessments, evaluating tests, streamlining the selection process, and reducing time delays (Köchling et al., 2020).

3.1 Recruitment and Its Ethical Significance

Although the term "recruitment" may appear straightforward, its process and significance are far more complex. As Cote (2002, p. 172) highlights, "If organizations are able to find and employ staff who consistently fulfill their roles and are capable of taking on increased responsibilities, they are immeasurably better placed to deal with the opportunities and threats arising from their operating environment". Effective recruitment demonstrates the significance of keeping organizations adaptable and competitive. Pfeffer (1998) recognizes selective hiring as a fundamental human resource practice, emphasizing its essential role in creating a capable workforce. Delery and Shaw (2001) further argue that organizational performance hinges on three key employee characteristics: ability, motivation, and empowerment, with ability and motivation being directly shaped by recruitment practices (Brewer, 2008). Collectively, these scholars affirm that recruitment is an integral factor in an organization's success, as it directly influences overall performance. Organizational performance is defined as the ability to achieve specific objectives, according to Daft (2000) and Kaplan (2001), who also suggest that goal achievement can be achieved efficiently by utilizing human resources. This results in recruitment being an ethically significant process that requires careful consideration at every stage. Conaboy and Richard (2005) highlight that transparent and fair human resource management practices are essential for fostering an ethical organization. Fairness in these processes is crucial since HRM includes hiring, appraising, and rewarding employees (N, Sang, & Ngure, 2018). Recruitment, as a key element of HRM, plays a pivotal role in establishing the ethical foundation of an organization (Conaboy & Richard, 2005). Brand (2008) argues that sustained ethical behavior in organizations depends on both an ethical mindset and practices, starting with recruitment—the first point of contact with potential employees. Ethical recruitment sets the tone for an organization's culture and fosters an ethical corporate environment. Vickers (2005) similarly emphasizes that recruitment is the foundation for ethical practices across the organization, promoting transparency in selection processes and contributing to an ethical workplace. Ethics is seen as critical in recruitment because human resources are a key source of competitive advantage. Nevertheless, given how efficient hiring and management help organizations maximize this advantage, there often is pressure to prioritize competitive gains over ethical standards (N, Sang, & Ngure, 2018). Therefore, Virovere et al. (2002) and Hatcher (2002) stress that business ethics require balancing economic goals, competition, and morality. In this context, Gert (2000) also stresses that in recruitment this balance is essential, while defining morality as a code of conduct upheld by individuals or groups. Organizations must achieve financial objectives while maintaining ethical dimensions such as fairness, transparency, and responsibility (Hatcher, 2002).

Finally, established ethical models further explain the importance of ethics in recruitment. For example, Jones's (1991) issue-contingent model introduces the concept of moral intensity, which underscores the need to evaluate the ethical implications of each decision. Rather than solely focusing on profits or candidate satisfaction, this model suggests that recruitment decisions should consider the ethical weight of each choice. Overall, it can be argued that the ethical significance of recruitment lies in its shaping of the organization's culture, influence on long-term success, and establishment of fairness as well as transparency in all human resource processes.

Nevertheless, by shaping the culture, AI can also create cultural misfits, which happen when people don't conform to an organization's current cultural norms or practices, making them feel like outsiders (Williamson & Perumal, 2018). Accordingly, the incorporation of ethical standards into recruitment practices can foster a culture of trust and integrity by creating aspiring talent who are capable and motivated.

3.2 Ethical vs. Unethical Recruitment

This chapter discusses ethical and unethical dimensions of AI-driven recruitment, including privacy, security, bias, accountability, and transparency. AI-driven recruitment systems handle large volumes of sensitive personal data, presenting risks within the privacy and security dimension if not properly safeguarded, leaving the data susceptible to misuse or breaches (Stuss & Fularski, 2024). Moreover, concerns regarding informed consent arise, as many candidates remain unaware of how their personal data is utilized by AI systems, particularly when it is collected from public domains like social media without explicit consent (Sánchez-Monedero et al., 2020). This leads to significant ethical concerns, highlighting the necessity of seeking informed consent in the recruitment process (Bîgu & Cernea, 2019; Sánchez-Monedero et al., 2020).

Furthermore, transparency and accountability are central ethical challenges in AI-driven recruitment, mainly due to the uncertainty regarding the responsibility. The blurred lines of AI systems make it difficult to determine responsibility, creating challenges in holding anyone accountable when issues, such as unfair hiring decisions, arise(Raghavan et al., 2020). In such cases, it is uncertain whether the developers, HR managers, or AI are accountable for these outcomes (Sánchez-Monedero & al, 2020; Chamorro-Premuzic et al., 2019). This lack of clear accountability is a significant ethical concern, especially as AI becomes more integrated into decision-making processes.

While AI systems are often praised for their ability to reduce human biases and enhance fairness, they may also create or perpetuate new biases, particularly those embedded in the training data (Stuss & Fularski, 2024; Hunkenschroer n. Laubach, 2020). These biases, stemming from historical or structural inequalities, can lead to biased outcomes, which disadvantage certain social groups (Hunkenschroer & Luetge, 2022). However, AI does have ethical opportunities to decrease human bias (Polli, 2019). By utilizing objective, quantifiable data instead of subjective judgments, AI is able to reduce unconscious human biases in recruitment decisions, potentially leading to fairer outcomes and diversity in the workforce (Polli, 2019). Thus, it also enhances process uniformity by applying identical standards to all candidates, resulting in more standardized and equitable evaluations (Hunkenschroer and Luetge, 2022).

AI systems do not operate in a vacuum; they are influenced by varying cultural values and societal impacts, which can shape how these systems function and are perceived. Consequently, AI systems can inadvertently establish biases regarding particular cultures, influencing recruitment practices and the treatment of candidates from diverse backgrounds (Stuss & Fularski, 2024). Given the global implementation of AI in recruitment, ethical concerns arise, as cultural biases may lead to unjust discrimination in diverse work environments (Stuart & Fularski, 2024).

After examining the ethical dimensions and their implications, the following chapter will primarily focus on the ethical dimensions: biases, fairness, accountability, inclusivity, justice, and transparency (see Figures 3 and 4 in the last column).

4. Ethical Risks and Opportunities in AI Recruitment

4.1 Introducing the Discussion

Reid Blackman's assertion that "there is no intrinsically biased data" (DataFramed Podcast, Episode #154, 2024, 10:45), highlights that the criteria for identifying an ethical risk should focus on whether "the AI obstructs people's access to basic goods of life" (DataFramed Podcast, Episode #154, 2024, 24:15). This includes access to jobs, as employment is essential for securing basic needs such as food and shelter. Moreover, the current literature indicates that the ethical implications of risks and opportunities in AI recruitment are not always thoroughly addressed—a gap that this thesis seeks to fill (see Figure 3). This chapter discusses both sides of the coin: the transformative potential of AI on recruitment and the challenges it creates. Thereby, setting the groundwork for comprehending the balance between optimization and ethics.

4.2 Opportunities of AI Recruitment

The following figure provides an overview of the opportunities presented by AI in recruitment while emphasizing their theoretical perspective (see Figure 3). Specifically, it highlights gaps in the literature where certain ethically significant risks are overlooked, with insufficient attention given to their ethical dimensions.

Opportunity	Theoretical perspective	Summarized argument	Supporting literature	Ethical risk in literature?	Ethical significance
Enhancing Recruiter Productivity Through Task Automation	Job Characteristics Model and AI-Human Collaboration	AI adoption enhances employee creativity and efficiency by freeing them from repetitive tasks, boosting satisfaction, performance, and process reliability.	Hackman et al., 1975; Davenport & Kirby, 2016; Wilson & Daugherty, 2018; Elsbach & Hargdon, 2006	No	Promotes fairness by minimizing human biases in repetitive tasks, while also supporting recruiters in fulfilling higher-level responsibilities that enhance meaningful work and professional growth. Aligns with principles of equity and dignity in labor.
Improving Problem Identification in HR with Design Thinking	Design Thinking Framework	AI improves the design-thinking process by enhancing abductive reasoning and supporting problem-finding through pattern recognition.	Beckman & Barry, 2007; Dong, Garbuio, & Lovallo, 2016	No	Allows for earlier / accurate identification of systemic issues, promoting safe practices. This leads to ethically responsible decisions that consider diverse perspectives and mitigate biases, aligning HR strategies with principles of justice.
Driving Innovative HR Solutions with Abductive Reasoning	Abductive Reasoning	AI aids in discovering innovative solutions by interpreting data, identifying root causes, and guiding creative responses to challenges like diversity and engagement.	Roozenburg, 1993; Dorst, 2011	No	Fosters inclusivity by addressing challenges like diversity / engagement with innovative and context-sensitive approaches. By guiding tailored interventions, it ensures ethical alignment with equitable and inclusive workplace practices.
Balancing Automation and Human Insight for Precise Candidate Selection	Paradox Theory (Automation vs Augmentation)	AI's automation and augmentation balance enhances recruitment by offering precise candidate evaluations and fostering collaboration through data centralization.	Schad et al., 2016; Smith & Lewis, 2011; Davenport & Kirby, 2016; Raisch & Krakowski, 2021	Yes	By centralizing and expanding organizational knowledge, it fosters collaborative decision-making and equitable hiring practices, ensuring ethical alignment with transparency and inclusivity in recruitment.
Building Organizational Knowledge through Collaborative Data Integration / Freeing Recruiters for Value-Adding Tasks	Skill-Biased Technology Theory & Resource-Based View (RBV)	AI enhances recruiter performance by automating routine tasks, allowing focus on strategic decisions, creating competitive advantage, and improving candidate experience.	Card & DiNardo, 2002; Barney, 1991; Raisch & Krakowski, 2021	No	Enhances workforce empowerment by enabling skilled recruiters to focus on strategic / creative tasks. Supporting meaningful professional development and well-being, aligning organizational practices with ethical principles of employee dignity and holistic growth.
Strengthening Recruitment Objectivity Through Cognitive Bias Mitigation	Cognitive Bias Mitigation	AI mitigates cognitive biases by analyzing large datasets, enhancing decision-making and recruitment outcomes by addressing biases in HR practices.	Garbuio & Lin, 2021	No	Promotion of ethical integrity in recruitment by ensuring actions are based on comprehensive and unbiased decisions, fostering a culture of accountability and rationality.

Figure 3: Overview of Opportunities, Theoretical Perspectives, and Ethical Salience

First, the job characteristics model and theories of AI-human collaboration (see Figure 3) explain how AI adoption in recruitment enhances employee creativity and efficiency by relieving recruiters from repetitive tasks (such as resume screening), allowing them to focus on more strategic, high-level tasks such as problem-solving and decision-making (Hackman et al., 1975; Davenport and Kirby, 2016). In addition to enhancing employee satisfaction and performance in the recruitment department, the improvement in job design and division of labor also results in faster and more reliable processes (Wilson and Daugherty, 2018; Elsbach and Hargdon, 2006).

The Design Thinking Framework (Beckman & Barry, 2007; Dong, Garbuio, & Lovallo, 2016) highlights how AI improves the design-thinking process, especially in HR. AI enhances abductive reasoning, supporting the generation of hypotheses from incomplete data, which helps identify problems early on (Dong, Garbuio, & Lovallo, 2016). By recognizing patterns, AI also uncovers less apparent issues, facilitating problem-finding (Beckman & Barry, 2007) (see Figure 3). Therefore, it helps identify challenges and enables HR practitioners to derive more innovative solutions from the generated hypotheses compared to traditional methods (Roozenburg, 1993; Dorst, 2011). This makes addressing challenges like diversity and engagement easier by identifying root causes and guiding effective responses.

