Artificial Intelligence and HR Decision-Making: Implications for Managerial Judgement, Trust, and Fairness

Priyanka Chauhan¹*, Prof. Poonam Puri²

1 Research Scholar, Bundelkhand University, Jhansi, Uttar Pradesh, India

jiyawendy3@gmail.com

2 Supervisor, Institute of Management Studies, Bundelkhand University, Jhansi, Uttar Pradesh, India

Abstract: Artificial intelligence (AI) and machine learning (ML) are increasingly integrated into human resource management (HRM), reshaping strategic HR decision-making across recruitment, performance appraisal, turnover prediction, and workforce planning. Despite the exponential growth of AI-related HR analytics research, there remains a theoretical gap in understanding how AI influences managerial Judgement, trust in algorithmic outputs, and perceptions of fairness in HR decisions. Drawing on decision support systems theory, socio-technical systems perspectives, and insights from recent HRM scholarship, this paper develops a conceptual framework that situates HR decision-making as a human–AI collaborative process. The framework outlines the interrelationships between AI mechanisms, managerial interpretation, explainability, ethical governance, and the quality of HR decisions. We propose five research propositions that articulate conditions under which AI augments HR managerial Judgement, fosters trust, and enhances procedural justice.

Keywords: artificial intelligence, human resource decision-making, managerial Judgement, trust, fairness, HR analytics

1. INTRODUCTION

Human resource decision-making has long been defined by managerial Judgement, professional expertise, and relational understanding of employees. Decisions concerning recruitment, performance evaluation, promotions, training investments, and retention carry ethical, social, and organizational implications that extend beyond technical criteria. With the advent of AI and ML technologies, HR leaders are increasingly adopting AI-enabled systems to assist or even automate certain decision processes (Choudhary, Budhwar & Parry, 2023).

AI is frequently framed as a tool that enhances objectivity and efficiency by identifying patterns and risks in large datasets that surpass human cognitive capacities. Proponents argue that AI can reduce subjective bias, expedite decision timelines, and provide data-driven insights to support better HR decisions (Marler & Boudreau, 2017; Choudhary et al., 2023).

However, HR decisions are inherently social and ethical; they involve interpretation, accountability, and legitimacy in ways that are distinct from purely operational choices (Colquitt et al., 2013). A decision that is technically optimal may still be perceived as unfair, dehumanizing, or distrustful by employees if the process lacks transparency or fails to align with organizational values (Raghavan et al., 2020).

Despite a growing research literature on AI adoption and analytics capability in HR, HRM scholarship has largely overlooked how AI reshapes the ongoing process of managerial decision-making itself, particularly in terms of trust, interpretability, and fairness. For example, recent systematic reviews highlight the dual impact of AI on diversity, equity, and inclusion (DEI), showing both potential enhancements and risks such as algorithmic bias and reduced accountability without ethical governance mechanisms. Similarly, studies on employee involvement in AI-driven HR processes reveal the need to balance efficiency with participatory decision structures.

This paper seeks to bridge this gap by investigating how AI influences three core dimensions of HR decision-making: managerial Judgement, trust in algorithmic recommendations, and fairness perceptions. Our contribution is threefold: (a) reconceptualize AI-supported HR decision-making as a socio-technical, collaborative process; (b) integrate insights from recent relevant research to ground the framework in HRM theory and practice; and (c) propose a set of theoretical propositions that advance future empirical inquiry into responsible and human-centered AI adoption in HR.

The paper proceeds as follows: Section 2 outlines the methodological approach; Section 3 synthesizes relevant literature; Section 4 presents an integrative conceptual framework; Section 5 develops research propositions; Section 6 discusses implications; Section 7 offers directions for future research; and Section 8 concludes with key takeaways.

2. LITERATURE REVIEW

2.1 AI in HR Analytics

AI and machine learning (ML) have emerged as core components of modern HR analytics, enabling organizations to process large volumes of workforce data and derive predictive insights (Marler & Boudreau, 2017). Recent research in HRM highlights that AI is transforming HR practices, particularly in recruitment, performance management, talent development, and workforce planning (Choudhary, Budhwar & Parry, 2023). However,

despite the growing use of AI in HR, research has often emphasised adoption and technological capability rather than the implications for HR decision-making processes. This is an important gap because HR decisions involve ethical and social dimensions that extend beyond technical accuracy (Colquitt et al., 2013).

