Marketing & Menedzsment
The Hungarian Journal of Marketing and Management

A Framework for Integrating AI-Powered Systems to Mitigate Bias Risk in HRM Functions

Rawia Naoum
University of Pécs

https://doi.org/10.15170/MM.2025.59.02.05

THE AIM OF THE PAPER

This paper investigates the dual role of artificial intelligence (AI) in human resource management (HRM), assessing its capacity to both perpetuate and mitigate biases, particularly within the framework of diversity, equity, and inclusion (DEI). The goal is to identify the sources of AI-induced biases in HRM and establish a strategic framework to effectively reduce these biases.

METHODOLOGY

This paper conducts a thorough literature assessment on the application of AI in HRM and its implications on DEI efforts. The review draws on a wide range of academic sources, including major databases like Web of Science and Google Scholar, to isolate the mechanisms through which AI tools introduce biases in HR functions such as recruitment, performance assessment, and compensation.

MOST IMPORTANT RESULTS

The investigation highlights that AI-induced biases in HRM are mainly attributable to three sources: human influence in the design and operation of AI tools, biases inherent in the datasets employed for training AI, and the algorithms' intrinsic biases. These elements collectively contribute to reinforcing discriminatory practices within organizations, thereby impeding DEI initiatives.

RECOMMENDATIONS

To combat these biases, the paper proposes a robust framework encompassing four key strategies: enhancing AI and DEI literacy among HR professionals, adopting inclusive design practices for AI tools, ensuring accountability in dataset management, and promoting transparency in AI implementations. These measures aim to equip organizations with the tools to integrate AI into HRM in a manner that fosters an inclusive, fair, and equitable workplace environment.

Keywords: Artificial intelligence, Human resources management (HRM), Diversity, Equity and Inclusion (DEI), Bias

INTRODUCTION

Artificial intelligence (AI) is revolutionizing human resource management (HRM), automating tasks from recruitment to performance management (Singh & Doval 2019). While AI-driven tools promise efficiency and strategic decision-making (Murugesan et al. 2023), their impact on diversity, equity, and inclusion (DEI) remains a critical concern. Though DEI initiatives are proven to boost innovation, productivity, and employee performance (Singha & Prakasam 2014), integrating DEI holistically remains challenging (Kochan et al. 2003; Shen et al., 2009 – a challenge compounded by AI's potential to perpetuate biases. Existing research highlights AI's role in HRM functions such as talent acquisition (Oswald et al. 2020), employee onboarding (Brown 2024), performance management (Charbonneau & Doberstein 2020; Cappelli et al. 2019b; West 2018), training and development (Chandar et al. 2017; Schweyer 2018), compensation (Cappelli et al. 2019b; Johnson et al. 2022), and turnover prediction (Schweyer 2018). While digital tools enhance operational effectiveness (Akter et al. 2020; Aström et al. 2022; Kumar et al. 2022), reduce HRM accuracy risks (Bresciani et al. 2021a, 2021b; Li et al. 2023) and optimize decision-making (Lindebaum et al. 2020; Rana et al. 2022), the domain of AI in HRM remains nascent, with a stark gap between its potential and practice (Cappelli et al. 2019a; Strohmeier & Piazza 2015). The gaps persist in understanding how this paper consequently narrows its focus to *AI-induced biases* rather than broader DEI challenges.

Research questions are:

(1) What bias sources arise from AI-HRM tools?

(2) How can organizations mitigate these biases?

By the end of the paper, we propose a conceptual framework addressing bias mitigation through four pillars: AI upskilling, inclusive design, dataset accountability, and transparency. Synthesizing literature on ethical AI and DEI, this work guides researchers and practitioners in aligning AI-HRM systems with equitable outcomes, ensuring technological advancements do not come at the cost of inclusivity.

METHODOLOGY

A structured narrative literature review methodology was adopted to explore how AI integration in Human Resource Management (HRM) may introduce biases, particularly in relation to Diversity, Equity, and Inclusion (DEI). This approach was chosen for its flexibility and suitability for conceptual and integrative research, allowing for a broad, theory-driven synthesis without the rigid constraints of a systematic review protocol. The review aimed to synthesize existing knowledge, identify conceptual and empirical gaps, and analyze how AI-powered HRM tools may bias or enhance organizational DEI practices. Searches were conducted across major academic databases, including Web of Science, Google Scholar, JSTOR, and IEEE Xplore, using combinations of keywords such as: "artificial intelligence", "human resource management", "AI bias", "DEI in HRM", "algorithmic fairness", and "bias mitigation in AI". Boolean operators were applied to refine results. Inclusion criteria focused on peer-reviewed journal articles and high-quality academic conference papers published between 2015 and 2024, addressing AI applications in HRM, DEI outcomes, or bias mitigation strategies. Studies that did not meaningfully address both AI and HRM contexts or were unavailable in full text were excluded. Non-scholarly sources (e.g., blogs, corporate booklets, etc.) were excluded to maintain academic integrity. This paper is conceptual in nature and does not aim to offer an exhaustive systematic review. Future research may empirically validate the framework or conduct a systematic synthesis of the literature on AI bias in HRM systems.

