A Survey on Human Resource Management Under the AI Act: Ethical, Practical, and Regulatory Perspectives

Nicola Alboré[1,*,†], Alessandro Castelnovo[1,†], Matteo Della Valle[2], Andrea Ermellino[1,†], Luca Puggini[1,†] and Silvia Tessaro[1,†]

[1]Data & Artificial Intelligence Office, Intesa Sanpaolo S.p.A., Italy
[2]Università Milano Bicocca, Milano, Italy

Abstract

This paper explores the integration of artificial intelligence into human resource management, focusing on its ethical, practical, and regulatory implications. As digital transformation reshapes human resources practices, artificial intelligence offers potential for increased efficiency and innovation, but also raises challenges related to fairness, transparency, and governance. Key areas such as fairness in decision-making, system explainability, and human oversight are examined to assess their impact on recruitment processes, employee well-being, and organizational performance. By critically analyzing these aspects, the study highlights the dual role of artificial intelligence in improving inclusion while exposing the risks of perpetuating bias and reducing accountability. Building on existing literature, this review discusses how organizations can balance technological advancements with ethical principles to promote trust and equity in the workplace. In addition, it calls for strengthened regulatory frameworks and collaborative efforts between policymakers, practitioners, and technologists to ensure responsible artificial intelligence deployment. By addressing these issues, the study aims to contribute to the development of sustainable human resource practices that align technological progress with organizational and social values.

1. Introduction

As organizations deal with increasingly complex workforce and societal dynamics, implementing well-suited human resource management (HRM) policies has become a key priority. The strategic resonance HRM practices yield within organizations, which impacts firm performance and employee satisfaction, has pivoted significant attention within both academic and industrial environments [1, 2, 3, 4].

Despite some criticism about the potential detrimental effects of HRM practices geared towards companies on employee wellbeing [2, 3], many consider HRM an instrumental approach that allows organizations to perform better [5, 6, 7, 8, 9]. Not only by pursuing a positive renewal of work environments, but also by allowing better engagement by employees, thus orienting HRM towards "common good" values, centered around the need for sustainability and progress [10, 11]. Some scholars suggested that both employers' productivity and employee wellbeing should not be viewed as competing objectives, but rather as complementary goals that could be reached together [12]. However, it is still not clear how this joint optimization could be carried out [13], and the ethical aspects surrounding the difficult balance between the prioritization of employees' needs and company goals [14, 15].

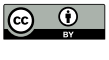
In today's economic landscape defined by a strong need for digitalization and a concurrent race for information harvesting [16, 17], data has become a critical catalyst for innovation [18]. At the forefront of this digitalization [19], the development of machine-based systems designed to operate with varying levels of autonomy, potentially exhibiting adaptiveness after deployment, and generating outputs—such

AIMMES 2025 Workshop on AI bias: Measurements, Mitigation, Explanation Strategies | co-located with EU Fairness Cluster Conference 2025, Barcelona, Spain

*Corresponding author.

†The views and opinions expressed are those of the authors and do not necessarily reflect the views of Intesa Sanpaolo, its affiliates or its employees.

nicola.albore@intesasanpaolo.com (N. Alboré); alessandro.castelnovo@intesasanpaolo.com (A. Castelnovo); m.dellavalle2@campus.unimib.it (M. D. Valle); andrea.ermellino@intesasanpaolo.com (A. Ermellino); luca.puggini@intesasanpaolo.com (L. Puggini); silvia.tessaro@intesasanpaolo.com (S. Tessaro)
0000-0003-1147-7026 (N. Alboré); 0000-0001-5234-1155 (A. Castelnovo); 0009-0005-1297-9185 (A. Ermellino)



as predictions, recommendations, or decisions—that influence physical or virtual environments, has found its perfect application [20]. However, the need to harness, process, and interpret data poses new challenges in understanding how companies should not only pursue firm return optimization, but also address human capital, especially for strategic decision-making [19, 21]. A larger commitment to technological funding is more feasible for tech-focused industries due to their substantial resources [22, 23].

Despite some mild skepticism [24, 25] regarding the negative effects that impact the management of human resources from such technologies, many have noted that a greater effort towards AI-based solutions boosts organizational growth [26], competitive advantage, and also reduces operational costs, thus leveraging a more attractive work environment [27, 28].

The need to align current HRM standards with recent advances in the field of AI development or computer science is generally justified by the impact on attracting and retaining motivated employees, due to the integration of a personalized, efficient and supportive workforce within organizations [29, 30]. In particular, many organizations have integrated AI-driven tools into their HRM systems to enhance various processes. In recruitment and onboarding, AI aids in pre-screening candidates, accelerating interviews, and identifying the best-fit applicants [31, 32]. For development and performance management, it helps track and predict learning needs, assess employee performance, and support managers in identifying strengths and areas for improvement [33, 34]. Additionally, in engagement and retention, AI analyzes employee sentiment, predicts turnover, and recommends resources to promote mental and physical well-being [35, 36]. Although these contributions are promising, the ethical implications of AI in HRM remain controversial [37]. The unprecedented insight and efficiency offered by AI is accompanied by the complexity of its underlying algorithms and methods, leading many to question whether such tools inadvertently reinforce biases, compromise privacy, or lack suitability for autonomous decision-making [38, 39, 40, 41]. This has sparked calls for deeper studies, particularly in HRM, given its significant influence on individual and organizational innovation [21, 42, 43].

The importance of addressing these concerns is underscored by the proactive efforts of the European Union to regulate AI [12]. In April 2021, the European Commission introduced the first proposal for a "Regulation laying down harmonized rules on artificial intelligence" (AI Act) [44], which, in its definitive version, officially came into force on 1 August 2024. An important focus of the AI Act is on high-risk AI systems, including those used in hiring processes and people management, both of which are pivotal to the HRM sector. It emphasizes the protection of fundamental human rights, from which ethical principles like fairness, transparency, explainability, and human oversight naturally emerge as essential requirements for AI systems developed in this context. Rather than restricting AI adoption, the Act provides a framework of essential guidelines to promote the ethical and responsible use of AI-driven solutions. By rooting these measures in the values and fundamental rights of the EU, the AI Act seeks to build trust in AI technologies while encouraging innovation and progress within organizations [45]. This paper responds to this request by conducting a literature review, offering an initial overview of ethical issues in the use of AI and its reflection on HR practices. After identifying key ethical opportunities and risks, we discuss and propose recommended practices to effectively manage these risks. Considering the profound impact recruitment decisions have on the lives of individuals [46], companies need to recognize both the benefits and the potential drawbacks of AI and understand how algorithmic decisions can sometimes conflict with their intended outcomes [47].

This paper contributes by providing a comprehensive literature review based on the key ethical pillars outlined in the AI Act—fairness, explainability, and human oversight—applied specifically to the HR sector. We evaluate 31 articles from top-tier journals, scoring and ranking them based on the level of detail provided. Additionally, we classify each article's perspective on AI adoption, categorizing it as presenting AI as an opportunity, a threat, or both. This analysis offers valuable insights into the ethical implications of AI in HR, helping organizations navigate the complexities of responsible AI implementation in recruitment and people management.

1.1. Key Ethical Pillars and Requirements in the AI Act

The Act aims to foster responsible artificial intelligence development and deployment in the EU, limiting risks of unintended societal impacts, such as reinforcing biases, compromising privacy, and making opaque or unaccountable decisions that affect individuals' lives and rights [48, 49]. The AI Act is a pioneering legislative framework designed to establish comprehensive standards for the ethical, safe, and human-centered use of AI across Member States. The regulation aims not only to protect citizens from potential risks but also to set a global standard, positioning Europe as a leader in responsible AI governance. The Act draws heavily on the EU's 2020 White Paper on Artificial Intelligence [50], which outlined the importance of upholding values such as human dignity, inclusion, and non-discrimination in AI applications. By instituting this framework, the EU underscores the importance of aligning AI innovation with fundamental rights, aiming to prevent a fragmented regulatory landscape within Europe and to establish a benchmark for AI standards worldwide. Building on a risk-based approach, the European Union has chosen to implement stringent regulations for high-risk systems, including, as previously mentioned, systems used for personnel evaluation and selection. Requirements for high-risk AI systems underline their essential role in safeguarding fundamental rights and fostering public trust. Three key ethical pillars to consider for mitigating the risks of introducing AI in delicate and highly sensitive sectors, such as HRM, are fairness, explainability, and human oversight.