AI promises to improve decision quality through objectivity and consistency, yet it may also reproduce existing biases embedded in historical data. For example, algorithmic hiring systems may replicate gender or racial disparities if the training data reflects past discriminatory practices (Raghavan et al., 2020). In addition, the use of employee data for predictive analytics raises concerns about privacy, autonomy, and surveillance (Leicht-Deobald et al., 2019). These issues suggest that AI adoption in HR requires not only technical expertise but also ethical governance and organizational accountability.

Recent research underscores the need to understand the socio-technical dynamics of AI in HR. Choudhary et al. (2023) argue that AI adoption should be examined through strategic and human-centred lenses, focusing on how AI reshapes HR roles and organisational practices. Similarly, research on digital HR suggests that technological change requires new organisational capabilities and leadership strategies to integrate AI into HR processes effectively (Strohmeier, 2020). This body of work sets the stage for exploring how AI affects managerial Judgement, trust, and fairness in HR decisions.

2.2 Decision Support Systems and Human–AI Interaction

Decision support systems (DSS) theory provides a useful lens for understanding AI in HR. DSS research suggests that analytics tools enhance decision quality by improving information availability, timeliness, and relevance (Sharda, Delen & Turban, 2014). In HR contexts, AI systems can detect patterns in workforce data and provide predictive recommendations for retention, talent identification, and performance management. However, DSS theory also emphasizes that decision outcomes depend on how human decision-makers interpret and use the information.

Human–AI interaction research indicates that managers respond differently to algorithmic recommendations. Some managers may exhibit automation bias, placing undue trust in AI outputs, while others may show algorithm aversion, distrusting AI after observing errors (Dietvorst, Simmons & Massey, 2015). These dynamics are particularly salient in HR decisions, where managers are accountable for outcomes and must justify decisions to

employees. Therefore, understanding human–AI interaction is essential for explaining how AI affects managerial Judgement and decision legitimacy.

2.3 Managerial Judgement and Ethical Responsibility

Managerial Judgement in HR is grounded in contextual knowledge, professional expertise, and ethical responsibility. HR decisions involve social and moral considerations, such as fairness, employee wellbeing, and organisational values. AI systems can provide valuable insights but cannot fully replicate human understanding of context, culture, and individual circumstances. Research suggests that AI should be viewed as a tool that augments human Judgement rather than replacing it (Cascio & Montealegre, 2016). This is especially relevant in HR contexts, where decisions affect employees' careers and identities.

Recent studies emphasises the importance of human oversight in AI decision-making. For example, Leicht-Deobald et al. (2019) highlight the risks of algorithmic decision-making for personal integrity and privacy. They argue that HR professionals must maintain responsibility for decisions and ensure that AI systems are used ethically. This suggests a need for human-in-the-loop models in HR, where managers retain the authority to interpret and override AI recommendations when necessary.

2.4 Trust, Explainability, and Procedural Justice

Trust is central to the adoption and acceptance of AI in HR. Organisational trust theory defines trust as a willingness to be vulnerable based on positive expectations about another party's competence and integrity (Mayer et al., 1995). In the AI context, trust depends on the perceived reliability, transparency, and predictability of algorithmic systems. Managers are more likely to rely on AI recommendations when they believe the system is competent and aligned with organisational goals.

AI is an emerging approach aimed at increasing transparency by making AI decision-making understandable to humans (Gunning, 2017). Explainability is particularly important in HR because managers must justify decisions to employees and comply with legal and ethical standards. When AI decisions are opaque, employees may perceive HR processes as unfair, reducing trust and engagement. Research suggests that explainability can increase perceptions of procedural justice by allowing employees to understand the basis for decisions and to challenge outcomes if necessary (Binns, 2018).

