Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 644–651
https://doi.org/10.54941/ahfe1008296

Governing Agentic AI Workflows: Balancing Automation and Human Oversight in Enterprise Practice

Natsumi Takahashi[1] and Indranil Bhattacharjee[2]

[1]Hitachi, Ltd., 1-1-2, Kashimada, Saiwai-ku, Kawasaki, Kanagawa, 212-8567, Japan
[2]Birla Institute of Technology and Science, Pilani, Rajasthan, India

ABSTRACT

Agentic AI is extending organizational automation beyond predefined tasks, requiring organizations to reconsider what should be delegated to AI and what forms of human involvement should remain. This study compares three enterprise AI development cases. One sought to reduce cost by substituting human work. Another made AI easier to use by letting practitioners assemble assistants through natural-language instructions. A third treated continued interaction as a way for an AI agent to become aligned with actual work. Greater AI capability did not lead to a stronger expectation of removing people from the workflow. In the learning-oriented case, evaluation, correction, and contextual input were not merely safeguards against AI failure. They became part of the mechanism through which the agent and the workflow were adapted. Human-centered design for agentic AI should therefore keep delegation adjustable and treat human correction as a productive component of AI-enabled work.

Keywords: Agentic AI, Human-centered design, Human–AI collaboration, Delegation, Human oversight, Co-adaptation

INTRODUCTION

Agentic AI is extending organizational automation beyond the execution of predefined tasks. By coordinating multiple processes, interpreting instructions, and taking actions within workflows, AI agents can assume activities that previously required substantial human involvement. This development raises a human-centered design question that goes beyond how much work can technically be automated. As AI systems become more autonomous, organizations must also determine what should be delegated to AI and what forms of human involvement should remain.

Previous research distinguishes automation from augmentation (Raisch and Krakowski, 2021), frames the human–technology relationship as delegation (Baird and Maruping, 2021), and argues that effective human–automation systems require a dynamic balance among ability, authority, control, and responsibility (Flemisch et al., 2012). These studies provide conceptual principles for balancing humans and automation. Less is known about how practitioners make these decisions in enterprise AI development

Received October 16, 2026; Revised November 24, 2026; Accepted December 3, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

projects: what they consider important when assigning work to AI agents, and how they attempt to preserve human roles as automation expands.

This study addresses this issue through a qualitative comparison of three enterprise AI development cases. The cases reflected different expectations of AI: automation and operational cost reduction, more flexible AI use through natural-language development, and continued adaptation of AI agents through interaction with practitioners. Drawing primarily on participant observation, the study examines how practitioners defined and adjusted the boundary between AI delegation and human involvement as the systems were developed and used.

The study is guided by the following research question: How do practitioners determine and adjust what to delegate to AI agents and what forms of human involvement to retain as enterprise AI workflows evolve?

LITERATURE REVIEW

From Automation to Human-Centered AI

How work should be divided between humans and technology has long been central to socio-technical and automation research. Socio-technical systems research emphasizes that technical systems and human work arrangements should be designed together (Trist and Bamforth, 1951; Baxter and Sommerville, 2011). Automation research made this allocation more explicit. Parasuraman, Sheridan and Wickens (2000) showed that automation can be distributed differently across information acquisition, analysis, decision making, and action implementation. Increasing automation does not require transferring every stage of work to machines.

Extensive automation can also create new demands for human judgment. Bainbridge (1983) noted that automation may remove routine human activity while leaving people responsible for difficult exceptions and failures. Human-Centered AI treats automation and human control as compatible objectives. Shneiderman (2020) argues that high levels of automation can coexist with high levels of human control when people retain opportunities to monitor, intervene, correct, and reject system outputs.

Automation, Delegation, and Dynamic Balance

The issue becomes sharper as AI systems become agentic. Raisch and Krakowski (2021) conceptualize the relationship between AI and human work as an automation–augmentation paradox. Automation refers to transferring activities from humans to machines, whereas augmentation refers to combining AI capabilities with human judgment. The two may coexist and influence one another within the same organization.

For agentic systems, the relationship can also be understood as delegation. Baird and Maruping (2021) distinguish agentic information systems from passive tools because they can accept delegated tasks and initiate actions. Their framework highlights three processes: appraisal of whether an activity should be delegated, distribution of work between actors, and coordination after delegation. The relevant question is what is delegated to AI and how human and AI activities are coordinated afterward.