According to paradox theory (Schad et al., 2016; Smiths & Lewis, 2011), there are advantages to both Automation and Augmentation. First, the vast amount of information processed by AI makes precise predictions more likely. This is especially the case for structured tasks like recruitment. By analyzing large datasets and comparing the performance of employees with similar backgrounds, AI can predict how well a candidate will perform (Davenport & Kirby, 2016). This balance between automation and augmentation facilitates candidate selection and enhances the overall hiring processes (Raisch & Krakowski, 2021). Paradox theory also emphasizes how AI increases collaboration within organizations by collecting available knowledge (Raisch & Krakowski, 2021). By utilizing data from various independent employees, AI can create a centralized knowledge repository that surpasses the capabilities of traditional HR departments (Raisch & Krakowski, 2021).

The skill-biased technology theory and resource-based view (RBV) (see Figure 3) emphasize AI's role in boosting the performance of skilled recruiters by automating routine tasks and empowering them to focus on complex, value-added activities like candidate evaluation (Card and DiNardo, 2002; Barney, 1991).
Furthermore, the RBV suggests that AI creates a competitive advantage by integrating human expertise with machine intelligence, enhancing organizational capabilities, and sustaining competitive differentiation (Card and DiNardo, 2002; Barney, 1991; Raisch & Krakowski, 2021). For example, AI can manage tasks like scheduling, streamlining the recruitment process, and enhancing the candidate experience. This, in turn, fosters positive interaction with prospective and current employees, strengthening the company's reputation.

Cognitive bias mitigation (Garbuio & Lin, 2021) underscores how AI can overcome human cognitive biases and constrained mental models, frequently hindering decision-making. By analyzing large datasets and uncovering patterns beyond human capability, AI enhances design thinking's ability to solve complex, ambiguous problems, particularly in HR practices where biases frequently distort judgment (Garbuio & Lin, 2021). This leads to more objective and effective recruitment decisions, further strengthening the recruitment process (Garbuio & Lin, 2021).

4.3 Challenges of AI Recruitment

The following figure provides an overview of AI opportunities in recruitment, highlighting literature gaps where ethically significant risks are overlooked, and their ethical dimensions are insufficiently addressed (see Figure 4).

Opportunity	Theoretical perspective	Summarized argument	Supporting literature	Ethical risk in literature?	Ethical significance
Risk of Bias in AI-Driven HR Decisions	Evidence-Based Management	AI systems in HR face risks from task complexity, potentially leading to biased decisions if not grounded in evidence-based management.	Barends & Rousseau, 2018; Pfeffer & Sutton, 2006; Rousseau, 2014	No	Reinforces systemic inequities, leading to unjust outcomes that disproportionately disadvantage certain groups, thereby challenging foundational principles of equity, fairness, and ethical responsibility in HR practices.
Perceived Random Decisions: The Risk of Randomization	Randomization	Randomization in AI decisions may enhance fairness but can also erode trust when outcomes appear arbitrary or lack transparency.	Denrell, Fang, & Liu, 2015; Liu & Denrell, 2018	No	Supports the ethical principle of fairness while simultaneously undermining it, as outcomes may feel arbitrary to employees, challenging foundational principles of transparency, accountability, and trust in management.
Bias Amplification Due to Limited Data	Small Data Sets and Causal Reasoning	Small data sets hinder AI's ability to predict rare events, leading to biased outcomes that disproportionately affect vulnerable groups.	Pearl, 2018; Tambe et al., 2019; Radford et al., 2019; Heaven, 2020	Yes	Weakens ethical principles of equity, inclusivity, and non-discrimination by perpetuating and exacerbating systemic biases through limited data, inadequate causal reasoning, and the reinforcement of existing stereotypes.
Risk of Reinforced Stereotypes	Bias in AI and Transformer-Based Models	AI models, especially language-based transformers, risk perpetuating societal biases present in the training data.	Meyer, 2018; Brown et al., 2020; Bouschery et al., 2023	Yes	Also weakens ethical principles of equity, inclusivity, and non-discrimination by perpetuating and exacerbating systemic biases through limited data, inadequate causal reasoning, and the reinforcement of existing stereotypes.
Disruption of Trait Balance: Risks from AI Replacing Human Judgment	Complementarity Theory	AI's takeover of tasks traditionally requiring human traits disrupts the balance employees seek between their traits and their work environment.	Carson, 1969; Kiesler, 1983; Grant et al., 2011; Barrick & Mount, 2012; Tang et al., 2021	No	By diminishing accountability and responsibility, it amplifies unchecked errors or biases, ultimately disrupting the ethical principle of balance in human-technology collaboration.
Blurred Role Expectations: Risks of Undervaluing Conscientiousness in HR	Role Theory and AI Impact on Job Expectations	AI automation in traditionally human-centric roles blurs job expectations and devalues traits like conscientiousness, crucial for job performance	Kahn et al., 1964; Tang et al., 2021; Barrick & Mount, 2012	No	Undermines the ethical principles of role integrity and professional standards, while also diminishing the principle of balance by blurring role expectations and undervaluing conscientiousness, a critical predictor of job performance.
Job Displacement Concerns	Perceived Job Displacement	AI-driven automation increases fear of job displacement, leading to employee anxiety, disengagement, and mistrust of AI feedback.	Acemoglu & Restrepo, 2020; Tong et al., 2021; Bughin & Manyika, 201	Yes	Decreases peace and security by fostering uncertainty and anxiety over role relevance, diminishing employee engagement and trust, which are foundational to ethical workplace practices.

Figure 4: Overview of Opportunities, Theoretical Perspectives, and Ethical Salience

First, relying on Evidence-Based Management (EBMgmt) as its theoretical framework, a significant risk arises from the inherent complexity of HR tasks, such as defining what constitutes a "good employee" or accurately measuring performance in team-based roles (Barends and Rousseau, 2018; Pfeffer and Sutton, 2006; Rousseau, 2014). Literature warns that AI systems can identify patterns without fully understanding their causes, resulting in biased and unfair decisions (Pfeffer and Sutton, 2006). This risk is particularly problematic in hiring and performance reviews, where ambiguous or incomplete information can intensify existing inequalities.

Additionally, the literature highlights risks associated with randomization as a theoretical construct in AI-driven HR decisions (Denrell, Fang, and Liu, 2015; Liu and Denrell, 2018). While randomization can promote fairness, it may also lead to decisions, such as promotions or layoffs, that feel arbitrary to employees (Liu and Denrell, 2018). This can erode trust in management, particularly when the decision-making process lacks transparency and the AI's reasoning is unclear (Liu and Denrell, 2018).

Another key risk with AI in recruitment is the small data sets typically used, which makes it difficult for machine learning models to predict rare events such as poor performances (Pearl, 2018; Tambe et al., 2019). The problem is related to causal reasoning, as AI systems that lack an understanding of the true causes of outcomes may become biased or ineffective, disproportionately impacting certain groups and reinforcing existing biases related to gender, race, or religion (Radford et al., 2019; Heaven, 2020; Tambe & al., 2019). As was seen with Amazon’s biased hiring algorithm, AI often fails to address these causal links, exacerbating discrimination (Meyer, 2018). An example is the transformer-based language models trained on large datasets scraped from the web, which can unintentionally encode and perpetuate stereotypes in the data (Brown et al., 2020). These biases may persist in AI-driven HR practices and influence decision-making, where customer data often reflects and reinforces societal biases (Bouschery et al., 2023).

In addition, the complementarity theory states that employees prefer a balance between their own traits and the traits of their work environment, including technology (Carson, 1969; Kiesler, 1983; Grant et al., 2011). This balance is disrupted when AI takes over tasks like CV screening that previously relied on human traits such as conscientiousness, defined by characteristics like diligence, organization, and reliability (Barrick & Mount, 2012). AI’s autonomous nature might overshadow human input, diminishing the need for oversight and possibly leading to unchecked errors or biases (Tang et al., 2021). Moreover, Role theory (Kahn et al., 1964) supports this argument by emphasizing that employees have expectations tied to their work roles. AI, which automates tasks traditionally requiring conscientious behavior, can blur the clarity of these expectations, leading to the undervaluation of conscientiousness (Tang et al., 2021). This is problematic because conscientiousness has long been considered a critical predictor of job performance (Barrick & Mount, 2012). The reliance on AI could undermine this trait's role, potentially affecting the quality and integrity of essential recruitment functions (Tang et al., 2021) (see Figure 4).

Role theory by Kahn et al. (1964) and the concept of perceived job displacement by Acemoglu & Restrepo (2020) provide insight into how AI automation of recruitment and job-related evaluations can make employees uncertain about their roles, hence raise concerns about their relevance and job security. Tong et al. (2021) confirmed that the fear of job displacement due to AI is increasing, with workers becoming apprehensive and disengaged, fearing that AI could replace their jobs. This aligns with the disclosure effect, where employees informed of AI use in evaluations may distrust its feedback and worry that AI will replace human oversight, impacting performance, especially when lacking transparency (Bughin & Manyika, 2019).

5. Identifying Gaps in the Literature

The current literature provides a thorough overview of AI in recruitment, outlining its key concepts, applications across recruitment stages, and ethical considerations. It also highlights substantial benefits along with considerable risks, which include the use of biased training data that leads to discriminatory outcomes (Tambe et al., 2019; Pearl, 2018), the undervaluation of essential human traits such as conscientiousness due to automation (Tang et al., 2021), and the limits of small datasets, which impedes AI's ability to predict rare events or understand causal relationships (Meyer, 2018; Heaven, 2020).

However, the analysis does reveal a critical and underexplored risk that neither of the previously mentioned literature discusses: AI choosing the candidate evaluation filters and its impact on recognizing human potential. Therefore, while existing studies focus on the more general risks, none examine the newly identified risk and its potential implications.

This thesis aims to add to the discussion of AI-driven recruitment by addressing this neglected risk and offering strategies that mitigate its resulting challenges. It is intended to be a call to action for the ethical implementation of AI technologies in recruitment processes, which is highly needed given the gap in the literature. Additionally, it is important to note that while existing studies discuss risks and biases, they often fail to provide practical and actionable recommendations for effectively addressing or resolving these concerns.

6. Revealing a New AI Recruitment Risk: Consequential Outcomes and Ethical Implications

6.1 Identifying Risk of AI-Defined Candidate Evaluation Filters Overlooking Human Potential

While existing literature acknowledges certain risks associated with AI in recruitment, specific gaps remain insufficiently explored. This chapter aims to revisit these research gaps and demonstrate how this thesis contributes to filling them.

Firstly, scholars like Pearl (2018), Tambe et al. (2019), Brown et al. (2020), and Meyer (2018) have highlighted that AI systems can inherit biases from historical data and biased training sets, leading to discriminatory outcomes and perpetuating existing inequalities. However, only little discussion is provided about how these biases arise concretely from the AI autonomously selecting the evaluation filters. These filters usually rely on arbitrary patterns the AI picks up, such as shared characteristics in past data, without questioning the ethics of the criteria. Two examples include considering gender as a factor in candidate evaluation or AI using university prestige as a criterion, influenced by past data showing most successful applicants came from certain schools. Additionally, AI relies on hard metrics and lacks human insights into how candidates could adapt to the company culture, potentially overlooking individuals who excel interpersonally or academically but lack high-profile brand names. This thesis, therefore, defines the concept of AI-defined evaluation filters as a distinct risk.

Secondly, while the benefits of AI in enhancing efficiency and objectivity are well-documented (Hackman et al., 1975; Davenport & Kirby, 2016; Wilson & Daugherty, 2018), less has been said about the negatives of overreliance on hard metrics without detailed human oversight (Barends & Rousseau, 2018; Pfeffer & Sutton, 2006; Rousseau, 2014). Additionally, the limitations of AI in capturing qualitative attributes such as attitude, human potential, and growth capacity are not extensively discussed. This research debates that the exclusive use of quantifiable metrics can lead to overlooking critical qualities, such as cultural fit, that human intuition would pick up on.

Thirdly, discussions around AI in recruitment often focus on the screening and selection stages. This thesis notes that less emphasis is placed on how AI affects the outreach stage and the initial composition of the candidate pool. Especially the role of AI-designed evaluation filters in shaping outreach efforts and potentially excluding under-networked communities is not mentioned and thereby underexplored. Extending the conversation to the outreach stage shows how AI-defined evaluation filters influence which candidates are targeted. Significantly, since AI-driven outreach can undermine inclusivity by potentially excluding talented individuals from diverse backgrounds, it also deepens the understanding of AI's impact across all stages of recruitment.