LITERATURE REVIEW

Recent decades have witnessed accelerated advancements in artificial intelligence (AI), with cross-disciplinary innovations like machine learning and big data analytics transforming business operations (Verma et al. 2021). In human resource management (HRM), AI promises enhanced efficiency and objectivity in talent decisions – from recruitment to performance evaluation (Hmoud & Laszlo 2019; Borry & Getha-Taylor 2018). However, this „*new service delivery regime*" (Giest & Klievink 2022, 2) introduces ethical paradoxes: algorithmic „black boxes" obscure decision logic (Chen 2023; Johnson et al. 2022), while claims of reduced human bias clash with evidence of amplified discrimination. For instance, Amazon abandoned gender-biased AI recruitment tools in 2018 after systematically excluding female applicants (Tambe et al. 2019). Such cases expose how AI-HRM systems risk codifying organizational prejudices (Drage & Mackereth 2022), complicating diversity, equity, and inclusion (DEI) efforts. Critical scholarship urges proactive identification of bias sources – whether in training data, design assumptions, or feedback loops – to mitigate discrimination risks in AI-powered HR tools. The subsequent sections examine the sources of biases and the strategies for their mitigation.

Source of Biases of AI-powered Systems

Human bias:

A recent systematic review (Kekez et al. 2025) across 64 papers found that many AI-HRM studies fail to define key terms like "bias" or "discrimination", and tend to focus disproportionately on gender and race bias, leaving intersectional and contextual forms underrepresented. However, based on Belenguer's 2022 definition, AI bias occurs when the outcomes of AI-powered systems discriminate against certain groups of people. Human cognitive biases fundamentally shape the development of AI-powered tools, embedding systemic inequities into systems marketed as „objective" (Letheren et al. 2020; Köchling & Wehner 2022). These biases manifest most critically through three groups of actors in the AI-HRM pipeline: (i) AI engineers, developers and designers; (ii) biases of domain experts; and (iii) biases from users and end users of AI-powered tools.

The first actors are AI developers or engineers, whose cognitive biases directly influence how data is processed and algorithms are coded, perpetuating discrimination (Chen 2023). For example, *status quo bias* compels designers to uncritically adopt historical datasets that reflect existing inequalities, reinforcing discriminatory patterns (Malin et al. 2024). Similarly, *in-group bias* prioritizes homogeneous team perspectives during development, stifling diverse input and innovation (Postmes et al. 2001). These biases intersect with hierarchical power dynamics: Bonilla-Silva's (2001) analysis of systemic exclusion – originally critiquing racialized hierarchies – highlights how elite minorities, even unconsciously, design strategies that restrict marginalized groups' access to privileges. Applied to AI-HRM, these dynamic risk encoding tools favor dominant demographics (e.g., privileging male-coded resumes in hiring algorithms), exacerbating inequity. Even well-intentioned developers, constrained by heuristics and flawed assumptions, create "short-sighted" AI solutions that amplify bias in HRM processes such as hiring or promotions (Myllylä 2022; Njoto 2020; Gulati et al. 2022).

The second actors are "domain experts", described as specialists possessing "*expert knowledge and experience in the field of application of the AI system*" (Lockey et al., 2021, 5468). In the context of AI-HRM, these experts are HR professionals tasked with aligning AI tools with organizational objectives (e.g., talent acquisition, performance management). While HR teams possess institutional knowledge of compliance and workforce dynamics, their limited technical literacy often hinders critical scrutiny of algorithmic outputs. For example, HR experts may prioritize operational efficiency (e.g., faster resume screening) over auditing how AI systems replicate historical biases (e.g., penalizing non-traditional career paths). This oversight entrenches inequity, as HR's validation of AI tools becomes a rubber stamp for biased design.

The third group of actors comprises the end-users. Here biases arise from the interactions between individuals and AI tools, where users – defined as those "*directly influenced by AI decisions*" (Lockey et al., 2021, 5468) – unintentionally skew system outputs. For example, social desirability bias compels users to submit responses aligned with societal expectations rather than authentic preferences, corrupting training data (Gulati et al. 2022). This manifests as content production bias: lexical, syntactic, or semantic distortions in user-generated data that entrenches stereotypes (Olteanu et al. 2019). User trust further complicates this dynamic. Inexperienced users may over-rely on AI recommendations (e.g., accepting flawed performance evaluations) or dismiss valid outputs due to skepticism (Adadi & Berrada 2018; Andras et al. 2018). When end-user behavior diverges from organizational intent – such as job applicants over-editing resumes to "game" AI screening tools – systems amplify disparities, disproportionately disadvantaging marginalized groups (Weller 2017).

Dataset Bias:

Dataset bias operates on a foundational axiom: biased inputs perpetuate biased outputs (Chen 2023; Huang & Rust 2021). When AI algorithms are trained on unrepresentative or historically skewed data – common in HRM systems – they codify and amplify societal inequities (Köchling et al. 2021). Historical bias entrenches past discrimination, such as talent acquisition tools favoring resumes of historically overrepresented groups (e.g., white males) and systematically excluding marginalized applicants (Chen 2023; Marabelli & Lirio 2025). Geographic or behavioral data that reflects classist or ableist standards can distort workforce analytics (Belenguer 2022). Representation bias compounds these risks through demographic over- or over-/underrepresentation. Facial recognition systems trained predominantly on lighter-skinned males misclassify darker-skinned females (Buolamwini & Gebru 2018), while personality assessments using Western-centric norms alienate non-Western candidates (Chen 2023; Suresh & Guttag 2021). These distortions create self-fulfilling prophecies: biased outputs (e.g., hiring recommendations) reinforce skewed training data, perpetuating exclusion (Köchling et al. 2021).

