ISSN 2395- 4981

JOURNAL OF MANAGEMENT AND INNOVATIVE INFORMATION TECHNOLOGY

Peer reviewed Annual Journal
Vol. 12 No.1 April 2026

Statement about Ownership and other particulars about Journal of Management and Innovative Information Technology

Form IV

Place of publication : MarThoma Institute of Information Technology,Chadayamangalam.P.O Ayur, Kollam (Dist.) Kerala PIN -691534

Periodicity of publication : Annually

Publisher's Name : Prof. Dr V Ambilikumar

Nationality : Indian

Address : MarThoma Institute of Information Technology, Chadayamangalam.P.O Ayur, Kollam (Dist.) Kerala PIN -691534

I, Prof. Dr. V. Ambilikumar, hereby declare that the particulars given above are true to the best of my knowledge and belief.

April 2026 *Sd/-*

INDEX

HUMAN AI COLLABORATION IN COMPENSATION DECISION MAKING

Arun Ignatius V

Research Scholar, Department of Management Studies,
Vels Institute of Science, Technology & Advanced Studies, Chennai;
Email: contactarun.vijayan@gmail.com;
Phone: +91 8056333638

Dr. P. Brindha

Assistant Professor, Department of Management Studies,
Vels Institute of Science, Technology & Advanced Studies, Chennai;
Email: brinda.mba@gmail.com; Phone: 9841126910

Abstract

Compensation decisions are simultaneously data-intensive, high-stakes, and deeply contextual. This paper develops a governance-ready framework for human–AI collaboration in compensation decision-making, positioning AI as an advisory system that augments rather than replaces professional judgment. The framework integrates five layers—Strategy & Guardrails, Data & Privacy, Models & Explainability, Decision Workflows (Human-in-the-Loop), and Governance & Assurance—linked by a closed-loop learning cycle for continuous improvement. We articulate expected impacts (decision quality, consistency, equity, and efficiency), managerial and regulatory implications (including pay transparency readiness), and a conceptual figure that translates design principles into practice. Research objectives and questions are specified to enable empirical evaluation via staged rollouts with fairness diagnostics and explainability embedded. We discuss limitations (data quality, subgroup sizes, change management, and cross-border comparability), propose a pragmatic methodology (stepped-wedge design, difference-in-differences, fairness metrics, power analysis, and measurement-error mitigation), and outline expected outcomes and contributions. The approach provides a scalable blueprint for global organizations to strengthen trust and performance while maintaining equity and compliance under evolving disclosure norms.

Keywords: Compensation decisions, human–AI collaboration, pragmatic methodology.

Introduction:

Compensation decision-making has changed from being solely a human judgment process to a collaborative system utilizing cutting-edge Artificial Intelligence (AI) capabilities in today's data-driven corporate context. The term "human-AI collaboration" in pay decision making refers to the combined use of AI's capacity to evaluate vast amounts of market and employee data with human knowledge, ethical reasoning, and contextual awareness. The goal of this partnership is to minimize prejudice and increase accuracy while creating compensation structures that are competitive, fair, and performance-aligned.

By evaluating elements including internal equity, market salary benchmarks, skill levels, employee performance measurements, and organizational budgets, AI systems assist in compensation decisions. These algorithms can quickly and accurately identify wage disparities, forecast retention concerns, and suggest salary adjustments. However, algorithms cannot handle the ethical, legal, and emotional ramifications of compensation judgments on their own.

Interpreting AI discoveries, ensuring labor law compliance, taking business culture into account, and applying empathy and moral judgment all require human engagement. In order to make final judgments that are consistent with company principles and strategic goals, managers and HR specialists verify AI recommendations and contextualize them according to specific situations.

As a result, human judgment and accountability are combined with AI's analytical capabilities in compensation decision-making. In an increasingly complex workplace, this balanced approach improves transparency, fairness, and strategic alignment, empowering employers to make well-informed, equitable, and long-lasting compensation decisions.

Research questions

RQ1: How does human–AI collaboration influence the consistency and accuracy of compensation decisions across teams and regions?

