Autonomous AI Agents as Co-Program Managers: A Conceptual Framework for Hybrid Intelligence Governance

N. Bingham[1], A. DeGroat[1], T. Hursa-Webster[1], S.Miller[1], B. Sanchez[1], R. Trujillo[1]

[1] Anderson School of Management, University of New Mexico, Albuquerque, NM, USA

Abstract

Program environments increasingly span interdependent projects, digital platforms, regulatory regimes, and geographically dispersed stakeholders. Those conditions stretch governance mechanisms that were built for episodic reporting and predominantly human decision cycles. This conceptual article examines whether bounded autonomous AI agents can function as co-program managers within contemporary program governance systems. Drawing on program management theory, AI governance research, human–AI collaboration scholarship, and explainable AI literature, the article develops a Hybrid Intelligence Governance Model that allocates continuous sensing, forecasting, anomaly detection, and scenario analysis to AI agents while reserving strategic judgment, stakeholder alignment, ethical interpretation, and formal decision authority to human program leaders. The paper makes three contributions. First, it defines the idea of the bounded co-program manager as an AI agent embedded within explicit governance thresholds rather than an unconstrained decision surrogate. Second, it specifies the division of labor, control architecture, and authority boundaries required for hybrid human–AI governance. Third, it develops a risk architecture for AI co-management centered on bias amplification, model drift, data-quality failure, over-automation, governance lag, and stakeholder trust erosion. This analysis argues that AI strengthens program

management only when autonomy is bounded, model behavior is auditable, escalation rights are explicit, and human accountability remains nondelegable. This article therefore shifts the discussion from AI adoption as a tooling question to AI deployment as a governance design problem.

Keywords

program management; artificial intelligence; AI governance; human–AI collaboration; hybrid intelligence; risk governance

Introduction

Program management increasingly unfolds in environments defined by interdependence, uncertainty, and institutional complexity. Programs often coordinate multiple related projects across digital infrastructures, organizational boundaries, regulatory regimes, and geographically distributed stakeholder groups. Under those conditions, governance cannot be reduced to schedule review and budget oversight. It must integrate strategic alignment, benefits realization, escalation discipline, and cross-project coordination in near real time (Brown, 2014; Martinsuo & Hoverfält, 2018; Project Management Institute [PMI], 2024; Roehrich et al., 2024).

Traditional program management tools were designed for slower reporting cycles and narrower data environments. Contemporary enterprise systems, by contrast, generate dense streams of performance, financial, operational, and risk data. Artificial intelligence promises to analyze those streams faster than manual governance processes typically allow, making earlier forecasting, anomaly detection, and scenario analysis possible (Davenport & Ronanki, 2018; Dwivedi et al., 2021; Fridgeirsson et al., 2021). However, much of the discussion still treats AI as

a useful tool rather than as a governance-embedded participant whose recommendations can materially shape program decisions.

That gap matters. Once AI systems begin recommending escalations, prioritizing interventions, drafting coordination actions, and surfacing strategic trade-offs, they are no longer peripheral dashboards, instead becoming part of the governance architecture itself. The practical question is therefore not whether AI can support program managers—it evidently can—but the conditions in which AI can operate as a bounded co-program manager without diluting human accountability or destabilizing governance.

This article addresses that question through a conceptual synthesis of program management, AI governance, explainability, and human–AI collaboration literature. It makes three contributions: First, it defines the bounded co-program manager as an AI agent permitted to perform delegated analytical and low-impact operational tasks within explicit authority thresholds; Second, it develops a Hybrid Intelligence Governance Model that clarifies how responsibility should be distributed between AI agents and human program leaders; Third, it offers a governance-centered risk architecture for AI co-management. The argument is straightforward: AI can strengthen program management, but only when organizations design for bounded autonomy, auditable reasoning, and nondelegable human accountability.

Conceptual Framing and Review Scope

This article is a conceptual paper rather than an empirical study. It uses an integrative review across four streams of scholarship: program management and governance, AI governance, human–AI decision collaboration, and explainable AI. The purpose is theory building, not exhaustive evidence aggregation. The objective is not to claim a definitive systematic review of the field, but to synthesize adjacent literature into a program-management-specific

governance framework.

The first stream treats program management as a governance discipline. Brown (2014) and PMI (2016, 2024) describe programs as vehicles for coordinating related initiatives in order to realize strategic benefits that exceed the sum of isolated projects. Traditional program management places authority, escalation, integration, and benefits accountability at the center of the program manager's role. Research on change programs and large interorganizational projects further shows that program environments are shaped by contextual volatility, interdependence, and governance tension rather than linear execution logic alone (Martinsuo & Hoverfält, 2018; Roehrich et al., 2024).

