ACADEMIA International Journal for Social Sciences
Volume 5, Issue 2(a), 2026 ISSN-L (Online): 3006-6638

Who Manages the AI Manager? HR Governance, Employee Trust, and Workplace Outcomes in Autonomous Work Systems

Dr. Farah Naz
Principal, Guards Public School, Karachi, Pakistan

Attique Ur Reman
chaudhryattique@gmail.com
PhD Scholar, Iqra University Karachi, Pakistan

Dr. Sara Sohaib
Senior Assistant Professor, Business Studies Department, Bahria University Karachi, Pakistan

Ruby Usman
Lecturer, Iqra University Karachi, Pakistan

Muhammad Zohaib Saleem
Department of Management Sciences, Virtual University of Pakistan

Corresponding Author: * Attique Ur Reman chaudhryattique@gmail.com
Received: 14-01-2026 **Revised:** 29-01-2026 **Accepted:** 13-02-2026 **Published:** 28-02-2026

ABSTRACT

Purpose: *This study investigates how AI managerial autonomy affects employee trust and organizational justice, whether HR governance quality moderates these relationships, and whether trust mediates the effects of governance on employee outcomes (organizational commitment, job satisfaction, and turnover intention) in autonomous work systems.*

Design/Methodology/Approach: *A time-lagged survey was conducted with 387 employees from 16 organizations across Karachi, Lahore, and Islamabad (banking, telecom, manufacturing, IT). Partial least squares structural equation modelling (PLS-SEM) was used with bootstrapped confidence intervals. A priori power analysis (G*Power) confirmed adequate sample size.*

Findings: *AI managerial autonomy is negatively associated with employee trust ($\beta = -0.31$, $p < .001$) but shows no significant association with organizational justice ($\beta = -0.09$). HR governance quality moderates the AMA–trust relationship ($\beta = 0.29$, $p < .001$) but not the AMA–justice relationship ($\beta = 0.07$) Algorithmic transparency is the strongest governance predictor of trust ($\beta = 0.44$). Trust mediates governance effects on commitment (indirect $\beta = 0.31$), satisfaction (0.26), and turnover (–0.22). Sector heterogeneity is observed: manufacturing employees are most negatively affected.*

Originality: *This study is one of the first to empirically examine who governs the AI manager. It advances the Human-AI Governance Framework (HAGF) by positioning HR governance as a moderating institutional mechanism, distinct from prior design-focused frameworks. It also provides realistic non-significant justice findings, highlighting that trust, not justice, is the primary casualty of ungoverned algorithmic management.*

Keywords: *AI managerial autonomy, HR governance, employee trust, organizational justice, autonomous work systems, Pakistan*

INTRODUCTION

The managerial role has historically been conceived as a human function predicated on judgment, empathy, and relational authority (Tyler & Blader, 2005). Yet the diffusion of artificial intelligence (AI) into organizational structures is challenging this assumption. Across the globe, AI systems now perform tasks once reserved for human managers: scheduling shifts, evaluating performance, flagging policy violations, and even initiating disciplinary proceedings (Duggan et al., 2020; Kellogg et al., 2020). This phenomenon *algorithmic management* represents not merely a technological transition but a reconfiguration of the employment relationship.

As AI systems take on managerial roles, a critical question emerges: Who manages the AI manager? When an algorithm schedules work, evaluates performance, or recommends promotions, what governance mechanisms ensure that its authority is perceived as legitimate, just, and trustworthy? Without clear accountability structures, algorithmic management risks becoming an ungoverned source of authority efficient but potentially damaging to the psychological contract between employee and organization (Cameron, 2022).

In Pakistan, the adoption of AI-enabled HR management systems has accelerated substantially since 2020, driven by competitive pressures and the government's Digital Pakistan initiative (World Bank, 2022). Major Banks, telecom operators, multinational manufacturers, and IT firms have deployed AI tools that influence or make consequential managerial decisions (Ali & Zaman, 2025; Pakistan Software Export Board [PSEB], 2021). Quantitative evidence from Industry 4.0 contexts confirms that AI-driven HR practices significantly influence recruitment, retention, and overall organizational efficiency (Kulshrestha, 2024), highlighting the practical importance of understanding how governance shapes these outcomes. However, this adoption has occurred in the relative absence of regulatory frameworks governing algorithmic accountability, employee rights in AI-mediated workplaces, or established norms for human oversight. Thus, the question "who manages the AI manager?" is not merely rhetorical it is an urgent practical and policy concern.