RQ2: Under what governance and workflow conditions do AI-assisted recommendations improve equity outcomes without introducing new risks?

RQ3: How do explainability and override protocols affect trust, perceived fairness, and adoption among managers and employees?

RQ4: What is the cost-effectiveness of the collaborative approach compared to traditional processes?

Objectives of the Study

1. To Design and validate a governance-ready architecture for human–AI collaboration in compensation decisions.
2. To evaluate effects on decision quality, speed, and consistency relative to traditional baselines.
3. To assess equity outcomes (parity, compression, unexplained variance) under HITL safeguards.
4. To examine trust and adoption among managers and employees with localized explanations and override protocols.
5. To determine global scalability under varying legal and cultural conditions, including pay-transparency requirements.

Need and Significance of the Study:

This work translates responsible AI principles into a governance-ready operating model for compensation. By uniting explainability, human oversight, fairness diagnostics, and constraint-aware optimization, it provides a replicable architecture for global organizations facing expanding pay-transparency and algorithmic accountability. Because it is framed as an implementable design and evaluation plan rather than a theory-anchored literature synthesis, it preserves flexibility for a later systematic review.

Literature Review

According to Kim, Schweitzer, Riedl, and De Cremer (2025), even in situations when performance levels are identical, employees who use AI tools are frequently paid less than those who do not due to a behavioral bias in compensation decisions. Their results demonstrate how important human judgment is in mitigating these biases and guaranteeing fair compensation outcomes in AI-augmented workplaces.

In thorough research of AI in HRM, Aksoy (2023) points out that applications of AI in remuneration enhance pay equity analysis, efficiency, and bias identification. The author warns that algorithmic bias and explainability issues could have a detrimental impact on compensation fairness in the absence of human monitoring. The study suggests a human-in-the-loop method for making ethical compensation decisions.

Employee perceptions of AI-supported HR choices are examined by Ayagoz et al. (2022), who discover that workers frequently view AI-driven remuneration decisions as less equitable and sympathetic than human decisions. The study emphasizes how human participation in compensation decisions enhances views of procedural and interactional justice, which boosts confidence in AI-assisted systems.

Jarrahi (2018) highlights how crucial human-AI cooperation is while making difficult managerial choices. The author claims that human intuition, ethical reasoning, and contextual awareness are necessary for high-stakes, value-laden compensation judgments that AI cannot make on its own. While people guarantee accountability and transparency, AI serves as a decision-support tool, improving accuracy and consistency.