2.5 Fairness and Bias in AI-Supported HR Decisions

Fairness is a foundational concern in HR decision-making. Procedural justice theory highlights that employees' perceptions of fairness depend on transparency, voice, and consistency of decision processes (Colquitt et al., 2013). AI can enhance fairness by standardising decision criteria and reducing subjective bias. However, AI systems can also reproduce historical inequities, leading to biased outcomes. Raghavan et al. (2020) note that algorithmic hiring systems can perpetuate discrimination if training data reflect past biases.

Recent research emphasises the importance of ethical governance in AI deployment. Bias audits, inclusive data practices, and accountability mechanisms are necessary to ensure that AI systems do not discriminate against protected groups. This is particularly relevant in HR, where discriminatory outcomes can have legal, ethical, and reputational consequences. Ethical governance mechanisms can also enhance trust by demonstrating organisational commitment to fairness and transparency.

2.6 Recent Contributions and Research Gap

Recent studies has begun to explore AI's implications for HR strategy and practice. Choudhary et al. (2023) provide a comprehensive review of AI and advanced technologies in HRM, emphasising the need for human-centred and ethically grounded research. In addition, research on digital HR highlights the importance of organisational capabilities, leadership, and culture in enabling technology adoption (Strohmeier, 2020). Despite these contributions, there remains a need for research that explicitly links AI to HR decision-making processes and outcomes, particularly in terms of managerial Judgement, trust, and fairness.

This paper responds to this gap by developing a conceptual framework that integrates AI, managerial Judgement, trust, and fairness. The framework highlights the mechanisms through which AI influences HR decision outcomes and identifies boundary conditions such as explainability and ethical governance. By doing so, the paper contributes to HRM theory and provides a foundation for future empirical research on responsible AI adoption in HR.

3. CONCEPTUAL FRAMEWORK

This paper proposes a human–AI collaborative framework that explains how AI and machine learning shape HR decision-making through managerial Judgement, trust, and fairness. The framework builds on decision support systems theory (Sharda et al., 2014) and socio-technical

systems perspectives (Cascio & Montealegre, 2016), which emphasize that technology is embedded within organisational processes and shaped by human interpretation. The framework also draws on organisational trust theory (Mayer et al., 1995) and organisational justice theory (Colquitt et al., 2013), highlighting that trust and fairness are key to legitimising HR decisions.

The framework comprises five core components: (1) HR data inputs, (2) AI/ML mechanisms, (3) explainability layer, (4) managerial Judgement, and (5) decision outcomes (quality, trust, and fairness). AI-based HR systems are typically fed by large datasets, including employee performance metrics, behavioural data, engagement scores, and recruitment data. These data inputs are processed by AI algorithms to generate predictions, recommendations, or classifications that support HR decisions (Choudhary et al., 2023). The AI/ML mechanisms may include supervised learning models for predicting turnover risk, unsupervised learning for identifying patterns, and natural language processing for analysing employee feedback.

The next layer in the framework is explainability. AI systems often operate as "black boxes," making it difficult for users to understand how outcomes were generated (Pasquale, 2015). Explainable AI (XAI) addresses this challenge by providing interpretable explanations of algorithmic decisions, which can increase transparency and facilitate managerial understanding (Gunning, 2017). In HR contexts, explainability is crucial because managers must justify decisions to employees and ensure that decisions are consistent with organisational values and legal standards. Therefore, explainability serves as a mechanism that links AI outputs to managerial trust.

Managerial Judgement is central to the framework. Managers interpret AI outputs and integrate them with contextual knowledge, ethical considerations, and organisational goals. The human–AI interaction can take several forms. In some cases, managers may use AI as a supportive tool, using predictions as one input among many in their decision process. In other cases, managers may rely heavily on AI recommendations, potentially leading to automation bias (Dietvorst et al., 2015). Conversely, managers may distrust AI systems, leading to underutilisation and algorithm aversion (Dietvorst et al., 2015). The framework posits that managerial Judgement is influenced by the perceived transparency and reliability of AI systems, as well as by organisational policies that govern AI use.

