

Training Every Party to Manage Agentic Outsourcing

Prove Capability Across the Delivery Relationship

Phil Hatch | Akholi

October 23, 2026

Managing agentic outsourcing requires people in both firms to authorize work, challenge results, accept outcomes, disclose interventions, and recover the service. Train participating roles for those duties and test the relationship through joint rehearsal before expanding consequential authority.

Train Both Firms to Manage the Delivery Relationship

Every party involved in agentic outsourcing needs preparation for managing the two-party delivery model. Application instruction cannot establish whether a person can authorize an action, challenge an output, accept a business result, communicate an intervention, or recover a service. Those responsibilities depend on decisions and evidence from the other firm. Training must therefore teach each role's duties and how to secure the response on which its work depends.

An illustrative correction exposes the management problem. A provider operator recognizes a defective result and knows how to repair it, but neither firm can identify who may approve the correction or notify the buyer. Additional tool instruction would leave that failure unresolved. The operator needs an agreed decision route, an authorized counterpart, and rules for recording and communicating the intervention. Joint preparation must establish whether participants understand those arrangements and whether the arrangements permit them to act.

Leadership should appoint an accountable buyer sponsor and provider counterpart, assess participating roles within 30 days, and fund common instruction, role-specific practice, and joint rehearsal. Require demonstrated readiness before assigning or expanding consequential authority. Critical existing gaps should remain subject to proportionate supervision or

restrictions while correction proceeds. The 30-day assessment and subsequent 90-day program are proposed planning targets, adjustable to exposure and resources.

The Study Identifies a Gap Between Tool Skills and Delivery Readiness

In the banking study, 380 of 804 valid ratings of how well vendor training built practical agentic AI skills, or 47.26 percent, were 3 or below on the 10-point scale. Respondents described theoretical courses, tools different from those in use, and insufficient preparation for variable outputs; others reported useful hands-on instruction. The ratings measure perceived preparation, not tested competence or a causal training effect.¹

A contributor in credit, underwriting, and suitability stated, “tool training was good, nobody taught us the suitability rules.” Accounts of immature delivery playbooks also describe escalation and production procedures being invented during live work. These findings make the service’s business judgment and operating relationship necessary subjects of preparation. A course program cannot supply an agreed delivery method where none exists.²

Define Competence Through Each Role’s Decisions

The common core should explain the outsourced outcome, each firm’s contribution, the division of authority, and the evidence required for acceptance. Participants need to understand consequential handoffs, decision deadlines, escalation, and the conditions for restricting work. Role-specific instruction then establishes what each person must do within that shared method. Executives should be able to challenge assurance and approve operating conditions; operators should be able to recognize exceptions and use the authorized route.

The program must reach the people whose decisions sustain the service: sponsors and business owners; provider operators and managers; procurement, legal, finance, technology, security, risk, compliance, internal audit, and learning teams. Include deputies, replacements, and relevant supporting-provider personnel under appropriate contractual arrangements. Switzerland’s financial supervisor, FINMA, examined broad AI training and whether outsourced

¹ Phil Hatch, *Agentic AI Outsourcing at Tier 1 Financial Institutions: Full Study Report*, version 1.8.2 (Akholi, forthcoming manuscript dated November 1, 2026), Part 2, “Practical Agentic Skills Built by Vendor Training,” <https://doi.org/10.2139/ssrn.7164559>.

² Hatch, *Agentic AI Outsourcing*, Part 5, “Adequacy of Regulatory Training,” individual-contributor account, CXM, credit, underwriting, and suitability; Hatch, *Agentic AI Outsourcing*, Part 2, “Maturity of the Delivery Framework.”

parties had the necessary skills and experience.³ Figure 1 translates this relationship-wide coverage into proposed demonstrations of competence.

Figure 1. A common delivery model, different demonstrations of competence

Role cluster in either firm	Demonstration in the actual service
Sponsors and commercial owners	Approve bounded delivery, truthful value claims and funded oversight.
Delivery and domain teams	Resolve handoffs, challenge output and record material human changes.
Technology, data and security	Demonstrate access limits, accepted changes, containment and recovery.
Risk, legal and compliance	Apply obligations and determine disclosure or notification escalation.
Independent assurance	Reperform controls and test closure independently of correction.
People and learning leaders	Identify role gaps; fund learning time; assess and refresh competence.

Buyer and provider share the service model. Include replacements and relevant subcontractor staff. Assess actual duties; these clusters are not a fixed organization chart.