Furthermore, while AI's tendency to undermine diversity is acknowledged (Pfeffer & Sutton, 2006; Tambe et al., 2019; Meyer, 2018), the tensions within AI practices and company inclusivity messaging remain underexplored. Resulting in a lack of literature on how AI-driven recruitment practices interact with organizational diversity initiatives. This study highlights how AI-driven outreach can work against inclusivity efforts from other departments, thereby wasting resources and decreasing the effectiveness of diversity initiatives. This interdepartmental perspective emphasizes the need to align AI systems and organizational values.

The inability of AI to assess specific soft skills is mentioned in existing literature (Tang et al., 2021; Barends & Rousseau, 2018); however, there is a limited exploration of how this affects candidates with unconventional career paths or unique achievements. This results in the exclusion of candidates who bring innovative perspectives and long-term value due to AI's rigid parameters not being sufficiently addressed. This research describes how the dependence on predefined criteria can exclude candidates, shrink the talent pool, and hence hinder organizational adaptability. Therefore, placing huge importance on considering soft skills and diverse experiences in recruitment.

Moreover, the risks associated with AI in recruitment have been explored either independently or within conceptual frameworks (Barends & Rousseau, 2018; Pfeffer & Sutton, 2006; Denrell et al., 2015; Liu & Denrell, 2018; Tambe et al., 2019), limiting the ability to fully capture their interconnected and multifaceted effects. This thesis thereby fills this gap by clustering the implications of the risk posed by AI-defined evaluation filters into three essential dimensions: Inclusivity & Accessibility, Depth & Accuracy in Candidate Evaluation, and Organizational Adaptability & Long-Term Impact. This comprehensive framework presents the directions of impact as well as the interconnections between the resulting risk implications for a more holistic understanding.

Overall, the findings of this thesis reveal the risk of AI-defined Candidate Evaluation Filters Overlooking Human Potential based on how AI prioritizes narrow or biased criteria. It also highlights the resulting implications and outcomes, such as failing to consider attributes that more accurately reflect an applicant's suitability for the role. Furthermore, qualities like attitude, often subtly conveyed in a CV through word choice, may be overlooked, or the AI might fail to recognize other valuable signs of a candidate's potential.

The risk of AI-Defined Candidate Evaluation Filters can be influenced by the moderator quality and scope of the sample data used to define screening criteria (see Figure 5). AI systems rely on training their algorithms using historical data, which tends to be a narrow sample that does not represent the diversity of the broader talent pool. Also, the data is moderated by two aspects. First, limited or unrepresentative sample groups may have AI identify trends that are not universal or current preferences by the company (see Figure 5). For instance, if the previous history of hiring indicates partiality to candidates from a small set of prestigious universities, then the AI may only pay attention to these few universities and exclude candidates with excellent skills and potential from other educational backgrounds. This narrow focus produces biased hiring outcomes, excluding qualified individuals whose strengths are just as valuable but who come from less traditional institutions.

Second, the inclusion of sensitive attributes, such as gender, race, or names, in training data can cause AI systems to adopt unintended biases, leading to discriminatory practices (see Figure 5). For instance, if the historical data indicates that males hold the majority of managerial positions, AI might start giving more preference to male candidates by considering such a trend as being positive and not discriminatory behavior. Similarly, the fewer people of color there are in senior positions, the more the AI could devalue candidates from other races and, therefore, continue to practice unequal hiring. Lastly, this thesis aims to avoid the negative case where a candidate is excluded for meeting all but one criterion deemed most important by the AI. As a result, the AI may fail to recognize that strengths in other areas could effectively compensate for this single shortfall.

Overall, the thesis makes a significant theoretical contribution by introducing the new risk and building upon and extending current theories on AI biases. It highlights that exclusionary practices and the undermining of inclusivity may result from biased training data and the particular design of the AI filters. Thereby offering an understanding of AI's impact beyond traditional data bias concerns. Additionally, by integrating new aspects, such as the inclusion of sensitive attributes and reframing moderating factors, the research enhances existing research on dataset limitations and bias amplification. This underlines that data quality, not quantity, is what shapes the outcome of AI. This sheds new light on challenges arising from AI-defined candidate evaluation filters that have previously been overlooked, offering fresh perspectives on the ethical and practical implications of AI in the recruitment process (see Chapter 6.2).

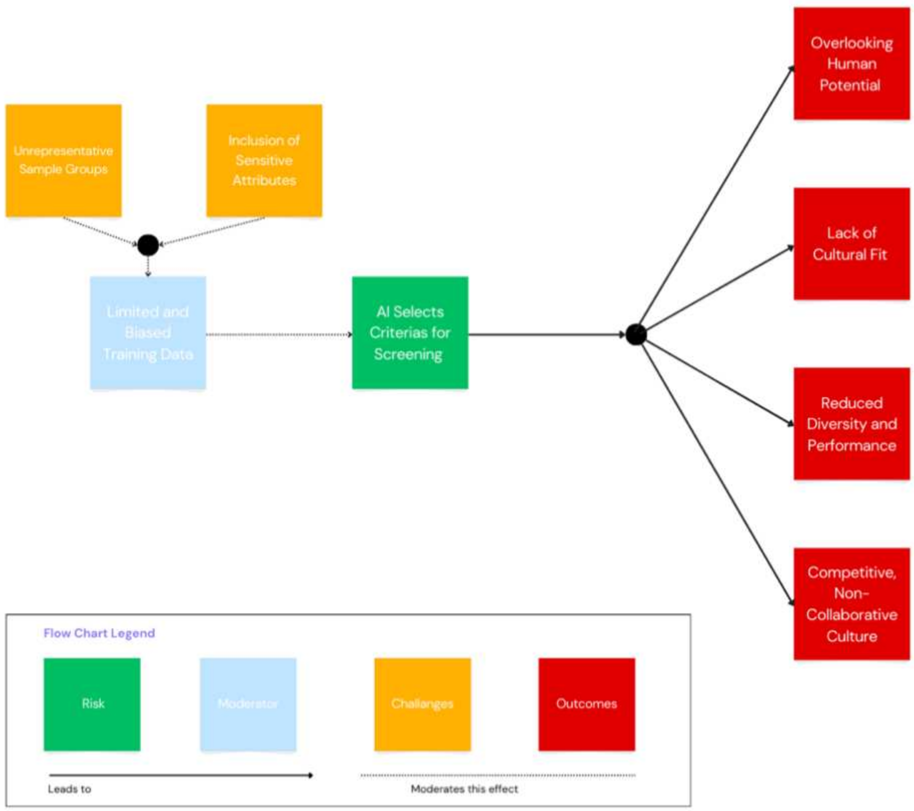

Figure 5: Organizational Effects of AI-Defined Evaluation Filters and Their Moderators

6.2 Impact of AI-Defined Candidate Evaluation Filters Across the Recruitment Process

The identified risks and outcomes across recruitment stages can be categorized into three dimensions: Inclusivity & Accessibility, which focuses on reducing barriers for diverse candidates and expanding the talent pool; Depth & Accuracy of Candidate Evaluation, emphasizing balanced and comprehensive assessments; and Organizational Adaptability & Long-Term Impact, which prioritizes fostering diversity and driving innovation for sustained success. This categorization aims to emphasize the interconnected nature of the 11 outcomes, providing a clear framework to categorize and understand their impact (see Figure 6).

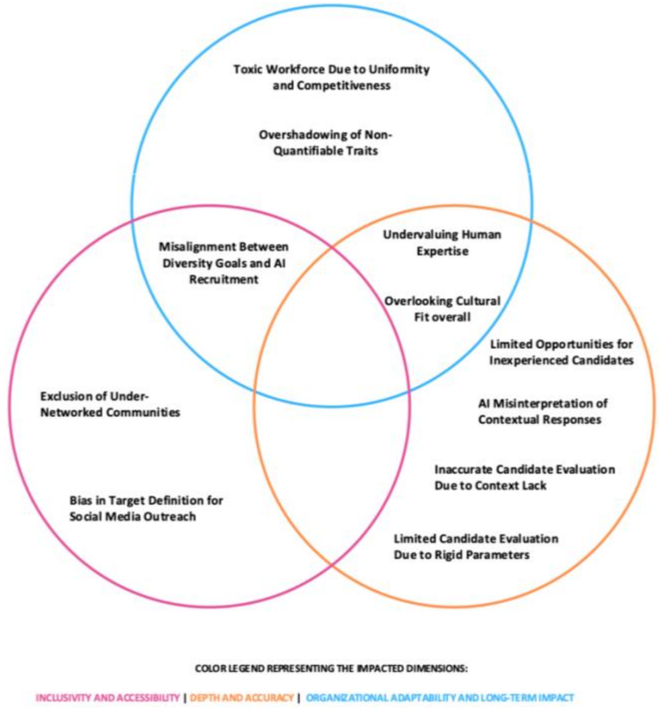

Figure 6: Visualization of the Impact of AI-Defined Evaluation Filters Across Stages

6.2.1 Outreach Stage

The outreach stage of the recruitment process is increasingly shaped by AI-defined evaluation filters that determine which candidates to target, particularly on platforms like LinkedIn or generally through social media (Olcot, 2024). In fact, as of 2019, 44% of companies have been using AI to source the right candidates from social media (Mujtaba & Mahapatra, 2024). Therefore, when placing more weight on university rankings or employer prestige, certain candidates are disproportionately placed as ideal prospects. Similarly, in personalized marketing and targeted social media advertisements—which select audiences based on factors such as demographics or interests (Babatunde et al., 2024)—the algorithm may direct opportunities exclusively toward profiles that align with these biased filters. The outcome is an Inclusivity & Accessibility issue, which creates a significant barrier for talented individuals from diverse backgrounds. This directly contradicts the purpose of outreach: to cast a wide net and engage a broad, representative pool of candidates.

In addition, outreach is not confined to online channels or targeted ads. It increasingly encompasses event sponsorships, whether professional or university events, a trend backed by Aptitude Research Partners' recent findings that 70% of companies recruit at on-campus events (Rivera, 2024; Team, 2024). The focus being cast primarily on those events that reflect the ones favored in AI-selected evaluation filters, such as the ones usually attended by well-connected candidates and top-rated institutions, fail to tap into under-networked communities, closing out competent candidates who would otherwise thrive given the chance. This again creates a focused initiative that moves away from, rather than enhances, Inclusivity & Accessibility in the recruitment process.

The third outcome is that AI-powered outreach might inadvertently undermine inclusion efforts from other departments of the company. According to McKinsey, HR and marketing teams spend significant time on gender-neutral, inclusive job descriptions, with U.S. companies investing around $8 billion annually in diversity and inclusion initiatives (King, 2023). Nevertheless, evaluation filters derived from biased historical data, such as the longstanding predominance of men in higher-ranking roles, reinforce these biases in targeting decisions. The outcome is a contradiction between the company's inclusivity messaging and outreach practices, wasting resources and diminishing the effectiveness of diversity initiatives. Therefore, presenting a dual challenge, impacting both Organizational Adaptability & Long-Term Impact and Inclusivity & Accessibility.

6.2.2 Screening Stage

When dealing with candidates who apply independently, AI-driven screening risks overlooking highly motivated individuals with a strong passion for the company or unique growth signals.
That is because the filtering criteria for the evaluation are one-dimensional metrics dependent on a strict skill match (Babatunde et al., 2024).
For instance, applicants without a certain skill listed on their CV might get rejected despite having transferable experience or a natural capacity for learning fast (Sankalana, 2023). Resumes with soft skills, unusual career progression, or exceptional life experiences are also downgraded for the same reason, as the pre-set filters do not cover their qualities. This approach ignores insights that HR specialists might recognize, such as patterns of success tied to non-linear career paths or traits like adaptability and resilience, which have proven valuable in past hires. Promising candidates with unconventional but high-potential profiles remain outside the selection fold. The talent pool is limiting its depth and the opportunity to include individuals who bring innovative perspectives

and significant long-term value to the organization. Thereby impacting the dimension of Depth & Accuracy of candidate evaluations.

