

# Testing Production Readiness of Outsourced Agentic Services

## *An Executive Gate for Live Authority*

Phil Hatch | Akholi

October 23, 2026

Executives should release an outsourced agentic service only when both firms can demonstrate accepted outcomes, bounded authority, effective intervention, and recoverable operation in the intended environment. A successful pilot establishes less than production readiness.

## Approve a Defined Production Service

The executive sponsor should require a precise account of what will enter production: tasks, users, tools, data, permissions, human roles, and operating limits. A demonstration can show useful capability while excluding difficult cases, live-system constraints, and the buyer decisions on which reliable delivery depends.

BCG's regulated-industry platform discussion emphasizes reusable technical controls and operating discipline.<sup>1</sup> These are architectural recommendations, not proof that a particular service is ready. The executive needs evidence from the assembled buyer-provider workflow under the conditions it will encounter.

Define the minimum acceptable outcome and the authority required to produce it. Identify actions that remain prohibited, cases that require review, and conditions that cause work to pause. Approval should attach to this scope and configuration rather than to a broad label such as an AI-enabled service.

---

<sup>1</sup> Daniel Martines et al., "Building Enterprise AI Agents in Regulated Industries," Boston Consulting Group, July 20, 2026, enterprise platform and identity control sections, <https://www.bcg.com/publications/2026/building-enterprise-ai-agents-in-regulated-industries>.

## Test the Buyer Alongside the Provider

The study describes processes and organizations insufficiently adapted to agentic delivery.<sup>2</sup> A supplier can meet its own readiness checklist while waiting on unavailable buyer approvers, inaccessible data, or incompatible tools. Production readiness therefore belongs to the joint operating model.

For a hypothetical procurement service, the provider may correctly identify an exception but need an authorized buyer to resolve it before an order deadline. The readiness test should establish who receives the case, what information is available, and whether the decision can be made in time.

Test normal, difficult, rejected, manually corrected, interrupted, and recovered work. Observe the resulting business state. A tool acknowledgement should not become acceptance of an outcome that neither firm has reconciled. Record the limitations of test data and any significant production condition not represented.

## Use a Joint Readiness Decision Record

Figure 1 proposes four evidence gates: result, authority, intervention, and recovery. Assign a buyer and provider owner for each. Record the tested configuration, acceptance criteria, findings, unresolved limitations, and the person authorized to accept residual risk.

Include the human work that makes the service function. Review staffing, training, review capacity, decision access, and escalation coverage. An apparently automated workflow may depend on specialists who were available during the pilot but are not funded or scheduled for live operation.

Stanford's selected successful deployment cases emphasize complementary organizational work.<sup>3</sup> The cases do not estimate a general success rate. They support evaluating the capabilities surrounding the technology rather than treating model performance as a complete production decision.

---

<sup>2</sup> 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, "The Shift From Managing People to Managing Agents"; Part 6, "The Executive Action Plan," <https://doi.org/10.2139/ssrn.7164559>.

<sup>3</sup> 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](https://digitaleconomy.stanford.edu/app/uploads/2026/03/EnterpriseAIPlaybook_PereiraGraylinBrynjolfsson.pdf).

Figure 1. The joint production readiness record

| Release gate        | Required demonstration                   | Executive decision                     |
|---------------------|------------------------------------------|----------------------------------------|
| <b>Result</b>       | Accepted end-to-end work                 | Is the service useful and complete     |
| <b>Authority</b>    | Permitted and prohibited action tests    | Is exposure bounded                    |
| <b>Intervention</b> | Timely qualified buyer-provider response | Can exceptions be governed             |
| <b>Recovery</b>     | Stop pending work and restoration test   | Can the service fail within tolerances |

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.

## Make Conditional Approval Operational

Conditional approval should specify enforceable limits: permitted volume, action scope, case types, review requirements, expiry, and retest triggers. An unresolved issue should have a named owner and decision date. A condition that no system or operating team implements does not constrain production exposure.

Applicable EU, UK, Swiss, US, and other obligations should be assessed for the actual activity and entity. Production authorization, contractual acceptance, regulatory approval where required, and internal risk acceptance are separate decisions. One should not be represented as automatically satisfying the others.

The provider should disclose material dependencies and change triggers. A later update to the model, instructions, tools, data route, or subcontractor can invalidate part of the readiness evidence. Agree which changes remain within approved scope and which require renewed acceptance.

## Release What the Evidence Supports

A reasonable objection is that waiting for complete certainty prevents useful deployment. Readiness does not require certainty. It requires an explicit, proportionate decision about the demonstrated service, remaining uncertainty, and the controls that limit consequences while learning continues.

Before launch, have both firms complete a bounded end-to-end exercise and an interruption scenario. Require operations to reconcile pending work and demonstrate escalation, stop, and recovery. The executive should see a concise recommendation with the evidence route and limitations, not an unqualified readiness score.

Approve the service at the scope the evidence supports. Expand only when the joint operating model can sustain the additional authority and workload. This approach makes production readiness a defensible decision about deliverable outcomes rather than a ceremonial final step after the commercial commitment.

## 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.<sup>4</sup>

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 joint production-readiness record that separates demonstrated scope from enforceable conditions and future expansion. 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

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>.

Martines, Daniel, Cullen Breithaupt, Drew Ingalls, Sawan Ruparel, and Prithvi Vadlamudi. “Building Enterprise AI Agents in Regulated Industries.” Boston Consulting Group. July 20, 2026. <https://www.bcg.com/publications/2026/building-enterprise-ai-agents-in-regulated-industries>.

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](https://digitaleconomy.stanford.edu/app/uploads/2026/03/EnterpriseAIPlaybook_PereiraGraylinBrynjolfsson.pdf).

---

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