Source: Author's proposed role-and-assessment framework; customize duties, criteria, and depth for the actual service. This is not a validated competency standard.

For each consequential duty, record the required decision or action, assessment criterion, supporting evidence, qualified assessor, and refresh trigger. Assessing an assurance reviewer, for example, should establish whether that person can determine what the report covers, what was tested, and which gaps require challenge. In material outsourcing, the UK PRA expects appropriate expertise, qualifications, and skills in both the auditing or certifying party and the firm's reviewer of its report.⁴ The same duty-based approach distinguishes demonstrated

³ Swiss Financial Market Supervisory Authority FINMA, *Governance and Risk Management When Using Artificial Intelligence*, FINMA Guidance 08/2024, December 18, 2024, sec. 2.1, 4, https://www.finma.ch/~media/finma/dokumente/dokumentencenter/myfinma/4dokumentation/finma-aufsichtsmittelungen/20241218-finma-aufsichtsmittelung-08-2024.pdf?sc_lang=en.

⁴ Prudential Regulation Authority, *Outsourcing and Third Party Risk Management*, Supervisory Statement SS2/21, November 2024, para. 8.9 and table 8, 29, <https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/supervisory-statement/2024/ss221-november-2024-update.pdf>.

readiness from confidence; the study's knowledge ratings were self-assessments, not independent tests.⁵

Jurisdiction-specific instruction should explain how the duty changes for the cases a person handles. In a Swiss, UK, EU, or US service, participants need the approved acceptance and authorization rules, record and access conditions, and escalation or notification route relevant to those cases. Where requirements differ, use paired exercises: present the same proposed correction under two approved case rules and require the participant to identify the authorized decision maker and required record or communication. Local specialists must supply and verify those differences. This prepares people to apply the service's rules rather than infer authority from a generic course.

Instruction should use the service's inputs, outputs, permissions, and acceptance rules. Participants must distinguish a valid result from uncertainty requiring domain judgment, record material human changes and their authority, and make the required buyer communication. Sponsors and commercial teams should practice reconciling accepted outcomes with activity and savings claims. NIST's voluntary AI Risk Management Framework includes training personnel and partners for their risk-management duties under relevant policies and agreements; the proposed curriculum gives those duties a service-specific form.⁶

Rehearse the Decisions That Cross the Firm Boundary

Joint rehearsal should test the moment when individual knowledge must produce a response from the other firm. The OECD's qualitative practitioner interviews describe calibrated decision checkpoints and concerns about overreliance and loss of expertise.⁷ Rehearsal should therefore require participants to exercise judgment at the points where the service depends on it. Use representative records, current permissions, and the agreed escalation route in a controlled environment. Introduce variable output, a delayed decision, a provider change, or an omitted correction, then observe whether both firms obtain an authorized decision, preserve evidence, communicate, and recover. Time expectations should follow the potential harm and decision requirement.

For example, give the provider an output that conflicts with an approved rule while the buyer's usual approver is unavailable. The operator must preserve the original and any proposed correction, reach the authorized deputy, and apply the agreed limit or fallback while waiting. The buyer must decide and record the decision; the engagement manager must reconcile

⁵ Hatch, *Agentic AI Outsourcing*, Part 4, "Self-Assessed Knowledge and Expertise."

⁶ Elham Tabassi, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1 (National Institute of Standards and Technology, January 26, 2023), GOVERN 2.2, <https://doi.org/10.6028/NIST.AI.100-1>.

⁷ Organisation for Economic Co-operation and Development, *Agentic AI in Organisations: Early Insights from Practitioner Interviews*, OECD Artificial Intelligence Papers 65 (September 16, 2026), secs. 4.1–4.2, <https://doi.org/10.1787/1257a26f-en>.

acceptance and reporting with the intervention. Safe test records allow the exercise to expose a defective handoff without creating live customer exposure.

Task-specific assessment matters because assistance can improve performance on one task and impair it on another. A Harvard-led experiment published in 2026 with BCG consultants found that uneven pattern even among skilled knowledge workers. It studied GPT-4-assisted consulting tasks, not an outsourcing curriculum. The relevant implication is to assess the decisions assigned in the service, including cases requiring participants to reject or correct plausible assistance.⁸

Correct decisions and participant satisfaction answer different assessment questions. In an experiment by Bućinca and colleagues, 199 online participants made meal-ingredient substitutions with simulated AI assistance. Designs requiring deliberation reduced inappropriate reliance without eliminating it, while participants liked the more effective designs less. The study evaluated an interface rather than management training, but it gives a reason to keep course satisfaction separate from observed performance when assistance is wrong.⁹

A qualified assessor should record the observed action against criteria agreed beforehand, including the decision record left by both firms. The finding should identify the exact decision or handoff that failed and the consequence for the service. That record allows the sponsor to distinguish a participant who needs further practice from a capable participant blocked by an unusable operating arrangement.