2025 / 2
Rawia Naoum: A Framework for Integrating AI-Powered Systems to Mitigate Bias Risk in HRM Functions
https://doi.org/10.15170/MM.2025.59.02.05

Algorithmic bias:
Algorithmic bias originates not only from input data but also from the design and logic of the algorithm itself (Baeza-Yates 2018). Defined as a "*finite, abstract, effective control structure*" (Hill 2016, 47), algorithms are inherently biased, though this opacity often remains hidden from users and organizations (Chen 2023; Walsh et al. 2020). Design choices codify exclusion into AI systems, such as variable selection, model architecture, or optimization goals. For instance, oversimplified relationships between variables create bias-variance trade-offs, where models prioritize familiar data patterns over adaptability, failing generalizability (Geman et al. 1992). This ambiguity perpetuates harm: algorithmic "black boxes" obscure whether bias stems from data or processing, leaving users unaware of how or why they face discrimination (Johnson et al. 2022). Compounding this issue, developers often lack frameworks to audit algorithmic outputs, necessitating human intervention (e.g., domain experts stress-testing tools against diverse scenarios) (Baeza-Yates 2018). Transparency – such as explainable decision pathways – is critical to remedying errors and rebuilding trust, yet remains aspirational in most AI-HRM applications (Shin & Park 2019).

Gaps in the literature

Despite growing awareness of AI-related bias, significant gaps remain in the literature. Many studies address technical aspects of bias (e.g., developing "fairness-aware" algorithms) without fully considering the organizational and human context (Selbst et al. 2019). Current frameworks and best practices often focus on narrow fixes in specific domains (such as bias mitigation in criminal justice algorithms or hiring tools) but remain reactive and siloed, neglecting deeper structural and cultural biases in organizations. For example, bias audits or adjustments are frequently performed after deployment – *after* harm has occurred – rather than embedding equity throughout the AI system lifecycle. Moreover, research explicitly linking AI ethics with DEI outcomes in HRM is scarce (Shams et al. 2023), especially in domestic contexts. This leaves a blind spot regarding how human biases interact with AI systems in workplaces (Gulati et al. 2022; Rastogi et al. 2022) and how organizations can proactively transform AI from a bias amplifier into an equity enabler.

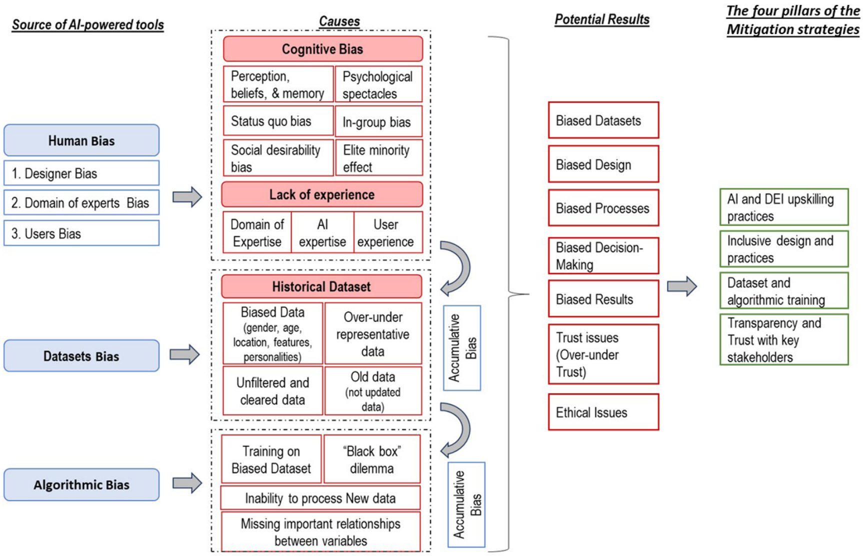
Our proposed framework addresses these gaps by integrating multidisciplinary strategies into a cohesive model. Unlike prior work that tends to tackle one dimension of the problem (for instance, improving data fairness or increasing algorithmic transparency in isolation), this framework concurrently targets multiple bias sources – human, data, and algorithmic – through four interlocking pillars. By combining technical solutions with HRM and DEI practices, the framework bridges AI ethics and HRM theory; it emphasizes not just algorithm performance, but also organizational processes (like upskilling and inclusive design) and ethical governance (like transparency and accountability) necessary for sustainable bias mitigation. The following section introduces these four pillars in detail and illustrates how each contributes to closing the identified gaps in the literature.

AI Bias mitigation practices in the context of HRM – Proposed Framework

Based on our review of the literature and identified gaps, we propose a four-pillar framework to mitigate AI-driven bias in HRM. The pillars – AI & DEI Upskilling, Inclusive Design, Dataset Accountability, and Transparency – address the human, technical, and organizational factors that contribute to AI bias. Figure 1. illustrates how these pillars interconnect to reduce bias across the AI-HRM pipeline.

