Generative AI Agent Governance in Business Processes

Hoang Vu[1]

[1]SAP, Berlin, Germany

Abstract

Generative AI agents are increasingly applied to business processes, yet organizations lack a structured management methodology to deploy and govern them. Existing BPM governance frameworks were designed for deterministic automation and cannot address agents that plan dynamically, act opaquely, and transform the roles of humans in processes. This research asks: *how can organizations govern generative AI agents in business processes to maintain alignment with organizational purpose, execution-level observability, and active runtime control?* Grounded in four completed studies covering the automation technology landscape, human factors of automation, practitioner governance requirements, and execution-level agent observability, we propose the *Agentic BPM Management Framework* as a design-science artifact: a two-layer instrument comprising four design-time governance dimensions (Technology, Agency, Purpose, Structure) and a runtime Active Governance cycle (Monitor–Diagnose–Intervene–Evaluate). Two sub-studies are proposed to operationalize the Purpose and Active Governance dimensions. Open questions concern how to scope and sequence the remaining studies and what evaluation designs would provide convincing evidence that the approach works in real organizational settings.

Keywords

AI Governance, Generative AI Agents, Business Process Management, Process Mining, Agent Behavior Mining

1. Introduction

As organizations deploy generative AI (GenAI) agents in business processes, they face an *autonomy-control paradox* [1]: granting agents autonomy increases operational flexibility, yet Business Process Management (BPM) governance depends on control, standardization, and auditable compliance [2]. Unlike Robotic Process Automation (RPA) [3] or rule-based workflow systems, GenAI agents plan dynamically, invoke tools autonomously, and behave in ways that weren't anticipated at design time [4]. Agents driven by large language models (LLMs) operate through opaque, non-deterministic reasoning that can be technically correct yet systematically violate business intent—and no current governance mechanism reliably detects this.

Process owners observe outputs but lack visibility into why decisions were made, whether policies were applied, and what LLM costs the agent behavior produces. Autonomous systems create a *contested terrain of control* [5], widening an accountability gap [6] that existing BPM constructs—built for deterministic, repeatable execution—cannot close. The EU AI Act [7] and the NIST AI Risk Management Framework [8] now require demonstrable auditability for AI in consequential processes, making the absence of a management methodology an immediate practical problem.

Prior work has begun to address AI's role in BPM. The AI-augmented BPM research manifesto [9] identifies monitoring, adaptation, and human-AI collaboration as key directions; recent work on Large Process Models [10] and LLM applications across the BPM lifecycle [11] demonstrate the technical potential of generative AI for process tasks. Most recently, the agentic BPM research manifesto [12] articulates the conceptual foundations for governing autonomous agents in business processes, identifying framed autonomy, explainability, and self-modification as essential capabilities—yet structured methods for runtime control, misalignment detection, and intervention remain unrealized. This gap motivates the central research question: *how can organizations deploy and govern GenAI agents in business processes such that agent behavior remains aligned with organizational purpose, observable at execution level, and subject to active runtime control?*

Proceedings of the BPM 2026 Best Dissertation Award, Doctoral Consortium, and Demonstration & Resources Forum (BPM 2026) , Toronto, Canada, September 27th to October 2nd, 2026.
h.vu@sap.com (H. Vu)
0009-0002-1656-7207 (H. Vu)

Table 1
Completed studies: publication status, method, and framework contribution.

Study	Status	Method	Framework dimension
P1	Published (HICSS 2023)	Vendor analysis	Technology
P2	Published (BPM Forum 2023)	Systematic review	Agency
P3	Published (BPM Forum 2025)	Thematic interviews	Purpose, Structure
P4	Accepted (BPM 2026)	Design & evaluation	Monitor (Active Governance)

2. Research Foundation

Four completed studies provide the empirical and theoretical foundation for the framework, building from diagnosing the governance challenge to delivering operational mechanisms for its resolution. Table 1 summarizes their publication status, method, and framework contribution.