Assess deputies and cross-shift handoffs as well as the best-prepared individual. A missed critical criterion should result in targeted practice and reassessment, with proportionate limits on the affected duty. The result must remain traceable to the observed failure so that the sponsor can select the appropriate remedy.

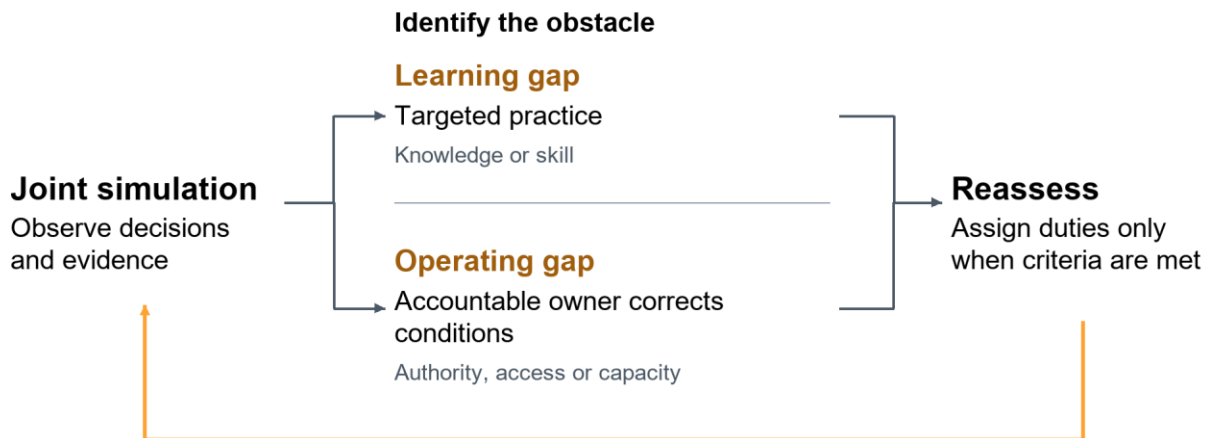
Use the Failure Evidence to Select the Remedy

A failed rehearsal must identify what prevented the correct response. Insufficient knowledge requires instruction and practice. An unavailable approver, absent permission, incompatible record, inadequate review capacity, or incentive discouraging disclosure requires management correction. Both defects can occur in the same event. Figure 2 separates the remedies while returning them to a shared reassessment.

⁸ Fabrizio Dell'Acqua et al., "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality," *Organization Science* 37, no. 2 (2026): 403–4 and 417–18, <https://doi.org/10.1287/orsc.2025.21838>.

⁹ Zana Bućinca et al., "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making," *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW1 (2021): article 188, secs. 3–4, <https://doi.org/10.1145/3449287>.

Figure 2. Use the simulation result to choose the right remedy



Repeat after correction and material change

Training cannot grant authority, access or capacity. Restrict affected duties while critical gaps remain.

Source: Author's proposed assessment and remediation loop; use role-specific criteria and qualified assessors. Training cannot grant authority, access, or capacity.

Attendance establishes participation, and a successful exercise establishes capability under tested conditions. Current authority and access determine whether the participant can act; operating evidence establishes whether the live arrangement performs. Leadership should keep these forms of evidence separate. A trained person cannot exercise ungranted authority or review more work than the available time permits, so a course certificate cannot close a critical operating gap.

Delivery managers should correct the operating arrangement, risk and compliance should discharge their respective challenge and monitoring duties, and internal audit should retain the objectivity needed to assess the result. An exercise that reveals a defective process should change that process before reassessment. Repeating instruction while leaving the obstacle in place would treat the symptom as the cause.

Fund Preparation and Maintain It as the Service Changes

The first 30 days should establish the participating roles, critical duties, priority gaps, and funded joint plan. The budget needs to cover domain specialists, protected learning time, backfill, simulations, assessment, and repeat work. Agree who pays and who provides evidence. Where a supporting provider cannot participate directly, establish another defensible way to verify its interface duties and escalation arrangements. Training repeatedly displaced by the delivery queue has not been implemented.

