IJC
International Journal of Computer (IJC)
Peer-Reviewed International Journal
ISSN: 2307-4523
https://ijcjournal.org

Responsible Governance of Agentic Artificial Intelligence within Human Resource Management Systems: A Lifecycle Framework for Accountability, Oversight, and Risk Control

Sambit Panigrahi*

Senior HRIT Analyst, VITAS Healthcare, Miramar, Florida, United States

* *Corresponding author: sambit.panigrahi@vitas.com*

Accepted: 23 August 2026 | Published: 5 October 2026

Abstract

Agents based on artificial intelligence (AI) are increasingly embedded within enterprise human resource management systems (HRMS) to automate recruitment, screening, performance, learning, and workforce analytics. Unlike static decision-support tools, agentic AI systems pursue goals, invoke tools, and execute multi-step workflows with limited human supervision, raising governance challenges that existing principles-based frameworks do not fully resolve. This article develops a conceptual governance framework for responsible agentic AI within HRMS. A structured review of peer-reviewed literature and authoritative standards, including the NIST AI Risk Management Framework, ISO/IEC 42001, the EU AI Act, and the OECD AI Principles, identifies the governance gap: existing instruments articulate principles and risk-management functions but are domain-agnostic and do not operationalize controls for agent autonomy, lifecycle dynamics, or HR-specific risks. The proposed Responsible Agentic AI Governance Framework (RAGF) integrates five lifecycle phases, design and authorization, deployment and provisioning, operation and monitoring, audit and assurance, and review and retirement, with six governance control layers and a four-tier risk-control model that maps risk levels to autonomy modes and concrete controls. The framework is analytically evaluated against existing approaches and HR governance requirements. The contribution is an integrated, operational governance architecture that translates abstract responsible-AI principles into lifecycle-oriented controls for agentic AI in HR systems, addressing accountability, transparency, human oversight, auditability, and risk mitigation.

Keywords: Agentic AI; AI governance; responsible AI; human resource management systems; AI accountability; AI risk management; AI lifecycle governance

1. Introduction

The integration of artificial intelligence (AI) into human resource management (HRM) has moved from isolated decision-support tools toward semi-autonomous agents that orchestrate workflows across the employee lifecycle. Modern enterprise human resource management systems (HRMS) increasingly incorporate AI agents that screen candidates, route learning recommendations, flag

performance concerns, and generate workforce analytics with limited step-by-step human supervision [1], [2]. This shift introduces a qualitatively different governance problem: the governing unit is no longer a static model invoked by a human, but an autonomous goal-directed system that invokes tools, accesses data, and executes multi-step actions within sensitive organizational processes [1].

Agentic AI systems differ from earlier predictive models in three governance-relevant respects. First, they exhibit goal-directed autonomy: agents pursue objectives over multiple steps without per-decision human authorization, which complicates accountability for outcomes [1], [3]. Second, they operate over time and across integrated data, so governance must address lifecycle dynamics, drift, and accumulation of risk rather than a single deployment event [3], [4]. Third, HRMS data is rights-sensitive, encompassing hiring, compensation, performance, and workforce analytics, so failures carry legal, ethical, and regulatory consequences under emerging AI regulation [5].

Several authoritative frameworks now shape responsible-AI practice, including the NIST AI Risk Management Framework (AI RMF) [6], ISO/IEC 42001:2023 for AI management systems [7], the European Union AI Act [5], and the OECD AI Principles [8]. These instruments establish essential principles, risk functions, and management-system requirements. However, they are deliberately domain-agnostic and process-level: they specify what governance should achieve but not how it should be operationalized for the specific characteristics of agentic AI within a defined enterprise domain such as HRMS. A systematic literature review of AI governance concludes that the field lacks concrete, operational mechanisms that translate abstract principles into controls for autonomous, tool-using systems [9].

This gap motivates the present study. The research question is: how can responsible governance of agentic AI be operationalized within HRMS so that accountability, oversight, auditability, and risk control are structurally embedded across the agent lifecycle? The article makes three contributions. First, it synthesizes the governance risks specific to agentic AI in HR systems. Second, it proposes the Responsible Agentic AI Governance Framework (RAGF), an integrated architecture coupling lifecycle phases with governance control layers and a risk-control model. Third, it analytically evaluates the framework against existing approaches and HR governance requirements. The framework is conceptual and analytically evaluated; it is not empirically validated and identifies the empirical validation required before operational claims can be made.

The remainder of the article is organized as follows. Section 2 reviews related work on AI governance and responsible AI. Section 3 establishes the research gap, questions, and methodology. Section 4 presents the proposed governance framework. Section 5 develops the governance control mechanisms. Section 6 evaluates the framework. Section 7 discusses implications, limitations, and future work. Section 8 concludes.