The second stream concerns AI governance. Recent reviews converge on recurring design requirements: accountability, lifecycle oversight, transparency, and structured organizational controls for AI deployment (Birkstedt et al., 2023; Batool et al., 2025). Official governance frameworks such as the OECD AI Principles and the NIST AI Risk Management Framework reinforce the same point. Trustworthy AI is not achieved by technical performance alone; it depends on governance mechanisms that connect design, monitoring, escalation, and human responsibility (High-Level Expert Group on AI, 2019; National Institute of Standards and Technology [NIST], 2023; Organisation for Economic Co-operation and Development [OECD], 2019).

The third stream examines human–AI collaboration in decision environments. Research shows that AI changes organizational decision structures by altering speed, replicability, search breadth, and interpretability, which in turn affects where authority should be located and how human and machine inputs should be combined (Shrestha et al., 2019). Trust also matters. Managers are more likely to assign decisional weight to AI recommendations when those systems

are perceived as reliable and intelligible (Rai, 2020; Wen et al., 2025). That is useful when trust is calibrated and dangerous when it becomes uncritical deference.

The fourth stream highlights a tension between automation and augmentation. Raisch and Krakowski (2021) argue that management practice should not imagine these two modes as neatly separable. AI can both automate tasks and augment judgment, and organizations that overcommit to one pole generate avoidable dysfunction. That insight is particularly relevant to program management, where analytical scale and strategic judgment must coexist. The implication is that the future of program management is not fully automated control or purely manual oversight. It is a designed hybrid in which human and AI capabilities are deliberately combined.

Hybrid Intelligence Governance Model

We use the term bounded co-program manager to denote an AI agent embedded inside formal governance constraints. The phrase does not imply equal authority between human and machine actors. It identifies a functional role within which the AI agent performs continuous sensing, forecast generation, anomaly detection, scenario analysis, and selected low-impact coordination tasks, while human leaders retain formal authority over strategy, ethics, stakeholder commitments, and benefits realization.

This framing matters because the existing literature often slips between two weaker positions. One treats AI as a passive tool with no governance significance. The other lapses into vague futurism in which AI is imagined as an autonomous substitute for managerial judgment. Both positions are analytically thin. If AI recommendations materially influence escalation, prioritization, or resource trade-offs, then AI has already entered the governance system. At the same time, because programs involve ambiguity, politics, legitimacy, and value conflicts, AI cannot

plausibly displace human responsibility. The relevant design challenge is therefore bounded agency, not managerial replacement.

The Hybrid Intelligence Governance Model developed here rests on three layered claims. First, AI is strongest where the problem is continuous sensing under conditions of data abundance. Second, human program managers are strongest where the problem is contextual interpretation under conditions of ambiguity, contestation, and reputational consequence. Third, governance quality depends on the controls that connect those two layers: autonomy thresholds, explainability, override rights, audit trails, and periodic model review. Figure 1 summarizes that architecture.

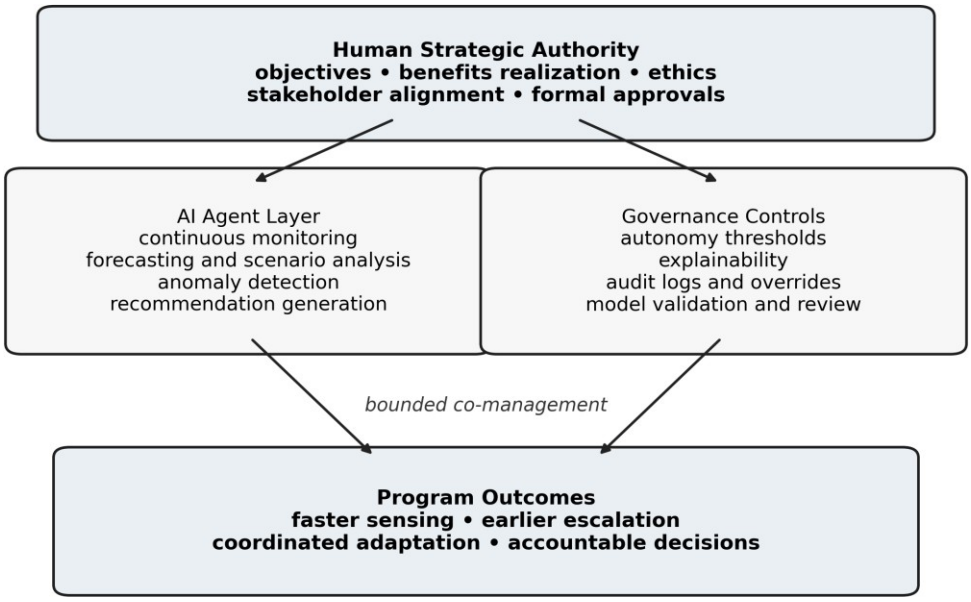

Figure 1. *Hybrid Intelligence Governance Model.*

Table 1. Functional allocation in hybrid intelligence governance.