During days 31–90, deliver the common core, rehearse assigned decisions, and assess priority personnel with their counterparts while correcting operating obstacles. Extend assessed coverage through the relationship as those prerequisites are met. McKinsey's October 2026

organizational analysis recommends deliberate practice, feedback, and progressively consequential judgment as agents absorb traditional learning tasks. Applying that discipline to the service preserves opportunities to develop expertise rather than assuming that access to an agent will develop it.¹⁰

Material changes, reassignment, turnover, incidents, and recurring failures should trigger refreshed instruction or reassessment. Participants need to recognize changes to models, tools, business rules, permissions, supporting providers, and delivery methods that affect their duties. Keep usable instructions and deputies current, and rehearse evidence preservation, suspension, fallback, notification assessment, and repair where relevant. Leadership reporting should show assessed coverage, unresolved gaps, restrictions, and operating failures; completion totals remain an administrative measure.

Make Demonstrated Competence a Condition of Delivery

Approve the management-training mandate with accountable counterparts, funded practice, and dated evidence returns. Each participating role should demonstrate the decisions its responsibilities require, and both firms should demonstrate that their arrangements permit those decisions. This provides a basis for continuing or expanding authority while keeping learning gaps, operating defects, and unresolved exposure visible. The standard is a manageable joint service supported by competence, authority, capacity, and controls.

Evidence and Source Note

The study records 850 vendor delivery-team respondents on 150 Tier 1 financial-institution engagements during January–June 2026 in a nonrepresentative, single cross-section, without bank or vendor corroboration. Its perceptions and associations do not establish tested competence, causation, or all-industry prevalence. The proposed curriculum and assessment framework require service-specific evaluation. The author-supplied version 1.8.2, dated November 1, 2026, was consulted as forthcoming material; external sources were checked through October 23, 2026.¹¹

¹⁰ McKinsey & Company, “AI Is Changing Work. Now It Has to Change the Organization,” October 5, 2026, “Talent: How to Build a System Around Work That Keeps Changing,” <https://www.mckinsey.com/capabilities/people-and-organization/our-insights/ai-is-changing-work-now-it-has-to-change-the-organization>.

¹¹ Hatch, *Agentic AI Outsourcing*, Introduction, “Evidentiary Basis and Its Limits.”

Bibliography

- Buçinca, Zana, Maja Barbara Malaya, and Krzysztof Z. Gajos. "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making." *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW1 (2021): article 188. <https://doi.org/10.1145/3449287>.
- Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, et al. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality." *Organization Science* 37, no. 2 (2026): 403–423. <https://doi.org/10.1287/orsc.2025.21838>.
- Hatch, Phil. *Agentic AI Outsourcing at Tier 1 Financial Institutions: Full Study Report*. Version 1.8.2. Akholi. Forthcoming manuscript dated November 1, 2026. Author-supplied manuscript consulted October 23, 2026. <https://doi.org/10.2139/ssrn.7164559>.
- McKinsey & Company. "AI Is Changing Work. Now It Has to Change the Organization." October 5, 2026. <https://www.mckinsey.com/capabilities/people-and-organization/our-insights/ai-is-changing-work-now-it-has-to-change-the-organization>.
- Organisation for Economic Co-operation and Development. *Agentic AI in Organisations: Early Insights from Practitioner Interviews*. OECD Artificial Intelligence Papers 65. OECD Publishing, September 16, 2026. <https://doi.org/10.1787/1257a26f-en>.
- Prudential Regulation Authority. *Outsourcing and Third Party Risk Management*. Supervisory Statement SS2/21. Bank of England. November 2024. <https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/supervisory-statement/2024/ss21-november-2024-update.pdf>.
- Swiss Financial Market Supervisory Authority FINMA. *Governance and Risk Management When Using Artificial Intelligence*. FINMA Guidance 08/2024. December 18, 2024. https://www.finma.ch/~/-/media/finma/dokumente/dokumentencenter/myfinma/4dokumentation/finma-aufsichtsmittelungen/20241218-finma-aufsichtsmittelung-08-2024.pdf?sc_lang=en.
- Tabassi, Elham. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. National Institute of Standards and Technology, January 26, 2023. <https://doi.org/10.6028/NIST.AI.100-1>.