Pillar 1: AI and DEI upskilling practices
Integrating AI into HRM demands radical upskilling of HR teams to mitigate systemic bias risks. Communication competencies are paramount: HR professionals must translate AI-derived metrics into actionable insights for stakeholders, while bridging gaps between technical vendors and organizational needs (Sakka et al. 2022; Margherita 2021). Without this dual fluency – negotiating vendor contracts *and* demystifying algorithmic outputs – miscommunication entrenches distrust and operational paralysis. Equally critical is strategic reskilling. HR teams require training in AI literacy to oversee tools ethically (Johnson et al. 2022; Brown 2024), while employees need foundational AI competencies to engage with systems confidently (Arslan et al. 2022; Gillath et al. 2021). Failure here risks organizational inertia, as underprepared workforces reject or misuse AI tools (Webster & Ivanov 2019). Finally, DEI-centric upskilling is non-negotiable. HR must lead training on cultural competencies, contextualizing how data demographics (e.g., geographic, racial) shape algorithmic biases (Dankwa-Mullan & Weeraratne 2022; Roopaei et al. 2021). Without this, AI-HRM systems risk automating exclusion under the excuse of objectivity. A 2022 survey of 191 Hungarian HR managers revealed generally positive attitudes toward AI in HRM, but noted mixed emotions and concerns about readiness – underscoring the importance of upskilling and trust-building in domestic HR units (Karácsony 2022).

Figure 1. Conceptual framework synthesizing literature on human, data, and algorithmic bias sources and the four pillars of mitigation (AI upskilling, inclusive design, dataset accountability, transparency)



Source: own illustration by synthesizing the literature

Pillar 2: Inclusive design and practices

Inclusive design practices are non-negotiable for mitigating systemic bias in AI-HRM systems. Diverse development teams – spanning race, gender, disability, and expertise – improve the detection of algorithmic inequities by integrating multifaceted perspectives on fairness (Bellamy et al. 2018; Selbst et al. 2019). This diversity is not optional: homogeneous teams replicate blind spots, while inclusive recruitment for AI roles disrupts unconscious biases entrenched in code (Dankwa-Mullan & Weeraratne 2022; Huang & Rust 2021; Jora et al. 2022; Roopaei et al. 2021). User-centered participatory design further disrupts exclusion. Involving marginalized groups (e.g., disabled individuals and elderly populations) in system development fosters trust and ensures training data reflects human diversity (Huang & Liem 2022; Srinivasan & Chander 2021). For HRM, this means co-designing tools with employees from underrepresented demographics to preempt discriminatory outcomes. Finally, algorithmic integration of inclusive practices – such as bias-aware machine learning frameworks – must align technical design with equity goals (Li et al. 2020; Nyariro et al. 2023). Without this, AI-HRM systems risk automating exclusion under the guise of neutrality.

Pillar 3: Dataset and algorithmic training

Mitigating algorithmic bias demands proactive dataset accountability. As Chen (2023, 7) asserts, "*unfair datasets are the root cause of bias*", necessitating rigorous audits to reconfigure skewed data and eliminate legacy prejudices (e.g., gender and racial markers). Organizations must prioritize transparent data collection, balancing representativeness with employee privacy safeguards while diversifying data points to counteract historical inequities (Chen 2023). Human-AI collaboration is critical; cross-functional teams – spanning HR, data scientists, and policymakers – must audit algorithms iteratively to identify and correct bias patterns (Budhwar et al. 2023; Köchling & Wehner 2022). For instance, Microsoft reduced facial recognition errors for darker-skinned women by 20-fold through dataset balancing (Grabovskyi & Martynovych 2019), illustrating how inclusive data practices yield measurable equity gains. Finally, bias-neutralizing technical interventions – such as masking demographic data in resumes (Jora et al. 2022) – must align with HRM ethics. These strategies are not optional but foundational to preventing AI-HRM systems from automating exclusion under the excuse of "neutrality".

Pillar 4: Transparency and Trust

A quantitative study across 580 employees in Hungarian service firms found that robust diversity management significantly enhances employee

2025 / 2
Rawia Naoum: A Framework for Integrating AI-Powered Systems to Mitigate Bias Risk in HRM Functions
https://doi.org/10.15170/MM.2025.59.02.05

engagement, with organizational trust acting as a strong mediator and reduced job insecurity reinforcing engagement outcomes (Alshaabani et al. 2021). Therefore, ethical AI integration in HRM demands moving beyond surface-level “fairness” to confront systemic inequities. While trust, transparency, and bias dominate discourse (Shams et al. 2023; Lee 2018; Lockey et al. 2021; Webster & Ivanov 2019; Yen & Chiang 2021), current frameworks inadequately address structural exclusion. “Fairness-aware” algorithms (Selbst et al. 2019) and sector-specific fixes (e.g., justice, recruitment) (Bellamy et al. 2018) remain reactive, neglecting diversity and inclusion (D&I) as ethical need (Saheb 2023; Zowghi & Rimini 2023). Stakeholder engagement is pivotal; HRM-AI systems require collaboration across employees, developers, and policymakers to anticipate harm and foster trust (Budhwar et al. 2023; Arslan et al. 2022). For instance, Microsoft’s AI ethics committee (Borry 2020) exemplifies corporate accountability, yet such initiatives are exceptions, not norms. Policymakers must enforce DEI-centric regulations – mandating transparency in hiring algorithms or bias audits (Huang & Rust 2021) – to dismantle the myth of “neutral” AI. Despite the urgency, research on D&I in AI ethics remains scarce (Shams et al. 2023), risking HRM’s complicity in automating inequity.