P1 — Automation landscape [13]. A structured analysis of 18 automation vendors across 21 capability dimensions produced a four-type taxonomy—RPA, workflow automation, Integration Platform as a Service (iPaaS), and integrated process automation—and established that how organizations structure tool access and oversight mechanisms directly shapes deployment risk and accountability. Fewer than half of vendors offered structured governance mechanisms, underscoring that oversight was already underserved in deterministic systems. P1 grounds the *Technology* dimension by providing the vocabulary to position GenAI agents as a categorically distinct deployment type requiring its own governance logic.

P2 — Theory of automation effects [14]. A systematic review of 52 human-factors studies synthesized a three-role model in which process participants face skill decay and automation complacency, process managers lose operational oversight, and software developers lose direct visibility into what the system does. Key interaction effects include expectation mismatch, automation bias, and reduced reactivity to failures. The central governance implication is that *interpretability is a precondition for oversight* [15]: practitioners cannot govern what they cannot understand. P2 grounds the *Agency* dimension: the three-role model shows what fails when the human-agent boundary is poorly designed, and the interpretability finding establishes what any workable boundary requires — agents must be understandable enough for humans to know when and how to intervene.

P3 — Practitioner perspectives [16]. Semi-structured interviews with 22 BPM practitioners across seven industries, analyzed using qualitative content analysis [17], surfaced governance structures as the top unmet need (cited by 8 of 22), followed by process integration and compliance mechanisms. Practitioners repeatedly raised the difficulty of translating organizational goals into agent instructions and the need for clear roles, accountability, and oversight when humans and agents share work. P3 distilled these findings into six adoption recommendations and four BPM implications, and concluded that organizations need a clear and structured management framework to integrate agents responsibly — a call this thesis takes up directly. P3 grounds the *Purpose* dimension (goal translation, performance metrics) and the *Structure* dimension (human-agent roles, accountability, process design).

P4 — Agent Behavior Mining [18]. P4 introduced an event data model extending the XES process mining standard [19] with LLM-specific attributes and event types—reasoning traces, token costs, model identity, tool invocations—to render GenAI agent execution observable at process level. Applied to 371 Order-to-Cash traces, Agent Behavior Mining (ABM) surfaced compliance violations, cost anomalies, and behavioral variants invisible to conventional process mining. An evaluation with 18 practitioners found 83% reported substantially improved governance transparency. P4 provides the observability infrastructure on which the Active Governance cycle is built, grounding the *Monitor* dimension.

After P1–P4, organizations have a taxonomy, a workforce effects theory, practitioner-validated requirements, and execution-level observability—but no management framework synthesizing these into structured guidance from deployment design through operational governance. Closing this gap is the proposed framework's contribution.

3. Research Direction

We propose the *Agentic BPM Management Framework* as a design-science artifact [20] providing structured guidance for deploying and governing GenAI agents in business processes. The framework comprises a design-time foundation of four governance dimensions and a runtime Active Governance engine. Figure 1 maps how P1–P4 ground each element; green elements are operational, purple elements constitute the proposed sub-studies.

Layer 1: Setup — Design-Time Governance. The *Setup* layer structures deployment decisions across four dimensions, configured before production and revisited when the deployment context changes materially.

Technology (P1) addresses the governance implications of architectural choices: a generalist agent with broad tool access presents fundamentally different accountability challenges than specialist agents with constrained interfaces. Architectural decisions—model selection, tool access, agent coordination—simultaneously determine the effective autonomy level, risk profile, and oversight requirements of a deployment [4]. P1's taxonomy of prior automation types — RPA, workflow, iPaaS, and integrated automation — provides the baseline for the *Technology* dimension and shows that GenAI agents constitute a distinct deployment category requiring its own governance logic.

Agency (P2) concerns the design of the *human-agent boundary*: which judgments are delegated to agents and which must retain human reasoning; what conditions trigger a handoff when an agent steps outside its approved boundaries; and how the agent is made interpretable enough that humans can meaningfully oversee it, rather than just watch it run. P2's three-role model structures these decisions across process participants, managers, and developers.

