A Work Project, presented as part of the requirements for the Award of a Master's degree in Management from the Nova School of Business and Economics.

ETHICAL IMPLICATIONS OF AI IN RECRUITMENT: AN EXAMINATION AND RISK MITIGATION FRAMEWORK

ANNIK CHRISTIN MÜLLER

Work project carried out under the supervision of:

Joana Corrêa Monteiro

20/12/2023

Abstract

The thesis examines ethical considerations surrounding the use of Artificial Intelligence (AI) in recruitment. It delves into how AI is integrated into various stages of the recruitment process: scouting, screening, assessment and coordination. While some advocate for AI's potential as an impartial third party, others question its reliability, highlighting concerns such as algorithmic bias and privacy rights. Through a systematic literature review, the paper outlines these ethical considerations, providing a normative guide and proposing a mitigation framework for organizations leveraging AI in recruitment.

Keywords: AI recruiting · Algorithmic hiring · Ethics of AI · Algorithmic bias · Employee selection

List of Abbreviations

AI	Artificial Intelligence
CV	Curriculum Vitae
HR	Human Resources
ML	Machine learning
SLR	Systematic Literature Review

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

In recent years, it has become clear that Artificial Intelligence (AI) has the potential to transform many areas of society and business (Fernández-Martínez & Fernández, 2020). This includes the development of AI software for recruitment to streamline and accelerate the hiring process, which has led to a new term: AI recruitment (Hunkenschroer & Luetge, 2022). AI recruitment is linked to so-called digital recruiting 3.0. In this regard, digital recruiting 1.0 refers to the stage when job advertisements could be found on the internet and digital recruiting 2.0 denotes the possibility of using several job boards and the development of professional as well as social networking platforms (e.g., LinkedIn). The shift to digital recruiting 3.0 is based on the incorporation of AI for selection and recruitment activities, where computers can take over tasks and make decisions at the level of a human beings (Black & van Esch, 2020). Companies in particular can take advantage of this development, as business success is inextricably linked to the skillful implementation of their strategies. Central to this dynamic is the identification of high potentials during the recruitment process who have exactly the right skills to meet the specific requirements of the company in question. This need underscores the compelling reasons for an increased commitment to advanced AI solutions within recruitment processes (Chamorro-Premuzic & Akhtar, 2019; França et al., 2023). Applications are used at various levels of the recruitment process, such as attracting the most promising candidates through job postings, reviewing resumes, or video analyzing interviews (Hunkenschroer & Luetge, 2022). In this emerging paradigm, many steps along the application process are heavily influenced by algorithms, with some human tasks getting completely replaced, justifying the need to consider the ethical impact on people (Polli, 2019). While many authors (e.g., Chamorro-Premuzic & Akhtar, 2019; Fernández-Martínez & Fernández, 2020; Florentine, 2016) see AI recruitment as having the potential to act as an objective third party to remove human bias, avoid “naïve” judgments, and thus create a fairer and more consistent application process, others argue that

AI software are not yet a valid recruitment tool. Such authors claim, from a business and technical perspective, that the issue of data collection needs to be examined more closely, as algorithms rely on data from societal or historical data sets that can lead to algorithmic bias by amplifying existing human bias (Frazzetto, 2023; Pena et al., 2020). Along with the collection of data comes the discussion of compliance with privacy rights, as well as the responsibility of organizations to ensure a transparent application process and to provide information about the use and limitations of AI (Hunkenschroer & Luetge, 2022). In addition, the use of AI solutions in the field of recruitment harbors the great risk of human autonomy being lost, because decisions and tasks are taken over automatically (Lee, 2018; Yarger et al., 2020). To this end, there are a few suggestions in the literature for a more ethical approach by testing and incorporating bias reduction techniques from the outset (e.g., Lin et al., 2021; Yarger et al., 2020).

In this context, the paper identifies three shortcomings: First and foremost, there is hardly any theoretical groundwork in the literature on the ethics of AI recruitment, so there is a lack of unified guidance. Second, ethical risks such as algorithmic bias are mainly addressed, while human autonomy and privacy rights take second place. Third, the mitigation techniques developed in this context are limited to a few. For these reasons, this paper attempts to present the various ethical perspectives identified in the literature by summarizing and ethically discussing the opposing viewpoints to provide a comprehensive overview of the state of research. Businesses need to develop a sophisticated understanding of the benefits and possible drawbacks of AI recruiting technologies. It is imperative that they understand how algorithmic results could deviate from their intended goals. For this reason, the overview primarily aims to serve as a normative guide for organizations and researchers, as well as to provide a more informed and enlightened approach to the topic of AI recruiting. Finally, a mitigation framework is developed to propose more precise regulatory possibilities for organizations using AI in recruitment. To this end, the following research question – *What are the key ethical considerations that arise* *when using Artificial Intelligence in recruitment and selection processes, and how can ethical risks be minimized?* – is answered by applying systematic literature review. Therefore, the examination begins with a description and explanation of the chosen research method including its selection criteria. Afterwards, the literature is systematically reviewed to identify application areas of AI recruiting, to compare the associated ethical opportunities and risks, and to outline possible solutions at technical, organizational, and regulatory levels. Finally, the results are discussed on a theoretical and empirical level including a definition of the limitations. The paper concludes with key conclusions and outlooks on further developments.

2. Methodology

2.1 Research method and design

A systematic literature review (SLR) is defined by Rojon et al (2021, 191) as a method-guided reproducible “manuscript for collecting, analyzing and synthesizing literature with clear audit trails about what is and what is not known regarding a research question or set of questions”. In contrast to a literature review as part of an empirical paper with the purpose to justify the objectives of the paper, the SLR is a stand-alone analysis that applies “a replicable, scientific and transparent process, intended to minimize bias through extensive searches and by providing an audit trail of the reviewers steps, strategies, procedures and decisions” (Andreini & Bettinelli, 2017, 2). The scientific gain of knowledge consists in illustrating the current state of art on a defined topic and/ or extending the theory-guided object of research through systematic collection, processing and analysis (Döring, 2023; Webster & Watson, 2002). By examining the research subject from multiple perspectives, Snyder (2019) argues that an SLR can answer research questions more precisely than single studies and can help clarifying an inconsistent or interdisciplinary research subject. Furthermore, a systematic literature review can be a useful tool for summarizing research findings, illuminating empirical evidence from a metatheoretical angle, and highlighting gaps in scholarly work – all of which are essential for developing both theoretical and conceptual frameworks (Snyder, 2019; Webster & Watson, 2002). The method can be subdivided into meta-analysis and narrative analysis (Frank & Hatak, 2014), whereas this thesis uses the latter to evaluate the theory of ethical AI recruiting embedded in the research question – *What are the key ethical considerations that arise when using Artificial Intelligence in recruitment and selection processes, and how can ethical risks be minimized?*. The method aims to identify relevant literature, to synthesize it, and to provide a structured review of key findings to create a comprehensive knowledge base for researchers and managers (Seuring & Gold, 2012; Tranfield et al., 2003). Synthesis can be useful for managers and organizations, as a full review of primary research is often too time-consuming due to resource constraints. Hence, the results provided through SLR can contribute to researcher-practitioner exchange and evidence-based practice (Petticrew & Roberts, 2006; Rojon et al., 2021). As the topic of this thesis deals with a relatively new branch of AI, namely the extension of the technology to the field of recruitment, and is thus characterized by a lack of explanatory power and inconsistency within the literature, the research subject is ideally suited for SLR. Moreover, as recruitment is a central part of any business, this thesis can also contribute to the basic understanding and ethical approach to AI recruitment by functioning as inspiration for organizations aiming to produce guidelines. Therefore, the literature is systematically categorized across three stages: the application of AI in recruitment, the resulting ethical considerations in terms of ethical opportunities and ethical risks, and the mitigation measures, divided into technical, organizational, and regulatory implications (see Appendix A). For this research design, an interpretive approach is adopted as this strategy is applicable to a wide range of literature and is often used for management topics (Kraus et al., 2020; Tranfield et al., 2003).

2.2 Quality criteria

To ensure that this systematic literature review can fulfill its scientific objectives, the method must follow specific quality standards. The SLR itself is characterized as replicable, transparent, and unbiased, which can only be achieved by applying the principles of trustworthiness, which are interpreted differently (Döring, 2023). Snyder (2019) argues that the quality of a systematic literature review depends on the *depth* and *rigor* of the search strategy for selecting and collecting data. The method should be *repeatable* as well as *transparent* so that a third party could perform the same analysis and find similar results, which Kraus et al. (2020) also underline. A more detailed list of principles that should be followed in addition to the criteria already mentioned is provided by Thorpe et al. (2005): 1. *Transparency*: establish inclusion and exclusion criteria and explain the rationale behind the selection; 2. *Clarity*: clearly outline the process and create an audit trail; 3. *Focus*: link the research question in a consistent way throughout the analysis; 4. *Accessibility*: ensure the availability of the databases; 5. *Synthesis*: incorporate diverse articles and research methods. To demonstrate the adherence to these principles and its rigorous critical examination of the research question with the greatest possible impartiality, a comprehensive disclosure of the data collection process is provided below.