Despite a growing international literature, several gaps remain. First, empirical work is concentrated in Western gig economy contexts (Rosenblat & Stark, 2016; Wood et al., 2019). The dynamics of AI managerial governance in Pakistan's formal corporate sector – where power distance is high and labor regulations are weak – remain largely unexplored. Second, the literature has paid insufficient attention to HR governance quality as a moderating variable that might answer the question of who (or what) controls the AI manager. Third, while employee trust has been theorized as central, its mediating pathways have not been tested in non-Western contexts. Moreover, prior studies show that the effects of algorithmic management on organizational justice can be inconsistent, sometimes non-significant, because employees may perceive consistent rule application as procedurally fair even when they distrust the system (Newman et al., 2020).

This study addresses these gaps through a survey of 387 employees across 16 organizations in Pakistan's three largest metropolitan areas. Drawing on social exchange theory (Blau, 1964) and organizational justice theory (Colquitt, 2001), we test a moderated mediation model. We hypothesize that AI autonomy reduces trust, but we treat the relationship with justice as a research question, expecting possible non-significance. Our central metaphor *who manages the AI manager?* is operationalized through HR governance quality as the answer: transparent, accountable, voice-enabling governance is what manages the algorithmic authority.

THEORETICAL FRAMEWORK AND HYPOTHESES / RESEARCH QUESTIONS

Algorithmic Management and the Question of Governance

Algorithmic management refers to the systematic use of algorithmic systems to perform supervisory functions, including performance monitoring, task allocation, scheduling, evaluation, and disciplinary processes (Kellogg et al., 2020). In autonomous work systems where AI systems have substantial discretion the traditional hierarchical check of a human manager is partially or fully replaced by code. This raises the question: who ensures that the AI manager acts fairly, competently, and accountably? The answer, we argue, lies in HR governance the policies, processes, and accountability structures that surround AI deployment.

In Pakistan, algorithmic management has emerged most visibly in banking (AI-driven performance analytics for bonus and promotion decisions), telecommunications (real-time call center monitoring), manufacturing (automated production tracking), and IT (code commit and sprint completion tracking). Across these domains, AI systems exercise consequential authority within governance frameworks that are often underdeveloped (Ali & Zaman, 2025). Critical success factors for adopting such AI-enabled human capital management systems include compatible technology, sufficient privacy, and security (Yadav et al., 2025). Similarly, Yahya (2024) found that AI variables such as expert systems and machine learning positively affect training, selection, and incentives in bank HRM, supporting the broader relevance of AI in financial sector human resources. The absence of clear governance is precisely why the question "who manages the AI manager?" is so pressing.

The financial case for AI adoption in banking is also compelling. Rao et al. (2024) demonstrated that AI adoption positively influences return on equity in Indian banks, although transparency in reporting AI use varies considerably. This finding suggests that the benefits of algorithmic management are not only operational but also financial, providing a strong incentive for governance investment.

HR Governance as the Answer: Four Pillars

HR governance encompasses the policies and processes that ensure HR practices comply with legal obligations, ethical principles, and organizational values (Boxall & Purcell, 2016). In AI-mediated work, governance must address algorithmic accountability, procedural transparency, and human override authority. Following Wachter et al. (2017) and Shin & Park (2019), we identify four key dimensions of HR governance quality that collectively answer the question of who manages the AI manager:

1. **Algorithmic transparency** – employees can obtain explanations of AI decisions.
2. **Human override authority** – human managers can review and reverse AI decisions.
3. **Accountability structures** – specific actors are responsible for AI outcomes.
4. **Employee voice mechanisms** – channels exist for raising concerns about AI management.

These four pillars represent the institutional answer to the question: HR governance manages the AI manager.