Governance domain	AI agent contribution	Human program manager contribution	Required controls
Environmental sensing and monitoring	Continuously ingest program data, detect anomalies, map cross-project dependencies, and flag threshold breaches.	Judge materiality, determine whether anomalies warrant escalation, and align responses with strategic intent.	Data-quality ownership; approved thresholds; audit log of alerts and overrides.
Forecasting and scenario analysis	Generate schedule, cost, risk, and dependency forecasts; simulate alternative scenarios and likely downstream effects.	Interrogate assumptions, weigh trade-offs, and decide which scenario matters for the organization's strategy and commitments.	Model validation; confidence reporting; periodic recalibration.
Routine coordination	Draft status summaries, recommend low-impact task adjustments, and surface coordination bottlenecks across teams.	Approve external communication, manage exceptions, and maintain stakeholder alignment across organizational boundaries.	Workflow permissions; communication approval rules; escalation triggers.
Stage-gate and resource decisions	Provide decision support by ranking options, surfacing projected consequences, and tracking expected benefits realization.	Authorize stage transitions, resource shifts, and structural changes; own benefits accountability.	Formal decision-rights charter; steering committee review.
Ethics, compliance, and legitimacy	Flag policy deviations, fairness concerns, and documentation gaps.	Adjudicate ambiguous cases, interpret stakeholder sensitivities, and absorb accountability for outcomes.	Human-in-the-loop approval; policy review; documented rationale.

From this model, four propositions follow. They are not empirical findings; they are theory-building statements intended to guide subsequent research and organizational design.

Proposition 1. In high-complexity programs, bounded AI agents improve program sensing and foresight when data inputs are sufficiently reliable, and model stewardship is continuous.

Proposition 2. AI-enabled decision velocity improves program responsiveness only when authority boundaries, approval rights, and escalation thresholds are explicit and institutionally enforced.

Proposition 3. The more opaque the AI agent's recommendations and reasoning processes, the greater the likelihood of responsibility diffusion, stakeholder resistance, and governance breakdown.

Proposition 4. Hybrid intelligence governance is more resilient than either purely manual governance or heavily automated governance when programs require both analytical scale and high-stakes contextual judgment.

Governance Design Requirements

If AI is to function as a co-program manager, organizations must redesign governance around explicit decision rights. The first requirement is bounded autonomy. Governance charters should specify which tasks the agent may execute independently, which tasks it may only recommend, and which tasks are categorically reserved for human decision-makers. Low-impact operational adjustments may be delegated when predefined thresholds are met. Strategic choices, stage-gate approvals, commitments to external stakeholders, and decisions with significant ethical or reputational implications should remain human-controlled. Without that architecture, the language of augmentation quickly becomes a cover for unmanaged delegation.

The second requirement is traceability. Hybrid governance fails when AI output enters the decision process without a visible chain of provenance. Program boards and PMOs therefore need more than performance dashboards, rather requiring decision logs that connect data sources, model versions, recommendations, overrides, approvals, and eventual outcomes. Explainability does not require that every model become trivially simple, but it does require that those accountable for decisions can inspect the rationale, assumptions, and confidence boundaries attached to AI-generated recommendations (Rai, 2020; NIST, 2023).

The third requirement is accountability architecture. One of the clearest lessons from AI governance literature is that organizations frequently define how AI is used more clearly than who remains answerable when its outputs shape consequential decisions (Batool et al., 2025; Birkstedt et al., 2023). In program environments, that ambiguity is unacceptable. Human program managers, sponsors, and steering committees must remain the accountable actors even when analytical responsibility is partially distributed to machines. A co-program manager can carry delegated work; it cannot carry ultimate fiduciary or ethical accountability.

The fourth requirement is ethical and regulatory governance. Program environments routinely involve stakeholder asymmetries, contractual obligations, public scrutiny, and cross-jurisdictional compliance demands. AI models trained on historical data can reproduce biased priorities, privilege measurable variables over politically salient ones, and create false confidence around uncertain forecasts. Governance must therefore include fairness reviews, policy checks, and mechanisms for surfacing when the model's preferred option is misaligned with stakeholder legitimacy or institutional values (High-Level Expert Group on AI, 2019; OECD, 2019).