2.3 Data collection process

The data collection process follows the guidelines of Frank and Hatak (2014) and is supplemented by Vom Brocke et al (2009). First, a research question was selected through an initial unsystematic literature review, which helped to identify initial gaps in the literature and formulate the overarching research question for the thesis. The central challenges and the theoretical background from the preliminary research are briefly presented in the introduction. After defining the research question, a systematic, keyword-based literature search was carried out in the databases: ScienceDirect, Scopus and Web of Science. These three databases were chosen as they are among the largest peer-reviewed repositories offering a comprehensive collection of diverse scholarly works, including ScienceDirect's wide range of books across various disciplines such as social sciences, technology studies and more. Emphasis was put on including a diverse range of articles with a variety of methods and perspectives to increase the validity and reliability of the literature sample (as recommended by Kraus et al., 2020; Seuring & Gold, 2012). Therefore, the above-mentioned reputable online databases made use of, as well as a hand search. As the area of AI recruitment remains recent and multidisciplinary, a comprehensive sampling approach was chosen without strict inclusion criteria, such as exclusive publication in reputable journals. Rather, this study embraced a holistic selection strategy, encompassing diverse sources that addressed the ethical implications of AI-facilitated recruitment within corporate environments, including scholarly peer-reviewed journals, conference reports, practical industry contributions (e.g., magazines), and individual chapters from books. In order to identify these references, exclusion criteria were also explicitly defined and a number of keyword combinations in the selected databases were used. In addition, all articles should have been published in English and in the period from 2010 to 2023 to narrow down the scope of the search. A complete list of criteria for the search strategy can be found in Appendix B.

The literature search for this paper commenced with an initial search pool of 271 sources. This pool was subjected to the exclusion criteria, dismissing certain articles that did contain the keywords and met the exclusion criteria but were from other scholarly such as medicine and engineering (e.g., Aquino et al., 2023; Tada et al., 2021). After this criteria-based curation and the removal of duplicates, the selection was reduced to 69 sources by screening the titles and abstracts. The refinement process was extensively expanded through the review of full texts and an additional investigation using backward searching techniques. This approach resulted in a limited number of 41 sources that met the criteria for study inclusion. In order to broaden the search horizon and include practical perspectives, the investigation went beyond traditional academic sources. A hand search was conducted in renowned business magazines such as Harvard Business Review, Forbes and CIO. In this way, a total of 52 sources were identified, each of which provided valuable insights into the ethical aspects of AI recruitment. Through this multistage literature search, a variety of sources were systematically evaluated to create a comprehensive and well-rounded foundation for the thesis, considering academic as well as business perspectives (see Appendix C). To this end, the synthesis consisted of both the summary and the analysis of the identified articles, where connections between them were identified, interpreted and used to create the following categorization system: AI applications; ethical considerations (divided into risks and opportunities); approaches to risk mitigation (divided into technical, organizational and legal aspects). As suggested by several authors (e.g., Frank & Hatak, 2014; Rojon et al., 2021; Tranfield et al., 2003; Vom Brocke et al., 2009), a data extraction form was developed to visualize the categorization as well as the different features of the articles and use it as a comprehensive repository (see Appendix D). Following the description of the procedures for data collection and extraction, which serve as an audit trail for quality assurance and reproducibility, the results obtained are presented in the next section.

3. Findings

3.1 Classification of the literature

The systematic analysis of 52 selected articles has led to the identification of different thematic areas within AI research. Through the categorization system applied to these articles specific areas of inquiry have been delineated that include AI applications, ethical opportunities, ethical risks, and risk mitigation strategies. A comprehensive breakdown of the individual authors, their respective topics, the place of publication and the methods used can be found in tabular form in Appendix D.

Seven articles addressed the applications and benefits of AI and identified four distinct phases: scouting, screening, evaluation, and coordination. These discussions used primarily descriptive methods to explain the role of AI in these areas and to highlight the different impacts of the technology in each phase. The ethical opportunities of AI were highlighted in 17 articles, with a focus on its ability to mitigate or eliminate human bias. The prevailing view in these articles emphasized the impartial and neutral nature of AI and pointed out its potential to reduce inherent bias in decision-making processes through programming. Authors used a variety of methods such as surveys, interviews, systematic literature reviews, and frameworks to shed light on the ethical possibilities of AI. The most widely discussed topic in the reviewed literature, with 34 articles, is the emerging ethical risks of AI, with a particular focus on algorithmic bias (see Appendix E). The discourse encompassed a variety of interpretations and implications of these risks and drew on a range of methodologies, including SLRs, descriptive analyses, interviews, case studies, frameworks, and experimental approaches. Furthermore, mitigation measures were discussed at three main levels: technical, organizational, and regulatory. Technical measures, discussed in 17 articles, focused primarily on removing bias in the training data to prevent bias at the root. At the organizational level, 19 articles proposed guidelines for companies, emphasizing the urgent need for data literacy and the establishment of audit teams to address the ethical challenges associated with AI. In contrast, nine articles stressed the lack of universally applicable legal regulations and emphasized the dependence on country-specific laws. The consensus within the discourses emphasized the need for further research to address the complexity of the regulatory aspects of AI.

It is important to acknowledge that most of the articles identified were published in the period from 2019 to 2023, with a notable concentration of nine articles published specifically in 2023 (see Appendix F). This observation suggests that the ethical considerations of AI recruitment are still at an early stage, indicating an emerging discourse that will lead to significant debate in the forthcoming period. Moreover, the geographical distribution is difficult to determine, as specific information on the location of publication is not available for each individual article. However, a prevailing trend indicates that most of the discussion takes place in the United States, as shown in Appendix G. After outlining the various classifications of the literature perspectives, a more detailed exposition of the findings is given below.

3.2 AI in recruitment

According to Hunkenschroer and Luetge (2022), *AI recruiting* can be defined as a new term that describes any process that uses AI to help assisting companies recruit and select candidates. Therefore, AI can be used in four main recruiting activities: scouting, screening, assessment, and coordination (Black & van Esch, 2020; Pena et al., 2020). In this context, AI is defined as the ability of a system “to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” (Kaplan & Haenlein, 2019, 15). The thesis makes reference to a wide definition of AI which encompasses advanced machine learning techniques such as deep neural networks and simple regression-based algorithms, as well as other categories including speech recognition or natural language processing (Hunkenschroer & Kriebitz, 2023).

AI-powered recruitment tools can be used to identify, attract or screen potential employees in a more efficient way and to guide them more intelligently through the interview process such as the overall coordination of the application (Fernández-Martínez & Fernández, 2020; Köchling et al., 2021). More specifically, AI tools have the ability to reconfigure the landscape of recruiting in a number of ways:

Scouting: Several articles highlight the great potential of identifying and predicting talent in the scouting phase (Black & van Esch, 2020; Bogen, 2019). By using predictive analytics, an applicant's suitability for the job can be predicted based on their CV, social media presence and online behavior (Kelly, 2023). In addition, potential candidates for a specific job can be identified who would not have applied themselves, but whose skills and qualifications are recognized by an algorithm, making it possible to attract active and passive candidates alike (Black & van Esch, 2020). Some articles also mention the challenges of finding the right candidates but highlight the ability of AI software to create tailored and engaging job descriptions (Köchling et al., 2021; Yarger et al., 2020).

Screening: This phase is the most discussed in literature. Here, algorithms are used to filter applications according to certain criteria and rank the most promising candidates (Fernández-Martínez & Fernández, 2020; Vasconcelos et al., 2018). Quickly recognizing whether an applicant's skills match the advertised requirements and the ability to screen many applications surpasses human performance (Kim-Schmid & Raveendhran, 2022). Various specialized tools exist for identifying semantic matches and terms that assess an applicant's qualifications. Some of these tools extend their functionality beyond mere qualification assessment and attempt to predict future job performance by analyzing productivity-related signals (Bogen, 2019).

Assessment: Few researchers are investigating job interviews conducted through AI-powered video interview tools. The primary aim of these tools is to reduce the time and costs associated with organizing and conducting in-person interviews (Kelly, 2023). In these AI-assisted interviews, candidates respond to standardized questions that are scored by audio and facial recognition software, eliminating the need for human intervention (Chamorro-Premuzic & Akhtar, 2019; Fernández-Martínez & Fernández, 2020; Polli, 2019). In addition to the interpretation of job interviews, AI-supported ability tests, simulations or video games are also used to record certain character traits such as willingness to take risks, ability to work in a team and resilience based on the variables required by a company (Polli, 2019; Raghavan et al., 2020).

Coordination: Furthermore, various articles mention that AI tools can speed up recruiters' routine tasks while increasing candidate engagement (Dattner, 2019; Hancock et al., 2023). In addition, they are able to guide candidates through the application process, automatically completing a company's application form based on the information contained in the uploaded CV and providing tailored responses to inquiries specifically related to the application procedure (Black & van Esch, 2020; Sánchez-Monedero et al., 2020).

In summary, the authors agree that the main potential of AI tools for talent management is to help companies find suitable candidates faster, develop and retain employees more effectively and streamline the recruitment process more than HR experts could (Kelly, 2023; Kim-Schmid & Raveendhran, 2022; Köchling et al., 2021; Polli, 2019). However, the question arises as to whether AI recruitment is or can be ethical, as the lack of human intervention threatens to reduce human autonomy.

