

Hospital AI Outsourcing Needs Proof of Completed Work

A Board Framework for Bounded Procurement

Phil Hatch | Akholi

October 23, 2026

Hospital boards should require evidence that an outsourced agent completes the purchased workflow under local conditions. Fluent documentation and strong component scores cannot establish that the necessary action occurred.

Purchase Demonstrated Completion

The hospital board should require management to define the completed business or clinical-support outcome before approving outsourced agentic delivery. The provider must show what changed in the relevant system and how the hospital verifies acceptance. A confident note or demonstration is insufficient when the promised action remains undone.

Stanford's October 2026 PhysicianBench report evaluated 100 multistep tasks in an electronic-health-record sandbox. The best tested model completed 46% on a single attempt. The report also identifies cases where agents documented an intended action without placing the order. These are benchmark findings, not a hospital-wide deployment failure rate.¹

The board implication is to require workflow evidence and qualified clinical oversight appropriate to the use. It is not to infer that every agent is unsuitable or that a published benchmark establishes safety. Management must demonstrate the proposed configuration in the hospital's own bounded context.

¹ Nigam Shah and Jonathan Chen, "Measuring What Medical AI Can Actually Do," Stanford University, October 19, 2026, "PhysicianBench and MedAgentBench" and "HealthAdminBench," <https://hai.stanford.edu/news/measuring-what-medical-ai-can-actually-do>.

Separate Administrative Assistance from Clinical Authority

Management should identify whether the service retrieves information, prepares documentation, coordinates administration, recommends care, or executes a consequential action. These functions carry different authority and assurance needs. An administrative contract should not become a route to clinical decisions that the hospital has not approved.

A hypothetical referral-support agent may assemble documentation and prepare an appointment request. Completion should require the agreed receiving-system state and escalation of unresolved cases. It should not include an unapproved clinical prioritization merely because that step would make the workflow appear complete.

The study's banking evidence describes manual corrections and reported completion inconsistent with actual results. Those mechanisms justify asking the hospital to reconcile work and reports. They do not establish healthcare prevalence, clinical harm, or a causal relationship between agent use and patient outcomes.²

Require Four Kinds of Procurement Evidence

Figure 1 proposes a board procurement framework: bounded authority, completed-work evidence, local safety assessment, and recoverable service. Management should provide a concise approval record for each. Unresolved evidence should constrain the scope of a pilot rather than disappear into a general supplier readiness score.

The test should represent the actual workflow, including missing information, failed interfaces, corrections, and handoffs. Inspect the result in the receiving system. The provider should disclose human assistance and retries so the hospital can distinguish reliable completion from a demonstration sustained by unreported intervention.

NHS England's June 2026 review explains the distinct clinical-risk responsibilities addressed by DCB0129 for manufacturers and DCB0160 for deploying organizations. The review is consultation context, not a new global standard. It reinforces the need to assess both product design and the hospital's local deployment responsibilities.³

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 5, "Five Manual Alteration Methods" and "Disclosure of Manual Alteration," table 5.8, <https://doi.org/10.2139/ssrn.7164559>; Hatch, *Agentic AI Outsourcing*, Part 3, AI Technology, sec. 4.1, "Accuracy of the Tools' Own Success and Failure Reports," indicator C172.

³ NHS England, "National Review of Clinical Risk Management Standards DCB0129 and DCB0160: Supporting Information," NHS England, June 29, 2026, "Why the standards were introduced"; consultation context, <https://www.england.nhs.uk/long-read/national-review-of-clinical-risk-management-standards-dcb0129-and-dcb0160-supporting-information/>.

Figure 1. The hospital board procurement decision

Approval area	Required management answer	Evidence before expansion
Authority	What the agent may and may not do	Bounded scope and enforced limits
Completion	How actual work is verified	Receiving-system result and handoffs
Local safety	Who owns clinical and privacy duties	Deployment-specific assessments
Recovery and value	How unfinished work is handled	Fallback staffing and full buyer cost

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.

Treat Local Safety and Privacy as Approval Conditions

For an NHS deployment, the hospital should obtain the relevant clinical safety assessment and confirm the applicable current standards. EU, Swiss, and US hospitals need separate analysis of their clinical, device where applicable, privacy, and institutional duties. An agent's commercial label does not settle its regulatory classification.