CONCLUSION

This paper presented a conceptual four-pillar framework designed to mitigate AI bias in HRM by addressing the interconnected sources: human, dataset, and algorithmic bias. Human bias stems from cognitive limitations and experiential gaps among designers, HR professionals, and end-users, which can perpetuate inequities throughout the design, deployment, and interpretation of AI systems (Gulati et al. 2022). Dataset bias reinforces historical discrimination through the use of unrepresentative or outdated data, while algorithmic bias – exacerbated by opaque logic and flawed variable relationships – intensifies exclusion (Rastogi et al. 2022; Rich & Gureckis 2019). These biases accumulate over time, leading to outcomes such as biased hiring practices, diminished trust in AI decisions, and ethical dilemmas that ultimately undermine DEI efforts.

Our proposed framework advocates a multi-faceted approach to counter these issues: AI and DEI upskilling to empower HR teams with technical and cultural competence; inclusive design practices co-developed with diverse stakeholders; dataset accountability via regular audits and data governance; and transparency mechanisms including policy-driven disclosure and ethical oversight of AI decision-making. Collectively, these strategies dismantle the misconception of AI neutrality and respond to the gaps identified in current literature. By implementing these pillars, organizations can transform AI-HRM systems from potential bias amplifiers into tools for equity promotion.

In doing so, this work advances the conversation at the intersection of AI ethics, HRM theory, and DEI practice. It provides a blueprint for HR professionals and organizations to follow, ensuring that technological advancements do not come at the cost of inclusivity and fairness. In essence, the framework positions HRM not just as a consumer of AI technologies, but as a guardian of ethical AI use in organizations, championing transparency, accountability, and continuous learning in the pursuit of equitable outcomes.

FUTURE RESEARCH DIRECTIONS

The literature highlights a research gap in understanding how human biases interact with AI systems – a blind spot often neglected by developers and researchers (Gulati et al. 2022; Rastogi et al. 2022; Rich & Gureckis 2019). Therefore, empirical research is urgently needed to explore these dynamics within HRM contexts where biased tools risk entrenching workplace inequities. Organizations should focus on understanding the causes of biases and transform AI-HRM systems from bias amplifiers into equity catalysts by adopting the strategies proposed in this framework – collaborative auditing, inclusive design, and transparency-driven processes. This approach aligns with ethical AI best practices to foster environments where stakeholders collectively refine tools to serve both organizational and societal goals.

This paper lays a theoretical foundation for future studies to question cognitive biases among designers, HR professionals, and end-users (e.g., applicants and employees), offering insights into how these biases shape AI integration and DEI outcomes. Further research is needed to explore AI’s potential to automate DEI efforts, ensuring that stakeholder roles, systemic barriers, and unintended results of AI automation are thoroughly examined. Without such a rigorous approach, AI-HRM tools risk replicating the inequities they promise to resolve, underscoring the necessity of treating algorithmic systems as socio-technical interventions that demand ethical care.

https://doi.org/10.15170/MM.2025.59.02.05

REFERENCES

Adadi, A., & Berrada, M. (2018), "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)", *IEEE Access*, 6, 52138–52160. DOI: 10.1109/access.2018.2870052

Alshaabani, A., Hamza, K. A., & Rudnák, I. (2021), "Impact of Diversity Management on Employees' Engagement: The Role of Organizational Trust and Job Insecurity", *Sustainability*, 14(1), 1–22. DOI: 10.3390/su14010420

András, P., Esterle, L., Guckert, M. et al. (2018), "Trusting Intelligent Machines: Deepening Trust Within Socio-Technical Systems", *IEEE Technology and Society Magazine*, 37(4), 76–83. DOI: 10.1109/MTS.2018.2876107

Arslan, A., Cooper, C., Khan, Z., Golgeci, I., & Ali, I. (2021), "Artificial intelligence and human workers interaction at team level: a conceptual assessment of the challenges and potential HRM strategies", *International Journal of Manpower*, 43(1), 75–88. DOI: 10.1108/ijm-01-2021-0052

Baeza-Yates, R. (2018), "Bias on the web", *Communications of the ACM*, 61(6), 54–61. DOI: 10.1145/3209581

Belenguer, L. (2022), "AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry", *AI and Ethics*, 2(2). DOI: 10.1007/s43681-022-00138-8

Bellamy, R., Dey, K., Hind, M. et al. (2023), *AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias*. ArXiv (Cornell University). DOI: 10.48550/arxiv.1810.01943

Bonilla-Silva, E. (2001), *White supremacy and racism in the post-civil rights era*. Lynne Rienner.