3.3 Ethical implications of AI recruiting

3.3.1 Ethical opportunities

In order to evaluate the question of ethical AI recruitment ethical possibilities are first examined. Many authors point out that there are numerous ethical opportunities where the use of AI can prevent implicit bias (also known as unconscious bias) in particular (Rodgers et al., 2023; Tambe et al., 2019). In this context, it is relevant to mention that explicit bias describes the phenomenon of being aware of one's thoughts, feelings, and actions toward a particular group (e.g., hate speech, active discrimination, etc.) while “implicit bias is an automatic form of stereotype or prejudice [that] people unconsciously act upon." (Lin et al., 2021, 67). More specifically, AI providers claim that AI software can make recruiting decisions more efficient and precise while being fairer and reducing implicit bias because it does not contain human cues (Köchling et al., 2021). These include, for example, anchoring bias, where an interviewer is influenced by information they see and hear first; confirmatory bias, where an interviewer only focuses on information that confirms their intuition; and similarity bias, where an interviewer gives preference to applicants that resemble them (Black & van Esch, 2020). In addition, there are personality traits of recruiters based on explicit biases such as sexism, racism, lookism and ageism, to name a few, that can also negatively impact the hiring process (Chamorro-Premuzic, 2019a). Each of these different forms of bias occur constantly in every step of the hiring process, especially in face-to-face job interviews, which is why they might not only be inefficient, but might also lead to biased decision making, e.g., selecting the most charismatic candidate (Chamorro-Premuzic & Akhtar, 2019; Köchling et al., 2021; Polli, 2019). Collectively, this means that traditional hiring processes are already prone to having bias (Polli, 2019). According to Mulcahy (2016), the biggest challenge in this context is that implicit bias occurs unconsciously, which is why the involvement of a third party, namely AI, may seem promising (Florentine, 2016). Similarly, Chamorro-Premuzic (2019) claims that AI can have the potential to be programmed in a way to avoid selected information that is irrelevant and focus only on decision-relevant criteria in order to make an unbiased decision. For example, when screening applications, the algorithm could be set to disregard gender, age, or nationality as irrelevant information, which could result in more diversity in the company (Fernández-Martínez & Fernández, 2020b; Florentine, 2016; Frazzetto, 2023; Yam & Skorburg, 2021). Additionally, AI-powered recruiting tools can screen hundreds of applications against these criteria, saving a company's HR department time and costs (Polli, 2019). Therefore, another benefit is overcoming cognitive biases, as AI technologies can make some human decisions at a scale and speed far beyond human capabilities (Polli, 2019). By some it is even hypothesized, that if all interviews were conducted through AI-powered video tools, bias and nepotism would be reduced while predictive accuracy and performance would increase (Chamorro-Premuzic, 2019b; Köchling et al., 2021). This is supported by several companies (e.g., Deloitte, 2018) which argue that they can use these algorithmic decision-making programs to save costs, avoid bias against different factions and thus guarantee consistency in the decision process (Köchling et al., 2021). Moreover, not only do organizations believe that “the consistency of administration in AI systems might serve as a heuristic indicating fairness, given their perceived lack of subjectivity”(Acikgoz et al., 2020, 401), but the fairness heuristics of applicants could also be affected. For example, a standard set of interview questions could be a signal to candidates of the company's commitment to equality (Acikgoz et al., 2020; da Motta Veiga et al., 2023).

In summary, the key ethical opportunities appear to be in reducing human bias, ensuring process consistency, and in efficiency gains by avoiding cognitive bias for organizations.

3.3.2 Ethical risks

Despite the potential of using AI to reduce bias in hiring practices, there is currently ample evidence that the AI recruitment process is not consistently efficacious and unbiased (Chamorro-Premuzic & Akhtar, 2019; Persson, 2016). Many studies highlight the existence of various types of algorithmic bias (Hunkenschroer & Luetge, 2022; Lin et al., 2021). Yarger et al. (2020) identify three potential dangers of bias that can occur in algorithm programming, namely the possibility of bias in model design, feature or label selection, or training data. As explained previously, AI recruiting can help to attract and proactively target potential applicants via algorithmic ad platforms. The issue in this case might be that the software can make superficial predictions if the model design of the algorithm focuses on who is most likely to view the job ad rather than who will succeed in the role, which "can lead to a reinforcement of gender and racial stereotypes" (Bogen, 2019, 3). Moreover, any machine learning algorithm for predicting data (test data) is backward-looking and uses data sets from the past (training data), which poses some vital ethical risks (Köchling et al., 2021; Persson, 2016; Tambe et al., 2019; Yarger et al., 2020). Using AI hiring tools that have been trained with real data from the past (e.g. a dataset of predominantly white applicants) would result in the algorithms reflecting human bias, making objectivity obsolete (Bogen, 2019; Frazzetto, 2023; Vasconcelos et al., 2018). A similar case occurred at Amazon in 2018, with the use of a hiring algorithm that was based on training data of current top performers, all of whom were men at the time, and as a result, the resumes of female applicants were ranked lower in the candidate rankings, thus unknowingly disadvantaging woman candidates (Mujtaba & Mahapatra, 2019; Pena et al., 2020; Ross, 2023). Moreover, companies that are only allowed to use customer data to train AI are bound by privacy, data consent, and compliance mandates. Here, collecting and leveraging this type of data can be cumbersome and conflict with the human right to privacy (Hunkenschroer & Kriebitz, 2023; Ross, 2023).

Further ethical risks include emotion recognition software unable to distinguish between different intonations or emotions that are expressed uniquely in various languages and therefore could disadvantage certain groups (of race, ethnicity, etc.), resulting in a loss of diversity in the workforce (Fernández-Martínez & Fernández, 2020; Williams et al., 2018). In addition, there is a trend toward AI video interview analysis, which uses speech and facial recognition software to evaluate applicants' word choice, tone of voice, body language, and facial movements, comparing them to the criteria companies have chosen for top performers (Fernández-Martínez & Fernández, 2020; Köchling et al., 2021; Langer et al., 2018). As a result, information about gender and sexual orientation might be collected without the consent of the applicant, which leads to discrimination and deprivation of privacy. (Langer et al., 2018). AI software also has difficulty recognizing the faces of non-white individuals, resulting in dark-skinned applicants not being properly recognized and therefore not fairly evaluated (Fernández-Martínez & Fernández, 2020; Köchling et al., 2021; Williams et al., 2018). In this context, Dattner (2019) points out that this can also apply to the external appearance of applicants in general. For example, thicker lips used by AI-assisted video software as an indication of homosexuality in men may lead to ethically questionable conclusions about the appropriateness of singling out people based on physical and immutable characteristics (Dattner et al., 2019). To summarize, “algorithms still lead to biases outcomes concerning gender, ethnicity, and personality traits if they build upon inaccurate, biased, unrepresentative, or unbalanced training data” (Köchling et al., 2021, 49), which later tend to be exacerbated by the prediction models (test data) (Chamorro-Premuzic et al., 2017; Köchling et al., 2021). This creates an additional ethical risk as companies begin to develop their own algorithms using their historical data from past employees and limited public databases, which can especially lead to a lack of diversity and autonomy (Fernández-Martínez & Fernández, 2020).

Besides algorithmic bias, AI-powered recruitment tools are unable to incorporate empathy or other human skills such as human intuition and cannot detect emotional intelligence to validate the recruitment process, yet (Hunkenschroer & Kriebitz, 2023; Lee, 2018). Acikgoz, et al (2020) also acknowledge that AI-assisted interviews may not be perceived by applicants as a fair chance of success in terms of procedural justice because of a lack of human intuition (supported by Lee, 2018; Newman et al., 2020). In addition, the absence of transparency and explicability of the various algorithms for applicants, as well as the fact that the use of AI is not always communicated clearly, also support the fear of dehumanization (Sánchez-Monedero et al., 2020; Schumann et al., 2020). In this regard, Raghavan points out that the large number of data points used makes it difficult even for programmers to reasonably explain the underlying properties of the algorithms' prediction and decision-making processes (Raghavan et al., 2020). To this end, Fritts and Cabrera (2021) draw attention to an ethical concern about the potential of dehumanization in the hiring process. This apprehension goes beyond debates about the bias of hiring algorithms and demonstrates that while issues of bias and discrimination may be addressed by ethical and technological advancements in the future, concerns about dehumanization will likely continue to exist separately from these different worries.

In summary, ethical risks most frequently cited in the literature are the occurrence of algorithmic bias, loss of privacy, and lack of transparency and explainability to all stakeholders, which justifies the importance of regulations and approaches to address these issues.

3.4 Approaches to attenuate bias

3.4.1 Technical considerations to exercise diligence

As outlined in the previous sections, the main source of algorithmic bias comes from datasets that are inherently characterized by partiality, which underlines the compelling need to develop preventive methods (Chen, 2023). Most researchers in data science argue that the most crucial requirement for the implementation of fair AI solutions is data literacy, which is described as the necessary capacity to check the accuracy and validity of algorithms created by programmers

(Fernández-Martínez & Fernández, 2020; Simbeck, 2019). In cases where companies purchase algorithms from external vendors rather than developing them themselves, it is strongly recommended to apply professional "testing standards and to obtain critical information about the tools: for example, evidence that informs psychometric reliability, criterion-related validity and bias implications." (Oswald et al., 2020, 523). Furthermore, Lin et al. (2021) and Yarger et al. (2020) make a case for the need for an ethical design of AI tools, whereby AI software should incorporate bias reduction techniques from the outset to de-bias the training data. As an illustrative example, the authors point to an AI software, which blends applicant profiles by cleaning up names and pictures and could thus foster equal opportunities by supporting underrepresented candidates. Exemplary endeavors by tech companies include Microsoft eliminating gender bias measures, Google using the What-if tool to detect bias, and Facebook correcting algorithmic bias by developing the Fairness Flow software (see Mishra, 2022; Xie, 2018).

