

Human Oversight Is a Capacity Commitment in the Deal

A Contract and Operating Method for Managers

Phil Hatch | Akholi

October 23, 2026

Sourcing managers should turn human-oversight promises into defined capacity, authority, response, and evidence commitments. A clause requiring a human in the loop is incomplete if neither firm has agreed who reviews what, when, and with which resources.

Define the Review the Contract Purchases

The sourcing manager should identify the exact judgment required at each material review point. State whether the person approves before action, handles exceptions, samples completed work, or collaborates throughout the task. These are different services with different capacity and authority requirements.

Stanford's selected enterprise cases describe oversight as dependent on task and consequence.¹ The contract should therefore avoid a universal review formula. It should specify the work necessary for the actual service and the evidence that the chosen review design operates.

For a hypothetical customer-remediation service, a reviewer may need transaction history, policy, prior communications, and authority to approve a correction. Providing only the agent's proposed answer and a short timer does not supply that capability. The schedule should identify the information and decision rights the reviewer needs.

¹ Elisa Pereira et al., *The Enterprise AI Playbook: Lessons from 51 Successful Deployments*, Stanford Digital Economy Lab, April 2026, methodology and chapters 1–3 and 8; purposive successful-case sample, https://digitaleconomy.stanford.edu/app/uploads/2026/03/EnterpriseAIPlaybook_PereiraGraylinBrynjolfsson.pdf.

Allocate Capacity Across Both Firms

The study reports human review mismatched to agent output and routine approval practices.² Managers should test their own staffing and workflow rather than infer a universal ratio from those accounts. The important question is whether qualified time can cover the required work at representative and peak demand.

Assign the buyer and provider contributions separately. Record who performs initial review, specialist escalation, quality assurance, and incident investigation. The same person should not be counted at full capacity in several commitments at once. Include absences, training, and the time needed to obtain missing evidence.

The commercial schedule should identify the capacity included in the price and the treatment of demand outside agreed assumptions. A provider should not silently return excess work to the buyer while continuing to report the service as fully managed. The buyer should not assume unpurchased specialist capacity is available on demand.

Use an Oversight Capacity Schedule

Figure 1 proposes four linked commitments: review scope, qualified resources, decision time, and overload response. Define the case types, evidence package, authority, normal capacity, and escalation path. State how the firms know a case is waiting and which clock applies while information or approval is outstanding.

Measure actual review effort and backlog during a pilot. Include difficult cases and clustered exceptions. The test should show what the reviewer inspected, decided, changed, or escalated. A completed approval field alone does not establish that meaningful review occurred.

The FCA's current practice review emphasizes authority, information, capability, and escalation in effective human oversight.³ That UK supervisory discussion does not prescribe a universal contract. It supports the manager's proposed test of whether the agreed operating resources match the oversight promise.

² 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, secs. 4.2–4.3, “Human Review Capacity Against Agent Output Volume” and “Routine Approval of Agent Output at Human Review Points,” <https://doi.org/10.2139/ssrn.7164559>.

³ Financial Conduct Authority, “AI in Financial Services: Good Practice and Areas for Improvement,” Financial Conduct Authority, October 19, 2026; updated October 20, 2026, secs. 1.2, 2, 3.4 and 4.2; illustrative practices, not new requirements, <https://www.fca.org.uk/publications/good-poor-practice/ai-financial-services>.

Figure 1. The oversight capacity service schedule

Commitment	Required definition	Acceptance evidence
Review scope	Judgment case types and evidence	Observed review and disposition
Qualified resources	Buyer and provider capacity	Skills availability and effort record
Decision time	Response and pending-work ownership	Timed case and escalation test
Overload	Restriction or added capacity	Backlog trigger and safe response

Author's proposed operating framework. Adapt to the service and applicable jurisdiction; not an empirically validated standard.

Source: Phil Hatch, proposed framework developed for this paper.

Specify the Response When Capacity Is Exceeded

Agree triggers for restricting agent volume, narrowing authority, routing to additional qualified staff, or suspending the affected work. The response should preserve case ownership and customer or business deadlines. An overload state should not automatically convert required review into unattended execution.

Changes in model behavior, volume, case mix, or tools can change the review burden. Require notice and a joint reassessment before carrying the old staffing assumption into the new service. A price reduction based on fewer human hours should be supported by evidence that the required oversight still operates.

EU, UK, Swiss, and US activities may impose different professional, conduct, privacy, and employment conditions. The schedule should reflect the relevant local requirements and authorized reviewer roles. A group service target cannot override a decision that requires a particular qualification or institutional authority.

Accept Capacity Before Accepting the Automation Claim

The practical objection is that a capacity schedule may become inflexible. Use workload assumptions, review bands, and a clear change route rather than a rigid staffing number alone. The schedule should allow improvement while preserving the required judgment and visibility of shortfalls.

Before production acceptance, reconcile the promised review with observed qualified time, case evidence, backlog, and overload behavior. Assign corrective owners where the test fails. Report unresolved conditions and the commercial effect rather than allow the buyer and provider to hold different assumptions about who will absorb the work.

The manager should sign an oversight commitment that can be staffed, performed, and evidenced. Human review is part of the delivery model and its economics. Treating it as a contractual phrase leaves the most important operating questions unanswered when the agent begins producing work at scale.

Evidence and Source Note

The study is a nonrepresentative cross-section of 850 vendor delivery-team respondents on 150 engagements serving Tier 1 financial institutions, collected during January–June 2026. Neither banks nor vendors corroborated the accounts. It establishes reported experience, not causation, legal breach, or market prevalence. Its application here is a mechanism to investigate, not an estimate of conditions in other industries. The author-supplied version 1.8.2 is dated November 1, 2026 and is cited as a forthcoming manuscript consulted October 23, 2026.⁴

This is an analytical decision paper based on the supplied study and selected primary public sources checked through October 23, 2026. The original contribution is a sourcing oversight-capacity schedule connecting required judgment, purchased qualified resources, decision-time obligations, and explicit overload behavior. The proposed method has not been independently validated. Legal requirements depend on the entity, activity, jurisdiction, and effective provisions; supervisory observations, research, and professional commentary have distinct authority.

Bibliography

Financial Conduct Authority. “AI in Financial Services: Good Practice and Areas for Improvement.” Financial Conduct Authority. October 19, 2026; updated October 20, 2026. <https://www.fca.org.uk/publications/good-poor-practice/ai-financial-services>.

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>.

Pereira, Elisa, Alvin Wang Graylin, and Erik Brynjolfsson. *The Enterprise AI Playbook: Lessons from 51 Successful Deployments*. Stanford Digital Economy Lab. April 2026. https://digitaleconomy.stanford.edu/app/uploads/2026/03/EnterpriseAIPlaybook_PereiraGraylinBrynjolfsson.pdf.

⁴ Hatch, *Agentic AI Outsourcing*, Introduction, “Executive Summary” and “Evidentiary Basis and Its Limits.”