Secondly, AI algorithms struggle to provide accurate candidate skills assessment, failing to contextualize their achievements or challenges (Wonchała, 2024). This can be noticed when a candidate who does well in a harsh socioeconomic environment or a candidate who successfully switched industries would not be noticed if AI simply looked at the raw metric for performance without context. Such lack of the dimension of Depth & Accuracy means that the organization's workforce does not benefit from characteristic values such as resilience during periods of economic downturn or adaptation to different areas.

6.2.3 Assessment Stage

Too much reliance on AI insights risks the human evaluator undervaluing his judgment and qualitative assessment of the candidates. For example, facial Expression Recognition (FER) technologies are hugely susceptible to the datasets on which they are trained and can easily introduce biases (Zahara et al., 2020; Raposo, 2023). An AI assessment might flag a candidate as 'unconfident' by merely assessing his facial signs. In contrast, the human evaluator would interpret such non-verbal hints as nervousness related to stress and not inadequacy in capability. Therefore, this process cuts out the HR professionals and their rich judgment, hence wasting a very valuable organizational asset. Economically, it represents a loss of investment in developing HR expertise without its practical use, ultimately reducing the overall effectiveness of the recruitment process.
Thereby, it results in a lack of accuracy in candidate evaluation and long-term organizational impact, as the human element—the most valuable resource for fair and nuanced assessments—is deprioritized in favor of AI-driven processes.

Furthermore, AI systems often rely on measurable outputs, such as facial expressions or word choice, to conclude complex, qualitative traits like creativity, empathy, or adaptability (Marr, 2024; Psico-smart Editorial Team, 2024). However, the software is unable to accurately capture an individual's entire personality (Psico-smart Editorial Team, 2024). These character traits tend to emerge in less measurable or observable ways for algorithms, although they are essential for specific roles. The overshadowing of non-quantifiable traits runs a significant risk of rejecting candidates with necessary but less measurable qualities, weakening diversity in character traits. This results in the screening stage becoming more meaningful due to AI's strength in hard metrics, while the assessment stage becomes secondary due to AI's weakness in this area. This outcome can be primarily clustered under long-term organizational impact and adaptability, as it limits the company's success by reducing the inclusion of critical qualitative traits essential for adaptability, innovation, and growth.

6.2.4 Selection Stage

In the selection stage, candidates are sometimes required to complete a final test. Suppose test answers are not completely clear or require nuanced understanding that might be accrued through many years in a field. In that case, they may be beyond the capabilities of AI to evaluate correctly. Lacking the contextual understanding that professionals develop over time, AI could overlook valuable insights demonstrated in the responses (Hoffman, 2023). The outcome is missing Depth & Accuracy of Candidate Evaluations by scoring highly suitable candidates inaccurately and ranking them lower. This again limits the assessment of a candidate to only the predefined hard metrics and does not allow areas beyond a traditional CV to be captured.

6.2.5 Broad Organizational Impacts Beyond Stage-Specific Risks

Finally, the cross-process impacts are multifaceted, with one important outcome being a lack of cultural fit (see figure 5). Suppose the focus is overly strong on past performance, especially when benchmarked against current high performers. In that case, it can completely overlook whether the high performers themselves are a good cultural fit for the company. This approach assumes that accomplishments alone reflect compatibility with the company's culture and work environment, which is often not the case. As a result, the AI may ignore candidates whose values, attitudes, and potential cultural alignment could make them an ideal fit. Therefore, it can be clustered into two perspectives: Depth & Accuracy of Candidate Evaluation and Organizational Adaptability & Long-Term Impact

Secondly, this approach lacks Depth & Accuracy of Candidate Evaluation by not offering limited opportunities for newer or younger candidates who may be seeking to build their accomplishments and gain experience. Many of these applicants are very motivated, driven, and willing to contribute but are filtered out because of the narrow focus. Thereby resulting in the outcome of overlooking human potential (see Figure 5)—an outcome that has been ignored in previous literature.

Finally, this recruitment process can generate a workforce of employees with the same qualifications, backgrounds, and strengths. Just as this may create a threshold of competence, it most often means a team lacking different perspectives, skills, and life experiences, which are vital for strong teamwork and driving innovation. Ultimately, such uniformity can limit organizational adaptability and long-term success and weaken its collaborative and problem-solving capabilities due to the absence of varied worldviews and unique strengths (Shrestha & Parajuli, 2021). In addition, it may have certain toxic side effects, including an over-competitive environment where employees may be out for self-promotion, creating a culture where one proves oneself rather than contributes to the goals shared by the organization.

To understand how existing literature on ethical risks and opportunities in AI recruitment completely overlooks the impact of AI-defined evaluation filters and how its implications fall on different recruitment phases, see Appendix C.

7. Mitigation Strategies for Risks Resulting from AI-Defined Evaluation Filters

This mitigation framework addresses the challenges stemming from the newly identified risk across the recruitment process by focusing on the previously defined dimensions. First, it focuses on Inclusivity & Accessibility by employing strategies to reduce barriers for diverse candidates and widen the talent pool. Second, it enhances the Depth & Accuracy by encouraging a balanced assessment of potential, integrating qualitative and quantitative factors. Finally, it emphasizes Organizational Adaptability & Long-term Impact to reach workforce diversity and adaptability that drives innovation and continued success.

The framework also considers two essential factors: resource intensity - the level of investment in terms of money and human resources involved (low, moderate, or high) - and Implementation Horizon, which takes into account strategies yielding short-term benefits versus contributing to long-term sustained results (See figure 7).

7.1 Outreach Stage

Current literature states that while the benefits of AI in improving recruitment efficiency and objectivity are well-documented (Hackman et al., 1975; Davenport & Kirby, 2016; Wilson & Daugherty, 2018), the focus remains on the screening and selection stages with less attention toward how AI shapes the initial outreach and composition of the candidate pool. For instance, narrowly focused outreach algorithms and overreliance on university prestige limit diversity by targeting only a specific group of candidates. Therefore, through the literature review, it has been uncovered that the lack of alternative outreach strategies perpetuates systemic exclusion, particularly of under-networked communities and candidates from non-traditional backgrounds.

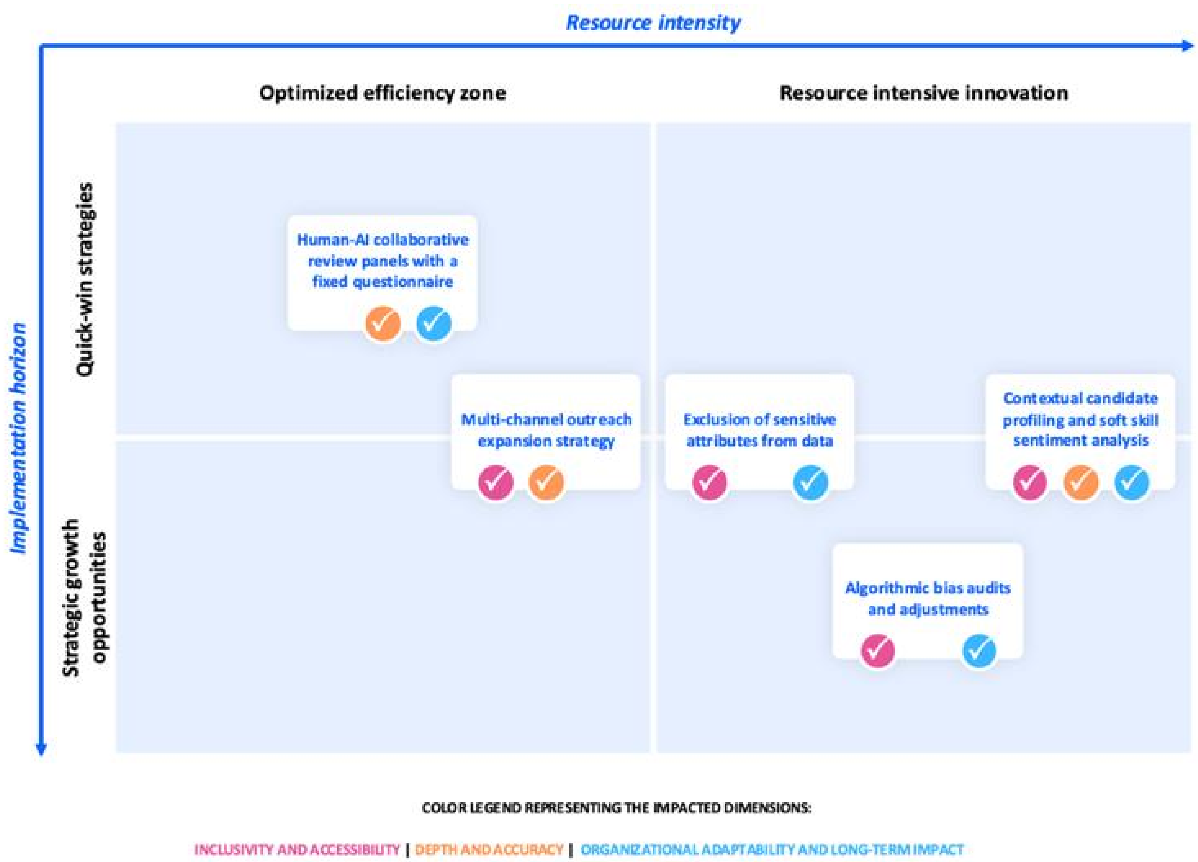

Figure 7: Matrix of Mitigation Strategies for AI-Driven Recruitment Biases

This gap undermines organizational efforts toward inclusivity and diversity while fostering workforce homogeneity. Therefore, recommending a mitigation strategy of a multi-channel outreach approach that reduces the risk of developing a homogeneous workforce. The strategy involves using alternative channels to reach underrepresented groups beyond just algorithm-driven

platforms. This includes engaging with in-company networks, hosting events in the local community, and directly connecting with diversity-focused professional networks. This strategy mitigates the risk of creating a workforce with similar backgrounds and strengths, which could lead to a toxic environment where diverse perspectives are lost. It also fosters positive effects on the dimensions of Inclusivity & Accessibility and Organizational Adaptability & Long-Term Impact.
The resource intensity of this strategy is moderate since it presupposes financial and human resources to research suitable organizations, build partnerships, and attend or sponsor various events. Still, it does not require a substantial initial investment, such as creating advanced technological infrastructure. Organizations can minimize costs by leveraging internal resources, such as employee ambassadors or virtual webinars, while adjusting the number of events or sponsorship levels to align with budgets without compromising impact. The impact of this strategy unfolds over the medium term due to the gradual changes it entails. Given how establishing outreach channels and partnerships takes time and effort, while nurturing new talent pipelines and raising awareness about opportunities (see Figure 7) also does. Candidates from these communities may need additional guidance, such as support with application processes, which delays immediate results. Furthermore, cultural changes within the organization take time to materialize since employees who have gotten used to doing things one way may resist these inclusive practices. With more time, this resistance eventually wanes, and as those benefits of the strategy unfold, so will the improved recruiting outcomes.