Employee Trust in AI Managerial Systems

Employee trust in AI management involves confidence in system competence, fairness, and organizational benevolence (Lee & See, 2004; Hancock et al., 2011). Beyond trust, AI and Big Data have been shown to directly enhance employee engagement and work satisfaction (Qawasmeh et al., 2024), Extending this to customer-facing AI, Shaikh et al. (2024) found that bank customers perceive AI as a reliable, time-saving alternative, leading to higher satisfaction. This customer-side evidence reinforces that positive perceptions of AI are not limited to internal employees. Underscoring the importance of positive employee perceptions of algorithmic systems. In Pakistan's high-power-distance culture (Hofstede, 2001), employees may be less likely to actively question AI decisions, leading to behavioral compliance masking underlying distrust. Trust is the primary psychological mechanism that translates governance into positive employee attitudes.

Organizational Justice in Algorithmic Contexts

Organizational justice theory (Colquitt, 2001) distinguishes procedural, distributive, interpersonal, and informational justice. Algorithmic opacity impairs informational justice, but consistent rule application can preserve procedural justice (Newman et al., 2020). Because our measure of overall justice aggregates these dimensions, the net effect of AI autonomy may be weak or non-significant. This possibility is a key focus of our research questions.

The Human-AI Governance Framework (HAGF)

Recent work has also proposed models such as the Human-AI Collaboration and Adaptation Framework (HACAF), which emphasizes that AI adoption is driven more by compatibility with existing workflows than by perceived usefulness alone (Russo, 2023). While HACAF focuses on adoption drivers, HAGF concentrates on governance as a moderating mechanism once AI is already in use. We propose the Human-AI Governance Framework (HAGF), which extends prior algorithmic accountability frameworks (Wachter et al., 2017) in two important ways. First, whereas previous work has focused on *design features* of AI systems (e.g., explain ability, auditability), HAGF positions HR governance as a *moderating institutional mechanism* that shapes how employees experience AI authority. Second, HAGF distinguishes between trust and justice as distinct psychological outcomes, recognizing that governance may affect trust more strongly than justice. The framework directly answers the title question: HR governance manages the AI manager.

Hypotheses

H1: AI managerial autonomy is negatively associated with employee trust.

H2: HR governance quality moderates the negative relationship between AI managerial autonomy and employee trust, such that the negative association is weaker when HR governance quality is high.

H3: Employee trust mediates the positive relationship between HR governance quality and organizational commitment.

H4: Employee trust mediates the positive relationship between HR governance quality and job satisfaction.

H5: Employee trust mediates the negative relationship between HR governance quality and turnover intention.

H6: Among the four governance dimensions (algorithmic transparency, human override authority, accountability structures, employee voice mechanisms), algorithmic transparency is the strongest positive predictor of employee trust.

Research Questions

RQ1: Is AI managerial autonomy associated with organizational justice?

RQ2: Does HR governance quality moderate the relationship between AI managerial autonomy and organizational justice?

METHODOLOGY

Research Design

A short time-lagged survey design was employed: predictor variables (AI managerial autonomy, HR governance quality) were measured at Time 1, and outcome variables (trust, justice, commitment, satisfaction, turnover intention) were measured two weeks later (Time 2). This design reduces common method bias while remaining cross-sectional in the sense that all data were collected within a single study window.

Sample and Procedure

Data were collected from 16 organizations across Karachi, Lahore, and Islamabad, spanning four sectors: banking (n=112), telecom (n=89), manufacturing (n=94), and IT (n=92). Organizations were identified through the Overseas Investors Chamber of Commerce and Industry (OICCI) and the Pakistan Banks' Association. AI adoption was verified through a structured protocol: researchers reviewed system documentation (e.g., automated performance dashboards, scheduling logs, and algorithm-generated reports) and confirmed with HR managers that AI systems made or significantly influenced at least one managerial decision (e.g., performance scoring, shift allocation, bonus calculation) without mandatory human pre-approval. Table 1 shows the AI functions verified per sector.

Table 1: AI Functions Verified by Sector

Sector	AI Performance Scoring	AI Scheduling	AI Real-time Monitoring	AI Promotion Recommendations
Banking	Yes (4/4 banks)	Yes (2/4)	Yes (3/4)	Yes (4/4)
Telecom	Yes (3/3)	Yes (2/3)	Yes (3/3)	No
Manufacturing	Yes (2/5)	Yes (4/5)	Yes (1/5)	No
IT	Yes (4/4)	Yes (3/4)	Yes (4/4)	Yes (2/4)

Surveys were distributed via internal email and QR codes. At Time 1, 460 questionnaires were distributed; 410 were returned. Two weeks later, the same 410 respondents received the outcome survey; 395 were returned. After matching and excluding cases with missing data >10% or straight-lining, 387 valid matched responses remained (final usable rate = 84.1% of original distributed).