Delegation also raises questions of authority, control, and responsibility. Flemisch et al. (2012) describe the relationship between humans and automation as a dynamic balance among ability, authority, control, and responsibility (A2CR). These dimensions need to remain aligned as system capabilities change, so the balance may need to be reconsidered during development.

The three perspectives are complementary. Automation–augmentation explains whether AI is expected to substitute for or complement human work. Delegation focuses on what practitioners transfer to AI and how work is subsequently distributed between practitioners and AI. Dynamic balance further draws attention to how authority, control, and responsibility need to be adjusted as the capabilities of automated systems change.

However, much of this literature remains conceptual or focuses on relatively stable forms of automation. Less is known about how these decisions are made in practice when organizations develop agentic AI systems whose outputs are imperfect, uncertain, or highly context-dependent. It remains unclear how practitioners revise their initial expectations of AI, learn what can and cannot be delegated, and accommodate imperfect outputs as part of actual work. This study addresses this gap by examining how practitioners negotiate delegation, human involvement, and the evolving balance between AI capability and human oversight during the development and use of enterprise agentic AI systems.

METHODOLOGY

The study uses a qualitative comparative case design. The role assigned to AI is shaped by the purpose of a project, expectations toward automation, the character of the work, and judgments made during development. Comparing cases within their organizational context can identify recurring patterns and meaningful differences (Eisenhardt, 1989; Eisenhardt and Graebner, 2007). The comparison was used to examine how different expectations toward AI shaped task allocation and the forms of human involvement that practitioners considered important.

The three cases differed in their starting points. The first was initiated to reduce operational costs through automation. The second emphasized easier and more flexible use of AI. Practitioners assembled assistants through natural-language development, without conventional programming. The third started from an expectation that an AI agent could adapt to work through continued interaction with practitioners.

The evidence came primarily from participant observation of enterprise AI development, supplemented by project documents, meeting records, design artifacts, and records of testing and adjustment. The unit of analysis was the episode in which participants discussed what should be automated, what could be delegated to an AI agent, where human confirmation or correction remained necessary, how incomplete or uncertain outputs were evaluated, and whether the expected role of AI changed as they gained experience with the system.

Each case was first described in terms of its objectives, task allocation, and observed forms of human involvement. The cases were then read through the three lenses above: whether AI was expected to substitute for or complement human work; what was delegated and what remained under

human control; and how practitioners accommodated imperfect AI outputs through a growing understanding of the system's limitations. The cross-case comparison focused on the logic of adoption. Particular attention was given to whether human oversight functioned only as a safeguard against imperfect AI, or also as a productive part of the workflow.

FINDINGS

Case 1: Automation as Labor Substitution

The first case was a top-down initiative aimed at reducing operational costs. AI optical character recognition (AI-OCR) was used to read quotations and application forms, extract and check required data, and prepare it for registration in an existing business system. The goal was to automate document receipt, data entry, and checking activities previously performed by employees.

Although AI was involved, the logic remained similar to earlier automation such as robotic process automation (RPA): existing human tasks were transferred to technology wherever possible. Usefulness was associated with lower labor and operating costs, and human involvement was treated mainly as something to be reduced.

Case 2: AI Adoption as Flexible Citizen Development

The second case involved an internal work assistant that practitioners assembled themselves in an agent-creation tool, using natural-language instructions rather than conventional programming. Practitioners could load internal manuals, standards, and organizational rules into the tool and retrieve relevant clauses or guidance through a chat interface.

The approach lowered technical barriers and allowed practitioners to incorporate their own knowledge into the assistant. In this form of citizen development, practitioners built assistants for their own work, so the main expectation was easier and more flexible AI development and use. Performance remained highly dependent on prompting and configuration, and the agents were typically designed for specific tasks.

Case 3: AI as a Learning Participant in Work

The third case involved an agentic AI application designed to analyze energy flows across factory equipment and recommend facility modifications to reduce CO_2 emissions and operating costs. The agent combined energy-flow data, equipment configuration diagrams, and physical calculations. However, practitioners recognized that technical optimization alone was insufficient because actual decisions also depended on business considerations, proprietary know-how, local practices, and changing information.

They therefore wanted the agent to incorporate knowledge that could not be fully represented in physical models and to reflect corrections and practical knowledge accumulated through use. The question was not simply whether AI could automate an existing task, but whether it could become better aligned with actual work through continued interaction.