The current literature also recognizes leading scholars such as Pearl (2018) and Tambe et al. (2019), stating that AI training data biases bring out discriminatory impacts. However, the literature review highlights no efforts in systematically auditing AI algorithms to address potential biases, such as those induced by parameter selection. This is a critical gap given how AI-driven systems, without frequent recalibration, amplify existing patterns, hindering long-term inclusivity. Consequently, the second mitigation strategy addresses the contradiction between inclusivity messaging and outreach practices, tackling the challenges that weaken broader diversity initiatives.
Resulting in the recommended approach is to implement Algorithmic Bias Audits and Adjustments. This involves regularly evaluating AI algorithms to identify and remove evaluation filters that perpetuate biases, such as those based on gender, institution type, or socioeconomic class. Additionally, developing dynamic algorithms ensures adaptability to evolving inclusion goals. Grounded in methodologies like End-to-End Socio-Technical Algorithmic Audits

(E2EST/AA) (Galdon Clavell, 2023), this recommendation evaluates AI systems comprehensively, addressing both technical and social dimensions. Focusing on the pre-processing and post-processing stages uncovers biases that technical evaluations might overlook. Tools such as Model Cards document algorithm design and training processes, while system maps could place AI decisions within organizational workflows, ensuring alignment with inclusivity goals (Galdon Clavell, 2023). This strategy aims to enhance the dimensions of inclusivity & Accessibility by reducing systemic biases in targeting candidates and Organizational Adaptability & Long-Term Impact by aligning outreach practices with the company's diversity and inclusion objectives.
This resource intensity is classified as high, given how it requires specialized expertise, ongoing evaluations, and substantial financial investment. Advanced technical capabilities and continuous integration into the recruitment system are required for dynamically developing algorithms. Moreover, the Impact is categorized as long-term since addressing systemic biases is very complex, and evolving algorithms to achieve diversity goals takes time (see Figure 7).
The final results from this approach depend on gradual alignment between AI recruitment practices and inclusivity initiatives, emerging over an extended period.

Finally, the current literature, as analyzed, acknowledges AI's potential to perpetuate biases, particularly through sensitive attributes like gender, race, or names. However, the discussion often focuses on theoretical implications rather than actionable solutions. The literature review revealed that removing sensitive attributes from training data is an underutilized strategy to reduce bias. This oversight unintentionally allows AI systems to perpetuate biased markers, undermining inclusivity, and equity in recruitment practices. As a result, the third mitigation strategy recommends excluding sensitive attributes—such as gender, race, or names—from training data to prevent the algorithm from inferring bias. This makes the decision-making process fairer, avoiding dependencies on attributes that may propagate systemic biases. By removing sensitive information, the algorithm can also focus on objective qualifications and skills that reduce the overrepresentation of patterns that are not representative of company-wide priorities or values. The approach directly contributes to improving the dimension of Inclusivity & Accessibility by furthering unbiased recruitment processes. It also adds to the dimension of Organizational Adaptability & Long-Term Impact by fostering a more diverse workforce, which enhances innovation and overall capabilities.

The resource intensity for this approach is low, given that it mainly involves excluding sensitive attributes rather than expanding datasets, hence minimizing the need for extensive data collection and curation. The timeframe of impact is medium-term since the exclusion of sensitive attributes and retraining processes can be completed within a foreseeable period (see Figure 7). However, realizing the full benefits depends on validating the updated models, implementing revised recruitment practices, and observing measurable workforce diversity and inclusion improvements over time.

Concluding to all three of these mitigation strategies offering actionable guidance for managers: multi-channel outreach to widen talent pipelines, algorithm audits to overcome systemic biases, and refined data practices to make decisions fair and representative.

7.2 Screening Stage

While current literature says that scholars such as Tang et al. (2021) and Barends & Rousseau (2018) recognize AI's limitations in assessing soft skills and nuanced traits, through the literature review, it has been uncovered that insufficient attention is given to actual measurements integrating qualitative, contextual measurements into AI recruitment systems. The existing frameworks only focus on hard metrics such as qualifications and work history, overlooking attributes such as resilience, adaptability, and personal motivations. This results in a systemic gap, which restraints the recognition of unconventional talent and results in a lack of high-potential candidates from non-traditional backgrounds. Therefore, in this case, the recommended action is a contextual Candidate Profiling and Soft Skill Sentiment Analysis involving integrating personal questions into the early application process, allowing AI to analyze CVs alongside candidate-provided statements about challenges overcome, achievements in non-traditional settings, or personal motivations.

By applying natural language processing (NLP), particularly natural language understanding (NLU), the AI could identify positive qualities such as teamwork, resilience, deep understanding, or interest in the company (DeepLearning.AI, 2023). In that way, it captures all those traits that static metrics usually miss and ensures the evaluation of a candidate's potential will be more thorough. The strategy improves all three dimensions: Inclusivity & Accessibility by considering diverse achievements and unconventional backgrounds, Depth & Accuracy in candidate evaluation by considering soft skills and contextual factors, and

Organizational Adaptability & Long-Term impact through fostering resilience and adaptability in the workforce. The resource intensity is classified as high, as this strategy involves enhancing the recruitment process by integrating qualitative questions alongside CV submissions, requiring advanced AI models, NLU capabilities, and extensive data processing. The timeframe of impact is medium-term, reflecting the effort needed for implementation, training, and fine-tuning. However, it ultimately delivers sustainable benefits in talent identification and workforce diversity (see Figure 7).

This results in overall actionable guidance for managers by suggesting the integration of contextual candidate profiling and soft skill sentiment analysis into the recruitment process. This could enable AI to evaluate nuanced qualities alongside conventional metrics, thereby achieving more accurate candidate assessments by recognizing soft skills and personality traits as key potential indicators.

7.3 Assessment & Selection Stage

Current literature documents that while AI excels in processing objective metrics, it cannot decipher nuances of traits or specialized knowledge acquired through field-specific experience. The literature review made it clear that little attention has been paid so far to integrating human judgment within the AI-driven evaluation process. This gap limits the ability to contextualize qualitative indicators, such as micro-expressions, that would be important for a holistic assessment of candidates. The Human-AI Collaborative Review Panels with a Fixed Questionnaire strategy establish a hybrid assessment process whereby AI provides initial insights while human reviewers conduct final evaluations based on a structured questionnaire. This questionnaire ensures standardization, fairness, and consistency across all candidates while reinforcing the understanding that AI outputs are not absolute. For instance, questions like, *"If flagged as nervous, is the candidate unsure of their answers or simply nervous?"* or *"Do the candidate's responses reflect creativity or empathy?"* prompt HR experts to contextualize traits in the assessment stage such as micro-expressions, word choice, and other qualitative indicators.

This approach strengthens contextual understanding in the assessment and selection stage, especially during final test evaluations. While the AI will process objective metrics, such as numerical data, human experts will review ambiguous responses or reflect on expertise acquired through many years of field-specific experience. This strategy ensures a comprehensive and balanced evaluation by addressing the risk of AI's inability to recognize specialized knowledge

and its tendency to overlook contextual information. Thereby reinforcing two key dimensions: Depth & Accuracy of Candidate Evaluation through integrating AI insights with human judgment and Organizational Adaptability & Long-Term Impact by valuing qualitative attributes along with quantitative metrics that foster diverse skills and perspectives for continued growth.
The resource intensity in this case is moderate, as it leverages existing HR competence, requiring only minor training and the development of the fixed questionnaire. The impact timeframe is immediate to medium-term, as the main effort is creating and implementing the questionnaire, which can, therefore, be integrated into current assessment processes quickly (see Figure 7).

In conclusion, the actionable guidance for managers here is to institute Human-AI Collaborative Review Panels with a set questionnaire. This way, fairness, contextual understanding, and thoroughness are assured by blending AI-driven insights with human judgment. With standardized questions, striking a balance between quantifiable metrics and specialized industry knowledge.

8. Conclusion

This thesis investigates the ethical implications of integrating Artificial Intelligence (AI) into recruitment processes, specifically focusing on how organizations can implement strategies to mitigate biases through evaluation filters. By addressing the research question: "*How can organizations deploy AI recruitment systems without risking overlooking human potential, especially among candidates with more individualistic or non-conforming backgrounds*?" the study seeks to bridge a critical gap in existing literature concerning AI's influence on recruitment ethics.

The analysis significantly contributes to the literature on the intersection of business ethics, human resource management, and AI ethics in various ways. Firstly, the systematic literature review outlines the current discussion regarding AI recruitment and its risks while critiquing its limitations. The discussion around AI-driven recruitment emphasizes positive aspects such as efficiency, cost-effectiveness, and the potential to reduce human biases while also addressing critical ethical concerns, including algorithmic bias, transparency, and accountability (Blackman, 2022; Tambe et al., 2019). However, this discussion reveals a critical gap in addressing AI-defined candidate evaluation filters and their significant impact on recognizing human potential. Specifically, the current literature has overlooked how the reliance on strict algorithmic criteria, rooted in narrowly defined skills and historical metrics, can undermine inclusivity and diversity. Thus, this thesis broadens the current understanding by identifying and addressing the dangers of AI-defined filters. It showcases how this risk decreases perceived human potential, cultural fit, teamwork, and performance, resulting in an overall toxic work environment. By considering the implications of this risk at every stage of the recruitment process, this research also shows how challenges produced by AI-defined evaluation filters can reduce inclusivity, diminish the depth and accuracy of candidate evaluations, and hurt organizational adaptability and long-term success. Furthermore, the study suggests five actionable mitigation strategies to blend AI effectiveness with ethical and inclusive practices: Multi-Channel Outreach, Algorithmic Bias Audits and Adjustments, Exclusion of Sensitive Attributes, Contextual Candidate Profiling and Soft Skill Sentiment Analysis, and Human-AI Collaborative Review Panels with Fixed Questionnaires. This set of strategies not only tackles the identified risk but also provides concrete steps for improving the ethics and inclusivity of recruitment processes.

This also results in concrete managerial implications, such as the need to integrate qualitative profiling and sentiment analysis into hiring procedures. Doing so enables a more comprehensive assessment that acknowledges candidates' full potential, counteracting the tendency of AI-defined evaluation filters to underestimate soft skills and individual traits. Furthermore, Biases ingrained in AI systems can foster discriminatory practices. The corresponding managerial implication is the necessity for mitigation strategies, such as multi-channel outreach, algorithmic bias audits, and the exclusion of sensitive attributes from training data. Upholding this implication could ensure that recruitment practices align with diversity and inclusion objectives. Moreover, overreliance on AI can diminish the importance of human judgment in hiring processes. Thus, the managerial implication being the incorporation of human-AI collaborative review panels in order to achieve a balance between AI effectives and human judgment. Additionally, AI-driven outreach frequently limits the variety of candidates by mainly focusing on traditional profiles, resulting in inhomogeneity that undermines adaptability and innovation. Therefore, the managerial implication is to broaden outreach efforts by incorporating marginalized groups and embracing unconventional career paths. This approach can enhance team collaboration, introduce new viewpoints, and help Equip organizations tackle obstacles more effectively.

This thesis is, however, limited by its explicit focus on the application of AI in recruitment and, by nature, constrains the transferability of insights into other organizational uses of AI. Additionally, AI is rapidly evolving, with new risks and ethical challenges continuously emerging.

Many of these risks have yet to be fully documented or studied, potentially providing fresh perspectives and critical information that could influence this work. Therefore, the fast-paced evolution of AI technologies emphasizes the need for continuous updates to research in this area. Moreover, future research could expand the scope to encompass diverse cultural, geographic, and industrial contexts, highlighting region-specific ethical concerns. Additionally, investigating the ethical implications of AI in broader HR functions could extend the applicability of these findings and result in innovative approaches to addressing AI's risks. The ethical challenges and risks of AI in recruitment addressed in this thesis create an avenue for a deepened understanding of the technology's organizational and societal implications. This thesis's goal is also to provoke critical reflection among readers, urging them to consider the potential consequences of AI before adopting it solely for its apparent benefits. This underlines the necessity of a thoughtful approach when working with AI, emphasizing that it must be deployed to mitigate risks while adhering to ethical principles. Thus, organizations must ensure that the means to mitigate potential risks are not only clearly defined but also available. This perspective benefits a broad audience—researchers, academics, organizations, and applicants alike—advocating for a more responsible and conscientious approach to AI integration. The findings have the potential to guide theoretical advancements in AI governance and reinforce the need to institutionalize AI audits within organizations.