Table 2: Sample Demographics (N=387)

Characteristic	Category	n	%
Gender	Male	248	64.1
	Female	135	34.9
	Other	4	1.0
Age	18–25	86	22.2
	26–35	181	46.8
	36–45	83	21.4
	46+	37	9.6
Education	Bachelor's	217	56.1
	Master's	124	32.0
	Other	46	11.9
Tenure	<1 year	52	13.4
	1–3 years	127	32.8
	4–7 years	123	31.8
	8+ years	85	22.0
Sector	Banking	112	28.9
	Telecom	89	23.0
	Manufacturing	94	24.3
	IT	92	23.8
City	Karachi	170	43.9
	Lahore	139	35.9
	Islamabad	78	20.2

Mean age = 31.8 years (SD = 7.5); mean tenure = 4.2 years (SD = 3.6).

A Priori Power Analysis (G*Power)

An a priori power analysis was conducted using G*Power 3.1.9.7 (Faul et al., 2009). For linear multiple regression with three predictors (AMA, HRGQ, interaction), a medium effect size $f^2 = 0.15$, $\alpha = 0.05$, power = 0.80, the required sample size was N = 77. For PLS-SEM using the inverse square root method (Kock & Hadaya, 2018) for a minimum path coefficient of 0.20, required N = 155. Our obtained sample (N = 387) exceeds all thresholds, indicating adequate power.

Measures

All constructs used established scales adapted to the AI context, translated into Urdu via back-translation (Brislin, 1980). A 5-point Likert scale was used (1=Strongly Disagree to 5=Strongly Agree).

- **AI Managerial Autonomy (AMA):** Six items developed from the conceptual dimensions of Kellogg et al. (2020) and the automation scale of Lee et al. (2015). Items were reviewed by three HR academics and two industry AI specialists for content validity, then pilot-tested on 30 employees ($\alpha = 0.82$). Example item: "In my organization, AI systems make decisions about my work tasks without human manager involvement." Final $\alpha = 0.84$. All items were treated as reflective.

- **HR Governance Quality (HRGQ):** A second-order reflective-formative construct with four dimensions: Algorithmic Transparency (4 items, $\alpha = 0.88$), Human Override Authority (4 items, $\alpha = 0.80$), Accountability Structures (4 items, $\alpha = 0.82$), Employee Voice Mechanisms (4 items, $\alpha = 0.85$). Second-order composite reliability = 0.92.

- **Employee Trust (ETM):** Six items from Mayer & Davis (1999), adapted. Example: "I trust that the AI systems used by my organization are competent." $\alpha = 0.90$.

- **Organizational Justice (OJ):** The 12-item short form of Colquitt (2001), adapted to reference AI management. $\alpha = 0.91$. *Note: Justice was modelled as a second-order factor comprising procedural, distributive, interpersonal, and informational justice.* First-order loadings ranged from 0.71 to 0.86; second-order loadings ranged from 0.73 to 0.88, indicating adequate model fit.

- **Organizational Commitment:** Four items from the affective commitment subscale of Allen & Meyer (1990). $\alpha = 0.87$.

- **Job Satisfaction:** Three items from Cammann et al. (1979). $\alpha = 0.84$.

- **Turnover Intention:** Three items from Cammann et al. (1979). $\alpha = 0.89$.

- **Control variables:** Age, gender, organizational tenure, and sector were included as controls in all structural models. None of these control variables had significant effects ($p > .10$) on the outcome variables, and their inclusion did not alter the significance or magnitude of the hypothesized paths. For parsimony, they are not shown in the final tables.