Swiss privacy guidance states that the FADP applies to AI-supported processing and addresses transparency and high-risk assessment. Other jurisdictions impose their own obligations. Management should explain how the service's data access, retention, and further use fit the hospital's actual purposes and patient responsibilities.⁴

Clinical and operational staff should be able to challenge the output and stop inappropriate progression. The staffing model must provide sufficient capacity for that role. Requiring a human reviewer in a contract is not evidence that a qualified person can intervene effectively at the volume and speed proposed.

⁴ Federal Data Protection and Information Commissioner, "AI and Data Protection," Federal Data Protection and Information Commissioner, September 24, 2025, "Transparent information" and "Protection of fundamental rights," <https://www.edoeb.admin.ch/en/ai-and-data-protection>.

Test Recovery and the Cost of Unfinished Work

A pilot should show how the hospital discovers incomplete work and returns it to a qualified owner. The provider should preserve case status and evidence rather than silently retry until a budget is exhausted. The hospital needs a safe process for urgent cases and for work whose completion is uncertain.

A pilot should include cases selected by the hospital, including difficult but permitted cases and known handoff problems. A provider-selected demonstration can show capability while underrepresenting local exceptions. Require the provider to disclose excluded cases, assistance, retries, and configuration changes during the test. Management should report those conditions alongside completion results so the board can judge whether the evidence supports the proposed deployment rather than only the demonstration.

The business case should include retained clinical or administrative effort, correction, assurance, integration, and continuity. McKinsey's agentic economics analysis supports examining total cost rather than tokens alone. Its examples do not determine hospital costs; local effort and accepted-work evidence must supply that comparison.⁵

A lower provider price can be outweighed by additional hospital work, while a well-designed hybrid service can still be worthwhile. Management should distinguish released capacity from cash savings and report improvements in quality or waiting time separately. Do not count unfinished work as an economic benefit.

Expand Only After Local Acceptance

The board should approve a bounded pilot with explicit success criteria, stop conditions, qualified ownership, and a review date. Management should report completion, safety findings, human contribution, unresolved cases, and full buyer cost. Evidence should support the particular authority being requested.

The procurement decision is whether the hospital can verify completed work and manage failure under local conditions. Expand only when the pilot supports that conclusion and the required clinical and legal assessments are complete. A persuasive demonstration should begin the inquiry, not close it.

⁵ Chandana Asif et al., "Where AI Agents Pay Off: A Practical Guide to the Economics of Agentic Workflows," McKinsey & Company, August 24, 2026, "What drives the unit economics of AI-based workflows?," <https://www.mckinsey.com/capabilities/quantumblack/our-insights/where-ai-agents-pay-off-a-practical-guide-to-the-economics-of-agentic-workflows>.

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, 2026.⁶

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 hospital board procurement framework linking bounded authority to execution-grounded acceptance and retained local duties. 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

Asif, Chandana, Dieter Kiewell, Lari Hämäläinen, Tom Kolaja, and Tunde Olanrewaju. “Where AI Agents Pay Off: A Practical Guide to the Economics of Agentic Workflows.” McKinsey & Company. August 24, 2026.

<https://www.mckinsey.com/capabilities/quantumblack/our-insights/where-ai-agents-pay-off-a-practical-guide-to-the-economics-of-agentic-workflows>.

Federal Data Protection and Information Commissioner. “AI and Data Protection.” Federal Data Protection and Information Commissioner. September 24, 2025. <https://www.edoeb.admin.ch/en/ai-and-data-protection>.

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>.

NHS England. “National Review of Clinical Risk Management Standards DCB0129 and DCB0160: Supporting Information.” NHS England. June 29, 2026. <https://www.england.nhs.uk/long-read/national-review-of-clinical-risk-management-standardsdcb0129-and-dcb0160-supporting-information/>.

Shah, Nigam, and Jonathan Chen. “Measuring What Medical AI Can Actually Do.” Stanford University. October 19, 2026. <https://hai.stanford.edu/news/measuring-what-medical-ai-can-actually-do>.

⁶ Hatch, *Agentic AI Outsourcing*, Introduction, “Executive Summary” and “Evidentiary Basis and Its Limits.”