Ultimately, while a utilitarian perspective, which advocates for actions that promote the greatest good for the greatest number, might view AI in recruitment as an ethical tool due to its efficiency, cost reduction, and expanded hiring reach (Mill, 1879; Mori et al., 2024), this viewpoint supports a limited managerial perspective that overlooks the critical importance of a diverse workforce. From an ethical standpoint, particularly concerning fairness, the presence of biased selection criteria raises serious concerns about equity. Such biases can unjustly exclude talented candidates simply because they do not align with conventional markers of "success," such as specific genders, races, or prestigious educational backgrounds—criteria that many capable individuals, including you, the reader, and your loved ones, may not meet. This results in such exclusion going beyond being purely a business issue and carrying significant social and ethical implications. It particularly affects socio-economic groups that are capable and willing yet deprived of opportunities.

Finally, this thesis is grounded in the ethical principle of equal opportunity in the workforce, aligning with Bowles' (1973) argument that income inequality can be traced back to mainly one factor: unequal opportunities.

9. Appendices

Appendix A: Systematic Literature Review - Summary of Initial Relevant Studies

Title	Journal / Publication	Database
Keywords: **AI recruitment opportunities & Ethics in AI recruitment & Artificial Intelligence recruitment Opportunities** *(11 relevant sources)*		
The power of artificial intelligence in recruitment: An analytical review of current AI-based recruitment strategies	Journal of Professional Business Review	Google Scholar
Innovating HRM recruitment: a comprehensive review of AI deployment.	Marketing and Management of Innovations	Google Scholar
Fairness, AI & recruitment.	Computer Law & Security Review	Google Scholar
Artificial intelligence–challenges and opportunities for international HRM: a review and research agenda	Journal of Human Resource Management	Google Scholar
The impact of artificial intelligence within the recruitment industry: Defining a new way of recruiting	Carmichael Fisher Executive Search	Google Scholar
Ethics guidelines for using AI-based algorithms in recruiting: Learnings from a systematic literature review	Proceedings of the 55th Hawaii International Conference on System Sciences	Google Scholar
Disability, fairness, and algorithmic bias in AI recruitment.	Ethics and Information Technology	Google Scholar
Analysis and issues of artificial intelligence ethics in the process of recruitment	2nd International Conference on Smart Electronics and Communication	Google Scholar
Artificial intelligence and the changing sources of competitive advantage.	Strategic Management Journal	Google Scholar
Machine learning and human capital complementarities: Experimental evidence on bias mitigation.	Strategic Management Journal	Google Scholar
Human resource management in the era of artificial intelligence: future HR work practices, anticipated skill set, financial and legal implications.	Strategic Management Journal	Google Scholar
Keywords: **AI recruitment ethical risks & AI recruitment risks** *(11 relevant sources)*		
Ethics and discrimination in artificial intelligence-enabled recruitment practices,	Humanities and Social Sciences Communications	Google Scholar
Ethics of AI-enabled recruiting and selection: A review and research agenda	Journal of Business Ethics	Google Scholar
Is AI recruiting (un) ethical? A human rights perspective on the use of AI for hiring	AI and Ethics	Google Scholar

Title	Journal / Publication	Database
Ethical concerns while using artificial intelligence in recruitment of employees	Business Ethics and Leadership	Google Scholar
The power of artificial intelligence in recruitment: An analytical review of current AI-based recruitment strategies	Journal of Professional Business Review	Google Scholar
When and how artificial intelligence augments employee creativity	Academy of Management Journal	Google Scholar
Human resource management in the era of artificial intelligence: future HR work practices, anticipated skill set, financial and legal implications	Strategic Management Journal	Google Scholar
Proposed strategic framework for effective artificial intelligence adoption in UAE.	Strategic Management Journal	Google Scholar
Bright and dark imagining: How creators navigate moral consequences of developing ideas for artificial intelligence.	Academy of Management Journal	Google Scholar
When conscientious employees meet intelligent machines: An integrative approach inspired by complementarity theory and role theory.	Academy of Management Journal	Google Scholar
Why do firms fail to engage diversity? A behavioral strategy perspective	Organization Science	Google Scholar
Keywords: AI recruitment *(4 relevant sources)*		
Recruiter's perception of artificial intelligence (AI)-based tools in recruitment	Computers in Human Behaviour Reports	ScienceDirect
AI-enabled recruiting in the war for talent	Business Horizons	ScienceDirect
Applicants' perception of artificial intelligence in the recruitment process	Computers in Human Behaviour Reports	ScienceDirect
Marketing AI recruitment: The next phase in job application and selection	Computers in Human Behaviour	ScienceDirect
Keywords: AI recruitment risks *(6 relevant sources)*		
The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work	Journal of Business Ethics	EBSCO
Using AI to make or influence workplace decisions: legal considerations and mitigating risk	Compliance & Risk	EBSCO
When Machines Make Hiring Decisions: Examining the Risks and Limitations of AI-Based Recruitment Tools	Florida State University Law Review	EBSCO
Disability discrimination in UK recruitment and the impact of AI.	Compliance & Risk	EBSCO

Title	Journal / Publication	Database
The Artificial Recruiter: Risks of Discrimination in Employers' Use of AI and Automated Decision-Making	Social Inclusion	EBSCO
Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda	Journal of Business Ethics	EBSCO
Keywords: AI recruitment ethics *(5 relevant sources)*		
Incorporating artificial intelligence (AI) into recruitment processes: ethical considerations	Vilakshan - XIMB Journal of Management	EBSCO
Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda	Journal of Business Ethics	EBSCO
The Role of Artificial Intelligence in Recruitment Process Decision-Making	International Conference on Decision Aid Sciences and Application	EBSCO
The Power of Artificial Intelligence in Recruitment: An Analytical Review of Current AI-Based Recruitment Strategies	Journal of Professional Business Review	EBSCO
Analysis and Issues of Artificial Intelligence Ethics in the Process of Recruitment	2nd International Conference on Smart Electronics and Communication	EBSCO
Keywords: Automation--Economic aspects *(2 relevant sources)*		
Only Humans Need to Apply: Winners and Losers in the Age of Smart Machines	Harper Business	EBSCO
Performance Persistence Through the Lens of Chance Models: When Strong Effects of Regression to the Mean Lead to Non-Monotonic Performance Associations	Academy of Management Proceedings	EBSCO
Keywords: Artificial Intelligence / Artificial Intelligence in business *(8 relevant sources)*		
Human + Machine: Reimagining Work in the Age of AI	Harvard Business Review Press	EBSCO
Language Models Are Unsupervised Multitask Learners	OpenAI	EBSCO
Language Models Are Few-Shot Learners	Proceedings of the 34th International Conference on Neural Information Processing Systems	EBSCO
Your AI efforts Won't succeed unless they benefit employees	Harvard Business Review	EBSCO
Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models	Journal of Product Innovation Management	EBSCO
Prediction machines: The simple economics of artificial intelligence	Ingram Publisher Services	EBSCO

Title	Journal / Publication	Database
A Literature Review: Artificial Intelligence Impact on the Recruitment Process	International Journal of Engineering and Management Sciences	EBSCO
Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring	AI and Ethics	EBSCO
Keywords: Personnel Management *(5 relevant sources)*		
Development of the job Diagnostic Survey	Journal of Applied Psychology	EBSCO
Artificial intelligence and management: The automation-augmentation paradox	Academy of Management Review	EBSCO
Innovative idea generation in problem finding: Abductive reasoning, cognitive impediments, and the promise of artificial intelligence	Journal of Product Innovation Management	EBSCO
Organizational stress: Studies in role conflict and ambiguity	Wiley	EBSCO
Artificial Intelligence in Human Resources Management: Challenges and a Path Forward	California Management Review	EBSCO
Keywords: Design Thinking / Design theory *(5 relevant sources)*		
Innovation as a Learning Process: Embedding Design Thinking	California Management Review	EBSCO
Generative Sensing in Design Evaluation	Design Studies	EBSCO
Enhancing creativity through "mindless" work: A framework of workday design	Organization Science	EBSCO
On the pattern of reasoning in innovative design	Design Studies	EBSCO
The core of 'design thinking' and its application	Design Studies	EBSCO
Keywords: Industrial robots *(1 relevant source)*		
Robots and jobs: Evidence from US labor markets	Journal of Political Economy	EBSCO
Category: Business & Economics / Human Resources & Personnel management *(1 relevant source)*		
Evidence-Based Management: How to Use Evidence to Make Better Organizational Decisions	Kogan Page	EBSCO

Title	Journal / Publication	Database
Category: Business & Economics / Management *(4 relevant sources)*		
Hard Facts, Dangerous Half-Truths, and Total Nonsense: Profiting from Evidence-Based Management	Harvard Business Review Press	EBSCO
The Oxford Handbook of Evidence-Based Management	Oxford University Press	EBSCO
Change Explanations in Management Science	Organization Science	EBSCO
Paradox research in management science: Looking back to move forward	Acade of Management Annals	EBSCO

Appendix B: Key AI Challenges Across Recruitment Stages

Risks idenditifed from literature

Risk	Rational for stage alignment	Supporting literature
Outreach		
Risk of Reinforced Stereotypes	AI used in outreach may reinforce societal biases embedded in its training data, leading to job ads that target only certain groups and potentially limiting diversity.	Meyer, 2018; Brown et al., 2020; Bouschery et al., 2023
Screening		
Disruption of Trait Balance	Reducing the emphasis on human judgment and traits like conscientiousness, which have traditionally played an important role in screening, could diminish the quality, logic, and cultural fit of candidates advancing through this stage.	Carson, 1969; Kiesler, 1983; Grant et al., 2011; Barrick & Mount, 2012; Tang et al., 2021
Blurred Role Expectations	AI automation in screening tasks can blur expectations around human oversight, reducing the level of attention and personal care dedicated to each candidate's evaluation.	Kahn et al., 1964; Tang et al., 2021; Barrick & Mount, 2012
Candidate assessment		
Perceived Random Decisions	Randomized elements in AI-driven assessments can make decisions appear unfair, reducing transparency and trust in the assessment process.	Denrell, Fang, & Liu, 2015; Liu & Denrell, 2018
Bias Amplification Due to Limited Data / Risk of Bias	AI systems relying on limited data may struggle to accurately assess candidates' expertise, potential, and other critical qualities, potentially leading to biased evaluations that rely on stereotypes and disproportionately affect vulnerable groups.	Pearl, 2018; Tambe et al., 2019; Radford et al., 2019; Heaven, 2020, Barends & Rousseau, 2018; Pfeffer & Sutton, 2006; Rousseau, 2014
Selection		
Bias Amplification Due to Limited Data / Risk of Bias	AI systems relying on limited data may struggle to accurately assess candidates' expertise, potential, and other critical qualities, potentially leading to biased evaluations that rely on stereotypes and disproportionately affect vulnerable groups.	Pearl, 2018; Tambe et al., 2019; Radford et al., 2019; Heaven, 2020, Barends & Rousseau, 2018; Pfeffer & Sutton, 2006; Rousseau, 2014

Appendix C: Impact of Risks Resulting from AI-Defined Evaluation Filters Across Recruitment Stages

Risks identified from literature

Risk	Rationale for assigned stage
Outreach	
Risk of Reinforced Stereotypes	AI used in outreach may reinforce societal biases embedded in its training data, leading to job ads that target only certain groups and potentially limiting diversity.
Screening	
Disruption of Trait Balance	Reducing the emphasis on human judgment and traits like conscientiousness, which have traditionally played an important role in screening, could diminish the quality, logic, and cultural fit of candidates advancing through this stage.
Blurred Role Expectations	AI automation in screening tasks can blur expectations around human oversight, reducing the level of attention and personal care dedicated to each candidate's evaluation.