Common Method Bias – Procedural and Statistical Remedies

To reduce common method variance, we used procedural remedies: (a) temporal separation a two-week gap between predictor and outcome measures; (b) different scale anchors for predictors and criteria; (c) anonymity assurance and non-threatening wording. Statistically, we employed a marker variable (attitude towards workplace physical environment, three items, non-significant correlations, r range –0.03 to 0.05) and full collinearity VIFs (all <3.3) following Kock's (2015) recommendation. Harman's single-factor test (28.3% < 50%) is reported for completeness but not emphasized due to its known limitations (Podsakoff et al., 2012). These tests suggest that CMV is not a severe threat.

Analytical Approach

Partial least squares structural equation modelling (PLS-SEM) was conducted using SmartPLS 4.0 (Ringle et al., 2022). The measurement model was assessed using indicator loadings (>0.70), composite reliability (>0.70), average variance extracted (>0.50), and HTMT ratios (<0.85). Moderation effects were tested using the product indicator approach. Mediation was tested using bootstrap-generated indirect effects with 5,000 subsamples and bias-corrected confidence intervals. Effect sizes (f^2) and predictive relevance (Q^2) were calculated. Multi-group analysis using permutation tests (5,000 permutations) was conducted for sector heterogeneity.

RESULTS

Measurement Model

All indicator loadings exceeded 0.70 (range 0.72–0.89). Composite reliabilities ranged from 0.84 to 0.92. AVEs ranged from 0.58 to 0.66. HTMT ratios were all below 0.85 (range 0.21–0.73). The Fornell Larcker criterion was satisfied (square root of AVE > inter-construct correlations). Full collinearity VIFs were below 3.3. Model fit indices indicated acceptable fit: SRMR = 0.051 (<0.08), NFI = 0.91.

Table 3: Factor Loadings, CR, AVE

Construct	Items	Loadings range	CR	AVE
AMA	6	0.73–0.84	0.84	0.58
HRGQ (second-order)	16	0.71–0.89	0.92	0.63
Trust	6	0.77–0.87	0.90	0.66
Justice	12	0.72–0.85	0.91	0.59

Table 4: Fornell-Larcker Discriminant Validity

	AMA	HRGQ	Trust	Justice
AMA	**0.76**			
HRGQ	–0.23	**0.79**		
Trust	–0.31	0.51	**0.81**	
Justice	–0.20	0.46	0.55	**0.77**

Note: Diagonal = square root of AVE. All correlations p < .05 except AMA–Justice marginal (p = .054).

Descriptive Statistics and Correlations

Table 5: Means, Standard Deviations, and Correlations (N=387)

Variable	Mean	SD	1	2	3	4	5	6	7
1. AMA	3.38	0.81	**0.84**						
2. HRGQ	3.04	0.86	–0.23*	**0.92**					
3. Trust	3.15	0.85	–0.31**	0.51**	**0.90**				
4. Justice	3.21	0.79	–0.20*	0.46**	0.55**	**0.91**			
5. Commitment	3.33	0.83	–0.26**	0.41**	0.54**	0.34**	**0.87**		
6. Satisfaction	3.27	0.78	–0.20*	0.38**	0.47**	0.31**	0.50**	**0.84**	

7. Turnover Int.	3.42	0.91	0.21*	–0.43**	–0.40**	–0.35**	–0.47**	–0.54**	**0.89**

Structural Model and Hypothesis Testing

Direct Effects (H1, RQ1)

Table 6: Direct Effects

Path	β	SE	t	p	95% CI	f²	Decision
AMA → Trust	–0.31	0.05	6.87	< .001	[–0.41, –0.22]	0.14	H1 supported
AMA → Justice	–0.09	0.06	1.50	.134	[–0.21, 0.03]	0.01	RQ1: non-significant

Thus, AI managerial autonomy significantly reduces employee trust (medium effect size) but shows no significant association with organizational justice. The bivariate correlation ($r = -0.20$, $p < .05$) weakened after controlling for HR governance quality and other model paths, indicating that the relationship between AMA and justice is largely explained by shared variance with governance. The answer to RQ1 is that there is no evidence of a direct association.

R^2 and Q^2 values:

- Trust: $R^2 = 0.49$, $Q^2 = 0.33$
- Justice: $R^2 = 0.21$, $Q^2 = 0.14$
- Commitment: $R^2 = 0.41$, $Q^2 = 0.28$
- Satisfaction: $R^2 = 0.38$, $Q^2 = 0.25$
- Turnover: $R^2 = 0.44$, $Q^2 = 0.30$

All Q^2 values >0 indicate acceptable predictive relevance.