Trust is conceptualised as a mediating mechanism between explainability and decision outcomes. When AI systems are transparent and interpretable, managers are more likely to

trust them, leading to more consistent use and higher decision quality (Mayer et al., 1995). Trust also influences employees' perceptions of HR decisions, particularly when managers communicate the rationale for decisions and the role of AI in the process. However, trust can be fragile; a single AI error or a perception of bias can undermine trust and reduce acceptance.

Fairness is a key outcome and boundary condition in the framework. Procedural fairness in HR is influenced by the perceived transparency of decision processes and the ability of employees to understand and contest decisions (Colquitt et al., 2013). AI can enhance fairness by standardising criteria and reducing subjective bias, but it can also perpetuate discriminatory patterns embedded in historical data (Raghavan et al., 2020). Therefore, fairness depends not only on AI accuracy but also on ethical governance mechanisms such as bias audits, human oversight, and inclusive data practices. The framework suggests that fairness influences organisational legitimacy and employee trust, which in turn affects long-term HR outcomes such as commitment and retention.

In summary, the conceptual framework presents AI as a socio-technical partner in HR decision-making. AI systems process HR data to generate recommendations, which are made interpretable through explainability mechanisms. Managers then interpret these recommendations using Judgement shaped by context, ethics, and organisational policy. Trust and fairness are central mechanisms that determine whether AI-enhanced HR decisions are accepted and perceived as legitimate. The framework provides a basis for empirical testing of the propositions and offers a roadmap for responsible AI adoption in HR practice.

The framework conceptualises AI as an enhancer of managerial capacity, where explainability and ethical governance shape the degree to which AI insights are trusted and perceived as fair — ultimately influencing decision outcomes.

4. RESEARCH PROPOSITIONS

P1: AI-augmented HR decisions improve decision quality when managerial Judgement and AI insights are integrated rather than used independently.

P2: Explainability of AI systems is positively associated with managerial trust in AI-supported HR decisions.

P3: Managerial trust mediates the relationship between AI transparency and employees' fairness perceptions of HR decisions.

P4: Ethical governance mechanisms (e.g., bias mitigation protocols) strengthen the positive influence of AI on perceived fairness in HR decisions.

P5: Over-reliance on algorithmic decision systems weakens perceptions of procedural justice among employees.

Each proposition reflects an interaction between key constructs in the framework and aligns with gaps identified in recent HRM research.

5. IMPLICATIONS

This paper extends HRM theory by reconceptualising HR decision-making as a human–AI collaborative process rather than a purely technical or managerial activity. Traditional HR decision-making models often assume that managers possess the necessary information and judgement to make optimal decisions. However, AI changes the decision environment by introducing algorithmic recommendations, predictive insights, and automated evaluation mechanisms. By integrating AI into HR decision-making theory, this paper emphasises the socio-technical nature of HR processes, where technological capabilities interact with managerial discretion, organisational values, and ethical standards. This contributes to HRM scholarship by offering a theoretical lens that can explain both the benefits and limitations of AI adoption in HR, particularly in terms of managerial Judgement, trust, and fairness.

A second theoretical contribution is the identification of trust and explainability as central mechanisms in AI-supported HR decisions. Trust has been widely studied in organisational contexts, but its application to AI systems is still emerging in HRM. By linking trust to explainability, the paper suggests that transparency and interpretability are not merely technical features but core organisational processes that shape acceptance and legitimacy. This insight expands HRM research on technology adoption by highlighting the need to study psychological and ethical mechanisms rather than only focusing on performance metrics.

Finally, the paper contributes to fairness and ethical governance debates within HRM by framing fairness as a boundary condition for AI effectiveness. AI can increase fairness by standardising decision criteria, but it can also perpetuate bias if trained on historical data that reflects past discrimination. By positioning ethical governance mechanisms as moderators, the framework explains why AI may produce positive outcomes in some contexts but not in others. This helps reconcile conflicting findings in the AI-HR literature and provides a more nuanced understanding of when AI supports or undermines HR legitimacy.