Conceptual Framework

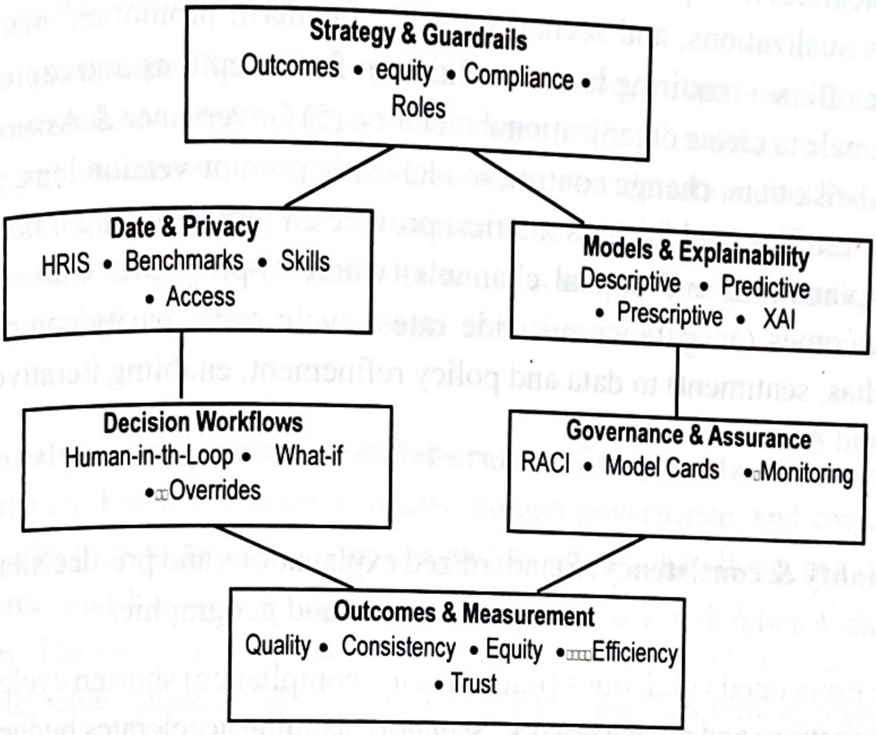

Figure 1. Source: Author's model on the Conceptual framework for human–AI collaboration in compensation

We treat compensation decisioning as a socio-technical system in which algorithms structure complexity while humans provide contextual judgment, accountability, and ethical stewardship. The proposed framework contains five interdependent layers. (1) Strategy & Guardrails: define outcomes (attraction, retention, internal mobility), equity principles (equal pay for equal work; compression safeguards), geographic compliance constraints, disclosure standards, and role boundaries for HR/COE, line leaders, and AI services; set minimum thresholds for model performance, stability, explainability, and fairness. (2) Data & Privacy: curate a minimum-necessary data catalog spanning HRIS (job architecture, ranges, compensation history), market benchmarks, performance and skills signals, recognition and benefits utilization, and contextual attributes (business unit, location); apply privacy-by-design controls—role-based access, de-identification for modeling, purpose limitation, retention schedules, and audit trails—anticipating pay-transparency norms. (3) Models & Explainability: combine descriptive equity baselines (internal parity, range penetration, compression) with predictive signals (attrition risk, market drift) and prescriptive aids (what-if tools; multi-objective optimization under fairness and compliance constraints). Each recommendation displays local explanations (top drivers, constraint checks, confidence bands) to support informed acceptance or override. (4) Decision Workflows (Human-in-the-Loop): integrate AI outputs as advisory inputs—proposed ranges, exception flags, equity/compliance visualizations, and scenario testing for merit, promotion, market correction, and retention offers—requiring human validation for exceptions and capturing a succinct override rationale to create organizational memory. (5) Governance & Assurance: formalize RACI, model-risk tiers, change control, model cards/prompt-version logs, and monitoring of drift, error profiles, and fairness metrics; provide employee-facing transparency (plain-language explanations and appeal channels) where appropriate. Closed-loop learning connects outcomes (acceptance/override rates, cycle time, parity/compression trends, retention deltas, sentiment) to data and policy refinement, enabling iterative co-adaptation of humans and AI.

Impacts

Decision quality & consistency: Standardized explanations and pre-decision checks reduce ad-hoc variance and policy violations across units and geographies.

Efficiency: Embedded validations (range, parity, compliance) shorten cycle times for merit and off-cycle actions and reduce rework. Scenario planning accelerates budgeting and market pricing, freeing experts for strategic dialogue.

Equity & risk mitigation: Early detection of compression and internal parity deviations enables proactive, targeted interventions; structured overrides build a defensible audit trail.

Trust & adoption: Human-readable explanations, visible policy checks, and manager override documentation reinforce perceived legitimacy of AI-assisted decisions and support sustainable adoption.

Operational KPIs & dashboards: Executive tiles track volume, percent reviewed, variance across comparable cases, equity indicators (unexplained variance, compression), confidence bands, and cycle times; deep-dive views show trends by business unit, job family, and geography, plus low-confidence cases routed to human review.

Implications

Operating model: HR shifts from episodic compensation events to a platform-like operating model with evergreen data governance, continuous monitoring, and iterative design changes.

Manager capability: Managers become stewards of transparent, AI-informed choices; they require training to interpret explanations, recognize bias, and document rationale.

Policy evolution: Policies should be machine-interpretable (clear thresholds, formulas) and human-legible (plain language), supporting defensibility under disclosure regimes.

Global localization: A configurable governance layer accommodates jurisdictional differences without fragmenting core standards.