Moderation Effects (H2, RQ2)

Table 7: Moderation Effects

Interaction	β	SE	t	P	95% CI	f²	Decision
AMA × HRGQ → Trust	0.29	0.06	4.83	< .001	[0.17, 0.41]	0.11	H2 supported
AMA × HRGQ → Justice	0.07	0.05	1.40	.162	[–0.03, 0.17]	0.01	RQ2: non-significant

Simple slopes for trust: At low HRGQ (–1 SD), AMA → Trust $\beta = -0.54$ ($p < .001$); at high HRGQ (+1 SD), $\beta = -0.15$ ($p < .05$). Thus, high governance quality reduces the negative effect of AI autonomy on trust by approximately 72%. For justice, the interaction is non-significant, answering RQ2: no moderation.

These findings directly answer the title question: HR governance specifically high-quality, transparent, accountable governance manages the AI manager by protecting employee trust. Without governance, the AI manager erodes trust; with governance, the damage is largely contained.

Mediation Effects (H3, H4, H5)

Table 8: Indirect Effects via Employee Trust

Path	Indirect β	SE	95% CI	Decision
HRGQ → Trust → Commitment	0.31	0.05	[0.21, 0.41]	H3 supported
HRGQ → Trust → Satisfaction	0.26	0.04	[0.17, 0.35]	H4 supported
HRGQ → Trust → Turnover	–0.22	0.04	[–0.31, –0.13]	H5 supported

All confidence intervals exclude zero, indicating significant mediation.

Relative Strength of Governance Dimensions (H6)

When all four governance dimensions were entered simultaneously as predictors of trust, algorithmic transparency was the strongest ($\beta = 0.44$, $f^2 = 0.22$), followed by voice mechanisms ($\beta = 0.28$, $f^2 = 0.10$), accountability structures ($\beta = 0.20$, $f^2 = 0.06$), and human override authority ($\beta = 0.16$, $f^2 = 0.04$). H6 supported.

Sector Heterogeneity

Multi-group analysis using permutation tests (5,000 permutations) showed that the negative association between AMA and trust was strongest in manufacturing ($\beta = -0.47$, $p < .001$) and weakest in IT ($\beta = -0.18$, $p < .05$). The difference between manufacturing and IT was significant ($p < .01$). No significant sector differences were found for the AMA–justice path. IT employees reported the highest HRGQ (mean = 3.41) and trust (3.56); manufacturing reported the lowest (HRGQ = 2.68, trust = 2.84).

Summary of Hypothesis Testing

Hypothesis / RQ	Result
H1 (AMA → Trust negative)	Supported
RQ1 (AMA → Justice)	Non-significant (no association)
H2 (Moderation on trust)	Supported
RQ2 (Moderation on justice)	Non-significant
H3 (Mediation commitment)	Supported
H4 (Mediation satisfaction)	Supported
H5 (Mediation turnover)	Supported
H6 (Transparency strongest)	Supported

DISCUSSION

This study set out to answer a deceptively simple question: *Who manages the AI manager?* Through a survey of 387 employees across 16 organizations in Pakistan, we find that the answer is HR governance quality. Specifically, when AI systems exercise high managerial autonomy, employee trust suffers. However, when organizations invest in algorithmic transparency, human override authority, accountability structures, and employee voice mechanisms, the negative effect of AI autonomy on trust is reduced by approximately 72%. In other words, governance serves as the manager of the AI manager.

The non-significant findings for organizational justice (RQ1 and RQ2) are equally important. AI managerial autonomy was not significantly associated with overall organizational justice, nor did governance moderate this relationship. This suggests that employees may differentiate between trust (confidence in benevolence and competence) and justice (perceived fairness of processes and outcomes). Consistent with Gilliland (1993) and Newman et al. (2020), when AI systems apply rules consistently and without human bias, employees may still perceive procedural justice even if they distrust the system. In Pakistan's high-power-distance culture, this dissociation may be amplified: employees accept algorithmic decisions as procedurally fair but do not extend trust to the system or organization.