Risks identified in this thesis

Risk	Description of risk
Outreach	
Bias in Target Definition in Social Media Outreach	AI-defined parameters prioritize factors like university rankings, employer prestige, or demographics, leading to biased targeting decisions. This creates barriers for diverse and underrepresented candidates, restricting access to opportunities and undermining inclusivity and accessibility in outreach efforts.
Exclusion of Under-Networked Communities	AI-driven outreach focuses on candidates from well-networked groups or prestigious institutions, such as through sponsored events or targeted recruitment campaigns. This excludes talented individuals without access to such networks, further marginalizing underrepresented communities and reducing the diversity of the talent pool.
Misalignment Between Diversity Messaging and AI Recruiting	While companies invest in diversity initiatives and gender-neutral job descriptions, AI outreach contradicts these efforts by relying on biased historical data. This diminishes the effectiveness of inclusivity initiatives, wastes resources, and creates inconsistencies in organizational messaging, impacting both adaptability and long-term impact.
Screening	
Limitation of Candidate Evaluation Due to Rigid Parameters	AI-driven screening relies on one-dimensional metrics, overlooking candidates with transferable skills, soft skills, or non-linear career paths. This approach excludes unconventional but high-potential profiles, limiting the depth of the talent pool and missing out on innovative candidates who could bring significant long-term value to the organization.
Inaccurate Candidate Evaluation Due to Lack of Context	AI algorithms fail to contextualize candidates' achievements and challenges, such as excelling in tough socioeconomic environments or transitioning between industries. This results in a lack of depth and accuracy in candidate evaluation, excluding diverse talent with valuable qualities like resilience and adaptability, which could enrich the organization's workforce.

Risks identified from literature		Risks identified in this thesis	
Risk	**Rationale for assigned stage**	**Risk**	**Description of risk**
Candidate assessment		**Candidate assessment**	
Perceived Random Decisions	Randomized elements in AI-driven assessments can make decisions appear unfair, reducing transparency and trust in the assessment process.	Undervaluing Human Expertise	Over-reliance on AI-generated insights sidelines HR professionals' nuanced judgment, undervaluing their ability to interpret qualitative candidate traits. This reduces the overall effectiveness of the recruitment process, wastes investments in HR expertise, and risks dismissing promising candidates based solely on hard parameters, impacting accuracy and long-term organizational success.
Bias Amplification Due to Limited Data / Risk of Bias	AI systems relying on limited data may struggle to accurately assess candidates' expertise, potential, and other critical qualities, potentially leading to biased evaluations that rely on stereotypes and disproportionately affect vulnerable groups.	Overshadowing Non-Quantifiable Traits	The focus on measurable outputs, such as facial expressions or word choice, fails to capture critical qualitative traits like creativity, empathy, or adaptability. This risks disregarding candidates with essential but less quantifiable qualities, limiting the richness of characteristics in the workforce and reducing adaptability, innovation, and long-term organizational impact.
Selection		**Selection**	
Bias Amplification Due to Limited Data / Risk of Bias	AI systems relying on limited data may struggle to accurately assess candidates' expertise, potential, and other critical qualities, potentially leading to biased evaluations that rely on stereotypes and disproportionately affect vulnerable groups.	Inaccurate Candidate Evaluation Due to Lack of Context	Struggling to accurately assess candidates' responses in final tests that require nuanced understanding or field-specific expertise. By relying on predefined hard metrics, AI risks scoring highly suitable candidates inaccurately and ranking them lower, missing depth and accuracy in candidate evaluations and disregarding their potential to excel in areas beyond what is measurable on their CV.

10. References

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from us labor markets. Journal of Political Econ- omy, 128(6), 2188–2244. https://doi.org/10.1086/705716
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

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction machines: The simple economics of artificial intelligence. Boston, MA: Harvard Business Review Press.

AI investment forecast to approach $200 billion globally by 2025. (2023, August 1). *Goldman Sachs*. https://www.goldmansachs.com/insights/articles/ai-investment-forecast-to-approach-200-billion-globally-by-2025

Argyres, N. S., & Zenger, T. R. (2012). Capabilities, transaction costs, and firm boundaries. Organization Science, 23(6), 1643–1657

Babatunde, N. S. O., Odejide, N. O. A., Edunjobi, N. T. E., & Ogundipe, N. D. O. (2024). THE ROLE OF AI IN MARKETING PERSONALIZATION: A THEORETICAL EXPLORATION OF CONSUMER ENGAGEMENT STRATEGIES. *International Journal of Management & Entrepreneurship Research*, *6*(3), 936–949. https://doi.org/10.51594/ijmer.v6i3.964

Barends, Eric and Denise M. Rousseau. 2018. *Evidence-Based Management: How to Use Evidence to Make Better Organizational Decisions.* Kogan Page.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120.

Barrick, M. R., & Mount, M. K. (2012). Select on conscientiousness and emotional stability (Handbook of Principles of Organizational Behavior, pp. 19–39). https://doi.org/10.1002/9781119206422.ch2

Beckman, Sara L., & Michael Barry. (2007). Innovation as a Learning Process: Embedding Design Thinking. California Management Review 50(1): 25–56.

Blackman, R. (2022). *Ethical machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI*. Harvard Business Press. (Blackman, 2022)

Blackman, R. (2024). *Building Ethical Machines with Reid Blackman, Founder & CEO at Virtue Consultants*. DataFramed Podcast, Episode #154. Retrieved from Spotify.

Blehar, M. (2024, May 2). *7 Companies Successfully Using AI in Their Recruiting Strategies*. Phenom. Retrieved October 14, 2024, from https://www.phenom.com/blog/examples-companies-using-ai-recruiting-platform

Bouschery, S. G., Blazevic, V., & Piller, F. T. (2023). Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. Journal of Product Innovation Management, 40(2), 139–153. https://doi.org/10.1111/jpim.12656

Bowles, S. 1973. “Understanding economic opportunity,” *Amer. Econ. Rev.* 63, 346-356

Brand, H. (2008). Proposing and evaluating a model for ethical recruitment and selection. *Journal of Contemporary Management*, *5*(1), 205–222. https://repository.up.ac.za/bitstream/2263/10010/1/Brand_Proposing%282008%29.pdf

Brewer, G. A. (2008). Employee and organizational performance. Motivation in public management: The call of public service, 136, 156.
Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, and Amanda Askell. 2020. “Language Models Are Few-Shot Learners.” ArXiv 2005: 14165.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & company.

Bughin, J., & Manyika, J. (2019). Your AI efforts Won't succeed unless they benefit employees. Harvard Business Review https://hbr.org/2019/07/your-ai-efforts-wont-succeed-unless-they-benefit-employees

Bîgu, D., & Cernea, M.-V,2019. "**Algorithmic Bias In Current Hiring Practices: An Ethical Examination**," Proceedings of the INTERNATIONAL MANAGEMENT CONFERENCE, Faculty of Management, Academy of Economic Studies, Bucharest, Romania, vol. 13(1), pages 1068-1073, November.

Card, D., & DiNardo, J. E. (2002). Skill-Biased technological change and rising wage inequality: some problems and puzzles. *Journal of Labor Economics*, *20*(4), 733–783. https://doi.org/10.1086/342055

Carson, R. C. 1969. ***Interaction concepts of personality***. Chicago, IL: Aldine.

Chatelaine, J. (2024, May 29). *How to create a recruitment strategy*. Cronofy. Retrieved October 5, 2024, from https://www.cronofy.com/blog/how-to-create-a-recruitment-outreach-strategy

Chamorro-Premuzic, T., & Akhtar, R. (2019). Should companies use AI to assess job candidates? *Harvard Business Review, May 17*.

Cheatham, B., Javanmardian, K., & Samandari, H. (2019). Confronting the risks of artificial intelligence. McKinsey Quarterly, 2(38), 1-9.

Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities:Experimental evidence on bias mitigation. Strategic Management Journal, 41(8), 1381–1411

Coté, J. E., & Levine, C. G. (2002). Identity formation, agency and culture. Mahwah, NJ: Lawrence Erlbaum.

Conaboy, k., & Richard P. (2005). "*Corporate Crime in America*: Strengthening the Good Citizen *Corporation* (Washington, DC; US Sentencing Commission), 1-2. Corporation," in *Corporate Crime in America: Strengthening the 'Good Citizenship'*

CSU Global. (2024, October 23). *How does artificial intelligence work? | CSU Global*. Colorado State University Global. https://csuglobal.edu/blog/how-does-ai-actually-work#:~:text=AI%20systems%20work%20by%20combining,performance%20and%20develops%20additional%20expertise.

Das, M. R. (2023, June 26). Fortune 500 companies are using AI to screen candidates. Here's how not to get rejected immediately. *Firstpost*. https://www.firstpost.com/world/fortune-500-companies-are-using-ai-to-screen-candidates-here-is-a-guide-to-not-get-rejected-immediately-12790602.html

Daft,R.L.(2000). Organization Theory and Design. (7th ed.) South-Western College Publishing, Thomson Learning. U.S.A.

Davenport, T. H., & Kirby, J. (2016). Only Humans Need Apply: Winners and Losers in the Age of Smart Machines. HarperBusiness.

Delery, John & Shaw, Jason. (2001). The strategic management of people in work organizations: Review, synthesis, and extension. Research in Personnel and Human Resources Management. 20. 165-197.

Denrell, Jerker, Christina Fang, Chengwei Liu. 2015. "Change Explanations in Management Science." *Organization Science* 26(3): 923-940.

DeepLearning.AI. (2023, January 11). *Natural Language Processing (NLP) [A complete guide]*. https://www.deeplearning.ai/resources/natural-language-processing/

Dong, Andy, Massimo Garbuio, & Dan Lovallo. (2016a). Generative Sensing in Design Evaluation. Design Studies 45: 68–91.

Dorst, K. (2011). The core of 'design thinking' and its application. Design Studies, 32(6), 521e532.

Elsbach, K. D., & Hargadon, A. B. (2006). Enhancing creativity through "mindless" work: A framework of workday design. *Organization Science, 17*(4), 470–483

Erickson, R. (2018) Infographic: Insights into a highly mature talent acquisition team.

FraiJ, J., & László, V. (2021). A Literature Review: Artificial Intelligence Impact on the Recruitment Process. *International Journal of Engineering and Management Sciences*, *6*(1), 108-119. https://doi.org/10.21791/IJEMS.2021.1.10.

Figueroa-Armijos, M., Clark, B. B., & da Motta Veiga, S. P. (2023). Ethical perceptions of AI in hiring and organizational trust: The role of performance expectancy and social influence. *Journal of Business Ethics*, *186*(1), 179-197.

Garbuio, M., & Lin, N. (2021). Innovative idea generation in problem finding: Abductive reasoning, cognitive impediments, and the promise of artificial intelligence. Journal of Product Innovation Management, 38(6), 701–725. https://doi.org/10.1111/jpim.12602

GALDON CLAVELL, G. (2023). AI Auditing - Checklist for AI auditing. In *SUPPORT POOL OF EXPERTS PROGRAMME*. https://www.edpb.europa.eu/system/files/2024-06/ai-auditing_checklist-for-ai-auditing-scores_edpb-spe-programme_en.pdf

Gasparyan, A. (2024, May 15). Candidate Assessment: Definition and Overview | DevsData. *DevsData LLC - Premium IT Recruitment Agency and Software Development Services*. Retrieved October 6, 2024, from https://devsdata.com/candidate-assessment-definition/#:~:text=Candidate%20assessment%20is%20the%20process,well%2Dsuited%20for%20the%20job.

Gert (2012), "The Definition of Morality," in Stanford Encyclopedia of Philosophy (Fall 2012 edition), ed. Edward N. Zalta, http://plato.stanford.edu/archives/fall2012/entries/morality-definition/.