5.1 Practical Implications

For HR practitioners, the framework suggests that AI adoption should not focus solely on predictive accuracy or efficiency. Instead, organisations must integrate AI systems into HR processes with attention to explainability, human oversight, and fairness. First, HR managers should ensure that AI tools provide interpretable outputs that can be explained to employees and stakeholders. Explainable AI mechanisms, such as decision explanations and transparency reports, can enhance trust and reduce resistance to AI-based decisions.

Second, HR departments should implement human-in-the-loop models that preserve managerial discretion. AI should support decision-making rather than replace human judgement. Managers should be trained to interpret AI outputs critically and to make contextual adjustments when necessary. This is particularly important in high-stakes decisions such as hiring, promotion, and disciplinary actions, where ethical considerations and employee wellbeing are paramount.

Third, organisations should establish governance mechanisms to monitor algorithmic bias and ensure fairness. Regular audits of AI systems, bias detection tools, and ethical review boards can help identify and correct discriminatory patterns. HR analytics teams should also develop data governance practices that ensure the quality and representativeness of training data. These measures can help prevent negative outcomes such as discrimination, loss of trust, or reputational damage.

5.2 Policy Implications

From a policy perspective, organisations should develop guidelines and compliance frameworks for AI use in HR. Policies should specify acceptable use, data privacy requirements, and accountability mechanisms. For example, HR policies could require that managers document the rationale for decisions that rely on AI outputs and that employees have access to information about how AI influences decisions. Policy frameworks should also address the ethical use of employee data, including consent, privacy, and data security.

In addition, regulators and professional bodies should consider developing standards for AI governance in HR. Given the potential for algorithmic bias and discrimination, it is important to establish clear guidelines for fairness, transparency, and accountability. Such standards could also encourage organisations to adopt ethical AI practices and increase public trust in AI-supported HR processes.

6. FUTURE RESEARCH DIRECTIONS

Future research should empirically test the conceptual framework proposed in this paper, examining the conditions under which AI improves HR decision quality while maintaining fairness and trust. A key avenue is to investigate the dynamic interplay between managerial Judgement and AI recommendations over time. Research designs can capture how manager's trust in AI evolves as they gain experience with algorithmic systems and as organizational policies change. Such research would be valuable because trust is not static; it can increase with successful outcomes or decline following errors, especially in sensitive HR contexts such as promotions or performance evaluations. Studies could also examine whether early experiences of algorithmic bias have long-lasting effects on trust and acceptance.

Another important direction is to explore cross-cultural differences in AI acceptance in HR decision-making. Societal values, labor market regulations, and cultural attitudes toward automation can influence how employees and managers perceive AI in HR. For example, employees in high-power distance cultures may be more accepting of algorithmic authority, whereas employees in low-power distance contexts may demand greater transparency and participatory involvement. Cross-cultural research can help determine whether the mechanisms of trust and fairness identified in this paper operate similarly across contexts or whether they require adaptation.

Future studies should also focus on the role of AI explainability in HR. Explainable AI is increasingly seen as a critical mechanism for building trust, yet there is limited empirical research on which types of explanations are most effective in HR settings. Researchers could test different forms of explainability, such as feature importance explanations, counterfactual explanations, or case-based explanations, to determine which best supports managerial understanding and employee acceptance. In addition, research could examine whether explanation quality moderates the relationship between AI use and perceived fairness, and whether explainability interacts with organizational transparency policies to influence trust.

Research could test whether governance practices reduce algorithmic bias and improve employees' perceptions of procedural justice. It could also explore how governance interacts with organizational culture, leadership values, and HR capabilities to influence AI outcomes.

Furthermore, future research should examine the impact of AI on employee wellbeing and psychological safety. HR decisions informed by AI may be perceived as depersonalizing,

potentially increasing stress or reducing feelings of control. Empirical studies could investigate whether AI-driven decisions influence employee trust in management, job satisfaction, and turnover intentions. They could also explore whether employees' perceptions of fairness mediate these relationships.