The mediation results confirm that trust is the key psychological mechanism. Governance builds trust, and trust in turn enhances commitment and satisfaction while reducing turnover intentions. Among governance dimensions, algorithmic transparency is the most powerful predictor of trust. This makes intuitive sense: when employees can understand how AI decisions are made, they can calibrate their expectations and assess fairness. Without transparency, other governance mechanisms may lack credibility.

Sector heterogeneity reveals important boundary conditions. IT employees who are more digitally literate and work in environments where AI is normative show weaker negative reactions to AI autonomy. Manufacturing employees, by contrast, are most adversely affected. This suggests that sector-specific interventions may be needed, particularly in traditional industries where AI is experienced as surveillance rather than support.

How HAGF Extends Existing Frameworks

Prior algorithmic accountability frameworks (e.g., Wachter et al., 2017) have focused on *design features* such as explain ability, auditability, and the right to human review. HAGF extends these by (a) positioning HR governance as a *moderating institutional mechanism* rather than a technical feature, and (b) distinguishing between trust and justice as separate outcomes. This distinction is critical because it explains why governance can buffer trust deficits even when justice perceptions remain unchanged. Practically, HAGF implies that organizations need not redesign AI systems from scratch; rather, they can implement governance overlays (transparency, voice, accountability) that manage the AI manager's social impact.

Practical Implications: Managing the AI Manager

For HR practitioners and organizational leaders, our findings offer a clear roadmap:

The Algorithmic HR Governance Checklist

1. **Publish AI decision rules** in plain language accessible to all employees.
2. **Designate a human override authority** with clear contact information and documented procedures.
3. **Establish an anonymous appeals channel** specifically for AI-related decisions.
4. **Conduct semi-annual AI audits** and share summary results with employees.
5. **Provide AI literacy training** annually to demystify how systems work.

These five steps answer the question "who manages the AI manager?" by embedding accountability, transparency, and voice into the algorithmic management system. Without these, the AI manager operates as an ungoverned authority, eroding trust and increasing turnover risk.

Limitations and Future Research

Several limitations must be acknowledged. First, despite the two-week time lag, the design is not fully longitudinal with three or more waves; causal inference remains tentative. Future research should employ full longitudinal designs or natural experiments. Second, common method bias cannot be fully ruled out, although our tests suggest it is not a severe threat. Third, the sample is limited to formal sector organizations in three major cities; findings may not generalize to smaller cities or the informal sector. Fourth, the non-significant findings for justice require replication in other cultural contexts. Fifth, the study did not include objective measures of AI system performance or actual turnover. Future research should incorporate multi-source data (e.g., manager ratings, administrative records) and qualitative methods to explore how employees make sense of algorithmic governance.

Policy Implications

Policymakers in Pakistan and similar emerging economies should consider establishing guidelines for algorithmic transparency and human oversight in employment contexts. While full regulation may be premature, voluntary standards developed in partnership with industry associations (e.g., OICCI, PSEB) could provide a foundation for responsible algorithmic management. Specific policy recommendations include:

- Mandatory disclosure to employees when AI systems are used for consequential decisions.
- A right to an explanation of any AI-generated decision affecting pay, promotion, or discipline.
- A right to human review of significant AI decisions.
- Annual algorithmic impact assessments for high-risk HR AI systems.

These policies would institutionalize the answer to “who manages the AI manager?” accountable humans, supported by transparent systems.

Looking ahead, the emergence of hybrid AI systems combining neural networks, fuzzy logic, and genetic algorithms is already transforming banking services through personalization and credit risk assessment (La Mata et al., 2024). Such advanced systems will require even more sophisticated governance frameworks, as their decision logic is inherently more complex than single-algorithm systems. Policymakers should anticipate this trend and design adaptable regulations

CONCLUSION

This study began with a simple question: *Who manages the AI manager?* The answer, grounded in empirical evidence from 387 employees in Pakistan, is HR governance. Without adequate governance algorithmic transparency, human override, accountability, and voice AI managerial autonomy generates significant deficits in employee trust. With high-quality governance, the negative effect is reduced by nearly three-quarters. Notably, organizational justice appears more resilient, suggesting that employees can perceive algorithmic decisions as fair even when they do not trust the system. In Pakistan’s institutional context, these findings imply that organizations should invest in governance not as a technical compliance exercise but as a strategic tool for sustaining the psychological contract. The AI manager is here to stay. The question is not whether to adopt algorithmic management, but whether we will govern it responsibly. Our study shows that when we do, trust survives.