Implementation roadmap: Phase 0 (6–8 weeks): inventory data, map decision flows, baseline parity/compression, define thresholds. Phase 1 (8–12 weeks): pilot for one job family/two geographies, embed explanations, collect override rationales. Phase 2 (12–20 weeks): extend optimization with fairness constraints, add scenario cockpit, expand to promotion/retention; publish model cards. Phase 3 (continuous improvement): quarterly fairness audits, threshold recalibration, manager refresher training.

Discussion

From assistive analytics to accountable collaboration: Compensation decisions sit at the intersection of market dynamics, internal equity, budget governance, and employee trust. Algorithms are effective at pattern discovery and constraint handling, but they do not substitute for contextual knowledge about career narratives, organizational priorities, or local legal norms. The collaborative paradigm therefore reframes AI as an advisory actor that proposes defensible options and surfaces risks, while humans remain responsible for outcomes.

Comparing HITL governance models: In practice, organizations adopt one of three archetypes: (A) Advisory only (AI produces recommendations; humans approve all actions); (B) Thresholded automation (AI can auto approve low risk, in policy cases; humans review exceptions or low confidence cases); and (C) Split routing oversight (AI routes to different reviewers based on risk tier, geography, or policy domain). Advisory only maximizes human control and trust but may slow throughput. Thresholded automation offers efficiency by allowing low impact, high confidence, in policy decisions to proceed quickly, provided that (i) policy codification is mature; (ii) confidence estimation is well calibrated; and (iii) random sampling audits ensure quality. Split routing oversight scales well in global settings where jurisdictional rules, disclosure norms, or works council practices differ; however, it requires robust metadata, routing logic, and reviewer training to prevent uneven standards. Across all three archetypes, documented overrides and periodic calibration are non negotiable.

Designing explanations that change decisions: Explanations must be local (for the specific employee/job), compact (3–5 salient drivers), and constraint aware (which policy limits were binding). Pairing textual rationales with micro visuals—range position, internal parity indicators, and budget impact—helps managers reason and justifies overrides. Counterfactual aids (e.g., "if range midpoint is moved +3%, internal parity would be violated for peer group X") discourage unintentional policy drift and make trade offs explicit.

Bias controls across the lifecycle: Bias risk arises in (i) data collection (historical inequities), (ii) modeling (feature selection, error asymmetries); and (iii) deployment (where and how recommendations are used). A layered control system combines: pre processing (reweighting, representation checks), in processing (fairness constraints/regularizers), and post processing (calibration and disparate error analysis). HITL checkpoints then focus reviewer attention where models are least certain or policy exposure is highest.

Targeting logic and equity guardrails: Where compensation adjacent interventions are considered (e.g., targeted retention adjustments), heterogeneous treatment effect (HTE) and uplift methods guide who is likely to benefit most. Deployment must be embedded in fairness constrained optimization to ensure that targeting efficiencies do not degrade internal parity or increase compression.

Failure modes and mitigations: Common pitfalls include: (1) Rubber stamping—reviewers approve AI outputs without scrutiny when caseloads are high. Mitigation: cap daily caseloads, use confidence threshold routing, and sample high confidence cases for audit. (2) Spec creep—policies are informally stretched in repeated exceptions. Mitigation: make exception templates explicit and report exception trends monthly. (3) Model drift—market shifts or organizational changes degrade recommendations. Mitigation: set drift monitors for input distributions and outcome gaps; trigger retraining windows and policy re baselining. (4) Explainability theater—verbose reports that do not influence behavior. Mitigation: co design explanation layouts with end users; A/B test which elements drive higher decision quality and parity compliance.

Global deployment and localization: Cross border rollouts benefit from a stable core (equity baselines, optimization goals, standard explanations) and pluggable localization modules handling currency, ranges, disclosure granularity, and stakeholder engagement. Works council settings may require early involvement and specific documentation. A design authority—an expert guild spanning Comp, Legal, HRIS, and Data Science—can approve policy encodings, prompts/model cards, and release trains.

Operating rhythm: Monthly fairness councils review metrics (unexplained variance, compression, disparity ratios), exception analytics, and override themes; quarterly model cards summarize performance, calibration, drift, and corrective actions; semi annual policy reviews recalibrate thresholds or ranges based on market and internal equity trends.