A recurring expectation was that the system should support *suriawase*, or iterative alignment through continued use. Practitioners did not expect all relevant knowledge and rules to be specified completely before deployment. Instead, they expected to judge recommendations, correct weak outputs, add knowledge from the work, and gradually change how tasks were shared with the AI.

Importantly, expectations also changed as practitioners gained experience with the system. Rather than continuing to expect fully reliable or near-perfect recommendations, practitioners developed a more realistic understanding of what the AI could and could not do. They learned which parts of the task could be delegated with confidence, where human judgment remained necessary, and how imperfect outputs could be accommodated within the workflow.

This distinguishes the case from robots and RPA. Human involvement was not treated only as a safeguard against AI error or as something to be removed as technical performance improved. It also supplied the contextual knowledge and feedback through which the agent and the workflow could be adapted to practice.

The expected relationship in Case 3 can be expressed as:

Human involvement → evaluation and correction → contextual feedback → AI and workflow adaptation → improved practical usefulness → revised delegation

In Case 1, the sequence ran in the opposite direction:

AI capability → increased automation → reduced human involvement

Case 1 sought to replace existing human activity through automation. Case 3 instead showed a process in which practitioners learned, through use, how much work could be delegated to AI and where continued human involvement remained valuable.

Table 1: Cross-case comparison of expectations toward AI and human involvement.

Analytical Dimension	Case 1: Automation-Driven	Case 2: AI-Enabled Development	Case 3: Learning-Oriented Agentic AI
Primary objective	Reduce operational cost	Make AI easier to use	Adapt AI to actual work
AI role	Substitute human work	Flexible development tool	Participant in an evolving workflow
Dominant logic	Automation / substitution	Enablement / delegation	Learning / co-adaptation
Human role	Target of automation	Citizen developer	Evaluator, corrector, and guide
Delegation	Transfer tasks to technology	Simplify delegation through natural-language instructions	Adjust delegation through experience
View of AI limitations	Technical problems to reduce	Managed through configuration	Limits learned through use
Human oversight	Exception handling	Configuration and use	Evaluation, correction, and adaptation

Cross-Case Comparison

The comparison shows that increased AI capability did not necessarily lead to a stronger expectation of removing humans from work. Expectations depended on the role practitioners assigned to AI. In Case 1, usefulness meant replacing existing human activity. In Case 2, it meant making AI-based development more accessible and flexible. In Case 3, usefulness emerged through continued interaction, as practitioners learned both the capabilities and the limitations of the agent and adjusted the boundary between delegation and human involvement accordingly. Table 1 summarizes the three logics.

DISCUSSION

Three Logics of AI Use

The cases were guided by different logics: substitution, making AI easier to use, and adaptation through continued human–AI interaction. Case 1 asked how much human work could be replaced. Case 2 asked how more people could delegate tasks to AI. Case 3 asked how continued interaction could bring the agent into closer alignment with the work. The rest of this section develops the implications of that third logic.

Extending the Automation–Augmentation Perspective

The findings support the automation–augmentation paradox (Raisch and Krakowski, 2021) and add a distinction that matters for agentic AI. Case 1 closely reflects automation, because activities previously performed by humans were expected to move to technology. Case 3 is closer to augmentation, because the value of AI emerged together with human knowledge and judgment.

Case 3 also extends augmentation as it is usually described. Augmentation generally emphasizes how AI enhances human capabilities. Here the other direction was equally visible: human evaluation, correction, and contextual input contributed to the practical adaptation of the agent. The working relationship was one in which humans and AI continuously adjusted the workflow. Practitioners revised how they used and delegated work, while the workflow was refined through their feedback. This is co-adaptation. Automation and augmentation can therefore belong to an evolving relationship, with the balance changing through use.

Delegation and Continued Human Involvement

The findings also extend the delegation perspective of Baird and Maruping (2021). Agentic AI makes it possible to delegate activities that previously required human execution. In these cases, delegation still left a place for human participation. In Case 1, transferring an activity to technology was expected to reduce that participation. In Case 3, tasks could be assigned to an agent while people continued to evaluate outputs, correct them, interpret exceptions, and reconsider the boundaries of delegation.

Two design questions follow. What can be delegated to AI? What forms of human involvement remain valuable after delegation? Delegation is an

adjustable arrangement. As practitioners gain experience with AI outputs, its scope may expand, contract, or be reconfigured.