2. Related Work

2.1 AI governance frameworks and standards

AI governance has matured from aspirational principles into formal standards and regulation. The OECD AI Principles, adopted in 2019 and updated in 2024, articulate five values-based principles for trustworthy AI, including human centred values, transparency, and accountability [8]. The NIST AI RMF organizes governance around four functions, Govern, Map, Measure, and Manage, and defines characteristics of trustworthy AI such as explainability, privacy, and safety [6]. ISO/IEC 42001:2023 establishes an auditable management-system standard for AI, specifying requirements for AI policy, roles, risk assessment, and lifecycle controls [7]. The EU AI Act introduces a risk-tiered regulatory

regime, classifying high-risk applications and imposing obligations on providers and deployers, including transparency, human oversight, and logging [5].

Scholarly work surveys and consolidates these developments. A systematic literature review of AI governance identifies accountability, transparency, fairness, privacy, and oversight as recurrent governance dimensions and notes that most frameworks remain abstract with limited operational guidance [9]. Camilleri examines AI governance frameworks and their intersection with corporate social responsibility, identifying accountability, transparency, explainability, reproducibility, fairness, privacy, safety, and risk prevention as key responsible-AI dimensions [10]. These contributions establish the conceptual vocabulary of governance but, as the review of autonomous-agent governance observes, they address autonomous systems only at the level of regulatory models and principles rather than operational controls [11].

2.2 Governance of autonomous and agentic AI

A distinct strand addresses governance of autonomous agents specifically. Anmol and Rana compare risk-based, principle-based, sector-specific, and audit-based governance models and propose the Multilevel Regulatory Model, organizing governance across legal, institutional, technical, and ethical levels to address accountability, adaptation, fairness, and lifecycle oversight [11]. Kshetri examines governance and security challenges in deploying agentic AI, including risks from non-human identities, autonomous behavior, and data security, and discusses zero-trust security, controlled agency, and human oversight [1]. A risk-management standards profile for agentic AI maps agent-specific risks, including autonomous goal pursuit, privilege escalation, and resistance to shutdown, onto the NIST AI RMF functions and identifies control levers such as human control and accountability, continuous monitoring, and containment [12]. Governance-by-design research, drawing on a case study of enterprise agentic AI rollout, emphasizes governing system permissions, tool and data access, and memory handling during operationalization [3].

Singapore's Infocomm Media Development Authority has issued a Model AI Governance Framework for Agentic AI, proposing guidelines on disclosure, human oversight, and accountability for AI agents [4]. These contributions signal that agentic AI demands governance mechanisms beyond those developed for predictive models, yet they remain largely at the level of principles, risk taxonomies, or sector-general guidance. None operationalizes a complete governance architecture integrated with the enterprise HRMS lifecycle.

2.3 Responsible AI in human resource management

Responsible AI in HRM has attracted substantial empirical attention. A systematic review of 107 empirical studies examines AI use across six HRM functions and how studies address bias, perceived justice and trust, privacy, explainability, and the human role [2]. This work establishes that HR AI raises fairness, transparency, and human-dignity concerns that governance must address. Research on ethical AI oversight in HRM finds that transparency, perceived fairness, and human-in-the-loop oversight shape employee autonomy through trust, psychological safety, and technostress [13]. These findings indicate that oversight and transparency are not only ethical requirements but determinants of employee outcomes, reinforcing the case for structured governance.

2.4 Human oversight, accountability, and transparency

Foundational work on automation defines a spectrum of human interaction with automation, from manual to fully automatic, establishing that oversight is a matter of degree that governance must specify explicitly [14]. The concept of meaningful human control argues that for autonomous systems

to remain legitimate, humans must retain the capacity to intervene in morally significant decisions [15]. Dignum frames responsible AI around agency, accountability, and the distribution of responsibility across human and artificial agents [16]. A systematic review of human-in-the-loop AI provides a taxonomy of loop placement, interaction granularity, and temporal characteristics and identifies challenges including the scalability of human oversight, trust calibration, and adversarial manipulation [17].

On accountability, Raji et al. propose an end-to-end framework for internal algorithmic auditing spanning the AI development lifecycle, arguing that audit integrity requires values-based assessment at each stage [18]. Cheong reviews legal, ethical, technical, and societal challenges of transparency and accountability in AI, discussing explainable AI, algorithmic audits, and data-protection regulation [19]. The TEUT framework argues that transparency and explainability support trust only when combined with uncertainty communication and governance frameworks such as the EU AI Act and the NIST AI RMF [20]. Collectively, this work supplies the building blocks, oversight modes, audit mechanisms, and transparency principles, that an integrated governance framework must assemble and operationalize for agentic AI in HR systems.

3. Research Gap and Methodology

3.1 Governance gap for agentic AI in HRMS

The literature reveals a governance gap along three dimensions. First, existing frameworks are domain-agnostic: the NIST AI RMF, ISO/IEC 42001, the EU AI Act, and the OECD Principles specify principles and risk-management functions but do not operationalize them for the specific characteristics of agentic AI (goal-directed autonomy, tool invocation, multi-step workflows) within a defined enterprise domain [6], [7], [5], [8]. Second, most governance guidance is principle-level or sector-general: systematic reviews note limited operational guidance for autonomous systems, and the agentic-AI frameworks that do exist provide risk taxonomies, principles, or regulatory models rather than an integrated, lifecycle-oriented control architecture [9], [11], [12]. Third, the HR-specific governance literature, while rich on fairness and oversight, does not address the governance of autonomous agents operating across integrated HR data over time [2], [13]. Table 1 summarizes the positioning of existing frameworks relative to agentic AI in HRMS.