Purpose (P3) addresses *semantic drift*: when agents optimize a measurable proxy, that proxy decouples from the original intent [21]. High-level business goals rarely translate directly into agent instructions, and once an agent is running against a measurable target, that target can drift from the outcome it was meant to capture. Organizations need ways to detect this gap and correct it. The Goal Translation sub-study (P5) develops a method for this dimension.

Structure (P2+P3) covers the *organisational design of the human-agent system*: role definitions, agent-to-agent coordination, escalation paths, context-passing protocols, and accountability assignments that prevent responsibility from becoming unclear across agent boundaries. Where Agency asks *when and how a human takes over from an individual agent*, Structure asks *how the whole system is organised and held accountable* before deployment begins.

Layer 2: Operate — Active Governance. The *Operate* layer runs as a continuous four-stage cycle. **Monitor** (P4, operational) uses ABM to observe agent execution and surface unwanted behavior—compliance violations, cost anomalies, and behavioral variants that deviate from expected execution. P4 does not, however, explain *why* these deviations occur. **Diagnose** (proposed) addresses *why* unwanted behavior occurs. Agent mining can surface several types of explanation — missing business or process context, model limitations, tool failures, or prompt issues — each calling for a different intervention. Diagnose classifies the cause so that Intervene can act at the right level. The *context gap* is the focus of this thesis: when agents lack sufficient business or process context, they resolve ambiguity through their own reasoning rather than organizational intent. Just as human workers need contextual grounding to operate within process boundaries, agents need the right context to constrain their decisions [22].

Intervene (core of P6, proposed) acts at two levels. *L1 Agent* adjusts a single agent — context infusion, guardrail activation, prompt enrichment — without changing the process. *L2 Process* restructures the work around the agent — task reassignment, approval steps, handoff redesign — and triggers the Redesign loop in Figure 1, feeding back into any of the four Setup dimensions.

Evaluate (proposed) compares ABM event logs before and after the intervention to check whether the change improved agent behavior against process and business objectives, and feeds the result into the next Monitor cycle. The framework as a whole will be evaluated through expert walkthroughs and instantiation in the two enterprise deployment contexts used in P5 and P6, following the design-science criteria of utility, rigor, and relevance [20].

Proposed Sub-Studies. P5 — Goal Translation. The semantic drift problem (P3) requires a method for translating organizational goals into agent objectives without loss of intent. The sub-study will develop a structured goal translation method grounded in P3's interview corpus, instantiated in two enterprise contexts using the ABM infrastructure from P4, and evaluated through structured walkthroughs assessing whether goal-objective alignment is maintained and practitioners judge decompositions as faithful to organizational intent [20]. The scope relative to the remaining PhD timeline is an open question discussed in Section 4.

P6 — Active Governance. This sub-study operationalizes the runtime engine through: (1) a conceptual artifact defining governable anomalies (conformance violations, variant deviations, business rule breaches) and L1–L2 intervention logic; (2) field instantiation against real customer O2C traces; (3) before/after ABM evaluation against process and business objectives [18]. The primary methodological threat is access to longitudinal enterprise trace data; a fallback using controlled synthetic traces from the P4 dataset is planned if field access is constrained.