Furthermore, Williams et al. (2018) propose another strategy: the intentional collection and use of social category data to uncover and address discriminatory behavior. The researchers contended that data scientists cannot identify, comprehend, or address discriminating trends unless the data are labeled with social categories such as gender, age or education. In addition, open-source tools and frameworks should be considered for data programmers to be inspected for systematic bias and help AI software vendors build more fairness and transparency into their algorithms (Chen, 2023; Fernández-Martínez & Fernández, 2020 Mujtaba & Mahapatra, 2019; Polli, 2019). In this context, three methods have emerged in the literature as the most widely used to mitigate biases in algorithms: pre-processing; in-processing; post-processing. Pre-processing is used to edit features in the data prior to classification according to fairness criteria, to delete obviously guarded variables that are not related to the position, or to change labels that would cause bias (Balayn et al., 2021; Bellamy et al., 2018; Mujtaba & Mahapatra, 2019; Zhang et al., 2017). In in-processing or optimization approaches, fairness is achieved after the model has made predictions by subsequently adjusting the classification results to fit the fairness definitions, which can correct for potential bias but may not ensure optimal classification accuracy (Balayn et al., 2021; Kamishima et al., 2018; Mujtaba & Mahapatra, 2019). Post-processing aims to increase the fairness of model predictions by applying thresholds to convert continuous results into binary labels, with methods customized to the specific fairness metrics and, in some cases, the potential to involve a human operator in the decision-making process (Balayn et al., 2021; Canetti et al., 2019; Mujtaba & Mahapatra, 2019; Raghavan et al., 2020). Overall, efforts are needed to enhance the reliability of AI tools on an ongoing basis (Dattner et al., 2019). External specialists, for instance, can be contracted to create an in-house inspection to run samples and analyze algorithmic judgments to ensure that organizations can provide transparency and causality to applicants during the hiring process (Chamorro-Premuzic et al., 2017; Chen, 2023; Köchling et al., 2021; Mann & O'neil, 2016).

3.4.2 Ethical considerations and responsibilities for organizations

For the ethical use of AI in recruitment processes and departments, the identified literature mainly refers to several organizational principles emphasizing validity, autonomy, non-discrimination, transparency and privacy.

First, it is essential to acknowledge the findings of the study by Al-Alawi et al. (2021) in which interviews were conducted with HR managers, consultants and academics. They claim significant differences in the transparent disclosure of the use of AI applications in current company reports. In particular, the study revealed a lack of clarity in describing the scope in which AI technologies are used. Furthermore, the research highlighted that these reports are predominantly published by larger high-tech companies as opposed to smaller firms.

In response to these concerns and with the aim of providing insight into improving the management of the applicant-employee relationship, numerous authors have described basic principles and made suggestions for improvement. Foremost, every candidate has the right to truthfulness and to be treated with dignity (Bird et al., 2020; Yam & Skorburg, 2021). Therefore, organizations should ensure that applicants give their consent to data use and diligently collect and secure only sensitive information required for job evaluation (Chamorro-Premuzic & Akhtar, 2019; Kriebitz & Lütge, 2020; Simbeck, 2019). Furthermore, since companies set the selection and success criteria for an algorithm during the recruitment process, it is their responsibility to offer a level of transparency and accountability in e.g., reports to inform about the data sets and algorithmic techniques used (Köchling et al., 2021; Sánchez-Monedero et al., 2020). Simbeck (2019) also stresses the need to inform candidates "whether they are communicating with another human or with AI" (Simbeck, 2019, 10). To ensure validity and autonomy, so that social intelligence and explainability are not compromised, many authors suggest that an experienced, algorithmically and statistically informed recruiter should supervise the AI decisions made during selection and assessment. This way, AI is used as a complementary recruiting tool (e.g., Köchling et al., 2021; Reicin, 2023). According to Simbeck (2019) and Tambe et al. (2019b), the above implications, as well as non-discrimination procedures (e.g., to prevent algorithmic bias), should even be included as standards in compliance, and an AI ethics committee should be established to discuss not only the testing but also the limitations of AI technology for the company. For instance, Google has implemented a model map function that clarifies the algorithm used, describes its strengths and limitations, and disseminates the operational results derived from several data sets (Mitchell et al., 2019 in Chen, 2023). In conjunction with the implications outlined previously, a closer examination of the programmers who code these algorithms is also imperative given the centrality of diversity in tech industries. Algorithms often reflect the views of their creators, so the underrepresentation of minority groups in the IT industry can also lead to distortions in the algorithmic results (Chen, 2023). In general, there seems to be a broad consensus in the literature that companies should be obliged to use audit tools and techniques for quality control and that human involvement in the hiring process is necessary to combat dehumanization and to ensure that AI is used ethically (Hunkenschroer & Kriebitz, 2023; Tambe et al., 2019; Yarger et al., 2020).

3.4.3 Legal and regulatory considerations

The use of algorithms in the recruitment sector may negatively impact a person's right to “work, equality and nondiscrimination, privacy, freedom of expression, and association” (Yam & Skorburg, 2021, 621) due to problems including prejudice and a lack of transparency (Williams et al., 2018). Therefore, it is essential for organizations to be keeping up to date with relevant hiring laws and regulations as well as to be declaring adherence to legal requirements (Tilmes, 2022). As each country or territory sets its own legal framework for Artificial Intelligence in recruitment, this paper specifically examines European legislation.

The European Commission created guidelines addressing the effects of algorithmic systems on human rights, while the Organization for Economic Cooperation and Development (OECD) released recommendations on Artificial Intelligence (Bird et al., 2020; Kriebitz & Lütge, 2020). An expert consortium on AI was formed by the European Commission in 2019, and it proposed ethical standards and self-regulatory actions related to AI and ethics. Legal protections such as nondiscrimination statutes and data privacy are essential in stopping biased methods used in algorithmic decision-making (Chen, 2023). Furthermore, laws that prohibit discrimination, especially those that deal with indirect discrimination, act as a barrier against several types of algorithmic bias. For instance, the EU's General Data Protection Regulation (GDPR), which came into force in May 2018, includes a *right to explanation* (articles 13-15) and covers the implications of using machine learning algorithms. This right enables individuals to demand actions to reducing discriminatory effects when working with sensitive data as well as clarifications for algorithmic decisions made by organizations (Pena et al., 2020). Organizations must perform a Data Protection Impact Assessment (DPIA) in accordance with GDPR, and every EU member state must have a separate data protection body armed with the competence to undertake investigations (Bird et al., 2020; Chen, 2023). Moreover, França et al (2023) emphasize articles 9 and 22 (GDPR), “which target the processing of specific categories of personal data, forbidding the processing of genetic data, biometric data used to uniquely identify a natural person, health data, or data about a natural person’s sex life or sexual orientation, as well as the processing of personal data which might reveal political opinions, religious or philosophical beliefs or racial or ethnic origin” (França, 2023, 21). In addition to regulations on the protection of privacy, transparency and declaration rights, the general need for more state regulation is also called for in literature. Fernández-Martínez & Fernández (2020) stress the need for government monitoring of recruitment processes to identify violations of basic employment laws or human rights. Government restrictions still leave room for unethical activities in the development and use of predictive recruitment technologies, which leads Bogen (2019) to declare that the responsibility of companies goes beyond the applicable laws. Raghavan et al (2020) discovered in a recent analysis that the diversity of bias reduction measures used by providers varies greatly. This implies that although prejudice is recognized in various industries, there is a lack of explicit instruction on how to respond correctly.

AI application	Ethical Opportunities	Ethical risks
1. Scouting	• Prevent human bias • Efficiency gains	• Algorithmic bias • Lack of transparency • Lack of explainability • Privacy concerns • Dehumanization
2. Screening	• Prevent human bias • Process consistency • Efficiency gains	• Algorithmic bias • Lack of transparency • Lack of explainability • Privacy concerns • Dehumanization
3. Assessment	• Prevent human bias • Process consistency • Efficiency gains	• Algorithmic bias • Lack of transparency • Lack of explainability • Dehumanization
4. Coordination	• Process consistency • Efficiency gains	• Lack of explainability • Dehumanization

Ethical risks	Mitigation approaches: Technical	Mitigation approaches: Organizational
Algorithmic bias	**Technical:** – Guarantee data literacy – Introduce test standards – De-bias designs of AI tools – Pre-processing; in-processing; post-processing	**Organizational:** – Ensuring diversity in IT teams – Anti-discrimination procedures as standards for compliance – Establish AI ethics committee
Lack of transparency	**Technical:** – Inhouse-inspections by external specialists	**Organizational:** – Provide reports about the data sets & algorithms used – Informing applicants about communicating with AI – Establish audit tools/teams
Lack of explainability	**Technical:** – Guarantee data literacy	**Organizational:** – Everyone has the right to truthfulness – Supervision of the process by AI-informed HR managers – Establish audit tools/teams
Privacy concerns	**Technical:** – No information	**Organizational:** – Obtaining consent for the collection of personal data – Supervision of the process by AI-informed HR managers – Establish audit tools/teams
Dehumanization	**Technical:** – No information	**Organizational:** – Supervision of the process by AI-informed HR managers – Establish audit tools/teams

Legal:
- Prohibition of any kind of discrimination
- Articles 13-15: Right to explanation (clarification for algorithmic decisions made by organizations)
- Articles 9 + 22: Prohibition: Processing of genetic and biometric data, health data, sexual orientation, data concerning political opinions, religious or philosophical beliefs, racial or ethnic origin

Table 1: Ethical analysis and risk mitigation framework

A complete list of technical, organizational and regulatory risk minimization approaches is presented along with a concise description of the ethical analysis summarized in Table 1 as a structured framework for risk mitigation.