1. Fairness: this pillar is prioritized to address the potential for AI systems to reinforce or even amplify existing societal biases and to ensure equality, promoting equitable outcomes for all individuals, regardless of demographic background. In applications like recruitment, credit scoring, or law enforcement, biased algorithms can lead to unjust outcomes, disproportionately affecting certain demographic groups. The AI Act mandates robust bias mitigation techniques, requiring developers and providers to implement mechanisms that continually monitor and reduce discriminatory effects throughout the AI lifecycle. This ensures that fairness is not just a design goal but an ongoing responsibility as systems evolve and adapt.
2. Explainability: another crucial factor, especially in complex AI applications where decisions affect individual rights or societal welfare. High-risk AI systems are required to provide sufficient transparency, enabling users and possibly affected individuals to understand how and why decisions are made. This is particularly significant in sectors such as employment and justice, where AI-driven decisions can have life-altering consequences. The Act's focus on explainability demands that AI systems are developed and used in a way that allows appropriate tracebility and transparency, making possible for users to grasp the systems' capabilities and, if necessary, challenge their outputs, in order to guarantee human autonomy and dignity.
3. Human Oversight: This pillar ensures control over AI systems and fixing accountability for the use of such systems. While AI is designed to operate independently, the Act stresses the importance of having human supervisors who can intervene if necessary. In high-risk areas, human oversight serves as a safeguard against potential errors or unforeseen consequences in AI behavior. By maintaining a human element in decision-making processes, the Act promotes a balanced approach that respects both technological autonomy and human judgment, particularly in critical sectors where the stakes are high.

These three ethical principles: fairness, explainability, and human oversight are essential to the ethical framework of the AI Act. They establish that high-risk AI systems must operate within clearly defined ethical and legal boundaries, ensuring that AI not only advances technological frontiers but does so responsibly, with a strong commitment to societal values. This approach highlights the commitment of the EU to maintaining ethical integrity in AI usage, emphasizing that regulatory measures are indispensable in high-stakes applications to protect public trust and individual rights. These three pillars form the foundation of the literature review presented in this paper.

Paper	Fairness	Explain.	Human Overs.	Orientation
[51]	2	1	2	T
[52]	3	3	4	O
[53]	3	3	4	O/T
[54]	4	3	4	O
[55]	2	2	3	O
[56]	2	3	3	O
[57]	3	3	4	O
[58]	2	2	2	O
[59]	4	3	3	O/T
[60]	4	3	3	O/T
[61]	3	3	3	O/T
[62]	4	3	3	O/T
[63]	2	3	3	O/T
[64]	3	3	4	O/T
[65]	3	3	2	O
[66]	4	3	3	O/T
[67]	2	2	2	O
[68]	4	3	3	O/T
[69]	3	2	3	O/T
[70]	3	2	3	O
[71]	4	3	3	O/T
[72]	3	3	4	O
[73]	4	3	4	O/T
[74]	3	3	3	O/T
[75]	2	3	3	O
[76]	4	4	4	O/T
[77]	4	3	4	O/T
[78]	3	3	3	O
[79]	4	3	3	O/T
[80]	4	3	4	O/T
[81]	4	3	3	O

Table 1
Summary of the papers along with their respective scores (ranging from 1 to 4) for each topic: fairness, explainability and human oversight, along side with the overall orientations (O opportunity, T threat or O/T both) in the context of AI-driven HRM.

2. Review

2.1. Research Methodology

To understand the perspective of the international scientific community on the aforementioned topics, articles published in relevant and high-impact scientific journals and articles were selected. These sources were chosen based on their relevance, impact on citations, and recognized authority in the fields of AI and HR. The selection process involved the identification of leading journals in the domains of artificial intelligence, computer science, and HR management. Extensive searches were conducted in various academic databases using keywords related to fairness, explainability, and human oversight in AI. The inclusion criteria involved the date of publication of the article to ensure the inclusion of recent and relevant studies, the number of citations as an indicator of influence, and the impact factor of the journal to ensure high-quality and credible sources. By focusing on these parameters, the research aimed to incorporate a diverse and comprehensive collection of scholarly works that reflect the current state of academic discourse on these critical issues. The selected articles were evaluated and classified according to their discussion of the three key topics according to the following criteria:

- Score 1: The article discusses the topic in general terms, without delving into specifics, and does not provide a detailed theoretical definition nor refer to hypothetical or real cases.
- Score 2: The article provides a theoretical definition of the topic, but does not present examples of hypothetical or concrete applications.
- Score 3: The article explores the topic through a hypothetical case or by theorizing a personal algorithm.
- Score 4: The article presents a real case of business or workplace application of the discussed topic, or a survey of the population to gather opinions on the subject.

After giving a score, we further classified the articles for each topic as follows:

- Opportunities (O): Articles in which the authors are in favor of introducing AI into the field of human resources, as they believe that it could promote greater equity compared to traditional HR office practices.
- Threats (T): Articles in which the authors are not entirely opposed to the introduction of AI in the field of human resources but still believe that it would not be the solution to ensure greater equity; on the contrary, these authors highlight the threats arising from its use, which due to its functioning could end up exacerbating and automating biases and unfair decisions.
- Both (O/T): Articles in which the authors highlight both the significant opportunities and risks associated with the use of AI.

2.2. Fairness

A substantial portion of the reviewed literature supports the integration of AI systems into routine HR practices, particularly in recruitment and selection processes. Currently, it has gained some traction, as companies have started to implement AI-based software [77], to leverage AI to correlate social media activity with long-term retention predictors, a factor previously unconsidered in traditional methods [69], or to obtain an effective bias reduction system, highlighting AI's ability to promote fairer hiring results [82].

Several scholars contend that AI's widespread adoption streamlines routine HR tasks while crucially mitigating the biases intrinsic to human judgment [83]. Given that human subjectivity in hiring and selection processes is inherently limited in its reliance on objective data alone, a meticulously designed AI algorithm can potentially mitigate many sources of bias, promoting fairness in these activities. At the same time, its unrestrained deployment could raise concerns by inadvertently reinforcing and institutionalizing algorithmic biases, underscoring the critical need to properly assess the structural integrity of software tools disposed of by HR operators. We note, for example, Amazon's algorithm-based hiring system, which operated from 2014 to 2018, but was abandoned after it was found to

discriminate against women in IT roles, having been trained on data favored male applicants. Despite Amazon's attempts to neutralize gender indicators, the inherent bias of the algorithm could not be sufficiently corrected, leading to its discontinuation [84]. Similarly, companies that use Facebook's targeted recruitment tools have faced criticism for excluding certain demographic groups based on social media data and targeted lookalike audiences. This practice, which restricts job advertisements to specific groups, is not only ethically problematic but also contravenes legal standards, including the European AI Act and Title VII of the U.S. Civil Rights Act [59, 85]. To examine the factors through which AI can foster but also threaten fairness in HRM, we highlight several themes.