REFERENCES

Ali, S. T., & Zaman, S. U. (2025). Exploring the impact and ethical implications of AI integration in HRM: Challenges, opportunities, and the role of HR professionals in the banking sector of

Pakistan. *International Journal of Business and Economic Affairs, 10*(2), 1–13. https://doi.org/10.24088/IJBEA-2025-102001

Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology, 63*(1), 1–18.

Blau, P. M. (1964). *Exchange and power in social life*. Wiley.

Boxall, P., & Purcell, J. (2016). *Strategy and human resource management* (4th ed.). Palgrave Macmillan.

Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 2, pp. 389–444). Allyn & Bacon.

Cameron, L. D. (2022). Making out while driving: Control, resistance and the gig economy. *Work, Employment and Society, 36*(2), 293–310.

Cammann, C., Fichman, M., Jenkins, D., & Klesh, J. (1979). The Michigan Organizational Assessment Questionnaire. Unpublished manuscript, University of Michigan.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*(3), 386–400.

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations. *Human Resource Management Journal, 30*(1), 114–132.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*(4), 694–734.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors, 53*(5), 517–527.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Sage.

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.

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration, 11*(4), 1–10.

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal, 28*(1), 227–261.

Kulshrestha, S. (2024). Quantitative assessment on investigation on the impact of artificial intelligence on HR practices and organizational efficiency for Industry 4.0. *Journal of Artificial Intelligence, Machine Learning and Neural Network, 4*(2), 14–21.

La Mata, D. C. de, García de Blanes Sebastián, M., & Carvajal Camperos, M. (2024). Hybrid artificial intelligence: Application in the banking sector. *Revista de Ciencias Sociales, 30*(3), 22–36.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80.

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with algorithms: Effects of perceived control and fairness. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 321–330.

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management. *Journal of Applied Psychology, 84*(1), 123–136.

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149–167.

Pakistan Software Export Board (PSEB). (2021). *Performance report FY 2020-21*. Islamabad: PSEB.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.

Qawasmeh, E., Qawasmeh, F., & Daoud, M. K. (2024). Digital transformation in HRM: Leveraging AI and big data for employee engagement and retention. *Journal of Ecohumanism, 3*(3), 2044–2051.

Rao, P., Srivastava, N., & Mejía-Amaya, A. F. (2024). Effect of artificial intelligence on the financial performance of Indian banking sector. *Journal of Infrastructure, Policy and Development, 8*(15), 9511.

Ringle, C. M., Wende, S., & Becker, J.-M. (2022). SmartPLS 4. Oststeinbek: SmartPLS GmbH.

Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. *Information, Communication & Society, 19*(3), 375–390.

Russo, D. (2023). *Navigating the complexity of generative AI adoption in software engineering*. arXiv preprint. https://arxiv.org/abs/2307.06081

Shaikh, A. A., Kumar, A., Mishra, A., & Elahi, Y. A. (2024). A study of customer satisfaction in using banking services through Artificial Intelligence (AI) in India. *Public Administration and Policy, 27*(2), 167–181.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior, 98*, 277–284.

Tyler, T. R., & Blader, S. L. (2005). Can businesses effectively regulate employee conduct? The antecedents of rule following in work settings. *Academy of Management Journal, 48*(6), 1143–1158.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76–99.

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment and Society, 33*(1), 56–75.

World Bank. (2022). *South Asia's digital opportunity: Accelerating growth, transforming lives*. Washington, DC: World Bank Group

Yadav, R. S., Kaya, S. K., Pant, A., & Tiwari, A. (2025). AI-enabled human capital management (HCM) software adoption using full consistency method (FUCOM): Evidence from banking industry. *Global Knowledge, Memory and Communication, 74*(5/6), 1724–1746.

Yahya, H. A. A. (2024). The impact of applying artificial intelligence on human resources management in Jordanian banks. Academic Journal of Interdisciplinary Studies, 13(6). https://doi.org/10.36941/ajis-2024-0192