4. Discussion

4.1 Critical evaluation of the ethical implications

Following the systematic organization and synthesis of the relevant findings on the ethical dimensions of AI recruitment, the subsequent discussion will focus on the classification and critical examination of these findings.

AI recruitment has a significant impact on various aspects of the application process, whereby authors agree that it offers tremendous opportunities to increase efficiency and reduce human bias in decision-making. Remarkably, in the literature reviewed, every academic article that recognizes the benefits of AI application also highlights the potential ethical risks associated with AI-enabled recruiting. A particular focus is on identifying and explaining algorithmic biases, with biased training data proving to be the most important causal factor. This view is supported by Polli (2019) and Florentine (2016), wo also point out that algorithmic biases are comparatively easier to detect and correct than human biases, hence it is possible to allocate validity to AI software. Although some articles have also addressed the privacy violation and transparency deficit, these concerns have largely been treated as side issues rather than serving as primary research focus areas. Therefore, there is a conspicuous gap in the literature regarding the potential dangers of dehumanization, validity testing of algorithms, and the establishment of accountability mechanisms. Quite the contrary, these underrepresented concerns are of great importance in ethical discourse, as principles such as autonomy and nonmaleficence are among the cornerstones of ethical considerations (Varkey, 2020). Beyond the lack of discourse on ethical reflections, some of the identified literature also addresses risk mitigation measures at a technical, organizational and legal level. Several of these measures are treated only cursorily and specific regulations and guidelines tailored to the recruitment sector are lacking. While regulations and laws to mitigate ethical risks are recognized, the prevailing discourse focuses on the ethical responsibility of companies when using AI in recruitment and suggests that monitoring should continue to be carried out by humans. Furthermore, although ethical considerations are examined at various stages of the recruitment process, the systematic analysis shows a conspicuous lack of theoretical underpinning, with only three articles in total adopting a theoretical perspective (Fernández-Martínez & Fernández, 2020; Simbeck, 2019; Yarger et al., 2020). This shortage makes it difficult to establish a basic theoretical framework that would provide a solid foundation for ethical considerations in AI recruitment. In contrast, a significant portion of the literature consists of complementary articles dealing with business considerations. While these perspectives offer practical insights into AI implementation, they widen the theoretical gap and hinder the development of a comprehensive ethical framework. Furthermore, the systematic literature review shows a preponderance of non-empirical papers – only 42% of the *corpus* account as empirical (see Appendix H). In these articles, studies are mainly created around the discussion of ethical risks associated with AI in recruitment and perceived fairness in decision-making (e.g., Acikgoz et al., 2020). In addition, it must be acknowledged that most of the articles in the business magazines also reported more on biased algorithms and how companies in general can take advantage of AI. This limited research leaves a critical gap, for instance, in understanding how AI decisions impact fairness, which deters the development of a specific implementation guide for attracting the most suitable talent. Moreover, empirical research could be very helpful in addressing issues such as the usage of personal data, the effect of AI on workforce diversity, and the reliability of assessments (Hunkenschroer & Luetge, 2022). Risk mitigation approaches are also underrepresented in the current state of science. As the guidelines presented by the authors are rather phrased as recommendations, the emphasis on responsibility lacks a theoretical basis, making it susceptible to subjective interpretation to provide clear guidelines for the ethical use of AI.

4.2 Limitations

Due to the given extensive scope of this work, a comprehensive explanation of terms and definitions was dispensed with to present the most important results. In particular, the lack of detailed technical definitions and the limited technical perspective may present gaps in the discourse. This limitation prevents a comprehensive understanding of the intricacies associated with the technical facets of the topic under discussion. Moreover, although the issue of dehumanization has been addressed, there is a window for further exploration, particularly in the context of a comparative analysis between human rights and the integration of Artificial Intelligence. This exploration could provide a nuanced understanding of the ethical implications associated with the potentially dehumanizing effects of AI technologies. Furthermore, limited attention has been paid to regulatory and legal perspectives aimed at mitigating ethical risks. Presenting only the key laws relevant to the ethical considerations outlined here is somewhat limiting and it could be argued that an exhaustive exploration of the regulatory landscape is needed in the future. Given the global reach of many companies, there is a need for a more comprehensive exploration of the legal frameworks in different jurisdictions in order to improve practical applicability for internationally operating companies.

5. Conclusion and recommendations for future research

In this study, a systematic literature review was conducted to explore the ethical perspectives arising from algorithmic recruitment and to describe strategies for dealing with the identified risks for companies as well as for future research.

Digital recruiting 3.0 has introduced technological innovations, namely AI, that have a significant impact on the recruitment process. Evidence shows that most authors share similar views on the potential of AI recruiting as in facilitating not only a more efficient but also a more ethical hiring and selection process. AI proves to be an important tool to mitigating human bias and to promoting fairness in hiring. However, it is critical to recognize that training data and criterion labels can inadvertently reflect and perpetuate human bias, with the primary source of algorithmic bias being historical data. A fundamental risk associated with AI in the hiring process lies in the challenge of explaining the underlying algorithms and the decisions generated by the software, which impact all stakeholders at all levels. Implementing technical measures, such as creating unbiased data sets and improving the transparency of algorithms, proves to be essential to reducing discrimination by algorithms in the hiring process.

However, further scientific research has the potential to improve the validity and accountability of AI-powered recruitment software. Subsequent research efforts aimed at exploring discrimination through algorithmic recruitment could benefit from the application of quantitative analysis (e.g., experiments) and from expanding their scope by covering different countries. These studies could look at the intricacies of algorithmic recruitment mechanisms and the technical rules that influence the recruitment process. In addition, an assessment of perceived fairness and of the attractiveness level of organizations using AI in recruitment processes by applicants could provide valuable insights. Within the existing literature, there was a predominant emphasis on discrimination factor within the hiring market of the traditional labor market. However, it is imperative for future theoretical research to broaden its perspective by including a comprehensive examination of how advanced technologies influence and contribute to considerations of fairness in hiring processes within the dynamic context of the digital economy. Moreover, even though the focus on the identified risks is of enormous importance, debates on the danger of dehumanization, validity testing and accountability are so far only dealt with to a limited extent although this is of today's fundamental concern. A more in-depth analysis is imperative to promote a holistic understanding of the ethical implications associated with AI, which could contribute to the development of responsible and accountable algorithmic systems in the corporate context.

References

Acikgoz, Y., et al. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, *28*(4), 399–416. https://doi.org/10.1111/ijsa.12306

Al-Alawi, A. I., et al. 2021). The Role of Artificial Intelligence in Recruitment Process Decision-Making. *2021 International Conference on Decision Aid Sciences and Application, DASA 2021*, 197–203. https://doi.org/10.1109/DASA53625.2021.9682320

Andreini, D., & Bettinelli, C. (2017). *International Series in Advanced Management Studies Business Model Innovation*. Springer International Publishing AG. http://www.springer.com/series/15195

Aquino, Y. S. J., et al. (2023). Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: A qualitative study of multidisciplinary expert perspectives. *Journal of Medical Ethics*. https://doi.org/10.1136/jme-2022-108850

Balayn, A., et al. (2021). Managing bias and unfairness in data for decision support: a survey of machine learning and data engineering approaches to identify and mitigate bias and unfairness within data management and analytics systems. *VLDB Journal*, *30*(5), 739–768. https://doi.org/10.1007/s00778-021-00671-8

Bellamy, R. K. E., et al. (2018). AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. *IBM*. http://arxiv.org/abs/1810.01943

Bird, E., et al. (2020). The ethics of artificial intelligence: Issues and initiatives. *European Parliament*. https://doi.org/10.2861/6644

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

Bogen, M. (2019). All the Ways Hiring Algorithms Can Introduce Bias. *Harvard Business Review*.

Canetti, R., et al. (2019). From Soft Classifiers to Hard Decisions: How fair can we be? In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 309–318). http://arxiv.org/abs/1810.02003

Chamorro-Premuzic, T. (2019a). Attractive People Get Unfair Advantages at Work. AI Can Help. *Harvard Buisness Review*.

Chamorro-Premuzic, T. (2019b). Will AI Reduce Gender Bias in Hiring. *Harvard Business Review*.

Chamorro-Premuzic, T., & Akhtar, R. (2019). Should Companies Use AI to Assess Job Candidates. *Harvard Business Review*.