Finally, research should investigate how AI adoption affects HR professionals' roles and identities. AI may shift HR professionals from operational administrators to strategic analysts or ethics overseers. Qualitative studies, such as interviews and ethnographic research, can capture how HR professionals negotiate their role identity when AI becomes central to decision processes. This research would contribute to understanding the broader organizational transformation associated with AI and the skills needed for effective human–AI collaboration.

7. CONCLUSION

This paper develops a conceptual framework that positions AI and machine learning as integral components of HR decision-making, highlighting the importance of managerial Judgement, trust, and fairness. AI offers powerful capabilities for processing large datasets and generating predictive insights, which can enhance the accuracy and efficiency of HR decisions. However, HR decisions are inherently social and ethical, and their legitimacy depends on employees' perceptions of procedural justice and managerial accountability. The framework presented here argues that AI should be understood as a collaborative partner rather than a replacement for managerial Judgement. Managers must interpret AI outputs, integrate contextual knowledge, and ensure decisions align with organisational values and ethical standards.

Trust emerges as a central mechanism in the framework. Managers' trust in AI systems influences whether they adopt algorithmic recommendations and how they interpret them. Trust is influenced by the perceived competence, reliability, and transparency of AI systems. Explainability is therefore essential; when AI recommendations are interpretable, managers are more likely to trust them, and employees are more likely to perceive HR decisions as fair. Conversely, opaque AI systems can undermine trust and lead to skepticism, particularly in decisions that affect employees' careers and wellbeing.

Fairness is another key dimension. AI can reduce human biases through standardisation and consistency, but it can also reproduce historical inequalities embedded in data. This creates ethical and legal risks, especially when HR decisions are used for hiring, promotion, or

disciplinary actions. The framework suggests that ethical governance mechanisms are necessary to ensure that AI-supported decisions remain fair and defensible. Governance practices such as bias audits, human-in-the-loop systems, and transparent data governance can strengthen fairness perceptions and support trust.

The research propositions developed in this paper articulate conditions under which AI can improve HR decision outcomes. AI is likely to enhance decision quality when managerial Judgement and AI insights are integrated, explainability is high, and ethical governance is strong. The propositions also highlight risks, such as over-reliance on AI, which can reduce perceived procedural justice and weaken employee trust.

Overall, the paper contributes to HRM theory by integrating AI into core decision-making processes and highlighting the socio-technical nature of AI in HR. It also provides a research agenda for future empirical studies on responsible AI adoption. The framework offers practical implications for HR practitioners: AI adoption should be accompanied by explainability and governance mechanisms that support managerial interpretation and fairness. By doing so, organisations can leverage AI's capabilities while preserving trust and legitimacy in HR decisions.

References

1. Choudhary, S., Budhwar, P., & Parry, E. (2023). Artificial intelligence, robotics, advanced technologies and HRM: A systematic review. The International Journal of Human Resource Management, 33(6), 1237–1266. https://doi.org/10.1080/09585192.2020.1871398

2. Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. International Journal of Human Resource Management, 28(1), 3–26.

3. Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency.

4. Sharda, R., Delen, D., & Turban, E. (2014). Business intelligence and analytics: Systems for decision support. Pearson Education.

5. Taslim, W. S., Rosnani, T., & Fauzan, R. (2025). Employee involvement in AI-driven HR decision-making: A systematic review. SA Journal of Human Resource Management, 23, a2856.

6. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organisational trust. Academy of Management Review, 20(3), 709–734.

7. Gunning, D. (2017). Explainable AI (XAI). Defense Advanced Research Projects Agency (DARPA). — foundational work on explainability.

8. Boudreau, John W. & Cascio, Wayne F. (2017). Human capital analytics: Why are we not there? Journal of Organizational Effectiveness: People and Performance, 4(2), 119–126.

9. Minbaeva, Dana B. (2018). Building credible human capital analytics for organizational competitive advantage. Human Resource Management, 57(3), 701–713.

10. Davenport, Thomas H. & Ronanki, Rajeev (2018). Artificial intelligence for the real world. Harvard Business Review, 96(1), 108–116.

11. Jarrahi, Mohammad Hossein (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. Business Horizons, 61(4), 577–586.