Borry, E. L., & Getha-Taylor, H. (2018), "Automation in the Public Sector: Efficiency at the Expense of Equity?", *Public Integrity*, 21(1), 6–21. DOI: 10.1080/10999922.2018.1455488

Brown, J. G. (2024), "The Impact of Artificial Intelligence in Employee Onboarding Programs", *Advances in Developing Human Resources*, 26(2–3). DOI: 10.1177/15234223241254775

Budhwar P., Chowdhury S., Wood G. et al. (2023), "Human Resource Management in the Age of Generative Artificial intelligence: Perspectives and Research Directions on ChatGPT", *Human Resource Management in the Age of Generative Artificial Intelligence: Perspectives and Research Directions on ChatGPT*, 33(3), 606–659. DOI: 10.1111/1748-8583.12524

Buolamwini, J., & Gebru, T. (2018), *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, in:* Proceedings; PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html [15.17.2025.]

Cappelli, P., Tambe, P., & Yakubovich, V. (2019a), "Artificial Intelligence in Human Resources Management: Challenges and a Path Forward", *SSRN Electronic Journal*, 61(4). DOI: 10.2139/ssrn.3263878

Cappelli, P., Tambe, P., & Yakubovich, V. (2019b), "Can Data Science Change Human Resources?", *The Future of Management in an AI World*, 93–115. DOI: 10.1007/978-3-030-20680-2_5

Chandar, P., Khazaeni, Y., Davis, M., Muller, M., Crasso, M., Liao, Q. V., Shami, N. S., & Geyer, W. (2017), "Leveraging Conversational Systems to Assists New Hires During Onboarding", *Human-Computer Interaction – INTERACT 2017*, 381–391. DOI: 10.1007/978-3-319-67684-5_23

Charbonneau, É., & Doberstein, C. (2020), "An Empirical Assessment of the Intrusiveness and Reasonableness of Emerging Work Surveillance Technologies in the Public Sector", *Public Administration Review*, 80(5), 780–791. DOI: 10.1111/puar.13278

Chen, Z. (2023), "Ethics and Discrimination in Artificial intelligence-enabled Recruitment Practices", *Humanities and Social Sciences Communications*, 10(1), 1–12. DOI: 10.1057/s41599-023-02079-x

Dankwa-Mullan, I., & Weeraratne, D. (2022), "Artificial Intelligence and Machine Learning Technologies in Cancer Care: Addressing Disparities, Bias, and Data Diversity", *Cancer Discovery*, 12(6), 1423–1427. DOI: 10.1158/2159-8290.cd-22-0373

Drage, E., & Mackereth, K. (2022), "Does AI Debias Recruitment? Race, Gender, and AI's 'Eradication of Difference', *Philosophy & Technology*, 35(4). DOI: 10.1007/s13347-022-00543-1

Geman, S., Bienenstock, E., & Doursat, R. (1992), "Neural Networks and the Bias/Variance Dilemma", *Neural Computation*, 4(1), 1–58. DOI: 10.1162/neco.1992.4.1.1

Giest, S. N., & Klievink, B. (2022), "More than a digital system: how AI is changing the role of bureaucrats in different organizational contexts", *Public Management Review*, 26(2), 1–20. DOI: 10.1080/14719037.2022.2095001

Gillath, O., Ai, T., Branicky, M., Keshmiri, S., Davison, R., & Spaulding, R. (2020), "Attachment and Trust in Artificial Intelligence", *Computers in Human Behavior*, , 0747-5632. DOI: 10.1016/j.chb.2020.106607

Grabovskyi, V., & Martynovych, O. (2019), "Facial recognition with using of the Microsoft face API service", *Electronics and Information Technologies*, *12*. DOI: 10.30970/eli.12.3

Gulati, A., Lozano, M. A., Lepri, B., & Oliver, N. (2022), *BIASeD: Bringing Irrationality into Automated System Design*, 1–14. DOI: 10.48550/arXiv.2210.01122 [15.17.2025.]

Hill, R. K. (2015), "What an Algorithm Is", *Philosophy & Technology*, 29(1), 35–59. DOI: 10.1007/s13347-014-0184-5

Hmud, B., & Lászlo, V. (2019), "Will Artificial Intelligence Take Over Human Resources Recruitment and Selection?", *Network Intelligence Studies*, 7(13*)*, 21–30. https://seaopenresearch.eu/Journals/articles/NIS_13_3.pdf [15.17.2025.]