Chamorro-Premuzic, T., et al. (2017). The datafication of talent: how technology is advancing the science of human potential at work. In *Current Opinion in Behavioral Sciences,* 18, 13–16. Elsevier Ltd. https://doi.org/10.1016/j.cobeha.2017.04.007

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. In *Humanities and Social Sciences Communications*, 10(1). *Springer Nature*. https://doi.org/10.1057/s41599-023-02079-x

da Motta Veiga, S. P., et al. (2023). Seeming Ethical Makes You Attractive: Unraveling How Ethical Perceptions of AI in Hiring Impacts Organizational Innovativeness and Attractiveness. *Journal of Business Ethics, 186*(1), 199–216. https://doi.org/10.1007/s10551-023-05380-6

Dattner, B., et al. (2019). The Legal and Ethical Implications of Using AI in Hiring. *Harvard Business Review*.

Döring, N. (2023). Forschungsmethoden und Evaluation in den Sozial-und Humanwissenschaften, 6. Auflage. *Springer-Verlag GmbH.*

Fernández-Martínez, C., & Fernández, A. (2020). AI and recruiting software: Ethical and legal implications. *De Gruyter*, *11*(1), 199–216. https://doi.org/10.1515/pjbr-2020-0030

Florentine, S. (2016). How artificial intelligence can eliminate bias in hiring Feature. *CIO*. https://www.cio.com/article/236913/how-artificial-intelligence-can-eliminate-bias-in-hiring.html

França, J. F. T., et al. (2023). Artificial intelligence applied to potential assessment and talent identification in an organisational context. *Heliyon*, *9*(4). https://doi.org/10.1016/j.heliyon.2023.e14694

Frank, H., & Hatak, I. (2014). Doing a research literature review. In *How to Get Published in the Best Entrepreneurship Journals*, 94–117. Edward Elgar Publishing. https://doi.org/10.4337/9781782540625.00012

Frazzetto, A. (2023). The Future Of Recruitment: How To Use AI To Make The Hiring Process More Inclusive. *Forbes Technology Council*. https://www.forbes.com/sites/amyfeldman/2023/08/15/how

Fritts, M., & Cabrera, F. (2021). AI recruitment algorithms and the dehumanization problem. *Ethics and Information Technology*, *23*(4), 791–801. https://doi.org/10.1007/s10676-021-09615-w

Hancock, B., et al. (2023). Generative AI and the future of HR. *McKinsey Company*.

Hunkenschroer, A. L., & Kriebitz, A. (2023). Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring. *AI and Ethics*, 3(1), 199–213. https://doi.org/10.1007/s43681-022-00166-4

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

Kamishima, T., et al. (2018). Model-based and actual independence for fairness-aware classification. *Data Mining and Knowledge Discovery*, *32*(1), 258–286. https://doi.org/10.1007/s10618-017-0534-x

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. In *Business Horizons*, 62(1), 15–25. Elsevier Ltd. https://doi.org/10.1016/j.bushor.2018.08.004

Kelly, J. (2023). How AI-Powered Tech Can Help Recruiters And Hiring Managers Find Candidates Quicker And More Efficiently. *Forbes*. https://www.forbes.com/sites/rashishrivastava/2023/09/01/ai-red-teams-google-nvidia-microsoft-meta/?

Kim-Schmid, J., & Raveendhran, R. (2022). Where AI Can — and Can’t — Help Talent Management. *Harvard Business Review*.

Köchling, A., et al. (2021). Highly Accurate, But Still Discriminatory: A Fairness Evaluation of Algorithmic Video Analysis in the Recruitment Context. *Business and Information Systems Engineering*, *63*(1), 39–54. https://doi.org/10.1007/s12599-020-00673-w

Kraus, S., et al. (2020). The art of crafting a systematic literature review in entrepreneurship research. *International Entrepreneurship and Management Journal*, *16*(3), 1023–1042. https://doi.org/10.1007/s11365-020-00635-4

Kriebitz, A., & Lütge, C. (2020). Artificial Intelligence and Human Rights: A Business Ethical Assessment. *Business and Human Rights Journal*, *5*(1), 84–104. https://doi.org/10.1017/bhj.2019.28

Langer, M., et al. (2018). Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection. *Computers in Human Behavior, 81*, 19–30. https://doi.org/10.1016/j.chb.2017.11.036

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

Lin, Y. T., et al. (2021). Engineering Equity: How AI Can Help Reduce the Harm of Implicit Bias. *Philosophy and Technology,* 34, 65–90. https://doi.org/10.1007/s13347-020-00406-7

Mann, G., & O'neil, C. (2016). Hiring Algorithms Are Not Neutral. *Harvard Business Review*.

Mujtaba, D. F., & Mahapatra, N. R. (2019). Ethical Considerations in AI-Based Recruitment. In M. Cunningham & P. Cunningham (Eds.), *IEEE International Symposium on Technology in Society (ISTAS) Proceedings*. IEEE International Symposium on Technology in Society (ISTAS) Proceedings.

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

Oswald, F. L., et al. (2020). Big Data in Industrial-Organizational Psychology and Human Resource Management: Forward Progress for Organizational Research and Practice. *Annu. Rev. Organ. Psychol. Organ. Behavior*, 7, 505–538. https://doi.org/10.1146/annurev-orgpsych-032117

Pena, A., et al. (2020). Bias in multimodal AI: Testbed for fair automatic recruitment. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2020-June*, 129–137. https://doi.org/10.1109/CVPRW50498.2020.00022

Persson, A. (2016). Implicit bias in predictive data profiling within recruitments. *IFIP Advances in Information and Communication Technology*, *498*, 212–230. https://doi.org/10.1007/978-3-319-55783-0_15

Petticrew, M., & Roberts, H. (2006). Systematic Reviews in the Social Sciences a practical guide. *Blackwell publishing.*

Polli, F. (2019). Using AI to Eliminate Bias from Hiring. *Harvard Business Review*.

Raghavan, M., et al. (2020). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. In *Conference on fairness, accountability, and transperancy*. Association for Computing Machinery. https://doi.org/10.1145/3351095.3372828

Reicin, E. (2023). Developing Principles And Protocols For Recruiting And Hiring With AI. *Forbes Nonprofit Council*. https://www.forbes.com/sites/isabelbekele/2023/08/15/

Rodgers, W., et al. (2023). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review*, 33(1). https://doi.org/10.1016/j.hrmr.2022.100925

Rojon, C., et al. (2021). Utilization and development of systematic reviews in management research: What do we know and where do we go from here? *International Journal of Management Reviews*, *23*(2), 191–223. https://doi.org/10.1111/ijmr.12245

Ross, C. (2023). Can AI Be As Unethical And Biased As Humans? *Forbes Business Council*. https://www.forbes.com/sites/amyfeldman/2023/08/15/how

Sánchez-Monedero, J., et al. (2020). What does it mean to "solve" the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 458–468. https://doi.org/10.1145/3351095.3372849

Schumann, C., et al. (2020). We Need Fairness and Explainability in Algorithmic Hiring. In N. An, A. Yorke-Smith, A. El Fallah Seghrouchni, & G. Sukthankar (Eds.), *19th international conference on autonomous agents and multiagent sstems (AAMAS 2020)*. www.ifaamas.org

Seuring, S., & Gold, S. (2012). Conducting content-analysis based literature reviews in supply chain management. In *Supply Chain Management*, 17(5), 544–555.

https://doi.org/10.1108/13598541211258609

Simbeck, K. (2019). HR analytics and ethics. *IBM Journal of Research and Developement* , *63*(4/5), 1–12.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Tada, H., et al. (2021). Effects of artificial intelligence-Supported Automated Nutritional Intervention on LDL cholesterol Control in Patients with Familial Hypercholesterolemia (iSTART-FH): Protocol for a randomized controlled trial. *BMJ Open*, *11*(12). https://doi.org/10.1136/bmjopen-2021-053453

Tambe, P., et al. (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

Thorpe, R., et al. (2005). Using knowledge within small and medium-sized firms: A systematic review of the evidence. In *International Journal of Management Reviews*, 7(4), 257–281. https://doi.org/10.1111/j.1468-2370.2005.00116.x

Tilmes, N. (2022). Disability, fairness, and algorithmic bias in AI recruitment. *Ethics and Information Technology*, *24*(2). https://doi.org/10.1007/s10676-022-09633-2

Tranfield, D., et al. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. In *British Journal of Management*, 14(3), 207–222. https://doi.org/10.1111/1467-8551.00375

Vasconcelos, M., et al. (2018). Modeling Epistemological Principles for Bias Mitigation in AI Systems: An Illustration in Hiring Decisions. In *AAAI/ACM Conference on AI, Ethics, and Society*, 323–329. https://doi.org/10.1145/3278721.3278751

Vom Brocke, J. et al. (2009), Reconstructing the giant: on the importance of rigour in documenting the literature search process. *ECIS 2009 Proceedings*. 161. https://aisel.aisnet.org/ecis2009/161

Williams, B. A., et al. (2018). How Algorithms Discriminate Based on Data They Lack Challenges, Solutions, and Policy Implications. *Journal of Information Policy*, *8*. https://doi.org/10.5325/jinfopoli.8.2018.0078/1611323/jinfopoli_8_1_78.pdf