- Inclusive Language in Job Advertisements: human-written job advertisements often unintentionally incorporate biased language that can exclude certain demographic groups. AI-powered software, while not entirely flawless, can flag such phrases to prevent exclusionary language in recruitment campaigns. For such reasons when algorithms are not appropriately developed, they may also introduce biases, underscoring the importance of continuous human oversight, as outlined by the AI Act [50, 85]. Thus the need for AI tools to assist companies in crafting unbiased job advertisements, fostering inclusivity by evaluating candidates based on skills rather than personal attributes.
- Implementation of Fair CV Screening Practices: AI applications in CV screening represent some of the earliest examples of AI's influence in HR. AI not only accelerates these processes but can also reduce the risk of bias. Literature suggests that without AI, human-only screening can lead to discrimination based on gender [86], minority status [87], and age [88], potentially restricting access for these groups in the labor market. By focusing on predefined criteria during the design phase, AI minimizes unconscious biases, filtering candidates objectively without regard to subjective human factors like physical appearance or first impressions, which are often difficult to avoid or identify in traditional evaluations.
- Evaluation of Candidates' Skills and Traits: AI-based assessment tools have introduced innovative approaches to evaluating candidates, particularly by using simulations and gamified assessments to capture diverse data on candidates' competencies. Traditional methods, such as tests and surveys, have faced criticism since the 1970s for their limitations in assessing complex human potential [82]. Recent AI-driven assessments allow candidates to engage with designed scenarios that mimic workplace tasks, capturing responses and strategies in real-time. Such methods alleviate candidates' typical interview anxiety, enabling full engagement and mitigating biases associated with identity by focusing solely on in-game performance data.
- Interviews for Screened Profiles: AI has transformed the interview process by facilitating structured and bias-resistant candidate evaluations. The reviewed studies critique the traditional reliance on human intuition in interviews, noting that intuition is inherently subjective and prone to biases, which can undermine fair hiring outcomes [69]. By automating structured interviews, AI can assess extensive data beyond human capacity, focusing on the qualifications necessary for reliable performance while disregarding irrelevant, bias-prone data points. However, researchers emphasize that algorithms must be carefully designed to prevent embedding human biases into AI-driven systems, as algorithmic output is highly dependent on initial design inputs and training data [89].
- Mitigation of Implicit Bias: Implicit bias, defined as an unconscious negative attitude toward certain social groups, poses a challenge in recruitment and selection, as it influences perceptions and behaviors beyond conscious awareness. Some have noted the role of AI as an external moderator of implicit bias, enabling organizations to identify and address such biases in real-time [90]. AI can reveal patterns or characteristics that contribute to successful job performance while eliminating biases related to protected characteristics. Others suggested strategies for reducing implicit bias, which require a combination of technical and procedural improvements. Effective reduction strategies include anonymizing demographic data and emphasizing cultural and social commitments to inclusivity.
- Bias from Historical Data: AI models trained on historical data can perpetuate existing inequities

if the underlying data reflect past discriminatory practices. For instance, if a company has historically favored certain demographic groups, AI models trained on such data are likely to replicate and amplify these biases in recruitment outcomes [73, 91]. To mitigate this, developers are urged to identify and rectify biases within training data, utilizing strategies such as removing category-linked attributes that are irrelevant to job performance [76]. Additionally, causal discovery techniques, which seek to link variables directly related to job performance rather than demographic attributes, are recommended to improve fairness. Although causal discovery remains a developing field, it holds promise for distinguishing individual-specific factors from general category traits [92, 93]. Explicit randomization during selection, wherein candidates with identical recommendation scores are chosen randomly, has also been suggested to prevent entrenched discriminatory practices [76, 91].

- Bias Based on Category Membership: Discrimination based on category membership, such as gender, ethnicity, or other demographic attributes, is a longstanding issue that algorithms may not effectively address. For example, studies report an under-representation of women, Latino-Americans, and African-Americans in technology roles in the United States, partly due to high turnover among these groups resulting from workplace biases [59]. Furthermore, selection algorithms are often designed by homogeneous teams, which may unintentionally embed biases within the algorithm's structure, favoring candidates from similar backgrounds [81]. Bias persists for three primary reasons: first, recruiters determine algorithmic outputs; second, designers embed specific performance-related factors that may inherently favor certain groups based on company data; and third, factors such as university affiliation or demographic membership, often linked to successful performance, can exclude diverse candidates from consideration [51]. AI systems, rather than eliminating these exclusions, may unintentionally highlight standout members from overrepresented groups [94]. Additionally, cultural differences in expressions and behaviors present challenges in video-interview AI tools, as models trained on culturally specific data may misinterpret candidates from different backgrounds [54].
- Proxy Discrimination: Proxy discrimination occurs when AI systems infer protected category information through correlated non-protected attributes, resulting in indirect yet substantial bias [180, 204]. Despite efforts to eliminate biases, implicit associations embedded in historical data are difficult to remove, leading to covert discriminatory outcomes [95, 66]. Mitigation strategies include anonymizing demographic information and utilizing neuroscience or chatbot-based data collection methods that focus exclusively on job-related skills and characteristics. The use of avatars to replace human interviewers is another technique proposed to reduce the impact of proxy discrimination [62, 80]. Although these methods represent considerable efforts toward eliminating bias, complete removal of implicit biases remains challenging, underscoring the continued need for advancement in bias-mitigation strategies [51].
- Disparate Impact: This phenomenon describes the unintended adverse effects on protected categories resulting from ostensibly neutral policies or algorithms. Techniques to reduce disparate impact, such as masking sensitive attributes, are proposed in the literature, along with adjustments to algorithmic language and labeling to ensure fairness [66]. Under United States law, managers can defend against disparate impact claims by proving that selection criteria serve a business necessity and that no reasonable alternative methods exist. This burden of proof then shifts to the complainant, who must demonstrate viable alternatives [46].

2.3. Explainability

The second topic we focus on is explainability. By this term, we mean the ability for humans to clearly understand how an artificially intelligent system works and is able to output decisions (Explainable AI, n.d.). The need to understand how and why an AI system made a specific decision is crucial not only to ensure transparency and fairness but also to adhere to the ethical and legal principles established by current regulations.

Before delving into the chapter, it is helpful to clarify the term "explainability." Although there is no

universally accepted definition of XAI (eXplainable Artificial Intelligence), terms such as "understanding," "interpreting," and "explaining" are often used interchangeably. Typically, interpretability refers to understanding how a predictive model works, while explainability is related to models that are inherently more complex and difficult to understand. From a regulatory perspective, both the GDPR and the EU AI Act outline important provisions and stringent norms concerning explainability. These regulations require that the providers and users of AI systems comply with specific requirements to ensure that their tools can provide clear explanations of their operations and decisions [85, 96]. The GDPR, for example, grants individuals the right to receive meaningful information about the logic involved in automated decision-making and profiling [96]. Similarly, the AI act places a strong emphasis on promoting transparency, accountability, and human oversight in AI systems, ensuring that they align with European values and fundamental rights [85]. In this chapter, we will explore the main criticisms raised by scholars on this topic, addressing the issue of "black-box" algorithms. We will then explain how the theme of explainability relates to the two core legislative frameworks in Europe.