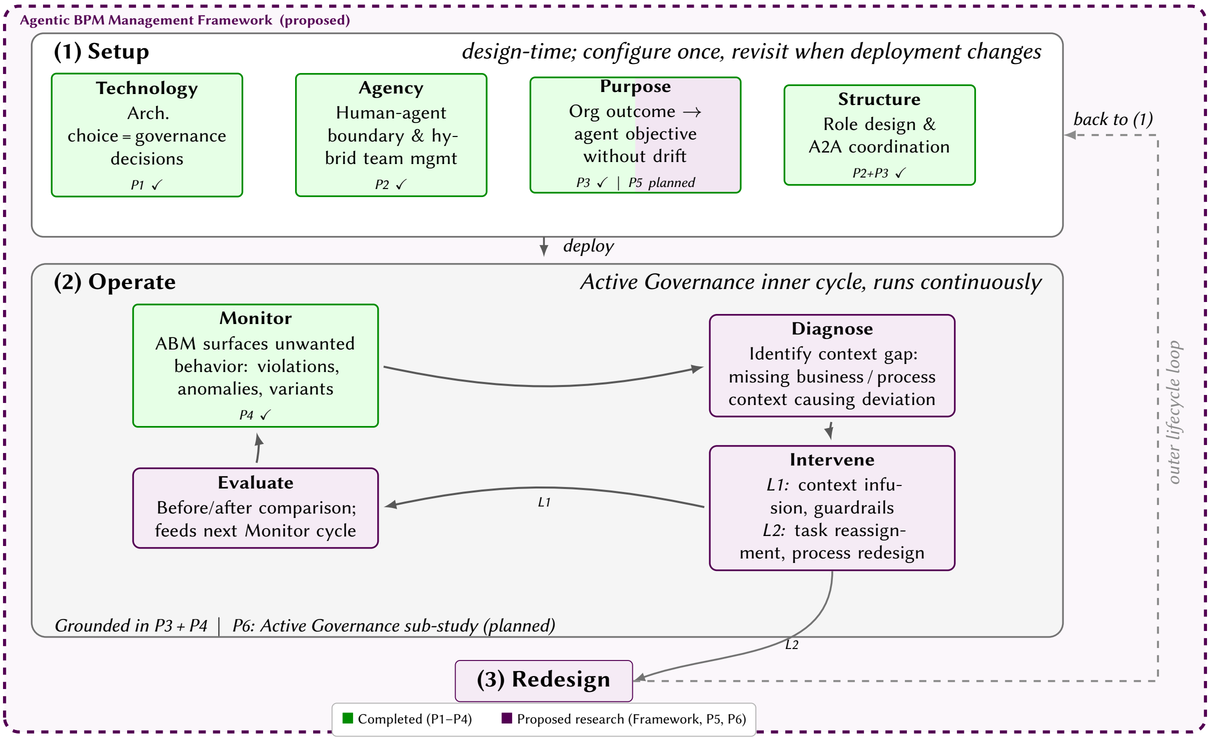

Figure 1: The Agentic BPM Management Framework. **Green** elements are grounded in completed studies (P1–P4); the Purpose box is split reflecting P3 (done) and the planned Goal Translation sub-study P5. **Purple** elements constitute the proposed research contribution. Within the Active Governance cycle, L1 interventions loop back to Evaluate; L2 interventions escalate to Redesign (3), which feeds back into any of the four Setup dimensions via the outer lifecycle loop (dashed).

4. Open Questions

The following questions represent core uncertainties in the research design, each shaping the framework's scope, methodology, or contribution.

Q1 — Does the two-layer structure reflect how governance actually works? The framework separates design-time from runtime. In practice, design-time decisions may need continuous revision as agent behavior evolves — meaning the two layers differ in *how often* they are revisited, not in *whether* they can change. Should governance be conceived as a fully integrated cycle from the outset [9], or does the separation usefully force organizations to treat architectural and operational decisions with different governance logic?

