

Building Responsible AI Frameworks for HRM Transformation

Dr. Ashish Mohture¹, Anu Sayal²

¹ Lincoln University College, Malaysia

² Centre for Artificial Intelligence, MITS, Gwalior, India

Email: ashishmohture@gmail.com, anu.sayal.07@gmail.com

Abstract:

The fast adoption of Artificial Intelligence (AI) into Human Resource Management (HRM) offers disruptive possibilities and severe ethical risks at the same time. AI-based applications improve recruitment, performance review, and workforce planning but unless they are properly implemented, they reinforce systemic discrimination, infringe upon employee privacy, and disempower accountability within organizations. The paper is a systematic literature review (SLR) based on thirteen peer-reviewed articles published within 2022-2026, selected by Web of Science (WoS), Scopus, and Q1-Q3 indexed journals, which summarizes the up-to-date picture of knowledge regarding responsible AI-HRM. It finds five thematic clusters, such as (1) responsible AI and ethical governance frameworks; (2) AI transformation of HRM functions; (3) algorithmic bias, fairness, and DEI; (4) bibliometric and systematic mapping studies; and (5) legal and regulatory perspectives. Based on these results, a new six-pillar Responsible AI-HRM Framework (RAIH-F) is suggested, which is mapped on six fundamental HRM functions and implementation mechanisms and quantifiable KPIs. Theoretical contributions, practical implications and a future research agenda are given.

Keywords:

Responsible AI; Human Resource Management; Algorithmic Bias; Ethical AI; Explainable AI

1. Introduction

The fourth industrial revolution has made Artificial Intelligence a key disruptor of the Human Resource Management. Talent acquisition, performance assessment, workforce planning, and employee engagement are now mediated by AI, which can guarantee a higher level of efficiency and a more objective approach to data. Nevertheless, this promise is tarnished with reported threats: an increase in the rates of algorithmic bias that discriminates against women and racial minorities, a lack of transparency in its decisions that undermines the rights of the employees, and legal liability due to the fact that the EU AI Act considers AI in the HR domain as a high-risk category (Chowdhury et al., 2023; Rodgers et al., 2023). Nevertheless, the field does not have a holistic, multi-functional, and practically implementable, responsible AI governance framework of HRM. The three research questions targeted to fill that gap in this paper are (RQ1) What responsible AI principles and governance mechanisms have been suggested or confirmed in the literature on HRM? (RQ2) What are the ways in which AI applications are changing the fundamental HRM functions and what are the ethical issues? (RQ3) What is an effective responsible AI in HRM? The paper adds a systematic literature review of thirteen WoS/Scopus-indexed articles and new six-pillar Responsible AI-HRM Framework (RAIH-F). THEORETICAL

BACKGROUND The history of HRM has been that of administrative personnel management, but the strategic human capital fit is now being followed by the third wave of transformative change, in which algorithmic mediation gains a place in the decision-making process, and the concept of HRM is now driven by algorithms rather than human discretion (Chowdhury, Budhwar et al., 2024). Nowadays AI systems, i.e. using machine learning to analyze historical HR data to predict the suitability of a candidate, the potential of their performance, and the probability of their attrition are qualitatively different to the previous administrator automation. Generative AI goes further and opens this frontier by allowing AI to write job descriptions, preliminary interviews, and custom learning streams. The notion of responsible AI is generally characterized by six principles that converge, namely fairness and non-discrimination, transparency and explainability, privacy and data protection, accountability and governance, human oversight, and safety (Chowdhury et al., 2023). These values are indicative of several ethical traditions: consequentialist, deontological and virtue ethics, and generate specific tensions in HRM: the efficiency imperative and the demands of equity are in conflict; the intensity of data and the expectation of privacy clash; the alleged objectivity of the algorithms and the empirical observations of bias. The two theoretical frames that will be used in this paper are Stakeholder Theory (Freeman, 1984), which helps understand the variety of interests of the employees, candidates, managers, shareholders, and the society; and Algorithmic Justice, which requires that the communities that are impacted by AI can exercise their voice in the process of its governance (Anonymous, 2026). The EU AI Act (2024) is the first global AI regulatory framework to operationalize these hypothetical concepts into legally enforceable rules, categorizing HRM AI systems as high-risk and imposing a transparency, human-supervision and conformity-assessment requirement.