2.3.1. The Problem of 'Black Box' Algorithms in HR Techniques

Our review of the modern literature focused on explainability has highlighted that all articles raise the issue of 'black box' algorithms. A 'black box' algorithm is defined as such because users cannot determine how the algorithm made a particular decision [97]; it is an algorithm where users can only know the input data provided to the algorithm and receive the output data from it without fully understanding why it is providing those data. Naturally, this inability on the users part raises significant issues in terms of opacity, non-transparency, and lack of clarity in decision-making. In specific HRM practices, such as selection process regulated by algorithms, it would be difficult to explain exclusion of candidates if the technology underlying such processes are of "black box" type [71]. If recruiters cannot provide an accurate explanation of the process through which the algorithm made that decision, it leads to significant issues not only from an ethical and social point of view, with unpredictable and destructive damage caused to individuals [94], but also legal problems, given the legislative boundaries imposed by GDPR and the AI Act. We highlight from our analysis various points:

- Obligation to interpretability: users of the algorithms have an obligation to fully understand the functioning of the software they are using, and companies themselves should draft ethical reports informing how their HR office, or any department using AI systems, utilizes such tools [51]. It is certainly not a simple operation. In fact, for "black box," we can also refer to those algorithms that are well understood and have impressive predictive capabilities, but whose predictive relationships are too complex to interpret [98]. Considering HR practices, even when recruiters obtain selection data using sound practices, and those data and the algorithms used to interpret them demonstrate stronger robustness compared to traditional methods, there still remain gray areas of interpretability. It is not guaranteed that the detected patterns of prediction will reflect important and interpretable constructs [55].
- Trade-off between complexity and explainability: the more complex an algorithm is, the more accurate it will be, but the less easy it will be to understand and explain [76]. Some argue that an automated algorithm that makes its decisions based on input provided and considers dozens of factors simultaneously cannot be as simple to explain to another which employs just a few. This is why most machine learning-based algorithms are capable of easily finding patterns through associations rather than causal explanations [76]. Patterns between data, despite some shade of reality in business contexts, they would still raise issues from an ethical and legal standpoint, as they would favour or exclude individuals based on sensitive and protected characteristics. Causal reasoning, on the other hand, can be a significant challenge for organisations using such tools, as it can meet the requirements of fairness and explainability [62, 76]. Methodologically speaking, causal discovery is a rapidly developing technique that automates the empirical verification of causal hypotheses, narrowing down plausible causal models for consideration and decision-making [92].

- Lack of employees training on AI policies: the public's lack of access to all information regarding the use, design, bias reduction methods, and transparency of algorithms used by companies in selection processes is a major drawback for companies [71]. Without access to comprehensive information about AI tools, it becomes challenging to determine if an algorithm is fair or explainable when information is kept private. Sometimes, machine learning models may be inaccessible to the public due to legal reasons or at the companies' discretion to protect user data privacy in their selection processes, but valuable insights can still be gleaned from what companies have made publicly available. This includes insights into bias reduction and transparency [71]. Such analysis also sheds light on the issue of explainability. Specifically, machine learning techniques excel in identifying correlations between certain factors and outcomes, predicting how a subject might behave, which can be beneficial or detrimental to what the company seeks. However, the challenge arises when experts cannot explain why the presence or absence of a specific factor may impact future performance. For instance, why a candidate's tone of voice might affect their job performance [71]. This circles back to the previously discussed problem where the more precise an algorithm becomes, the less understandable its correlations are to users and experts alike due to the increasing number of variables considered [76]. Moreover, if experts cannot explain these correlations and outputs provided by the algorithm, it raises concerns that the algorithm may unknowingly use sensitive factors or data for its analysis, potentially compromising the fairness of the selection process. International public opinion has voiced numerous criticisms regarding the use of facial, vocal, and emotional analyses [99, 60] This makes it impossible for a human recruiter to determine if the AI-driven algorithm inadvertently learned sensitive characteristics, thereby compromising the fairness of its usage.
- Privacy concerns around employees data: employers today already possess a lot of user data, including mailing addresses, bank account details for salaries, CVs for the hiring process, and even medical details for requesting sick leave. Having all this data available allows employers to perform HR analytics without directly involving employees; for example, a group of employees might be considered at low risk of leaving the company, and consequently, policies for this group would be influenced by these analyses, such as smaller salary increases or less expensive training [100]. These operations should be condemned both ethically and legally as they obscure transparency in the employment relationship. Those who use AI systems to make predictions must always communicate which data they use, the purpose of such operations, and how the algorithm makes decisions, providing a comprehensible explanation for users. This principle allows employees to be protagonists in their own careers rather than passive subjects to company policies [100].

2.4. Human Oversight

This chapter of the review analyzes in detail the role of human control over AI technologies in HRM, especially its risks and positive outcomes. We focused on the reasons behind such intervention and we also explored the European AI Act as an innovative approach to strengthening this control.

2.4.1. The importance of human in the loop

Public fear often centers on AI-driven job displacement. However, AI in HR aims to foster human-machine collaboration [57, 83], not replacement. This collaboration improves HR processes such as recruitment, learning, and development by performing both "augmentation" and "automating" tasks [57]. More specifically, augmentation refers to human-machine collaboration in strategic decisions, while automation denotes full machine replacement of routine tasks [57]. Although this could theoretically displace some workers, it can also create new opportunities [101]. By easing the role of recruiters [78], AI allows them to focus on higher-value activities, proving how essential data-driven recruitment has become. Despite promising efforts, algorithms are not inherently neutral and can sometimes reflect the biases of their human creators, which justifies why automated decision-making in recruitment should

be avoided; human review, especially by bias-trained experts, is crucial [51, 69]. Although complete algorithmic neutrality is impossible, developers and recruiters must actively work to identify and mitigate biases through software testing [51]. Simply removing sensitive input data is insufficient due to proxy variables [102]. More complex debiasing methods compromise predictive effectiveness. Algorithmic software should support, not replace human recruiters, making human oversight crucial to protecting organizations and employees' rights [51, 78, 103]. We note the case of a multinational company's AI-based trainee selection application, which, despite being designed for fairness, incorporated human oversight at every stage. This prevented fully automated decisions and allowed for corrections [104]. Human intervention proved paramount, as the HR department identified repeated violations, such as the creation of multiple accounts to manipulate scores [78]. Various techniques to reduce implicit bias in recruitment have been presented [62]. Creating control and experimental groups with identical qualifications but different demographics allows biometric data collection, revealing recruiter biases through body language, speech, and stress levels [105, 106]. Technologies such as natural language processing can predict and mitigate biased decisions in real time [62], but are far from being reliably employed. A possibility would be to combine diverse human recruiters with AI agents through various algorithms [62] or even by providing concise explanations in ranking systems [103], showing why one item is ranked higher than another [107]. Although this helps decision-making, it is the fact that final decisions remain with human experts who consider context and nuances not captured by the algorithm that ensures ethical alignment [103]. As many have pointed out, despite some privacy and confidentiality concerns [62, 69], the importance of extensive data collection and analysis, not just for candidate evaluation [76], becomes evident for nuanced decision-making, to ensure diversity in hiring outcomes [79], as oversight processes that identify strong correlations between decisions and sensitive attributes (e.g., race) help mitigate bias before algorithm implementation [79].

3. Discussion

The intersection of AI and HRM is not only a technological advancement, but also an ethical and organizational challenge. The three pillars of fairness, explainability, and human oversight provide the framework to evaluate AI integration into HR processes. Although these pillars represent important ethical commitments, their enforcement raises critical questions about AI feasibility, implementation, and broader implications for both employees and employers. We argue that by critically analyzing these elements, both the strengths and blind spots in the literature and the practices surrounding AI-driven HRM can be revealed. From the reviewed studies, 42% primarily view AI in HRM as an opportunity, while 55% adopt a dual perspective, acknowledging both its potential and associated risks. A small minority of 3% focus exclusively on the threats AI poses to HR processes (Tab. 1).

Fairness is often heralded as a solution to systemic human biases in hiring, promotions, and performance assessments. The literature emphasizes the ability of AI to filter candidates objectively and minimize discriminatory tendencies inherent in human judgment [77, 62]. However, this optimism may be overstated. Bias in AI is not an exception, but an outcome of the systems it reflects. In particular, more than half of the reviewed studies emphasize the dual nature of AI in fairness, recognizing its ability to address systemic bias while simultaneously perpetuating inequities through historical data dependencies (Tab. 1). The reliance on historical data, which can encode past inequities, challenges the premise that AI is inherently more equitable than human decision-makers. For example, while the review acknowledges the value of inclusive language and CV screening algorithms [69, 76], it fails to adequately address how these tools may perpetuate structural biases in less visible forms, such as proxy discrimination [46]. Even seemingly objective criteria, such as education or experience, may inadvertently disadvantage underrepresented groups if these attributes correlate with historical inequities [66]. Moreover, the emphasis on fairness presumes that it is a static property to be designed into AI systems. In reality, fairness is a dynamic and context-dependent value. Organizations may prioritize fairness differently depending on their objectives, such as diversity, meritocracy, or efficiency [59, 73]. However, AI systems lack the adaptive capabilities to reconcile these competing values without

human intervention. For employees, this rigidity risks reducing complex identities to predefined metrics, while for employers, it limits the adaptability of HR strategies to address unique workforce challenges. Fairness, then, is not a guaranteed AI output, but a contested and negotiated process requiring sustained human oversight [82].