What good looks like: Mature programs demonstrate: (a) high coverage of explanation review; (b) low reclassification rates after review (indicating good model policy alignment); (c) stable or improving parity/compression indices; (d) reduced variance in like for like cases; (e) short, well documented cycle times; and (f) clear employee facing narratives aligned with pay transparency requirements.

Limitations

1) Data availability and quality. Sparse or noisy features (skills, performance) reduce reliability of recommendations and fairness checks; data quality programs should run in parallel with model deployment.

2) Subgroup sample sizes. Small populations limit statistical power for fairness monitoring; pooled analysis and longer observation windows can help.

3) Change management. Without training and incentives, HITL devolves into rubber-stamping; leadership sponsorship and periodic calibration are critical.

4) Context dependence. Localization demands may limit cross-country comparability and require tailored workflows and disclosures.

Expected Outcomes

1) Higher decision quality/consistency with fewer policy violations and narrower dispersion in like-for-like cases.

2) Shorter cycle times and reduced rework through embedded validations and scenario tools.

3) Stabilized or improved parity/compression metrics due to proactive flagging and fairness constraints.

4) Increased trust and adoption among managers and employees, supported by concise explanations and transparent override pathways.

5) Improved budget efficiency via targeted, constraint-aware allocations.

Conclusion

Human–AI collaboration in compensation is best understood as disciplined augmentation: algorithms structure complexity and surface options; humans apply context, ethics, and accountability. The five-layer framework, HITL workflows, and closed-loop learning together enable higher-quality, more consistent, and fairer pay decisions at scale. With staged evaluation, embedded explainability, and fairness-constrained optimization, the approach is well positioned for empirical testing and responsible global scaling.

References

1. Aksoy, M. S. (2023). Artificial intelligence in human resource management: A literature review. Journal of Organizational Change and Culture.
2. Ayagoz, G., et al. (2022). AI decision-making with dignity?. Information Systems Frontiers.
3. Cullen, Z. (2024). Is Pay Transparency Good? Journal of Economic Perspectives, 38(1), 153–180. https://doi.org/10.1257/jep.38.1.153
4. Hu, J., de Haan, E., & Skiera, B. (2024). Uplift modeling with quasi-loss-functions. Data Mining and Knowledge Discovery. https://doi.org/10.1007/s10618-024-01042-x
5. Jabagi, N., Croteau, A.-M., Audebrand, L. K., & Marsan, J. (2025). Do algorithms play fair? Analysing the perceived fairness of HR-decisions made by algorithms and their impacts on gig-workers. The International Journal of Human Resource Management, 36(2), 235–274. https://doi.org/10.1080/09585192.2024.2441448
6. Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in HR recruitment and development. Business Research, 13, 795–848. https://doi.org/10.1007/s40685-020-00134-w
7. Kim, T., Schweitzer, M. E., Riedl, C., & De Cremer, D. (2025). People reduce workers' compensation for using artificial intelligence. Organizational Behavior and Human Decision Processes.
8. McKay, M. H. (2024). Realizing the promise of AI governance involving humans-in-the-loop. In HCI International 2024 – Late Breaking Papers (LNCS 15382, pp. 107–123). https://doi.org/10.1007/978-3-031-76827-9_7.
9. Stanford Report (Mello, M., et al.). (2026, January 6). AI-driven insurance decisions raise concerns about human oversight. Health Affairs (news summary). https://news.stanford.edu/stories/2026/01/ai-algorithms-health-insurance-care-risks-research.
10. WTW (Wakelin, R., & Rakyan, S.). (2025, January 14). The impact of AI on the 'how' and 'who' of employee pay practices. Retrieved from WTW Insights: https://www.wtwco.com/en-us/insights/2025/01/the-impact-of-ai-on-the-how-and-who-of-employee-pay-practices.
11. XAI in Employment Decision-Making (Fabeyo, S.). (2025). Issues in Information Systems, 26(3), 127–135. https://doi.org/10.48009/3_iis_2025_2025_110.