The fifth requirement is managerial capability renewal. In a hybrid governance environment, the program manager's role becomes more interpretive rather than less important.

Program leaders must understand what the model is optimizing, what assumptions drive its recommendations, what signals are absent from the data, and when automated output should be challenged. AI literacy in this context does not mean coding fluency. It means disciplined skepticism, probabilistic reasoning, and the ability to convert machine-generated insight into strategically legitimate action. The future program manager is not replaced by AI; the role becomes more governance-intensive because intelligence itself must now be managed.

Risk Architecture for AI Co-Management

AI co-management changes the program risk profile in two directions at once. It can reduce traditional risks by improving early detection, forecast quality, and cross-project visibility. It can also create new risks tied to model behavior, data integrity, authority ambiguity, and organizational dependence on automated judgment. The relevant question is not whether AI introduces risk. It does. The real question is whether governance can transform those exposures into manageable and reviewable forms.

The risk architecture developed here centers on six recurrent categories: bias amplification, model drift, data-quality and integration failure, over-automation and skill erosion, governance lag, and stakeholder trust erosion. These are not exhaustive, but they are central because each one threatens program performance through a different governance failure mode. Figure 2 organizes those risks using a qualitative probability–impact logic adapted to the needs of hybrid program governance.

Table 2. Qualitative risk matrix for AI co-management.

Probability ↓ / Impact →	Low impact	Medium impact	High impact
Low probability	Monitoring risks	Monitoring risks	Monitoring risks
Medium probability	Monitoring risks	Over-automation and skill erosion	Governance lag; stakeholder trust erosion
High probability	Monitoring risks	Bias amplification	Model drift; data-quality and integration failure

The matrix is intentionally conservative. Model drift and data-quality failure are placed in the highest exposure zone because they can silently degrade forecast quality while preserving the appearance of analytical precision. Bias amplification is treated as a high-probability, medium-impact exposure in most program contexts because it is likely to emerge whenever historical data encode organizational blind spots; its impact, however, can become severe when decisions affect sensitive stakeholder groups or compliance obligations. Governance lag and stakeholder trust erosion occupy the medium-probability, high-impact zone because they often emerge later than technical failures but can destabilize the broader legitimacy of the governance system once they do.

Risk categorization is useful only if it leads to disciplined response design. Table 2 therefore translates the qualitative ratings into governance-aligned treatments. The point of response planning is not to eliminate uncertainty. It is to make clear which controls, owners, and review routines are necessary for each exposure.

Table 3. Governance-aligned response matrix for AI co-management.

Risk	Primary response	Key controls	Oversight owner
Bias amplification	Mitigate	Representative training data, fairness testing, explanation requirements, and human approval for consequential recommendations.	Program manager + AI governance lead
Model drift	Mitigate	Performance monitoring, challenger models, retraining schedule, and documented thresholds for model retirement or rollback.	PMO / model owner
Data-quality and integration failure	Avoid + mitigate	Data ownership, pipeline redesign, quality controls, and refusal to automate decisions when upstream data are incomplete or contradictory.	Data steward + program manager
Over-automation and skill erosion	Mitigate	Human checkpoints, rotation of manual review duties, periodic no-AI exercises, and review of override behavior.	PMO leader
Governance lag	Avoid + mitigate	Updated governance charter, stage-gate rules for AI use, sandbox pilots, and change-control review before scope expansion.	Steering committee
Stakeholder trust erosion	Mitigate	Transparent communication, explanation design, documented human ownership of final decisions, and monitoring of stakeholder feedback.	Program sponsor + program manager

Not all risks in a program environment are materially reduced by AI. Macroeconomic shocks, geopolitical disruption, legal change, and stakeholder politics remain fundamentally exogenous to the intelligence system. AI may surface some of those developments earlier, but it cannot remove their underlying causal force. Recognizing that limit is part of responsible governance. It prevents organizations from attributing strategic resilience to automation that in fact depends on leadership, negotiation, and institutional judgment.

For that reason, AI should be treated as a second-order governance technology. It changes how programs sense, interpret, and respond to uncertainty, but it does not abolish uncertainty itself. Mature governance acknowledges both the analytical advantages and the structural limits of AI co-management.