3.2 Research questions

Four research questions emerge from the gap:

RQ1: What governance risks emerge when agentic AI operates within HRMS?

RQ2: What governance mechanisms are required to ensure the responsible operation of agentic AI in HRMS?

RQ3: How can accountability and human oversight be structurally incorporated into an agentic HR governance framework?

RQ4: How does an integrated, lifecycle-oriented framework address the limitations of existing AI governance approaches for agentic AI in HRMS?

3.3 Methodology

The study employs design-science and conceptual framework development grounded in a structured literature review, consistent with established research approaches for developing and evaluating artifacts in information systems. The methodology proceeds in four stages. Stage 1

conducted a structured review of peer-reviewed literature on AI governance, responsible AI, autonomous-agent governance, and responsible AI in HRM, supplemented by authoritative standards and frameworks. Sources were identified through database and web search and verified for authorship, venue, and publication details; only verified sources were retained. Stage 2 synthesized the governance gap and risk taxonomy. Stage 3 developed the framework by integrating lifecycle phases with governance control layers and a risk-control model, drawing on theoretical foundations for automation levels, meaningful human control, and responsibility [14], [15], [16]. Stage 4 analytically evaluated the framework against existing approaches and HR governance requirements.

The study is non-empirical. No primary data, participants, surveys, experiments, or validation results were collected. Where empirical validation is necessary to support operational claims, the article identifies the validation required rather than asserting empirical findings. The framework is evaluated analytically through comparison with existing frameworks and assessment against derived governance requirements; it is not claimed to be empirically validated.

4. Proposed Responsible Agentic AI Governance Framework

The RAGF is an integrated architecture comprising three components: (1) a governance lifecycle of five phases that structure how agentic AI is governed over time; (2) six governance control layers that span the lifecycle and translate principles into operational controls; and (3) a four-tier risk-control model that maps risk levels to autonomy modes and concrete controls. The framework is designed to be domain-applicable to HRMS, where agents operate on rights-sensitive data and affect employment outcomes, but its architecture generalizes to other enterprise domains.

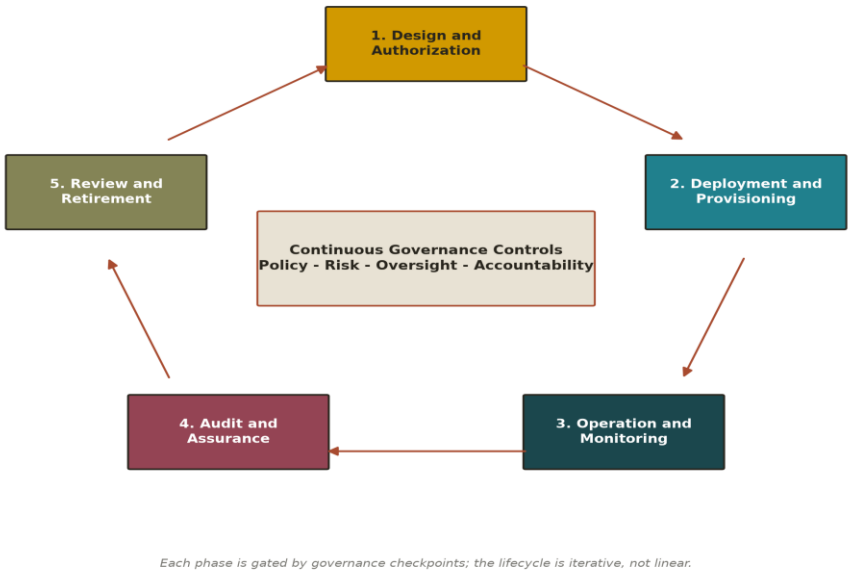


Figure 1. Agentic AI governance lifecycle in HR management systems.

4.1 Governance lifecycle

Governance is organized as an iterative lifecycle of five phases, each gated by governance checkpoints (Figure 1). Phase 1, design and authorization, establishes the agent's purpose, scope, data access, decision boundaries, and risk classification before development. Phase 2, deployment and

provisioning, controls the introduction of the agent into the HRMS environment, including identity, access provisioning, and pre-deployment authorization. Phase 3, operation and monitoring, governs live operation through continuous monitoring, logging, and oversight checkpoints. Phase 4, audit and assurance, provides independent assessment of agent behavior, decision provenance, and compliance. Phase 5, review and retirement, governs re-evaluation, modification, or decommissioning of the agent. The lifecycle is iterative rather than linear: monitoring findings feed back into design and authorization, and audit findings may trigger re-provisioning or retirement.