Huang, H.-Y., & Cynthia. (2022), "Social Inclusion in Curated Contexts: Insights from Museum Practices", in: *2022 ACM Conference on Fairness, Accountability, and Transparency*. 300–309. DOI: 10.1145/3531146.3533095

Huang, M.-H., & Rust, R. T. (2021), "A Strategic Framework for Artificial Intelligence in Marketing", *Journal of the Academy of Marketing Science*, 49(1), 30–50. DOI: 10.1007/s11747-020-00749-9

Johnson, B. A. M., Coggburn, J. D., & Llorens, J. J. (2022), "Artificial Intelligence and Public Human Resource Management: Questions for Research and Practice", *Public Personnel Management*, 51(4). DOI: 10.1177/00910260221126498

Jora, R. B., Sodhi, K. K., Mittal, P., & Saxena, P. (2022), *Role of Artificial Intelligence (AI) In meeting Diversity, Equality and Inclusion (DEI) Goals*, IEEE Xplore. DOI: 10.1109/ICACCS54159.2022.9785266

Karácsony, P. (2022), "Analysis of the Attitude of Hungarian HR Professionals to Artificial Intelligence", *Naše Gospodarstvo/Our Economy*, 68(2), 55–64. DOI: 10.2478/ngoe-2022-0011

Kekez, I., Lode Lauwaert, & Ređep, N. B. (2025), "Is Artificial Intelligence (AI) Research Biased and Conceptually Vague? A Systematic Review of Research on Bias and Discrimination in the context of using AI in Human Resource Management", *Technology in Society*, *81*. DOI: 10.1016/j.techsoc.2025.102818

Kochan, T., Bezrukova, K., Ely, R., Jackson, S., Joshi, A., Jehn, K., Leonard, J., Levine, D., & Thomas, D. (2003), "The effects of diversity on business performance: Report of the diversity research network", *Human Resource Management*, 42(1), 3–21. DOI: 10.1002/hrm.10061

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2021), "Highly Accurate, But Still Discriminatory", *Business & Information Systems Engineering*, 63(1), 39–54. DOI: 10.1007/s12599-020-00673-w

Köchling, A., & Wehner, M. C. (2022), "Better explaining the benefits why AI? Analyzing the impact of explaining the benefits of ai-supported selection on applicant responses", *International Journal of Selection and Assessment*, 31(1), 45–62. DOI: 10.1111/ijsa.12412

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

Letheren, K., Russell-Bennett, R., & Whittaker, L. (2020), "Black, white or grey magic? Our future with artificial intelligence", *Journal of Marketing Management*, 36(3–4), 216–232. DOI: 10.1080/0267257x.2019.1706306

Li, F., Dong, H., & Liu, L. (2020), "Using AI to Enable Design for Diversity: A Perspective", in: *Advances in Intelligent Systems and Computing*, 77–84. DOI: 10.1007/978-3-030-51194-4_11

Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021), "A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions", in: *Proceedings of the 54th Hawaii International Conference on System Sciences*. DOI: 10.24251/hicss.2021.664

Malin, C. D., Fleiß, J., Seeber, I., Kubicek, B., Kupfer, C., & Thalmann, S. (2024), "The application of AI in digital HRM – an experiment on human decision-making in personnel selection", *Business Process Management Journal*, 30(8), 284–312. DOI: 10.1108/bpmj-11-2023-0884

Marabelli, M., & Lirio, P. (2025), "AI and the metaverse in the workplace: DEI opportunities and challenges, *Personnel Review*, 54(3), 844–853. DOI: 10.1108/pr-04-2023-0300

Margherita, A. (2022), Human resources analytics: A systematization of research topics and directions for future research", *Human Resource Management Review*, 32(2). DOI: 10.1016/j.hrmr.2020.100795

Murugesan, U., Subramanian, P., Srivastava, S., & Dwivedi, A. (2023), "A Study of Artificial Intelligence Impacts on Human Resource Digitalization in Industry 4.0", *Decision Analytics Journal*, 7. DOI: 10.1016/j.dajour.2023.100249

Myllylä, M. (2022), "Psychological and Cognitive Challenges in Sustainable AI Design", in M.

Rauterberg (ed.), *Culture and Computing: 10th International Conference: Vols. C&C 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings*. Issue 13324, 426–444. DOI: 10.1007/978-3-031-05434-1_29

Njoto, S., Cheong, M., Lederman, R., McLoughney, A., Ruppanner, L., & Wirth, A. (2022), "Gender Bias in AI Recruitment Systems: A Sociologicaland Data Sciencebased Case Study", *2022 IEEE International Symposium on Technology and Society (ISTAS)*, 1, 1–7. DOI: 10.1109/ISTAS55053.2022.10227106

Nyariro, M., Emami, E., Caidor, P., & Rahimi, S. A. (2023), "Integrating equity, diversity and inclusion throughout the lifecycle of AI within healthcare: a scoping review protocol", *BMJ Open*, 13(9). DOI: 10.1136/bmjopen-2023-072069

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). "Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries", *Frontiers in Big Data*, 2(13). DOI: 10.3389/fdata.2019.00013

Oswald, F. L., Behrend, T. S., Putka, D. J., & Sinar, E. (2020), "Big Data in Industrial-Organizational Psychology and Human Resource Management: Forward Progress for Organizational Research and Practice", *Annual Review of Organizational Psychology and Organizational Behavior*, 7(1). DOI: 10.1146/annurev-orgpsych-032117-104553

Patrick, H. A., & Kumar, V. R. (2012), "Managing Workplace Diversity: Issues and Challenges", *SAGE Open*, 2(2), 1–15. DOI: 10.1177/2158244012444615

Postmes, T., Spears, R., & Cihangir, S. (2001), "Quality of decision making and group norms", *Journal of Personality and Social Psychology*, 80(6), 918–930. DOI: 10.1037/0022-3514.80.6.918

Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A., & Tomsett, R. (2022), "Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making", *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1), 1–22. DOI: 10.1145/3512930