The second pillar of ethical AI usage, explainability, is positioned as a mechanism for transparency and trust in AI systems. The review rightly critiques the "black box" algorithms for their opacity, but its proposed solutions, such as ethical reporting and causal discovery techniques, fail to address the systemic barriers to explainability [97, 46]. The trade-off between algorithmic complexity and interpretability is more than a technical challenge; it reflects a broader tension between efficiency and accountability. Advanced machine learning models often produce outputs that are difficult to rationalize, not because of a lack of effort, but because their inner workings are designed to optimize predictive power, not clarity [98, 76]. This raises critical ethical concerns. Can HR decisions be considered fair or justifiable if their underlying processes are incomprehensible, even to experts? [100, 55]. Moreover, explainability efforts often focus on post hoc rationalizations rather than proactive transparency. This reactive approach risks reducing the explainability to a formality, satisfying regulatory requirements without truly addressing the ethical implications of opaque AI systems. For employees, this creates a disempowering environment in which decisions about their careers are made by systems they cannot question. For employers, it undermines the legitimacy of AI-driven decisions, exposing organizations to reputational and legal risks. The explainability must then be reimagined not as a technical feature but as an integral part of AI development, requiring a cultural change in how organizations design, deploy, and evaluate AI systems [72, 103].

Lastly, human oversight, presented as a safeguard against the excesses of autonomous AI systems, is itself a contested concept. The review lauds human intervention as a necessary counterbalance to algorithmic biases, but this perspective risks oversimplifying the complexities of human-AI interaction [78, 51]. Human oversight is not a panacea; it introduces its own challenges, including the potential for bias, fatigue, and overreliance on AI recommendations. For example, the concept of "automation bias" suggests that human reviewers can respect the output of AI even when it is flawed, affecting the very purpose of the oversight [57, 62]. Furthermore, the review's focus on technical oversight mechanisms, such as bias detection and audit processes, neglects the organizational and cultural dimensions of effective human-AI collaboration. Oversight is not merely about identifying errors, but about embedding ethical deliberation into decision-making processes. The prevalence of dual-oriented studies also reflects the critical role of human oversight in balancing AI efficiency with ethical considerations (Tab. 1. This requires training HR professionals not only in technical competencies but also in ethical reasoning and critical thinking. The absence of such training risks reducing oversight to a superficial practice, where human involvement is nominal rather than substantive [79, 76]. For employees, ineffective oversight erodes trust in the system, while for employers, it compromises the ethical and operational integrity of HRM practices. In addition, investing in comprehensive training and development programs, pursuing a stronger technological orientation [13, 108, 109] could reduce workforce stress [110]. More broadly, the emphasis on human oversight raises fundamental questions about the division of labor between humans and machines in HRM. Although the review highlights the potential for AI to augment rather than replace human decision making, it underestimates the organizational restructuring required to achieve this balance. Human oversight is not simply an add-on to existing HR processes; it requires a rethinking of roles, responsibilities, and workflows. For employees, this could mean greater participation in decision-making processes, fostering a sense of agency and inclusion. For employers, it offers an opportunity to align AI-driven processes with broader organizational values, but only if oversight is integrated thoughtfully and strategically [103, 100].

4. Conclusions

This paper contributes to the literature by providing a comprehensive review of articles discussing the implementation of AI-based decision systems in the HRM sector. Specifically, we structure our review

around the three main ethical pillars of fairness, explainability, and human oversight, ensuring a focus on the ethical considerations that arise when AI is applied to high-risk contexts such as recruitment and people management. We evaluate 31 articles from top-ranked journals, scoring them based on the level of detail and classifying their treatment of AI as either an opportunity, a threat, or both. Additionally, we highlight the necessity of fostering a critical understanding of AI-driven decision systems among HR professionals and stakeholders, ensuring that ethical considerations, transparency, and human oversight remain central to AI deployment in HRM.

Fairness, explainability, and human oversight are essential principles for the ethical deployment of AI in HRM, but their application is complex and requires careful consideration of their limitations. Fairness is a contested value shaped by organizational priorities, explainability demands cultural commitment beyond technical solutions, and human oversight necessitates a fundamental rethink of human-machine decision-making. While AI-driven HR systems hold the potential to improve efficiency and reduce biases, this promise can only be realized with robust governance, continuous monitoring, and cultural alignment within organizations to prevent reinforcing inequities.

Several critical gaps remain in the application of these principles. The operationalization of fairness requires further exploration to develop dynamic, context-sensitive frameworks [111]. Participatory design should be integrated into AI systems to enhance explainability, ensuring transparency is meaningful for all stakeholders [112]. Lastly, human oversight needs clearer practical guidelines, including how to effectively train HR professionals and balance human judgment with AI recommendations. Addressing these gaps through future research will help organizations more effectively integrate AI in HR while ensuring ethical practices and societal trust [113].

To address these shortcomings, future research should prioritize several key directions. First, longitudinal studies are needed to examine the long-term impacts of AI-driven HRM systems on organizational fairness, transparency, and employee trust. Such studies would provide empirical evidence to validate or challenge current assumptions about the efficacy of these systems. Second, interdisciplinary research involving computer science, organizational psychology, and ethics could yield innovative approaches to integrate fairness, explainability, and oversight into AI design. Third, comparative analyzes across industries and cultural contexts could illuminate how different organizational environments influence the ethical challenges and opportunities associated with AI in HRM. Finally, the regulatory landscape for AI in HRM requires further strengthening. Although frameworks such as the EU AI Act provide a starting point, research should examine how regulatory compliance can be balanced with innovation, particularly in resource-constrained settings. In addition, there is a need to investigate how global variations in regulatory standards affect the adoption and ethical deployment of AI in multinational organizations.

In summary, enforcing an ethical usage of AI in HRM is not merely a technical or procedural endeavor; it is a transformative process that requires organizations to reimagine their values, workflows, and relationships with technology. By addressing the critical gaps identified in this review and advancing new research directions, scholars and practitioners can ensure that AI serves as a tool for inclusion, accountability, and shared prosperity in the workplace. Only through such efforts can the full potential of AI-driven HRM be realized while protecting the dignity and rights of employees and the integrity of organizations.

Declarations

During the preparation of this work, the authors used GPT-4o and Writefull in order to: Grammar and spelling check. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. Guest, Human resource management and performance: Still searching for some answers, Human Resource Management Journal 21 (2010) 3 – 13. doi:10.1111/j.1748-8583.2010.00164.x.

[2] T. Keenoy, Human Resource Management, OUP, 2009, pp. 454–472. doi:10.1093/oxfordhb/9780199595686.013.0022.

[3] C. Ogbonnaya, J. Messersmith, Employee performance, well-being, and differential effects of human resource management subdimensions: Mutual gains or conflicting outcomes?, Human Resource Management Journal 29 (2018). doi:10.1111/1748-8583.12203.

[4] J. Liu, Y. Zhu, H. Wang, Managing the negative impact of workforce diversity: The important roles of inclusive hrm and employee learning-oriented behaviors, Frontiers in Psychology 14 (2023). doi:10.3389/fpsyg.2023.1117690.