4.2 Governance architecture

Six governance control layers span the lifecycle (Figure 2). The governance and policy layer establishes principles, authorization policy, risk-classification rules, and compliance obligations. The oversight and intervention layer implements human-in-the-loop checkpoints, escalation, override, and suspension capabilities. The agentic AI layer comprises the agents and their orchestrator, bounded by the controls above and below. The accountability and auditability layer captures decision provenance, logging, and the audit trail. The data, privacy, and security layer enforces access control, HR data protection, and fairness controls. The HRMS and enterprise data layer provides the HR master data, payroll, identity and single sign-on, and application programming interfaces that agents act upon. The layered structure separates concerns: policy is defined once and enforced across all agents, oversight is centralized, and accountability is captured uniformly.

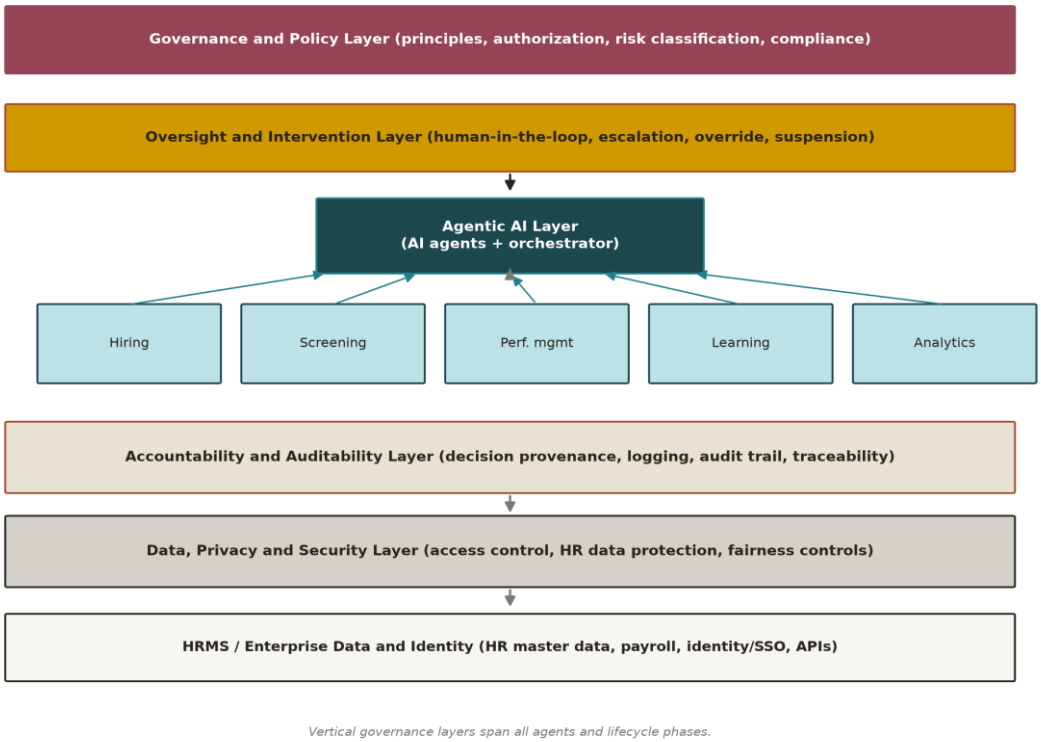


Figure 2. Responsible agentic AI governance architecture for HRMS.

4.3 Governance control flow

Within the architecture, each agent request follows a governance control flow (Figure 3). A request is subject to authorization, risk classification, and policy enforcement before action execution. High-risk or rights-affecting actions route through a human checkpoint before execution. All actions are

logged to an audit trail, monitored for drift and incidents, and escalated or overridden when thresholds are exceeded. The control flow ensures that no rights-affecting HR action is executed by an agent without authorization, logging, and, where risk warrants, human review.

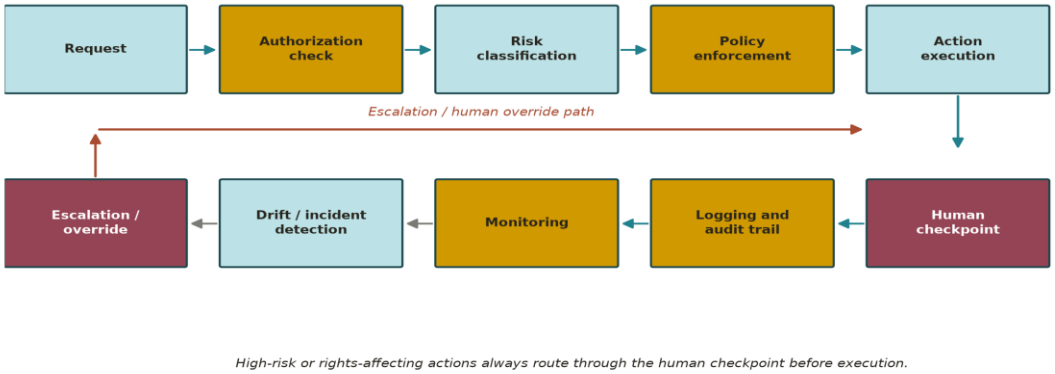


Figure 3. Governance control flow for HR AI agents.