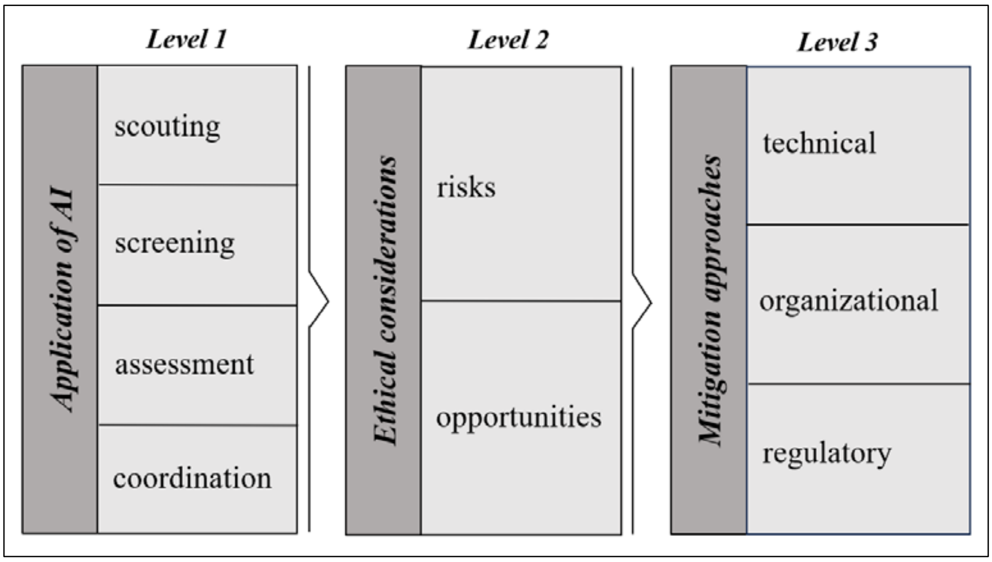
Yam, J., & Skorburg, J. A. (2021). From human resources to human rights: Impact assessments for hiring algorithms. *Ethics and Information Technology*, *23*(4), 611–623. https://doi.org/10.1007/s10676-021-09599-7

Yarger, L., et al. (2020). Algorithmic Equity in the Hiring of Underrepresented IT Job Candidates. *Online Information Review*, *44*, 383–395.

Zhang, L., et al. (2017). Achieving non-discrimination in data release. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F129685*, 1335–1344. https://doi.org/10.1145/3097983.3098167

Appendix

Appendix A: Research Design



Appendix B: Search criteria

Criteria	***Description:***
Search terms	“AI”; “Artificial Intelligence”; “Algorithm”; “Machine Learning” AND “Ethical”; “Ethics”; “Responsibility”; “Fairness”; “Bias”; AND “Recruitment”; “Recruiting”; “Selection”; “Hiring”; “Job interview”; “Applicant screening”
Search procedure	Initial keyword search; backward search
Databases	ScienceDirect, Scopus, Web of Science
Language	English
Time frame	2010-2023
Inclusion	Articles from journals or databases, conference reports, books or books chapters, practitioner-oriented articles, open access to full text
Exclusion	Interviews, commentaries, conference abstracts, letters, book reviews, studies without direct relation to the topic
Quality criteria	Peer-reviewed
Screening process	1. Title; 2. Abstract; 3. Full text
Analysis process	Interpretive and descriptive approach: Review and categorize the identified literature and outline a risk mitigation framework

Appendix C: List of academic, conference and business journals used

JOURNAL	TOTAL
Academic	***27***
AI and ethics	1
Annual rev. Organ. Psychol. Organ. Behavior	1
Big data and society	1
Business and human rights journal	1
Business and information systems engineering	1
Business horizons	1
California management review	1
Computers in human behavior	1
Current opinion in behavioral science	1
Data mining and knowledge discover	1
De Gruyter	1
Ethics and information technology	3
Heliyon	1
Human resource management review	1
Humanities and social sciences communications	1
IBM journal of research and development	2
International journal of selection and assessment	1
Journal of business ethics	2
Journal of information policy	1
Online information review	1
Organizational behavior and human decision processes	1
Philosophy and technology	1
VLDB journal	1
Conference	***11***

19th international conference on autonomous agents and multiagent systems (AAMAS)	1
AAAI/ACM conference on ai, ethics, and society	1
European parliament conference paper	1
FAT conference on fairness, accountability, and transparency	2
IEEE computer society conf. on computer vision and pattern recognition workshops	1
IEEE international symposium on technology in society (ISTAS) proceedings	1
IFIP advances in information and communication technology	1
International conference on decision aid sciences and application	1
Proceedings of the ACM international conf. on knowledge discovery & data mining	1
Proceedings of the conference on fairness, accountability, and transparency	1
Business	***14***
CIO magazine	1
Forbes magazine	3
Harvard business review	9
McKinsey company	1
	52

Appendix D: Summary of articles used for the systematic literature review

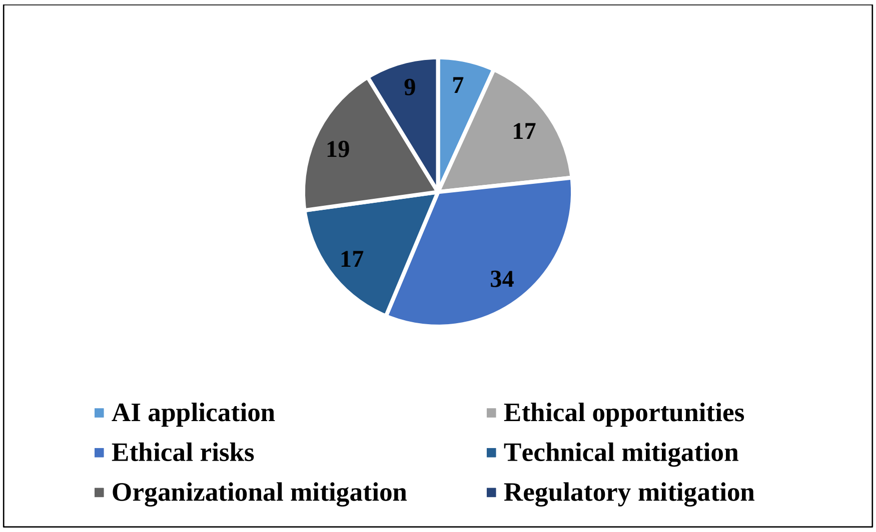
Author	Year	Country	Topic	Method	AI Application	Ethical Opportunities	Ethical risks	Technical mitigation	Organizational mitigation	Regulatory mitigation
Acikgoz, Y., et al.	2020	USA	Justice perceptions of artificial intelligence in selection	Survey		X	X			
Al-Alawi, A.I., et al.	2021	Saudi-Arabia	The role of AI in recruitment process decision-making	Semi-structured Interview			X		X	
Balayn, A., et al.	2021	Nether-lands	Managing bias and unfairness in data for decision support	SLR				X		
Bellamy, R., et al.	2018	USA	AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias	Descriptive				X		
Bird, E., et al.	2020		The ethics of artificial intelligence: Issues and initiatives	Descriptive and case study			X		X	X
Black, J. and van Esch, P.	2020	USA	AI-enabled recruiting: What is it and how should a manager use it?	Descriptive	X	X			X	
Bogen, M.	2019		All the Ways Hiring Algorithms Can Introduce Bias	None			X			
Canetti, R., et al.	2019	USA	From Soft Classifiers to Hard Decisions: How fair can we be?	Experiment				X		
Chamorro-Premuzic, T. and Akhtar, R.	2019		Should Companies Use AI to Assess Job Candidates	None	X	X	X		X	
Chamorro-Premuzic, T.	2019		Will AI Reduce Gender Bias in Hiring	None		X	X			
Chamorro-Premuzic, T.	2019		Attractive People Get Unfair Advantages at Work. AI Can Help	None		X				
Chamorro-Premuzic, T., et al.	2017	USA	The datafication of talent: how technology is advancing the science of human potential at work	Descriptive	X	X				
Chen, Z.	2023	China	Ethics and discrimination in artificial intelligence-enabled recruitment practices	SLR			X	X	X	X
Dattner, B., et al.	2019		Legal and Ethical Implications of Using AI in Hiring	None			X	X		