[5] J. E. Delery, D. H. Doty, Modes of theorizing in strategic human resource management: Test of universalistic, contingency, and configurational performance predictions, Academy of Management Journal 39 (1996) 802–835. doi:10.2307/256713.

[6] J. P. Guthrie, High-involvement work practices, turnover, and prod-uctivity: Evidence from new zealand, Academy of Management Journal 44 (2001) 180–190. doi:10.2307/3069345.

[7] M. Huselid, The impact of human resource management practices on turnover, productivity, and corporate financial performance, Academy of Management Journal 38 (1995) 635–872. doi:10.5465/256741.

[8] M. Koch, R. McGrath, Improving labor productivity: Human resource management policies do matter, Strategic Management Journal 17 (1996) 335–354. URL: 10.1002/(SICI)1097-0266(199605)17:5<335::AID-SMJ814>3.0.CO;2-R.

[9] P. M. Wright, T. M. Gardner, L. M. Moynihan, The impact of HR practices on the performance of business units, Human Resource Management Journal 13 (2003) 21–37. doi:10.1111/j.1748-8583.2003.tb00096.x.

[10] I. Aust, B. Matthews, M. Muller-Camen, Common good hrm: A paradigm shift in sustainable hrm?, Human Resource Management Review 30 (2020) 100705. doi:https://doi.org/10.1016/j.hrmr.2019.100705, sustainable Human Resource Management and the Triple Bottom Line: Multi-Stakeholder Strategies, Concepts, and Engagement.

[11] N. T. Pham, T. H. Tuan, T. D. Le, P. N. D. Nguyen, M. Usman, G. T. C. Ferreira, Socially responsible human resources management and employee retention: The roles of shared value, relationship satisfaction, and servant leadership, Journal of Cleaner Production 414 (2023) 137704. doi:https://doi.org/10.1016/j.jclepro.2023.137704.

[12] P. Boselie, J. Paauwe, In search of balance - managing the dualities of hrm: An overview of the issues, Personnel Review 38 (2009) 461–471. doi:10.1108/00483480910977992.

[13] H. Ho, B. Kuvaas, Human resource management systems, employee well-being, and firm performance from the mutual gains and critical perspectives: The well-being paradox, Human Resource Management 59 (2019). doi:10.1002/hrm.21990.

[14] D. L. Deephouse, To be different, or to be the same? it's a question (and theory) of strategic balance., Strategic Management Journal 20 (1999) 147–166. doi:0.1002/(SICI)1097-0266(199902)20:2<147::AID-SMJ11>3.0.CO;2-Q.

[15] D. N. Den Hartog, P. Boselie, J. Paauwe, Performance management: A model and research agenda, Applied Psychology 53 (2004) 556–569. doi:10.1111/j.1464-0597.2004.00188.x.

[16] R. Kitchin, Big data, new epistemologies and paradigm shift, Big Data & Society 1 (2014) 1–12. doi:10.1177/2053951714528481.

[17] A. Matthews, Review of mark andrejevic (2020), Automated Media. Postdigit Sci Educ 4 (2022).

[18] V. Mayer-Schönberger, K. Cukier, Big Data: A Revolution that Will Transform how We Live, Work, and Think, Houghton Mifflin Harcourt, 2013.

[19] S. Jackson, R. Schuler, K. Jiang, An aspirational framework for strategic human resource management, The Academy of Management Annals 8 (2014). doi:10.1080/19416520.2014.872335.

[20] European Parliament and Council of the European Union, Regulation (EU) 2024/1689, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689, 2025. Article 3.

[21] H. Shipton, M. A. West, J. Dawson, K. Birdi, M. Patterson, Hrm as a predictor of innovation, Human Resource Management Journal 16 (2006) 3–27. doi:https://doi.org/10.1111/j.1748-8583.2006.00002.x.

[22] H. Ruël, T. Bondarouk, J. K. Looise, E-hrm: Innovation or irritation. an explorative empirical study in five large companies on web-based hrm, Management Revue - The international Review of Management Studies 15 (2004). doi:10.5771/0935-9915-2004-3-364.

[23] U. Association, E. Galanaki, L. Panayotopoulou, Adoption and Success of E-HRM in European Firms, Encyclopedia of Human Resources Information Systems: Challenges in e-HRM, 2011, pp. 948–955. doi:10.4018/978-1-60960-587-2.ch404.

[24] D. Oglesby, D. Boudreaux, K. Manix, D. Serviss, D. Hair, Ai in hr: Perception is reality, in: SIGMIS-CPR '24: 2024 Computers and People Research Conference, 2024, pp. 1–2. doi:10.1145/3632634.3655879.

[25] M. Groß, Yes, AI Can: The Artificial Intelligence Gold Rush Between Optimistic HR Software Providers, Skeptical HR Managers, and Corporate Ethical Virtues, Springer, 2021, pp. 191–225. doi:10.1007/978-3-030-66913-3_10.

[26] K. Voorde, R. Peccei, M. Veldhoven, HRM, well-being and performance: A theoretical and empirical review, Wiley-Blackwell, 2013, pp. 15–46.

[27] A. Malik, P. Thevisuthan, T. De Silva, Artificial Intelligence, Employee Engagement, Experience, and HRM, Springer, Cham, 2022, pp. 171–184. doi:10.1007/978-3-030-90955-0_16.

[28] M. Manfrino, Ai and employee wellbeing: Navigating human-centric integration, interactive coaching, and turnover mitigation, SSRN Electronic Journal (2024). doi:10.2139/ssrn.4821351.

[29] A. Gélinas, Daniel; Sadreddin, R. Vahidov, Artificial intelligence in human resources management: A review and research agenda, Pacific Asia Journal of the Association for Information Systems: 14: (2022). doi:DOI:10.17705/1pais.14601.

[30] S. Nishar, The role of artificial intelligence in transforming human resource management: A literature review, Journal of Artificial Intelligence & Cloud Computing (2023) 1–4. doi:10.47363/JAICC/2022(1)155.

[31] R. Tsiskaridze, K. Reinhold, M. Järvis, Innovating hrm recruitment: A comprehensive review of ai deployment, Marketing and Management of Innovations 14 (2023) 239–254. doi:10.21272/mmi.2023.4-18.

[32] E. Sipahi, E. Artantaş, Artificial intelligence in hrm, Handbook of Research on Innovative Management Using AI in Industry 5.0 (2022) 1–18.

[33] A. Alsaif, M. Sabih Aksoy, Ai-hrm: artificial intelligence in human resource management: a literature review, Journal of Computing and Communication 2 (2023) 1–7.

[34] M. El-Ghoul, M. M. Almassri, M. F. El-Habibi, M. H. Al-Qadi, A. Abou Eloun, B. S. Abu-Nasser, S. S. Abu-Naser, Ai in hrm: Revolutionizing recruitment, performance management, and employee engagement, International Journal of Academic Applied Research (Ijaar) (2024).

[35] S. Rao, J. Chitranshi, N. Punjabi, Role of artificial intelligence in employee engagement and retention, Journal of Applied Management-Jidnyasa (2020) 42–60.

[36] H. A. Alrakhawi, R. Elqassas, M. M. Elsobeihi, B. Habil, B. S. Abunasser, S. S. Abu-Naser, Transforming human resource management: The impact of artificial intelligence on recruitment and beyond, International Journal of Academic Applied Research (Ijaar) (2024).

[37] E. Farndale, J. Paauwe, Shrm and context: why firms want to be as different as legitimately possible, Journal of Organizational Effectiveness: People and Performance 5 (2018). doi:10.1108/JOEPP-04-2018-0021.

[38] S. Zuboff, Surveillance capitalism and the challenge of collective action, New Labor Forum 28 (2019) 10–29. doi:10.1177/1095796018819461.

[39] D. Boyd, K. Crawford, Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon, Information, Communication & Society 15 (2012) 662–679.