Figure 4. Agent governance risk-control model.

4.4 Risk-control model

The risk-control model (Figure 4) classifies agent actions into four tiers and maps each to an autonomy mode and concrete controls. Low-risk actions, such as routing an internal query, may be autonomous within rules with monitoring and logging. Medium-risk actions, such as screening shortlisted candidates, operate in authorized-execution mode requiring approval, monitoring, and logging. High-risk actions, such as performance-flagging or termination recommendations, require human decision, audit, and escalation. Critical or irreversible actions, such as dismissal or

compensation decisions, are prohibited for agents and require mandatory human authority and full audit. The model makes the autonomy-decision explicit and bounded: as risk increases, autonomy decreases and controls intensify.

Table 1. Positioning of existing AI governance frameworks relative to agentic AI in HRMS.

Framework	Type	Lifecycle integration	Agentic-AI specificity	HR operationalization	Limitation
NIST AI RMF [6]	Risk mgmt. framework	Process-level functions	General	No	Domain-agnostic, not agent-operational
ISO/IEC 42001 [7]	Mgmt. system std	Lifecycle requirements	General	No	Auditable but not agent-specific
EU AI Act [5]	Regulation	Risk-tiered obligations	High-risk only	Partial	Compliance-oriented, not architecture
OECD Principles [8]	Principles	None	No	No	Aspirational, non-operational
Agentic AI risk profile [12]	Risk taxonomy	RMF-aligned	Yes	No	Risk list, no control architecture
Governing autonomous agents [11]	Regulatory model	Conceptual	Yes	No	Principle-level, no HR controls
Responsible AI in HRM [2]	Empirical review	No	No	Fairness/oversight	Not agent governance
Proposed RAGF	Governance framework	Five-phase lifecycle	Yes	Yes	Conceptual, not empirically validated

4.5 Framework components

Table 6 (Section 6) consolidates the framework components. The three components are mutually reinforcing: the lifecycle provides temporal structure, the control layers provide functional coverage, and the risk-control model provides the decision rule linking risk to autonomy. Together they ensure that governance is not a single gate but a continuous, layered, risk-calibrated process.

5. Governance Control Mechanisms

This section develops the operational control mechanisms that instantiate the framework, addressing each governance requirement derived from the literature and standards. Table 2 summarizes the governance risks in agentic HR systems, and Table 3 maps governance principles to control mechanisms.

Table 2. Governance risks in agentic HR systems.

Risk	Description	HR example	Primary control
Unauthorized autonomy	Agent acts beyond authorized scope	Agent rejects candidate without approval	Authorization scope, boundaries
Drift	Behavior changes over time	Screening criteria shift after retraining	Continuous monitoring, audit
Accountability gap	Outcome not attributable to a human role	Rejection reason unclear	Provenance, audit trail

Risk	Description	HR example	Primary control
Bias / discrimination	Disparate impact on protected groups	Disparate screening by gender or age	Fairness monitoring, human review
Privacy breach	Unauthorized access to HR data	Agent reads compensation data	Access control, data minimization
Security vulnerability	Privilege escalation, tool misuse	Agent invokes unauthorized tool	Least privilege, containment
Regulatory non-compliance	Failure to meet obligations	No logging for high-risk system	Lifecycle compliance artifacts

5.1 Authorization and decision boundaries

Authorization governs which actions an agent may initiate, on which data, and under what conditions. Each agent receives an authorization scope defining permitted actions, data classes, tool access, and decision boundaries before deployment, consistent with governance-by-design guidance on system permissions and tool and data access [3]. Decision boundaries specify the outcomes an agent may recommend versus execute; in HRMS, rights-affecting outcomes such as hiring rejection, performance rating, or termination remain human decisions unless explicitly authorized within the risk-control tier. Authorization is reviewed at each lifecycle gate and revocable on policy violation or drift.

5.2 Human oversight and intervention

Oversight is structured through human-in-the-loop checkpoints calibrated to risk, drawing on the spectrum of automation and meaningful human control [14], [15]. Low-risk actions use human-on-the-loop monitoring; medium-risk actions use human authorization; high-risk actions require human decision; and critical actions are prohibited for agents. Intervention capabilities include pause, override, correction, and suspension, implemented so that oversight is scalable, as the systematic review of human-in-the-loop AI identifies scalability of human oversight as a key challenge [17]. Oversight is logged so that each intervention is attributable and auditable.

5.3 Accountability and auditability

Accountability requires that every agent decision is attributable to a human or governance role and traceable to its inputs, reasoning, and authorization. The accountability layer captures decision provenance, including the request, authorization decision, risk classification, model version, data accessed, and human review, if any. This implements the end-to-end algorithmic auditing approach, in which each lifecycle stage produces auditable records [18], and supports the transparency and accountability mechanisms reviewed by Cheong [19]. The audit trail is tamper-evident, retained per regulation, and accessible to authorized assurance and compliance functions.