Rich, A. S., & Gureckis, T. M. (2019), "Lessons for artificial intelligence from the study of natural stupidity", *Nature Machine Intelligence*, 1(4), 174–180. DOI: 10.1038/s42256-019-0038-z

Roopaei, M., Horst, J., Klaas, E., Foster, G., Salmon-Stephens, T. J., & Grunow, J. (2021), *Women in AI: Barriers and Solutions*, IEEE Xplore. DOI: 10.1109/AIIoT52608.2021.9454202

Saheb, T. (2022), "Ethically Contentious Aspects of Artificial Intelligence surveillance: a Social Science Perspective", *AI and Ethics*, 2(2), 369–379. DOI: 10.1007/s43681-022-00196-y

Schweyer, A., & Advisor, A. (2018), *Predictive Analytics and Artificial Intelligence in People Management Predictive Analytics and Artificial Intelligence in People Management*. https://theirf.org/wp-content/uploads/2018/08/2018-ai-study-white-paper-pdf-updated.pdf [15.17.2025.]

Selbst, A. D., Boyd, D., Friedler, S., Venkatasubramanian, S., & Vértesi, J. (2019), "Fairness and Abstraction in Sociotechnical Systems", FAT*, 19, in: *Proceedings of the Conference on Fairness, Accountability, and Transparency* Atlanta, GA, USA. ACM, New York, NY, USA, 59–68. DOI: https://doi.org/10.1145/3287560.3287598 [15. 07. 2025.]

Shams, R. A., Zowghi, D., & Bano, M. (2023), "AI and the quest for diversity and inclusion: a systematic literature review", *AI and Ethics*, (3). DOI: 10.1007/s43681-023-00362-w

Shen, J., Chanda, A., D'Netto, B., & Monga, M. (2009), "Managing diversity through human resource management: an international perspective and conceptual framework", *The International Journal of Human Resource Management*, 20(2), 235–251. DOI: 10.1080/09585190802670516

Shin, D., & Park, Y. J. (2019), "Role of fairness, accountability, and transparency in algorithmic affordance", *Computers in Human Behavior*, 98, 277–284. DOI: 10.1016/j.chb.2019.04.019

Singh, E. P., & Doval, J. (2019), "Artificial Intelligence and HR: Remarkable Opportunities, Hesitant Partners", in: *Proceedings of the 4th National HR Conference on Human Resource Management Practices and Trends*. https://ssrn.com/abstract=3553448 [15.17.2025.]

Singha, S., & Prakasam, K. (2014), "Exploring the Factors That Facilitate Workforce Diversity Management in ITES Organizations", *International Journal of Management and Humanities (IJMH)*, (3)1, 11–15. https://www.ijmh.org/portfolio-item/c0014121314/ [15.17.2025.]

Srinivasan, R., & Chander, A. (2021), "Biases in AI Systems", *Communications of the ACM*, 64(8), 44–49. DOI: 10.1145/3464903

Strohmeier, S., & Piazza, F. (2015), "Domain driven data mining in human resource management: A review of current research", *Expert Systems with Applications*, 40(7), 2410–2420. DOI: 10.1016/j.eswa.2012.10.059

Suresh, H., & Guttag, J. (2021), "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle", in: *Proceedings* *of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–9. DOI: 10.1145/3465416.3483305

Tambe, P., Cappelli, P., & Yakubovich, V. (2019), "Artificial Intelligence in Human Resources Management: Challenges and a Path Forward", *California Management Review*, 61(4), 15–42. DOI: 10.1177/0008125619867910

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021), "Artificial intelligence in marketing: Systematic review and future research direction", *International Journal of Information Management Data Insights*, 1(1), DOI: 10.1016/j.jjimei.2020.100002

Walsh, C. G., Chaudhry, B., Dua, P., Goodman, K. W., Kaplan, B., Kavuluru, R., Solomonides, A., & Subbian, V. (2020), "Stigma, biomarkers, and algorithmic bias: recommendations for precision behavioral health with artificial intelligence", *JAMIA Open*, 3(1), 9–15. DOI: 10.1093/jamiaopen/ooz054

Webster, C., & Ivanov, S. (2019), "Robotics, Artificial Intelligence, and the Evolving Nature of Work", in: *Digital Transformation in Business and Society*, 127–143. DOI: 10.1007/978-3-030-08277-2_8

Weller, A. (2019), "Transparency: Motivations and Challenges", in: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning,* 23–40. DOI:10.1007/978-3-030-28954-6_2 [15.17.2025.]

West, D. M. (2018), *Future of Work: Robots, AI, and Automation*. Rowman & Littlefield.

Yen, C., & Chiang, M.C. (2020), "Trust me, If You can: a Study on the Factors That Influence Consumers' Purchase Intention Triggered by Chatbots Based on Brain Image Evidence and self-reported Assessments", *Behaviour & Information Technology*, 40(11), 1–18. DOI: 10.1080/0144929x.2020.1743362

Zowghi, D., & Rimini, F. (2023), "*Diversity and Inclusion in Artificial Intelligence", AI and Ethics,* 3(4), 873–876. DOI: 10.48550/arxiv.2305.12728