[40] G. George, A. Pentland, Big data and management, Academy of Management Journal 57 (2014) 321–326. doi:10.5465/amj.2014.4002.

[41] A. Castelnovo, Towards responsible ai in banking: Addressing bias for fair decision-making, arXiv preprint arXiv:2401.08691 (2024).

[42] H. Shipton, P. Sparrow, P. Budhwar, A. Brown, Hrm and innovation: looking across levels, Human Resource Management Journal 27 (2017) 246–263. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1748-8583.12102. doi:https://doi.org/10.1111/1748-8583.12102. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/1748-8583.12102.

[43] M. Biron, C. Boon, P. A. Bamberger, Human Resource Strategy: Formulation, Implementation, and Impact (2nd ed.), Routledge, 2014.

[44] The European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, 2021. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

[45] L. Nishii, D. Lepak, B. Schneider, Employee attributions of the "why" of hr practices: Their effects on employee attitudes and behaviors, and customer satisfaction, Working Papers 61 (2008). doi:10.1111/j.1744-6570.2008.00121.x.

[46] M. Raghavan, S. Barocas, J. Kleinberg, K. Levy, Mitigating bias in algorithmic hiring: evaluating claims and practices, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 469–481. URL: https://doi.org/10.1145/3351095.3372828. doi:10.1145/3351095.3372828.

[47] A. Hunkenschroer, C. Lütge, Ethics of ai-enabled recruiting and selection: A review and research agenda, Journal of Business Ethics 178 (2022). doi:10.1007/s10551-022-05049-6.

[48] S. K. Katyal, Private accountability in the age of artificial intelligence, UCLA L. Rev. 66 (2019) 54.

[49] G. Malgieri, F. Pasquale, From transparency to justification: Toward ex ante accountability for ai, SSRN Electronic Journal (2022). doi:10.2139/ssrn.4099657.

[50] European Commission, White Paper on Artificial Intelligence – a European approach to excellence and trust, https://eur-lex.europa.eu/EN/legal-content/glossary/white-paper.html, 2020. Archived from the original on 5 January 2024. Retrieved 6 January 2024.

[51] D. Bîgu, M.-V. Cernea, Algorithmic bias in current hiring practices: An ethical examination, in: 13th International Management Conference (IMC) on Management Strategies for High Performance, Bucharest, Romania, October, 2019.

[52] S. Chowdhury, P. Dey, S. Joel-Edgar, S. Bhattacharya, O. Rodriguez-Espindola, A. Abadie, L. Truong, Unlocking the value of artificial intelligence in human resource management through ai capability framework, Human Resource Management Review 33 (2023) 100899. URL: https://www.sciencedirect.com/science/article/pii/S1053482222000079. doi:https://doi.org/10.1016/j.hrmr.2022.100899.

[53] R. Deepa, S. Sekar, A. Malik, J. Kumar, R. Attri, Impact of ai-focussed technologies on social and technical competencies for hr managers – a systematic review and research agenda, Technological Forecasting and Social Change 202 (2024) 123301. URL: https://www.sciencedirect.com/science/article/pii/S0040162524000970. doi:https://doi.org/10.1016/j.techfore.2024.123301.

[54] C. Fernandez-Martinez, A. Fernandez, Ai and recruiting software: Ethical and legal implications, Paladyn, Journal of Behavioral Robotics 11 (2020) 199–216.

[55] M. F. Gonzalez, J. F. Capman, F. L. Oswald, E. R. Theys, D. L. Tomczak, "where's the io?" artificial intelligence and machine learning in talent management systems, Personnel Assessment and Decisions 5 (2019) 5.

[56] N. Haefner, J. Wincent, V. Parida, O. Gassmann, Artificial intelligence and innovation management: A review, framework, and research agenda, Technological Forecasting and Social Change 162 (2021) 120392. URL: https://www.sciencedirect.com/science/article/pii/S004016252031218X. doi:https://doi.org/10.1016/j.techfore.2020.120392.

[57] A. Hemalatha, P. B. Kumari, N. Nawaz, V. Gajenderan, Impact of artificial intelligence on

recruitment and selection of information technology companies, in: 2021 international conference on artificial intelligence and smart systems (ICAIS), IEEE, 2021, pp. 60–66.

[58] M. Jatobá, J. Santos, I. Gutierriz, D. Moscon, P. O. Fernandes, J. P. Teixeira, Evolution of artificial intelligence research in human resources, Procedia Computer Science 164 (2019) 137–142. URL: https://www.sciencedirect.com/science/article/pii/S1877050919322045. doi:https://doi.org/10.1016/j.procs.2019.12.165, cENTERIS 2019 - International Conference on ENTERprise Information Systems / ProjMAN 2019 - International Conference on Project MANagement / HCist 2019 - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2019.

[59] P. T. Kim, S. Scott, Discrimination in online employment recruiting, . Louis ULJ 63 (2018) 93.

[60] J. Kleinberg, J. Ludwig, S. Mullainathan, C. R. Sunstein, Discrimination in the age of algorithms, Journal of Legal Analysis 10 (2018) 113–174.

[61] M. K. Lee, Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management, Big Data & Society 5 (2018) 2053951718756684.

[62] Y.-T. Lin, T.-W. Hung, L. T.-L. Huang, Engineering equity: How ai can help reduce the harm of implicit bias, Philosophy & Technology 34 (2021) 65–90.

[63] M. Madanchian, H. Taherdoost, N. Mohamed, Ai-based human resource management tools and techniques; a systematic literature review, Procedia Computer Science 229 (2023) 367–377. URL: https://www.sciencedirect.com/science/article/pii/S187705092302029X. doi:https://doi.org/10.1016/j.procs.2023.12.039, 12th International Young Scientists Conference in Computational Science, YSC2023.

[64] P. Martín-Hernández, Artificial intelligence: The present and future of human resources recruitment and selection processes, Engineering Proceedings 56 (2023) 188.

[65] P. Mikalef, M. Gupta, Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance, Information & Management 58 (2021) 103434. URL: https://www.sciencedirect.com/science/article/pii/S0378720621000082. doi:https://doi.org/10.1016/j.im.2021.103434.

[66] D. F. Mujtaba, N. R. Mahapatra, Ethical considerations in ai-based recruitment, in: 2019 IEEE International Symposium on Technology and Society (ISTAS), IEEE, 2019, pp. 1–7.

[67] A. B. P.R. Palos-Sánchez, P. Baena-Luna, J. Infante-Moro, Artificial intelligence and human resources management: A bibliometric analysis, Applied Artificial Intelligence 36 (2022) 2145631. URL: https://doi.org/10.1080/08839514.2022.2145631. doi:10.1080/08839514.2022.2145631. arXiv:https://doi.org/10.1080/08839514.2022.2145631.

[68] A. Pena, I. Serna, A. Morales, J. Fierrez, Bias in multimodal ai: Testbed for fair automatic recruitment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 28–29.

[69] A. Persson, Implicit bias in predictive data profiling within recruitments, Privacy and Identity Management. Facing up to Next Steps: 11th IFIP WG 9.2, 9.5, 9.6/11.7, 11.4, 11.6/SIG 9.2. 2 International Summer School, Karlstad, Sweden, August 21-26, 2016, Revised Selected Papers 11 (2016) 212–230.

[70] V. Prikshat, M. Islam, P. Patel, A. Malik, P. Budhwar, S. Gupta, Ai-augmented hrm: Literature review and a proposed multilevel framework for future research, Technological Forecasting and Social Change 193 (2023) 122645. URL: https://www.sciencedirect.com/science/article/pii/S004016252300330X. doi:https://doi.org/10.1016/j.techfore.2023.122645.

[71] M. Raghavan, S. Barocas, J. Kleinberg, K. Levy, Mitigating bias in algorithmic hiring: Evaluating claims and practices, in: Proceedings of the 2020 conference on fairness, accountability, and transparency, 2020, pp. 469–481.