2. Methodology

This paper uses PRISMA-based systematic literature review. WoS Core Collection and Elsevier Scopus were searched using search terms with the following form: (artificial intelligence OR machine learning OR algorithmic decision-making) AND (human resource management OR HRM) AND (responsible AI OR ethical AI OR algorithmic bias OR fairness). The limit was constrained to January 2022 - March 2026. The inclusion criteria included peer-reviewed articles included in WoS or Scopus (Q1-Q3 per SCImago/JCR), that had an ethical or governance component in the context of AI in employment. N=847 records were yielded at first, which were narrowed down to 13 eligible articles following the elimination of duplicates (n= 213), removal of titles and abstracts (n= 512), and quality evaluation of full-text articles using MMAT (n= 13). Author(s), year, journal, quartile, methodology, HRM function, responsible AI principles, and key findings were extracted using data and analyzed using thematic synthesis to form cross-study analytical themes.

4.1 Conscientious AI Principles and Ethical Frameworks.

Chowdhury et al. (2023) present the empirical mapping of the field, stating the principles of responsible AI as six core ones in HRM (fairness, transparency, privacy, accountability, human oversight, beneficence) and showing that the principles interact in complicated and even antagonistic manners (in particular, transparency requirements may contradict privacy obligations in cases when sensitive information about employees is disclosed). Rodgers et al. (2023) do not ignore the accountability aspect, but according to the authors, the mediation of algorithmic recommendations by HR decision-makers is provided by the Throughput model that explains how ethical positions, professional norms, and culture of the organization are mediated.

The model explains how the responsibility is shared among AI systems, developers, organizations and human decision-makers - which is a crucial factor when biased AI advice is implemented. Anonymous (2025b) build upon this with a legal-technical assessment framework of equity and transparency in AI-HRM systems, systematic deficiencies in commercial AI system in respect to indirect discrimination and aligning the proposals with the requirements of the EU AI Act. 4.2 AI Revolutionization of HRM Foundational functions. In a study on 312 articles in Scopus, Anonymous (2024) utilizes PRISMA and BERTopic modeling, and it is concluded that the majority of AI-HRM studies involve recruitment (47 percent), then performance management (21 percent) and L&D (15 percent) and workforce planning (11 percent) indicate severe gaps in research on compensation, employee relations, and organization culture. The critical point made by Chowdhury, Budhwar et al. (2024) is that the current theories of HRM (Resource-Based View, Dynamic Capabilities, AMO) are no longer applicable to the generative AI era because a new strategy to HRM is needed, and its power is based on the quality of human-AI collaboration and the ability to control AI responsibly. Anonymous (2025c) distinguish between three AI knowledge-transformation processes, including automation of routine, enhancement of complex judgment and the issue of creating new organizational knowledge via pattern recognition.

3. Key Contribution

This study proposes a six-pillar Responsible AI-HRM Framework integrating ethical principles with legal compliance and practical implementation mechanisms.

RAIH-F Applied Across Core HRM Functions

HRM Function	AI Application	Primary Risk	RAIH-F Pillar(s)	Mitigation Strategy
Recruitment & Selection	CV screening, video interview AI	Historical bias in training data	Pillars 1, 2, 5	Fairness-aware algorithms; human review mandate
Performance Management	Predictive analytics, sentiment AI	Opaque algorithmic decisions	Pillars 3, 5	XAI dashboards; mandatory manager override
Learning & Development	Personalized training, skill-gap AI	Data surveillance, over-reliance	Pillars 4, 5	Privacy-by-design; employee consent protocols
Workforce Planning	Attrition prediction, demand forecasting	Model drift, dynamic environments	Pillars 1, 6	Continuous validation; compliance audit

Employee Engagement	Sentiment chatbots, pulse surveys	Psychological safety concerns	Pillars 3, 4	Ethical data governance; anonymization
Compensation & Benefits	Market benchmarking, pay equity AI	Gender/racial pay gap perpetuation	Pillars 2, 6	Algorithmic pay equity audits; legal review

4. Method, Experiments and Results

A PRISMA-based systematic literature review was conducted using WoS and Scopus databases. Key risks identified include algorithmic bias, lack of transparency, and privacy concerns. The RAIH-F framework includes six governance pillars.

5. Discussions

The findings highlight that responsible AI in HRM is a **socio-technical challenge**, not merely a technical issue.

- **Efficiency vs Equity Trade-off:** AI improves efficiency but may compromise fairness
- **Data vs Privacy Conflict:** Increased data usage raises ethical concerns
- **Global Imbalance:** Research is heavily Western-centric
- **Lack of Employee Voice:** Limited inclusion in AI governance

The RAIH-F framework addresses these gaps by integrating governance, ethics, and legal compliance into a unified model.