George, G., Lakhani, K. R., & Puranam, P. (2020). What has changed? The Impact of Covid Pandemic on the Technology and Innovation Management Research Agenda. *Journal of Management Studies*, *57*(8), 1754–1758. https://doi.org/10.1111/joms.12634
Guan, H., Dong, L., & Zhao, A. (2022). Ethical risk factors and mechanisms in artificial intelligence decision making. *Behavioral Sciences*, *12*(9), 343. https://doi.org/10.3390/bs12090343

Grant, A. M., Gino, F., & Hofmann, D. A. 2011. Reversing the extraverted leadership advantage: The role of employee proactivity. ***Academy of Management Journal***, 54: 528-550

Guntzburger, Y., Pauchant, T. C., & Tanguy, P. A. (2016). Ethical Risk Management Education in Engineering: A Systematic Review. *Science and Engineering Ethics*, *23*(2), 323–350. https://doi.org/10.1007/s11948-016-9777-y

Hackman, J. R., & Oldham, G. R. (1975). Development of the job Diagnostic Survey. Journal of Applied Psychology, 60(2), 159–170. https://doi.org/10.1037/h0076546

Hatcher, T. 2002. Ethics and HRD: a new approach to leading responsible organizations. Perseus: Cambridge.
Helfat, C. E., & Peteraf, M. A. (2015). Managerial cognitive capabilities and the microfoundations of dynamic capabilities. Strategic Management Journal, 36(6), 831–850

Hoffman, B. (2023, October 5). Leaders looking to leverage AI need to think about context. *Forbes*. https://www.forbes.com/sites/brycehoffman/2023/03/31/leaders-looking-to-leverage-ai-need-to-think-about-context/

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-Enabled Recruiting and Selection: A Review and Research agenda. *Journal of Business Ethics*, *178*(4), 977–1007. https://doi.org/10.1007/s10551-022-05049-6

Hunt, S. D., & Vasquez-Parraga, A. (1993). Organizational Consequences, Marketing Ethics, and Salesforce Supervision. JMR, Journal of Marketing Research, 30(February), 78–90. https://doi.org/10.2307/3172515.

Integrating AI into HR recruitment processes. (2024). Development in Learning Organizations an International Journal, 38(4), 45–46. https://doi.org/10.1108/dlo-05-2024-0129

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. Academy of Management Journal, 67(1), 5-32.

Jones, T. M. (1991). Ethical Decision Making by Individuals in Organizations: An Issue Contingent Model. Academy of Management Review 16(February), 366–395. https://doi.org/10.2307/258867.

Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. A. 1964. ***Organizational stress: Studies in role conflict and ambiguity***. New York, NY: Wiley.

Kaplan, A., & Haenlein, M. (2018). 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

Kaplan, Robert & Norton, David. (2001). The Strategy-Focused Organization : How Balanced Scorecard Companies Thrieve in the New Business Environment / R.S. Kaplan, D.P. Norton.

Kaushik, A. (2024, May 30). *How AI helps these 9 companies hire Better*. Retrieved October 14, 2024, from https://www.wecreateproblems.com/blog/companies-using-ai-for-recruitment

Kellogg, K. C., Valentine, M. A., & Christin, A. (2019). Algorithms at Work: the new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kiesler, D. J. 1983. The 1982 interpersonal circle: A taxonomy for complementarity in human transactions. ***Psychological review***, 90: 185-214.

King, M. P. (2023, May 17). Who benefits from diversity and inclusion efforts? *Forbes*. https://www.forbes.com/sites/michelleking/2023/05/16/who-benefits-from-diversity-and-inclusion-efforts/

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2020). Highly Accurate, But Still Discriminatory. Business & Information Systems Engineering, 63(1), 39–54. https://doi.org/10.1007/s12599-020-00673-w

Krakowski, S., Luger, J., & Raisch, S. (2022). Artificial intelligence and the changing sources of competitive advantage. Strategic Management Journal, 44(6), 1425–1452. https://doi.org/10.1002/smj.3387

Liu, Chengwei and Jerker Denrell. 2018. "Performance Persistence Through the Lens of Chance Models: When Strong Effects of Regression to the Mean Lead to Non- Monotonic Performance Associations."

Liu, J. Y., Zou, P. X. W., & Meng, F. Y. (2014). Does expectation match reality? Examination of risk management education in China. *Journal of Professional Issues in Engineering Education and Practice*, *140*(3). https://doi.org/10.1061/(asce)ei.1943-5541.0000200

Marr, B. (2024, November 29). AI won't replace humans – Here's the surprising reason why. *Forbes*. https://www.forbes.com/sites/bernardmarr/2024/11/28/ai-wont-replace-humans--heres-the-surprising-reason-why/

Meyer, David. 2018. *Amazon Reportedly Killed an AI Recruitment System Because It Couldn't Stop the Tool from Discriminating Against Women.* Fortune. October 10th. http://fortune.com/2018/10/10/amazon-ai-recruitment-bias-women-sexist/

Mill, J. S. (2016). Utilitarianism. In *Seven masterpieces of philosophy* (pp. 329-375). Routledge.

Mori, M., Sassetti, S., Cavaliere, V., & Bonti, M. (2024). A systematic literature review on artificial intelligence in recruiting and selection: a matter of ethics. *Personnel Review*. https://doi.org/10.1108/pr-03-2023-0257

Mujtaba, D. F., & Mahapatra, N. R. (2024). Fairness in AI-Driven Recruitment: Challenges, metrics, methods, and future directions. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2405.19699

Mori, M., Sassetti, S., Cavaliere, V., & Bonti, M. (2024). A systematic literature review on artificial intelligence in recruiting and selection: a matter of ethics. Personnel Review

Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and technology: Forms of conjoined agency in organizations. Academy of Management Review, 46(3), 552–571

N, M. S., Sang, A., & Ngure, S. (2018). Ethical issues in recruitment, selection and employee performance in public universities in Nyeri County, Kenya. *International Journal of Academic Research in Business and Social Sciences*, *8*(5). https://doi.org/10.6007/ijarbss/v8-i5/4090

Olcot, L. (2024, September 17). *AI and Social Media: Revolutionising Recruitment for a Digital Era*. https://www.linkedin.com/pulse/ai-social-media-revolutionising-recruitment-digital-era-lewis-olcot-sndgc/

Peteraf, M. A., & Bergen, M. E. (2003). Scanning dynamic competitive landscapes: A market-based and resource-based framework. Strategic Management Journal, 24(10), 1027–1041

Pearl, Judea. 2018. *The Book of Why: The New Science of Cause and Effect*. Basic Books

Pfeffer, J. (1998). The Human Equation: Building Profits by Putting People First. Harvard Business School Press, Boston.

Pfeffer, Jeffrey and Robert I. Sutton. 2006. *Hard Facts, Dangerous Half-Truths and Total Nonsense: Profiting from Evidence-Based Management.* Harvard Business Review Press.

Polli, F. (2019). Using AI to eliminate bias from hiring. *Harvard Busi- ness Review, October 29*.

Psico-smart Editorial Team. (2024, August 28). *The accuracy of AI algorithms in predicting personality traits*. https://psico-smart.com/en/blogs/blog-the-accuracy-of-ai-algorithms-in-predicting-personality-traits-171559

Radford, Alec, Wu Jeffrey, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. "Language Models Are Unsupervised Multitask Learners." OpenAI Working Paper, San Francisco.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. Academy of Management Review, 46(1), 192–210.

Rąb-Kettler, K., & Lehnervp, B. (2019). Recruitment in the times of machine learning. Management Systems in Production Engineering, 27(2), 105–109. https://doi.org/10.1515/mspe-2019-0018

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigat- ing bias in algorithmic hiring: Evaluating claims and practices. In *Conference on fairness, accountability, and transparency, Barcelona, Spain, January 27–30*. New York: Association for Computing Machinery.

Read set, fail? Avoiding setbacks in the intelligent automation race. (2018). In KPMG Research Study. KPMG.

Rest, J. R. (1986). Moral Development: Advances in Research and Theory. Praeger, New York.
Recruitment and selection Process: Best practices & Overview. (2022, December 1). Personio. Retrieved October 6, 2024, from https://www.personio.com/hr-lexicon/recruitment-and-selection/#

Reinecke, J., Arnold, D. G., & Palazzo, G. (2016). Qualitative Methods in Business Ethics, Corporate Responsibility, and Sustainability Research. Business Ethics Quarterly, 26(4), xiii–xxii. doi:10.1017/beq.2016.67

Rivera, L. (2024, April 2). *Firms are wasting millions recruiting on only a few campuses*. Harvard Business Review. https://hbr.org/2015/10/firms-are-wasting-millions-recruiting-on-only-a-few-campuses

Roozenburg, N. (1993). On the pattern of reasoning in innovative design. Design Studies, 14(1), 4e18.

Rousseau, Denise (Editor). 2014. *The Oxford Handbook of Evidence-Based Management.* Oxford University Press.

Sattu, R., Das, S., & Jena, L. K. (2024). Should I adopt AI during talent acquisition? Evidence from HR professionals of Indian IT organisations. Journal of Organizational Effectiveness People and Performance. https://doi.org/10.1108/joepp-05-2023-0186

Sankalana, N. (2023, September 16). The resume mistakes instantly rejected by employers, thanks to AI. Medium. https://medium.com/@nirmalsankalana/the-resume-mistakes-instantly-rejected-by-employers-thanks-to-ai-eb9952cd344e

Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to 'solve' the problem of discrimination in hiring? Social, technical, and legal perspectives from the UK on automated hir- ing systems. In *Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27–30* (pp. 458–468). New York: Association for Computing Machinery. https://doi. org/10.1145/3351095.3372849.

Schad, J., Lewis, M. W., Raisch, S., & Smith, W. K. 2016. Paradox research in management science: Looking back to move forward. ***Academy of Management Annals***, 10: 5–64.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. California Management Review, 61(4), 66–83
Shrestha, P., & Parajuli, D. (2021). Impact of workforce diversity on employee performance. *International Journal of Management*, *12*(2), 86-95.

Smith, W. K., & Lewis, M. W. 2011. Toward a theory of paradox: A dynamic equilibrium model of organizing. ***Academy of Management Review***, 36: 381–403.

Statista. (2024). *Artificial Intelligence - Global | Statista Market Forecast*. https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide

Stuss, M.M., Fularski, A. (2024). Ethical Considerations of Using Artificial Intelligence (AI) in Recruitment Processes. Education of Economists and Managers, 71 (1), p. 53–67.Retrieved from: https://econjournals.sgh.waw.pl/EEiM/article/view/4550

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

Tang, P. M., Koopman, J., McClean, S. T., Zhang, J. H., Li, C. H., De Cremer, D., Lu, Y., & Ng, C. T. S. (2021). When Conscientious employees meet intelligent machines: an integrative approach inspired by complementarity theory and role theory. Academy of Management Journal, 65(3), 1019–1054. https://doi.org/10.5465/amj.2020.1516

Team, Y. (2024, April 1). *Measuring the Effectiveness of Campus Recruiting: Four Strategies for success*. Yello. https://yello.co/blog/measuring-the-effectiveness-of-campus-recruiting/

Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. Strategic Management Journal, 42(9), 1600–1631. https://doi.org/10.1002/smj.3322

The state of AI in 2023: Generative AI's breakout year. (2023, August 1). McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year#/

Van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS quarterly*, *45*(3).

Vickers, M. (2005). *Business ethics and the HR role: Past, present, and future. Human Resource Planning*, Wiley-India.

VIROVERE A., KOOSKARE M. & VALLER M. 2002. Conflict as a tool for measuring ethics at the workplace.
Journal of Business Ethics, 39 (1/2):75-81

What is candidate screening? Process & Benefits | peopleHum. (2023). Peoplehum. Retrieved October 6, 2024, from https://www.peoplehum.com/glossary/candidate-screening#:~:text=Screening%20is%20the%20process%20of,%2C%20experience%2C%20and%20skill%20sets.

Wilson, H. J., & Daugherty, P. R. (2018). Human + Machine: Reimagining Work in the Age of AI. Boston, MA: Harvard Business Review Press.
Williamson, M. K., & Perumal, K. (2018). Perceptions of misfit: A study of South African employees. *Journal of Contemporary Management*, *15*(1), 551-579.

Wonchała, K. (2024, September 26). Exploring the obstacles of AI integration in recruitment processes - simply talented. *Simply Talented*. https://wearesimplytalented.com/exploring-the-obstacles-of-ai-integration-in-recruitment-processes/

Yudkowsky, E. (2008). Artificial Intelligence as a Positive and Negative Factor in Global Risk.

Zahara, L., Musa, P., Wibowo, E. P., Karim, I., & Musa, S. B. (2020). The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi. *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1–9. https://doi.org/10.1109/icic50835.2020.9288560