5.4 Transparency and explainability

Transparency requires that stakeholders can understand what an agent does, on what basis, and with what limits. The framework requires disclosure of agent presence, purpose, and autonomy scope to affected employees where consistent with regulation, consistent with HRM findings that transparency shapes trust and psychological safety [13]. Explainability is provided through decision rationales, risk-tier labels, and audit-accessible provenance, aligned with the TEUT framework's argument that transparency and explainability support trust only with uncertainty communication and governance [20].

5.5 Risk management and continuous monitoring

Risk management follows the structure of the NIST AI RMF functions, Govern, Map, Measure, and Manage, operationalized for agents [6]. The Map function classifies agent actions by risk tier; the Measure function monitors performance, drift, and incidents; the Manage function implements controls, escalation, and remediation; and the Govern function maintains policy, roles, and accountability. Continuous monitoring detects drift, unauthorized tool use, and outcome disparities, triggering escalation under the control flow. The agentic-AI risk profile informs the risk taxonomy, addressing agent-specific risks such as privilege escalation and resistance to shutdown [12].

5.6 Data, privacy, and security governance

The data, privacy, and security layer enforces access control aligned with the HRMS identity model, data minimization, and protection of HR data classes such as compensation, health, and performance. Fairness controls monitor for disparate impact in hiring, screening, and performance outcomes, addressing the bias and discrimination concerns documented in the HR AI literature [2]. Security controls address the non-human identity and autonomous-behavior risks that Kshetri identifies for agentic AI [1], including least-privilege provisioning, tool-access governance, and containment for unauthorized or unsafe behavior.

5.7 Lifecycle compliance

Lifecycle compliance aligns the framework with the management-system requirements of ISO/IEC 42001 and the risk-tiered obligations of the EU AI Act [7], [5]. Each lifecycle phase produces compliance artifacts: authorization records, risk assessments, monitoring logs, audit reports, and review decisions. The framework supports the high-risk-system obligations of the EU AI Act, including logging, transparency, human oversight, and risk management, where an HR AI system qualifies as high-risk. Compliance is not a one-time certification but a lifecycle process maintained through the review and retirement phase.

Table 3. Governance principles mapped to control mechanisms.

Governance principle	Source	Control mechanism in RAGF
Accountability	[8], [18]	Decision provenance, audit trail, attributable roles
Transparency / explainability	[19], [20]	Disclosure, decision rationales, risk-tier labels
Fairness / non-discrimination	[2]	Fairness monitoring, human-required modes for outcomes
Human oversight	[14], [15], [17]	Risk-calibrated checkpoints, intervention capabilities
Privacy and data protection	[5]	Access control, data minimization, HR data classes
Safety / security	[1]	Least privilege, containment, tool-access governance
Risk management	[6], [12]	RMF functions, risk tiers, continuous monitoring
Lifecycle compliance	[7], [5]	Phase artifacts, management-system alignment

6. Evaluation

The framework is evaluated analytically against existing approaches and derived governance requirements. This evaluation is conceptual, not empirical; it assesses whether the framework addresses identified gaps and requirements, and it does not constitute empirical validation.

6.1 Comparison with existing governance approaches

Table 4 compares the framework with representative existing approaches across lifecycle integration, agentic-AI specificity, HR-domain operationalization, accountability mechanisms, oversight modes, and auditability. Existing standards provide principles and management-system requirements but limited lifecycle-operational controls for agents; agentic-AI guidance provides risk taxonomies and principles but not an integrated control architecture; and HR responsible-AI research addresses fairness and oversight but not agent governance. The proposed framework integrates these contributions into lifecycle-oriented control architecture.

Table 4. Comparison of governance approaches.

Approach	Lifecycle	Agentic-AI	HR-domain	Accountability	Oversight	Auditability
NIST AI RMF [6]	Partial	General	No	Principles	Principles	Recommended
ISO/IEC 42001 [7]	Yes	General	No	Management	Required	Auditable
EU AI Act [5]	Risk-tiered	High-risk	Partial	Obligations	Mandated	Logging
Agentic AI risk profile [12]	RMF-aligned	Yes	No	Levers	Levers	Recommended
Governing autonomous agents [11]	Conceptual	Yes	No	Model	Model	Conceptual
Responsible AI in HRM [2]	No	No	Fairness	Empirical	Studied	Not addressed
Proposed RAGF	Five-phase	Yes	Yes	Provenance	Risk-calibrated	Tamper-evident

6.2 Addressing the research questions

RQ1 is addressed by the risk taxonomy (Table 2), which identifies governance risks specific to agentic AI in HRMS, including unauthorized autonomy, drift, accountability gaps, bias, privacy breaches, and security vulnerabilities. RQ2 is addressed by the control mechanisms (Section 5), which translate principles into authorization, oversight, accountability, transparency, risk management, and security controls. RQ3 is addressed by the oversight and intervention layer and the risk-control model, which structurally embed accountability and human oversight calibrated to risk. RQ4 is addressed by the comparison (Table 4), which shows that the framework integrates lifecycle phases, agentic-AI specificity, and HR operationalization more completely than existing approaches.