Author	Year	Country	Topic	Method	AI Application	Ethical Opportunities	Ethical risks	Technical mitigation	Organizational mitigation	Regulatory mitigation
Da Motta Veiga, S.P., et al.	2022	France	Unraveling How Ethical Perceptions of AI in Hiring Impacts Organizational Innovativeness and Attractiveness	Survey		X	X			
Fernandez-Martinez, C. and Fernandez A.	2020	Spain	AI and recruiting software: Ethical and legal implications	Theoretical			X	X		
Florentine, S.	2016		How artificial intelligence can eliminate bias in hiring Feature	None		X		X		
Frazzetto, A.	2023		How To Use AI To Make the Hiring Process More Inclusive	None		X				
França, T., et al.	2023	Portugal	Artificial intelligence applied to potential assessment and talent identification in an organizational context	SLR		X	X			X
Fritts, M. and Cabrera, F.	2021	USA	AI recruitment algorithms and the dehumanization problem	Descriptive			X			
Hancock, B., et al.	2023		Generative AI and the future of HR	Interviews		X				
Hunkenschroer A. and Luetge C.	2022	Germany	Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda	SLR	X	X	X		X	X
Hunkenschroer A. and Kriebitz, A.	2023	Germany	Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring	SLR			X	X	X	
Kamishima T., et al.	2018	Japan	Model-based and actual independence for fairness-aware classification	Model-based Datamining				X		
Kelly, J.	2023		How AI-Powered Tech Can Help Recruiters and Hiring Managers Find Candidates Quicker and More Efficiently	None	X					
Kim-Schmid, J. and Raveendhran, R.	2022		Where AI Can — and Can't — Help Talent Management	None	X	X	X			
Köchling, A., et. al.	2020	Germany	A Fairness Evaluation of Algorithmic Video Analysis in the Recruitment Context	Statistical data analysis	X	X	X		X	
Kriebitz, A. and Luetge C.	2020	Germany	Artificial Intelligence and Human Rights: A Business Ethical Assessment	Descriptive			X		X	

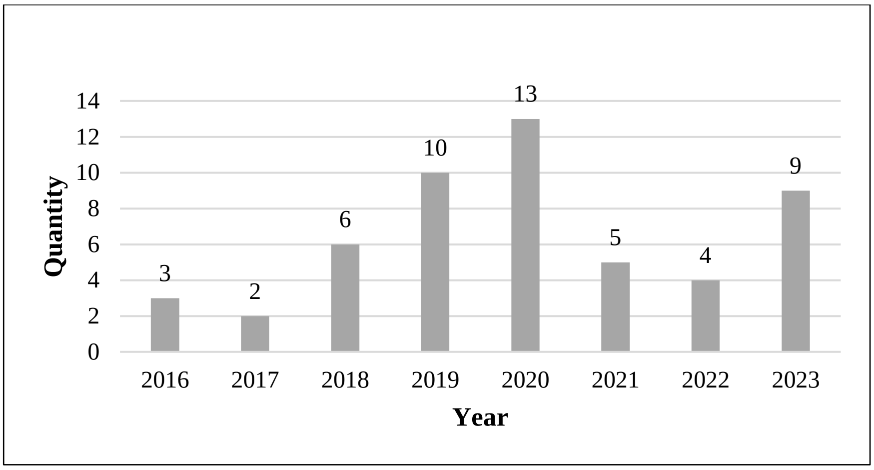
Author	Year	Country	Topic	Method	AI Application	Ethical Opportunities	Ethical risks	Technical mitigation	Organizational mitigation	Regulatory mitigation
Langer, M., et al.	2018	Germany	The role of computer experience and information on applicant reactions towards novel technologies for personnel selection	2x2 study design			X			
Lee, M.	2018	USA	Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management	Online experiment			X			
Lin, Y., et al.	2021	Taiwan	Engineering Equity: How AI Can Help Reduce the Harm of Implicit Bias	Two-dimensional framework		X	X	X		
Mann, G. and O`Neil, C.	2016		Hiring Algorithms Are Not Neutral	None				X		
Mujitaba, D. and Mahapatra, N.	2019	USA	Ethical Considerations in AI-Based Recruitment	Descriptive			X	X	X	
Newman, D., et al.	2020	USA	Algorithmic reductionism and procedural justice in human resource decisions	Laboratory experiments			X			
Oswald, F., et al.	2020	USA	Big Data in Industrial-Organizational Psychology and Human Resource Management	Descriptive				X		
Pena, A., et al.	2020	Spain	Bias in multimodal AI: Testbed for fair automatic recruitment	Model-based experiment			X			X
Persson, A.	2016	Sweden	Implicit bias in predictive data profiling within recruitments	Descriptive			X			
Polli, F.	2019		Using AI to Eliminate Bias from Hiring	None		X				
Raghavan, M., et al.	2020	USA	Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices	Systematic review of assessment vendors				X	X	X
Reicin, E.	2023		Developing Principles and Protocols for Recruiting And Hiring With AI	None				X	X	
Rodgers, W., et al.	2023	USA	An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes	Model framework			X		X	

Author	Year	Country	Topic	Method	AI Application	Ethical Opportunities	Ethical risks	Technical mitigation	Organizational mitigation	Regulatory mitigation
Ross, C.	2023		Can AI Be as Unethical and Biased As Humans?	None			X			
Sanchez-Monedero, J., et al.	2020	United Kingdom	Social, technical, and legal perspectives from the UK on automated hiring systems	Descriptive			X		X	X
Schumann, C., et al.	2020	USA	We Need Fairness and Explainability in Algorithmic Hiring	Descriptive			X			
Simbeck, K.	2019	USA	HR analytics and ethics	Theoretical				X	X	
Tambe, P., et al.	2019	USA	Artificial intelligence in human resources management: Challenges and A path forward	None			X		X	
Tilmes, N.	2022	USA	Disability, fairness, and algorithmic bias in AI recruitment	Descriptive			X		X	
Vasconcelos, M., et al.	2018	Brasil	Modeling Epistemological Principles for Bias Mitigation in AI Systems: An Illustration in Hiring Decisions	Two-dimensional framework			X			
Williams, B., et al.	2018	USA	How Algorithms Discriminate Based on Data They Lack Challenges, Solutions, and Policy Implications	Descriptive			X			X
Yarger, L., et al.	2020	USA	Algorithmic Equity in the Hiring of Underrepresented IT Job Candidates	Conceptual			X		X	
Yam, J. and Skorburg, J.A.	2021	Canada	From human resources to human rights: Impact assessments for hiring algorithms	Descriptive		X	X		X	X
Zhang, L., et al.	2017	Canada	Achieving non-discrimination in data release	Experiment				X		
Subtotal					**7**	**17**	**34**	**17**	**19**	**9**
Total										**52**

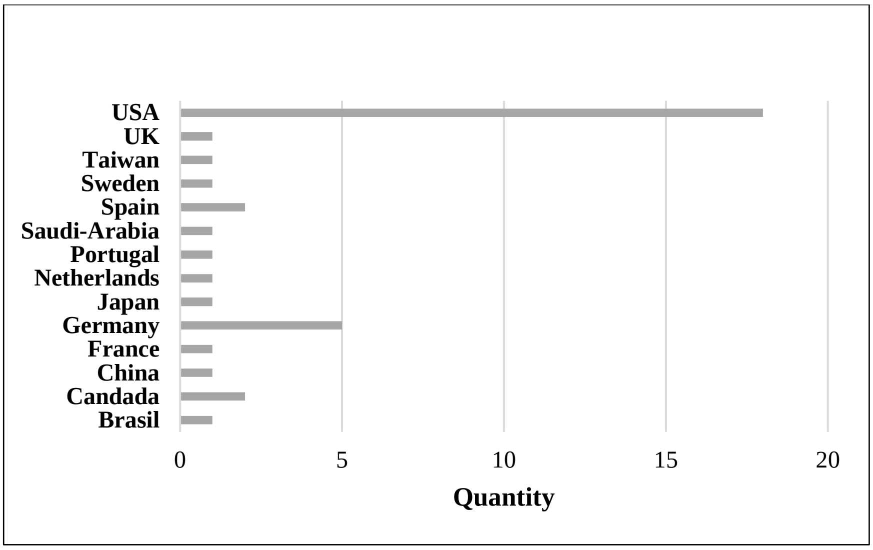
Appendix E: Distribution of published articles per category



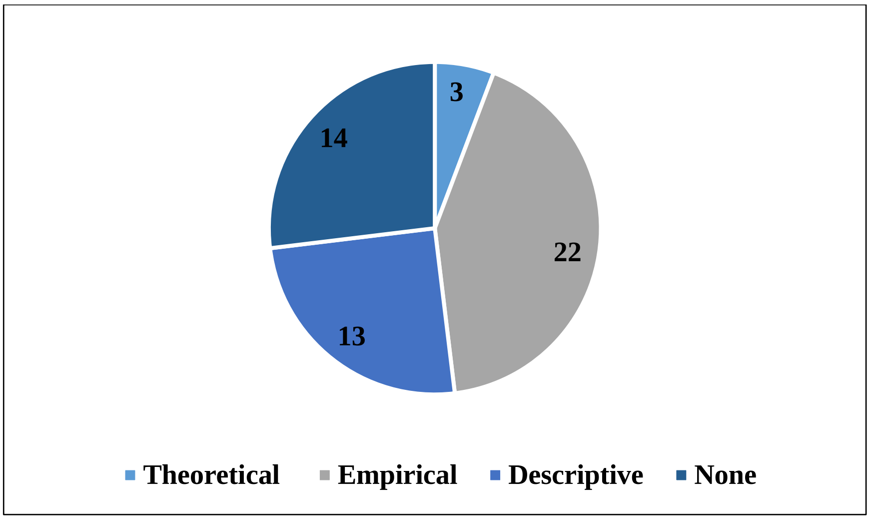
Appendix F: Distribution of published articles per year



Appendix G: Geographical distribution of published articles



Appendix H: Distribution of published articles by research design



**The terms include the following methods, which are listed in appendix C:*

Theoretical	Theoretical; conceptional
Empirical	Case Study; (online) experiment; interviews; SLR; survey; frameworks
Descriptive	Literature review
None	None