Q2 — What evaluation design convincingly validates Active Governance? P6's before/after comparison raises the question: what counts as proof that governance worked? Conformance rate improvements are measurable but may not capture whether interventions served organizational purpose. Practitioner judgment provides ecological validity but risks confirmation bias. Field deployment offers high external validity but low experimental control. Whether the synthetic fallback provides sufficient validity for the claims P6 needs to make remains an open question.
Q3 — Is Goal Translation within scope of the PhD? P5 addresses semantic drift in the Purpose dimension. Three options exist: a full sub-study with a validated Goal Translation method; a prototype establishing the problem and a candidate solution without full validation; or deferral to post-doctoral work, with this paper providing theoretical framing only. The choice turns on timeline feasibility and whether reviewers regard an unvalidated Purpose dimension as a critical gap or an acceptable scope boundary.
Q4 — How much of P1 and P2 transfer to GenAI agents? Both foundational studies predate commercial GenAI agents. P1's taxonomy covers RPA, workflow automation, iPaaS, and integrated automation — none exhibiting probabilistic, goal-directed reasoning. P2's three-role model and interpretability principle were derived from human interaction with deterministic automation. The open question is whether GenAI agents require an *extension* of these frameworks — refining existing categories — or a *replacement*, because probabilistic reasoning is different enough in kind to require new conceptualizations of the automation landscape and the human-agent boundary [23]. P6 field data offers a test: where P2's roles or P1's categories fail to fit GenAI deployments, replacement rather than refinement is the answer.
Q5 — What are the primary threats to the research program? Three threats are identified. *Data access*: P5 and P6 depend on longitudinal enterprise GenAI agent traces; restricted access shifts evaluation to synthetic data and limits external validity. *Generalizability*: both sub-studies are grounded in a single domain (Order-to-Cash) at SAP; this paper must explicitly delimit the framework's applicability. *Evaluation validity*: the before/after comparison in P6 cannot fully control for concurrent process changes or model updates, which must be addressed in the instrumentation design to support causal claims.

5. Conclusion

This paper has presented a research agenda for governing GenAI agents in business processes. Drawing on four published studies — a landscape analysis (P1), a theory of automation effects (P2), practitioner governance requirements (P3), and execution-level agent observability (P4, accepted) — we propose the *Agentic BPM Management Framework*: a two-layer instrument with four design-time governance dimensions and a runtime Active Governance engine. Two sub-studies (P5, P6) develop the Purpose and Active Governance dimensions, closing the loop P4 left open.

P1–P4 are complete; P5–P6 are in design with field instantiation planned for 2026–2027. The proposed journal paper will synthesize all six studies into a structured management methodology for governing GenAI agents — grounded in practitioner requirements, validated observability infrastructure, and a design-science evaluation across two enterprise contexts. The open questions in Section 4 will shape this contribution: whether the two-layer structure is the right organizing logic, how P5 and P6 should be scoped, what evaluation design provides convincing evidence, and how much theoretical development the Technology and Agency dimensions require.

Declaration on Generative AI

During the preparation of this work, the author used Claude (Anthropic) in order to: proofread draft text, correct grammar and spelling, and assist with sentence-level reformulation. After using these tools, the author reviewed and edited all content as needed and takes full responsibility for the publication's content.

References

[1] M. Mohlmann, L. Zalmanson, Hands on the wheel: Navigating algorithmic management and uber drivers' autonomy, in: Proceedings of the International Conference on Information Systems (ICIS) 2017, AIS, 2017. URL: https://aisel.aisnet.org/icis2017/DigitalPlatforms/Presentations/3, paper 3.

[2] M. Rosemann, J. vom Brocke, The six core elements of business process management, in: Handbook on business process management 1: introduction, methods, and information systems, Springer, 2014, pp. 105–122.

[3] W. M. Van der Aalst, M. Bichler, A. Heinzl, Robotic process automation, Business & information systems engineering 60 (2018) 269–272.

[4] D. Acharya, K. Kuppan, D. B. Ashwin, Agentic ai: Autonomous intelligence for complex goals – a comprehensive survey, IEEE Access PP (2025) 1–1.

[5] K. C. Kellogg, M. A. Valentine, A. Christin, Algorithms at work: The new contested terrain of control, Academy of management annals 14 (2020) 366–410.

[6] I. D. Raji, A. Smart, R. N. White, M. Mitchell, T. Gebru, B. Hutchinson, J. Smith-Loud, D. Theron, P. Barnes, Closing the ai accountability gap: Defining an end-to-end framework for internal algorithmic auditing, in: Proceedings of the 2020 conference on fairness, accountability, and transparency, 2020, pp. 33–44.