6.3 Addressing HR governance requirements

Table 5 maps HR governance requirements, derived from the responsible-AI-in-HRM literature and regulation to framework components. The framework addresses fairness through monitoring and human-required modes for rights-affecting outcomes, transparency through disclosure and explainability, oversight through calibrated checkpoints, privacy through access control and data

minimization, and accountability through provenance and audit. Table 6 summarizes the complete framework, listing each component, its governance purpose, and the control mechanism.

Table 5. HR governance requirements mapped to framework components.

HR governance requirement	Source	Framework component
Fairness in hiring and screening	[2]	Fairness monitoring, human-required modes
Transparency to affected employees	[13]	Disclosure, explainability
Human oversight of decisions	[13], [14]	Risk-calibrated checkpoints
Privacy of HR data	[5]	Access control, data minimization
Accountability for outcomes	[18]	Provenance, audit trail
Risk management and logging	[5], [6]	RMF functions, audit trail
Lifecycle and management-system compliance	[7]	Phase artifacts, ISO alignment

Table 6. Summary of the RAGF.

Component	Governance purpose	Control mechanism
Lifecycle phase 1: Design and authorization	Bound the agent before deployment	Authorization scope, risk classification
Lifecycle phase 2: Deployment and provisioning	Control introduction to environment	Identity, access provisioning
Lifecycle phase 3: Operation and monitoring	Govern live operation	Monitoring, logging, checkpoints
Lifecycle phase 4: Audit and assurance	Independent assessment	Audit trail, provenance review
Lifecycle phase 5: Review and retirement	Re-evaluate or decommission	Review decisions, retirement records
Policy and authorization layer	Define and enforce policy	Principles, authorization rules
Oversight and intervention layer	Embed human oversight	Checkpoints, escalation, override
Accountability and auditability layer	Ensure attributability	Provenance, tamper-evident logging
Data, privacy, and security layer	Protect HR data	Access control, fairness, containment
Risk-control model	Link risk to autonomy	Four tiers, autonomy modes, controls

7. Discussion

7.1 Theoretical implications

The framework contributes to AI governance by operationalizing abstract principles into lifecycle-oriented controls for agentic AI, addressing the operationalization gap identified in the governance literature [9], [11]. By coupling lifecycle phases with governance control layers and a risk-control model, it shows how principles such as accountability and oversight can be structurally embedded rather than asserted. The integration of meaningful human control and automation-level theory [14], [15] with audit and transparency mechanisms [18], [20] offers a reusable architecture for governing autonomous systems in sensitive enterprise domains.

7.2 Practical implications

For HRMS and enterprise AI practitioners, the framework provides a concrete architecture for governing agentic AI in HRMS: authorization scopes, risk-tiered oversight, audit trails, and lifecycle gates that map to management-system and regulatory requirements. The risk-control model offers a

decision rule for when agents may act autonomously and when human authority is mandatory, supporting responsible adoption without requiring full empirical validation before deployment. The layered architecture aligns with enterprise identity, data, and audit infrastructures, supporting implementation within existing HRMS environments.

7.3 Limitations

The framework is conceptual and analytically evaluated; it is not empirically validated. The risk-control tiers and control mechanisms are derived from literature and standards and require empirical calibration to organizational context. The framework does not specify technical implementations, such as particular logging, access-control, or monitoring technologies, which are domain- and platform-dependent. The HR-domain scope is illustrative; applicability to other domains requires adaptation. The evaluation is conceptual and does not demonstrate that the framework improves governance outcomes in practice.

7.4 Required empirical validation

Before operational claims can be made, the framework requires empirical validation, including: (1) case studies implementing the framework in live HRMS environments to assess feasibility and control effectiveness; (2) quantitative evaluation of whether risk-tiered oversight reduces unauthorized or harmful agent actions; (3) assessment of audit-trail and provenance mechanisms in supporting accountability and compliance; and (4) stakeholder studies evaluating whether oversight and transparency mechanisms affect employee trust and autonomy, extending the findings of HRM oversight research [13]. These validations are necessary before the framework can be claimed to improve governance outcomes.

7.5 Future work

Future work should develop reference implementations of the control mechanisms, including authorization scopes, audit trails, and monitoring for representative HR agents; empirically calibrate the risk-control tiers using incident and outcome data; and extend the framework to adjacent enterprise domains such as finance and customer operations to test generalizability. Integrating the framework with emerging agentic-AI governance guidance [4] and agent risk profiles [12] as they mature is a further direction.