6. Conclusions

Problem Addressed: Lack of a comprehensive responsible AI framework in HRM

Method Used: Systematic literature review (PRISMA)

Key Findings:

- AI introduces ethical risks in HRM, Existing research is fragmented, Need for integrated governance framework.

Contribution: Development of RAIH-F with six pillars

Limitations & Future Work:

- Requires empirical validation, Limited Non-Western perspectives, Future research should focus on longitudinal and organizational studies.

References

1. Aguinis, H., Villamor, I., & Ramani, R. (2021). MTurk research: Review and recommendations. *Journal of Management*, 47(4), 823–837. <https://doi.org/10.1177/0149206320969787>
2. Bogen, M., & Rieke, A. (2018). Help wanted: An examination of hiring algorithms and bias. *Technology Science*.
3. Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence. *Harvard Business Review*.
4. Cappelli, P., Tambe, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15–42. <https://doi.org/10.1177/0008125619867910>
5. Chamorro-Premuzic, T., Winsborough, D., Sherman, R., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology*, 9(3), 621–640. <https://doi.org/10.1017/iop.2016.6>
6. Chowdhury, S., Budhwar, P., Wood, G., & Dey, P. (2024). Generative Artificial Intelligence in Business: Towards a Strategic Human Resource Management Framework. *British Journal of Management*, Wiley. <https://doi.org/10.1111/1467-8551.12824>
7. Davenport, T. H., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48, 24–42. <https://doi.org/10.1007/s11747-019-00696-0>
8. Dwivedi, Y. K., Hughes, L., Ismagilova, E., et al. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research. *International Journal of Information Management*, 57. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
9. European Parliament. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence (AI Act). Official Journal of the European Union.
10. Floridi, L., Cows, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
11. George, G., Lakhani, K. R., & Puranam, P. (2020). What has changed? The impact of artificial intelligence on management research and practice. *Academy of Management Discoveries*, 6(4), 537–556. <https://doi.org/10.5465/amd.2020.0144>
12. Glikson, E., & Woolley, A. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
13. Huang, M. H., & Rust, R. T. (2021). Artificial intelligence in service. *Journal of Service Research*, 24(1), 3–13. <https://doi.org/10.1177/1094670520902266>
14. Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586. <https://doi.org/10.1016/j.bushor.2018.03.007>
15. Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366–410. <https://doi.org/10.5465/annals.2018.0174>

16. Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing AI into the organization. *Journal of Business Research*, 120, 262–273. <https://doi.org/10.1016/j.jbusres.2020.07.045>
17. Minbaeva, D. (2021). Disrupted HR? Human resource management in the digital age. *Human Resource Management Review*, 31(1). <https://doi.org/10.1016/j.hrmr.2019.100820>
18. Newell, S., Marabelli, M., & Galliers, R. (2021). A knowledge-centric perspective on AI. *Journal of the Association for Information Systems*, 22(3). <https://doi.org/10.17705/1jais.00660>
19. Palos-Sanchez, P. R., Reyes-Menendez, A., & Aguenza-Velazquez, M. (2022). Artificial Intelligence and Human Resources Management: A Bibliometric Analysis. *Applied Artificial Intelligence*, Taylor & Francis. <https://doi.org/10.1080/08839514.2022.2145631>
20. PRISMA Group. (2020). PRISMA 2020 Explanation and Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews. *BMJ*, 372, n160. <https://doi.org/10.1136/bmj.n160>
21. Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1), 192–210. <https://doi.org/10.5465/amr.2018.0072>
22. Rana, G., & Kumar, R. (2025). AI in Human Resource Management: An Interdisciplinary Review and Bibliometric Analysis Using SPAR-4-SLR. *SAGE Journals*. <https://doi.org/10.1177/01672533251365116>
23. Ransbotham, S., Gerbert, P., Reeves, M., et al. (2017). Reshaping business with artificial intelligence. *MIT Sloan Management Review*.
24. Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66–83. <https://doi.org/10.1177/0008125619862257>
25. Stahl, B. C., Antoniou, J., Ryan, M., et al. (2021). Organisational responses to the ethical issues of artificial intelligence. *AI & Society*, 36, 23–37. <https://doi.org/10.1007/s00146-019-00958-4>
26. Vrontis, D., Christofi, M., Pereira, V., et al. (2022). Artificial intelligence, robotics, advanced technologies and HRM. *The International Journal of Human Resource Management*, 33(6), 1237–1266. <https://doi.org/10.1080/09585192.2020.1871398>