[7] European Union, Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence, Official Journal of the European Union L (2024). ELI: http://data.europa.eu/eli/reg/2024/1689/oj.

[8] N. AI, Artificial intelligence risk management framework (ai rmf 1.0), URL: https://nvlpubs. nist. gov/nistpubs/ai/nist. ai (2023) 100–1.

[9] M. Dumas, F. Fournier, L. Limonad, A. Marrella, M. Montali, J. Rehse, R. Accorsi, D. Calvanese, G. D. Giacomo, D. Fahland, A. Gal, M. L. Rosa, H. Völzer, I. Weber, Ai-augmented business process management systems: A research manifesto, ACM Trans. Manag. Inf. Syst. 14 (2023) 11:1–11:19.

[10] T. Kampik, C. Warmuth, A. Rebmann, R. Agam, L. Egger, A. Gerber, J. Hoffart, J. Kolk, P. Herzig, G. Decker, H. van der Aa, A. Polyvyanyy, S. Rinderle-Ma, I. Weber, M. Weidlich, Large process models: A vision for business process management in the age of generative ai, KI - Künstliche Intelligenz (2024).

[11] B. Estrada-Torres, A. del-Río-Ortega, M. Resinas, Mapping the landscape: Exploring llm applications in business process management, in: Enterprise, Business-Process and Information Systems Modeling, 2024, pp. 22–31.

[12] D. Calvanese, A. Casciani, G. De Giacomo, M. Dumas, F. Fournier, T. Kampik, E. La Malfa, L. Limonad, A. Marrella, A. Metzger, et al., Agentic business process management: A research manifesto, Information Systems 140 (2026) 102738.

[13] H. Vu, H. Leopold, H. van der Aa, What is business process automation anyway?, in: Hawaii International Conference on System Sciences, 2023, pp. 5462–5471.

[14] H. Vu, J. Haase, H. Leopold, J. Mendling, Towards a theory on process automation effects, in: International Conference on Business Process Management, Springer, 2023, pp. 285–301.

[15] S. J. Russell, P. Norvig, Artificial Intelligence: A Modern Approach (4th Edition), Pearson, 2020. URL: http://aima.cs.berkeley.edu/.

[16] H. Vu, N. Klievtsova, H. Leopold, S. Rinderle-Ma, T. Kampik, Agentic business process management: Practitioner perspectives on agent governance in business processes, in: International Conference on Business Process Management, Springer, 2025, pp. 29–43.

[17] P. Mayring, T. Fenzl, Qualitative Inhaltsanalyse, Springer Fachmedien Wiesbaden, Wiesbaden, 2019, pp. 633–648.

[18] H. Vu, M. Körner, A. Rebmann, G. Kevorkian, M. Perscheid, G. Berg, T. Kampik, Agent behavior mining: Generative ai agent governance in business processes, in: International Conference on Business Process Management, 2026. Conditionally accepted.

[19] W. M. P. van der Aalst, Process Mining - Data Science in Action, Second Edition, Springer, 2016.

[20] A. Hevner, S. Chatterjee, Design science research in information systems, in: Design research in

information systems: theory and practice, Springer, 2010, pp. 9–22.

[21] T. Heindel, I. Weber, Incentive alignment of business processes, in: D. Fahland, C. Ghidini, J. Becker, M. Dumas (Eds.), Business Process Management - 18th International Conference, BPM 2020, Seville, Spain, September 13-18, 2020, Proceedings, volume 12168 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 93–110. URL: https://doi.org/10.1007/978-3-030-58666-9_6. doi:10.1007/978-3-030-58666-9_6.

[22] K. J. Feng, D. W. McDonald, A. X. Zhang, Levels of autonomy for ai agents, arXiv preprint arXiv:2506.12469 (2025).

[23] A. Simkute, L. Tankelevitch, V. Kewenig, A. E. Scott, A. Sellen, S. Rintel, Ironies of generative ai: Understanding and mitigating productivity loss in human-ai interaction, International Journal of Human–Computer Interaction 41 (2025) 2898–2919.