8. Conclusion

Agentic AI in HRMS presents a governance challenge that existing principles-based frameworks do not fully resolve: autonomous, tool-using systems operating on rights-sensitive data over time require governance that is operational, lifecycle-oriented, and domain-aware. This article developed the Responsible Agentic AI Governance Framework, integrating five lifecycle phases, six governance control layers, and a four-tier risk-control model that maps risk levels to autonomy modes and concrete controls. The framework translates abstract responsible-AI principles into operational controls for authorization, oversight, accountability, transparency, risk management, and security in agentic HR systems. Analytical evaluation against existing approaches and HR governance requirements shows that the framework integrates lifecycle, agentic-AI, and HR operationalization more completely than prior work. The contribution is an integrated governance architecture; the framework is conceptual and requires empirical validation before operational claims are warranted.

Acknowledgements

The author acknowledges the foundational contributions of the AI governance, responsible-AI, and HR-AI research communities whose published work and standards informed this framework. No external funding was received. The author declares no conflicts of interest.

References

[1] N. Kshetri, "Governing agentic AI: Security, identity, and oversight in the age of autonomous intelligent systems," IEEE Computer, vol. 58, no. 8, pp. 123–129, Aug. 2025, doi: 10.1109/MC.2025.3632053.

[2] A. Bujold, I. Roberge-Maltais, X. Parent-Rocheleau, J. Boasen, S. Senecal, and P.-M. Leger, "Responsible artificial intelligence in human resources management: A review of the empirical literature," AI and Ethics, vol. 4, pp. 1185–1200, 2024, doi: 10.1007/s43681-023-00325-1.

[3] N. Dux, C. Alaimo, P. Roussiere, and A. K. Mishra, "Governance by design: Architecting agentic AI for organizational learning and scalable autonomy," arXiv:2605.20210, 2026, doi: 10.48550/arXiv.2605.20210.

[4] IMDA, "Model AI governance framework for agentic AI (Version 1.5)," Infocomm Media Development Authority, Singapore, 2026.

[5] European Parliament and Council, "Regulation (EU) 2024/1689 (Artificial Intelligence Act)," Official Journal of the European Union, L 2024/1689, 2024.

[6] NIST, "Artificial intelligence risk management framework (AI RMF 1.0)," NIST.AI.100-1, National Institute of Standards and Technology, 2023.

[7] ISO/IEC, "ISO/IEC 42001:2023 - Artificial intelligence - Management systems," International Organization for Standardization, 2023.

[8] OECD, "Recommendation of the Council on Artificial Intelligence (OECD AI Principles)," OECD/LEGAL/0449, adopted May 2019, updated May 2024.

[9] A. Batool, D. Zowghi, and M. Bano, "AI governance: A systematic literature review," AI and Ethics, vol. 5, pp. 3265–3279, 2025, doi: 10.1007/s43681-024-00653-w.

[10] M. A. Camilleri, "Artificial intelligence governance: Ethical considerations and implications for social responsibility," Expert Systems, vol. 41, no. 7, e13406, 2024, doi: 10.1111/exsy.13406.

[11] A. Anmol and C. Rana, "Governing autonomous AI agents: Regulatory models for ethical and safe deployment," AI and Ethics, vol. 6, Art. 292, 2026, doi: 10.1007/s43681-026-01153-9.

[12] N. Madkour, J. Newman, D. Raman, K. Jackson, E. R. Murphy, and C. Yuan, "Agentic AI risk-management standards profile," Center for Long-Term Cybersecurity, UC Berkeley, 2026.

[13] T. Patnaik, R. I. Gnankob, and N. S. Nanda, "Governing algorithms, empowering people: How ethical AI oversight shapes trust, technostress, and employee autonomy in AI-enabled HRM," Frontiers in Artificial Intelligence, vol. 9, Art. 1911545, 2026, doi: 10.3389/frai.2026.1911545.

[14] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," IEEE Transactions on Systems, Man, and Cybernetics, Part A, vol. 30, no. 3, pp. 286–297, 2000.

[15] F. Santoni de Sio and J. van den Hoven, "Meaningful human control over autonomous systems: A philosophical account," Frontiers in Robotics and AI, vol. 5, Art. 15, 2018, doi: 10.3389/frobt.2018.00015.

[16] V. Dignum, Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Cham, Switzerland: Springer, 2019.

[17] K. Lazaros, A. G. Vrahatis, and S. Kotsiantis, "Human-in-the-loop artificial intelligence: A systematic review of concepts, methods, and applications," Entropy, vol. 28, no. 4, Art. 377, 2026, doi: 10.3390/e28040377.

[18] I. D. Raji, A. Smart, B. White, B. Hutchinson, D. Theron, J. Smith-Loud, M. Mitchell, P. Barnes, and T. Gebru, "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing," in Proc. 2020 Conf. Fairness, Accountability, and Transparency (FAT*), 2020.

[19] B. C. Cheong, "Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making," Frontiers in Human Dynamics, vol. 6, Art. 1421273, 2024, doi: 10.3389/fhumd.2024.1421273.

[20] H. Guo and P. Polak, "Building trustworthy artificial intelligence through transparency, explainability, uncertainty and trust calibration," Discover Artificial Intelligence, vol. 6, Art. 620, 2026, doi: 10.1007/s44163-026-01219-x.