Human Oversight as a Productive Component

The dynamic balance framework (Flemisch et al., 2012) helps explain why this adjustment continues. As AI capabilities change, the allocation of ability, authority, control, and responsibility may also need to change. These dimensions cannot be determined by technical capability alone. Even when an agent could execute a task, practitioners kept authority and responsibility with humans when contextual judgment, correction, or accountability remained important.

Human oversight was a source of adaptation as well as a defense against error. In the learning-oriented case, review, correction, and interaction provided the information needed to bring the workflow closer to actual work requirements. Oversight therefore had two functions. It was protective, by preventing inappropriate outputs from directly affecting work, but it was also productive, by providing the feedback and contextual judgment through which the AI-enabled workflow improved.

Figure 1 summarizes this cycle.

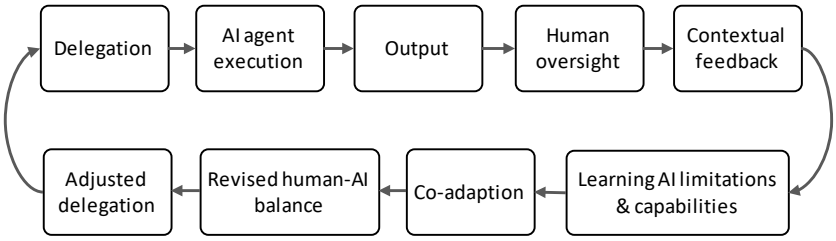


Figure 1: Human oversight as a mechanism for learning and continuous adaptation in agentic AI workflows.

Implications for Human-Centered Design

Human-centered design for agentic AI should support continued adjustment of the human–AI relationship, allowing the level of human control to change as the workflow evolves.

Three implications follow. First, delegation should remain visible and adjustable. Practitioners need to see which activities have been delegated and be able to change those boundaries as they learn more about the system. Second, correction and intervention should be treated as normal parts of the workflow. With generative and agentic AI, working with imperfect outputs can help practitioners understand the system's limits and adapt it to organizational requirements. Third, responsibility and control should remain aligned. Where humans remain accountable for outcomes, they need the authority to inspect, question, modify, or reject AI-generated actions.

Governing agentic AI is therefore not simply a matter of increasing autonomy as technical capability improves. It also requires conditions that

allow practitioners to learn where AI can be trusted, where human involvement still adds value, and how the balance between the two should change over time.

CONCLUSION

This study compared three enterprise AI development cases to examine how practitioners decide what to delegate to AI agents and what forms of human involvement to retain. The cases showed three orientations: substitution aimed at cost reduction, making AI easier to use through natural-language development, and learning-oriented co-adaptation. Increasing AI capability did not imply decreasing human involvement. In the learning-oriented case, practitioners developed a more realistic understanding of what the AI could and could not do as they gained experience with the system. Human oversight was both protective and adaptive. Evaluation, correction, and contextual input prevented inappropriate outputs from passing directly into work while also providing feedback through which the workflow could be adapted. The design question is therefore not simply how much autonomy AI should have, but how organizations can maintain an adaptable relationship between human participation and machine autonomy.

REFERENCES

Bainbridge, L. (1983). Ironies of automation, Automatica Volume 19 No. 6, pp. 775–779.

Baird, A. and Maruping, L.M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts, MIS Quarterly Volume 45 No. 1, pp. 315–341.

Baxter, G. and Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering, Interacting with Computers Volume 23 No. 1, pp. 4–17.

Eisenhardt, K.M. (1989). Building theories from case study research, Academy of Management Review Volume 14 No. 4, pp. 532–550.

Eisenhardt, K.M. and Graebner, M.E. (2007). Theory building from cases: Opportunities and challenges, Academy of Management Journal Volume 50 No. 1, pp. 25–32.

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A. and Beller, J. (2012). Towards a dynamic balance between humans and automation: Authority, ability, responsibility and control in shared and cooperative control situations, Cognition, Technology & Work Volume 14, pp. 3–18.

Parasuraman, R., Sheridan, T.B. and Wickens, C.D. (2000). A model for types and levels of human interaction with automation, IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans Volume 30 No. 3, pp. 286–297.

Raisch, S. and Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox, Academy of Management Review Volume 46 No. 1, pp. 192–210.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe and trustworthy, International Journal of Human–Computer Interaction Volume 36 No. 6, pp. 495–504.

Trist, E.L. and Bamforth, K.W. (1951). Some social and psychological consequences of the longwall method of coal-getting, Human Relations Volume 4 No. 1, pp. 3–38.