Implications for Research and Practice

The framework developed here has immediate implications for practice. Organizations should not begin AI co-management with high-stakes autonomy. They should begin with sensing, monitoring, and recommendation tasks where the benefits of scale are large and the consequences of error are bounded. As reliability, transparency, and governance discipline improve, organizations may cautiously expand the agent's scope. Even then, expansion should be earned through evidence, not declared through enthusiasm.

Organizations also need structural preparation, not merely software adoption. Governance charters should be rewritten to specify decision rights, approval thresholds, and escalation rules for AI-supported actions. Risk registers should include model-specific exposures. PMOs should oversee not only project and program performance but also the quality of the intelligence infrastructure that informs managerial judgment. In practical terms, that means model review cadence, data stewardship responsibilities, override reporting, and periodic ethical review becoming ordinary features of program governance rather than exceptional controls.

The framework also generates a research agenda. The propositions advanced here can be tested through comparative case studies, field experiments, and surveys of program leaders operating in different governance regimes. Future research should examine whether hybrid governance improves benefits realization, whether decision quality changes under different levels of autonomy, how override behavior affects trust calibration, and which governance routines most effectively prevent responsibility diffusion. The most useful scholarship will move beyond the generic question of AI adoption and focus instead on institutional design under conditions of bounded machine agency.

Conclusion

The future of program management will not be defined by the disappearance of the program manager. It will be defined by the redesign of governance around hybrid intelligence. As programs become more data-rich, interdependent, and time-sensitive, AI agents can legitimately assume a larger share of continuous monitoring, forecast generation, and coordination support. But that expansion is defensible only when it is paired with tighter human accountability.

The central contribution of this article is therefore conceptual but practical: it reframes AI in program management as a governance design problem rather than a mere tooling upgrade. A bounded co-program manager can strengthen sensing, speed, and coordination. It cannot absorb strategic accountability, stakeholder legitimacy, or ethical judgment. Those remain irreducibly human obligations. The organizations most likely to benefit from AI co-management will be the ones disciplined enough to preserve that distinction.

References

Batool, A., Zowghi, D., and Bano, M. (2025). AI Governance: A Systematic Literature Review. AI and Ethics, 5, 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Birkstedt, T., Minkkinen, M., Tandon, A., and Mäntymäki, M. (2023). AI Governance: Themes, Knowledge Gaps and Future Agendas. Internet Research, 33(7), 133–167. https://doi.org/10.1108/INTR-01-2022-0042

Brown, J. T. (2014). The Handbook of Program Management: How to Facilitate Project Success with Optimal Program Management (2nd ed.). McGraw-Hill Education.

Davenport, T. H., and Ronanki, R. (2018). Artificial Intelligence for the Real World. Harvard Business Review, 96(1), 108–116.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., and Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy. International Journal of Information Management, 57, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Fridgeirsson, T. V., Ingason, H. T., Jonasson, H. I., and Jonsdottir, H. (2021). An Authoritative Study on the Near Future Effect of Artificial Intelligence on Project Management Knowledge Areas. Sustainability, 13(4), Article 2345.

High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. European Commission.

Martinsuo, M., and Hoverfält, P. (2018). Change Program Management: Toward a Capability for Managing Value-Oriented, Integrated Multi-Project Change in Its Context. International Journal of Project Management, 36(1), 134–146.

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce.

Organisation for Economic Co-operation and Development. (2019). OECD Principles on Artificial Intelligence. OECD.

Project Management Institute. (2016). Governance of Portfolios, Programs, and Projects: A Practice Guide. Project Management Institute.

Project Management Institute. (2024). The Standard for Program Management (5th ed.). Project Management Institute.

Rai, A. (2020). Explainable AI: From Black Box to Glass Box. Journal of the Academy of Marketing Science, 48(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Raisch, S., and Krakowski, S. (2021). Artificial Intelligence and Management: The Automation–Augmentation Paradox. Academy of Management Review, 46(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Roehrich, J. K., Davies, A., Tyler, B. B., Mishra, A., and Bendoly, E. (2024). Large Interorganizational Projects (LIPs): Toward an Integrative Perspective and Research Agenda on Interorganizational Governance. Journal of Operations Management, 70(1), 4–21. https://doi.org/10.1002/joom.1280

Shrestha, Y. R., Ben-Menahem, S. M., and von Krogh, G. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. California Management Review, 61(4), 66–83. https://doi.org/10.1177/0008125619862257

Wen, Y., Wang, J., and Chen, X. (2025). Trust and AI Weight: Human–AI Collaboration in Organizational Management Decision-Making. Frontiers in Organizational Psychology, 3, Article 1419403. https://doi.org/10.3389/forgp.2025.1419403