[72] W. Rodgers, J. M. Murray, A. Stefanidis, W. Y. Degbey, S. Y. Tarba, An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes, Human resource management review 33 (2023) 100925.

[73] J. Sánchez-Monedero, L. Dencik, L. Edwards, What does it mean to'solve'the problem of discrimination in hiring? social, technical and legal perspectives from the uk on automated hiring

systems, in: Proceedings of the 2020 conference on fairness, accountability, and transparency, 2020, pp. 458–468.
[74] C. Schumann, J. Foster, N. Mattei, J. Dickerson, We need fairness and explainability in algorithmic hiring, in: International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2020.
[75] S. Strohmeier, F. Piazza, Artificial Intelligence Techniques in Human Resource Management—A Conceptual Exploration, Springer International Publishing, Cham, 2015, pp. 149–172. URL: https://doi.org/10.1007/978-3-319-17906-3_7. doi:10.1007/978-3-319-17906-3_7.
[76] P. Tambe, P. Cappelli, V. Yakubovich, Artificial intelligence in human resources management: Challenges and a path forward, California Management Review 61 (2019) 15–42.
[77] E. Van Den Broek, A. Sergeeva, M. Huysman, Hiring algorithms: An ethnography of fairness in practice, in: 40th international conference on information systems, ICIS 2019, Association for Information Systems, 2020, pp. 1–9.
[78] P. Van Esch, J. S. Black, Factors that influence new generation candidates to engage with and complete digital, ai-enabled recruiting, Business horizons 62 (2019) 729–739.
[79] M. Vasconcelos, C. Cardonha, B. Gonçalves, Modeling epistemological principles for bias mitigation in ai systems: an illustration in hiring decisions, in: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 2018, pp. 323–329.
[80] B. A. Williams, C. F. Brooks, Y. Shmargad, How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications, Journal of Information Policy 8 (2018) 78–115.
[81] L. Yarger, F. Cobb Payton, B. Neupane, Algorithmic equity in the hiring of underrepresented it job candidates, Online information review 44 (2020) 383–395.
[82] S. Delecraz, L. Eltarr, M. Becuwe, H. Bouxin, N. Boutin, O. Oullier, Responsible artificial intelligence in human resources technology: An innovative inclusive and fair by design matching algorithm for job recruitment purposes, Journal of Responsible Technology 11 (2022) 100041.
[83] O. Ore, M. Sposato, Opportunities and risks of artificial intelligence in recruitment and selection, International Journal of Organizational Analysis 30 (2022) 1771–1782.
[84] J. Dastin, Amazon scraps secret ai recruiting tool that showed bias against women, in: Ethics of data and analytics, Auerbach Publications, 2022, pp. 296–299.
[85] European Parliament and Council of the European Union, Regulation (EU) 2024/1689, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689, 2024. Official Journal of the European Union.
[86] C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, J. Handelsman, Science faculty's subtle gender biases favor male students, Proceedings of the national academy of sciences 109 (2012) 16474–16479.
[87] M. Bertrand, S. Mullainathan, Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination, American economic review 94 (2004) 991–1013.
[88] D. Neumark, I. Burn, P. Button, Age discrimination and hiring of older workers, Age 6 (2017) 1–5.
[89] E. Rosenbaum, Silicon valley is stumped: Even ai cannot always remove bias from hiring. cnbc, 2018.
[90] J. C. Meister, K. Mulcahy, The Future Workplace Experience: 10 Rules For Mastering Disruption in Recruiting and Engaging Employees, 1st edition ed., McGraw Hill, 2016.
[91] E. A. Lind, K. Van den Bos, When fairness works: Toward a general theory of uncertainty management, Research in organizational behavior 24 (2002) 181–223.
[92] D. Malinsky, D. Danks, Causal discovery algorithms: A practical guide, Philosophy Compass 13 (2018) e12470.
[93] J. R. Loftus, C. Russell, M. J. Kusner, R. Silva, Causal reasoning for algorithmic fairness, arXiv preprint arXiv:1805.05859 (2018).
[94] C. O'neil, Weapons of math destruction: How big data increases inequality and threatens democracy, Crown, 2017.
[95] M. Brownstein, The implicit mind: Cognitive architecture, the self, and ethics, Oxford University Press, 2018.

[96] E. Parliament, C. of the European Union, Regulation (eu) 2016/679, general data protection regulation (gdpr), Official Journal of the European Union, 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679.

[97] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Comput. Surv. 51 (2018). URL: https://doi.org/10.1145/3236009. doi:10.1145/3236009.

[98] A. Adadi, M. Berrada, bipeeking inside the black-box: a survey on explainable artificial intelligence (xai), IEEE access 6 (2018) 52138–52160.

[99] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, S. D. Pollak, Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements, Psychological science in the public interest 20 (2019) 1–68.

[100] K. Simbeck, Hr analytics and ethics, IBM Journal of Research and Development 63 (2019) 9–1.

[101] M. R. Frank, D. Autor, J. E. Bessen, E. Brynjolfsson, M. Cebrian, D. J. Deming, M. Feldman, M. Groh, J. Lobo, E. Moro, et al., Toward understanding the impact of artificial intelligence on labor, Proceedings of the National Academy of Sciences 116 (2019) 6531–6539.

[102] M. Veale, F. Zuiderveen Borgesius, Demystifying the draft eu artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach, Computer Law Review International 22 (2021) 97–112.

[103] A. Castelnovo, R. Crupi, N. Mombelli, G. Nanino, D. Regoli, Evaluative item-contrastive explanations in rankings, Cognitive Computation (2024) 1–16.

[104] G. S. Levanthal, What should be done with equity theory, Social exchange: Advances in theory and research (1980) 27–55.

[105] C. Clabaugh, M. Matarić, Robots for the people, by the people: Personalizing human-machine interaction, Science robotics 3 (2018) eaat7451.

[106] C. K. Lai, M. R. Banaji, The psychology of implicit intergroup bias and the prospect of change, Difference without domination: Pursuing justice in diverse democracies (2020) 151–173.

[107] T. Miller, Explainable ai is dead, long live explainable ai! hypothesis-driven decision support using evaluative ai, in: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, ACM, 2023, pp. 333–342.

[108] D. Mihail, P. Kloutsiniotis, The impact of high-performance work systems on greek hospital employees' work-related well-being and job burnout, in: British Academy of Management 2016 (BAM2016), 2016.

[109] D. L. Stone, D. L. Deadrick, K. M. Lukaszewski, R. Johnson, The influence of technology on the future of human resource management, Human Resource Management Review 25 (2015) 216–231. URL: https://www.sciencedirect.com/science/article/pii/S1053482215000030. doi:https://doi.org/10.1016/j.hrmr.2015.01.002, human Resource Management: Past, Present and Future - Volume 2.

[110] A. Jaiswal, S. Sengupta, M. Panda, L. Hati, V. Prikshat, P. Patel, S. Mohyuddin, Teleworking: role of psychological well-being and technostress in the relationship between trust in management and employee performance, International Journal of Manpower 45 (2022). doi:10.1108/IJM-04-2022-0149.

[111] A. Castelnovo, N. Inverardi, G. Nanino, I. G. Penco, D. Regoli, Fair enough? a map of the current limitations of the requirements to have" fair"algorithms, arXiv preprint arXiv:2311.12435 (2023).

[112] T. A. R. Sure, Human-computer interaction techniques for explainable artificial intelligence systems, Research & Review: Machine Learning and Cloud Computing 3 (2024) 1–7. doi:10.46610/RTAIA.2024.v03i01.001.

[113] S. Bankins, The ethical use of artificial intelligence in human resource management: a decision-making framework, Ethics and Information Technology 23 (2021). doi:10.1007/s10676-021-09619-6.