

Measuring ROI in Agentic Outsourcing

Functional Baselines, Full Costs, and Recovered Value

Phil Hatch | Akholi

October 23, 2026

Measure the outsourced function before and after transition on a comparable price-and-performance basis. Establish full buyer cost and captured benefits, account for the investment, and repair the delivery constraints that consume value. A cheaper agent task or released capacity alone does not establish ROI.

Require a Functional Baseline Before Accepting the ROI Claim

A defensible return on agentic outsourcing requires a functional price-and-performance baseline. The buyer must know what the business function cost and delivered before transition, then compare the redesigned service under equivalent conditions. Without that starting point, finance cannot distinguish incremental value from a changed scope, lower acceptance standard, or cost transferred elsewhere. The CFO should commission a baseline audit, reconstruct missing evidence, and require reconciled costs and benefits before approving further scale or savings commitments.

The comparison must survive a change in delivery method. Agentic AI can combine tasks, eliminate intermediate deliverables, move human work, and change handoffs. The price of an old process phase may therefore have no equivalent after transition. The stable unit is the accepted business outcome at its required quality, control, and timing. Faster execution contributes value only when the complete function delivers that outcome under acceptable conditions.

The study demonstrates why reported return needs examination. Seventeen of 30 Production engagements had positive return entries; 13 were recorded at zero. A zero could mean no return, no return yet, or unavailable information. The field had no established common investment denominator, measurement period, or gross-versus-net convention, and no

financial institution record was available to verify it. These vendor-reported percentages cannot be treated as standardized, independently verified ROI or a causal estimate of agentic savings.¹

Establish a Comparable Baseline and Preserve Its Evidence

Define the function through what the buyer receives. An invoice service might produce valid invoices ready for authorized payment, with exceptions resolved or explicitly assigned. Extracted fields, agent actions, and invoices routed onward measure different work. The baseline should specify the completion rule, scope, service hours, quality, control obligations, and timing, with separate performance gates. Lower cost should not qualify as improvement when the service no longer meets those conditions.

Use a representative pre-transition period and record demand, case complexity, rejected and reopened work, seasonality, and exclusions. Normalize legacy cost to the comparison workload using documented drivers and assumptions. Identify scope, wage, price, and policy changes separately. A group comparing services in Switzerland, the United Kingdom, the European Union, and the United States should retain the relevant local performance and cost conditions and disclose its currency-conversion basis. Exchange-rate movements or a shift to easier cases should not be attributed to AI.

The FinOps Foundation distinguishes resource-efficiency measures from business-unit metrics and emphasizes documented definitions and sources. Cost per token can inform technical efficiency; cost per accepted outcome informs the functional decision.² The contract should use compatible performance definitions. Mayer Brown's agentic-services analysis proposes operational outcome measures rather than platform availability alone. This makes the promised service assessable, although an outcome service level does not itself establish a financial return.³

When the historical baseline is missing, recover contemporaneous records before substituting estimates. Reconcile invoices, general-ledger entries, retained buyer staffing, technology charges, accepted volumes, service times, quality, incidents, and correction effort to the same functional boundary. Staff recollection may help locate omitted work but should not become a precise historical cost without support.

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), Introduction, "Measured Return and Indicators Distressed"; Part 1, "Measured Return in Production," <https://doi.org/10.2139/ssrn.7164559>.

² FinOps Foundation, "Unit Economics," FinOps Framework, accessed October 23, 2026, "Define Unit Metrics Which Support Organizational Goals" and "Functional Activities," <https://www.finops.org/framework/capabilities/unit-economics/>.

³ Rohith P. George et al., "Contracting for Agentic AI Solutions: Shifting the Model from SaaS to Services," Mayer Brown, February 17, 2026, sec. 3, <https://www.mayerbrown.com/en/insights/publications/2026/02/contracting-for-agentic-ai-solutions-shifting-the-model-from-saas-to-services>.

Identify each reconstructed component as observed, allocated, estimated, or unknown, and show how plausible alternatives affect the decision. A matched service, controlled pilot, or limited parallel run may inform an estimate if differences are explicit. It is not an observed pre-transition counterfactual. Material unresolved gaps require a qualified range or an incomplete investment case, rather than definitive measured ROI.

The controller should approve the baseline, allocation rules, and version history, with the buyer and provider agreeing the outcome definition. Preserve that approved starting point. Legitimate scope revisions require dated approval and reconciliation; weak results do not justify rewriting earlier assumptions. A study account describes a vendor ROI tracker changing its baseline savings assumption to match results. The audit should test the buyer's records for that failure mode without treating the account as a prevalence estimate.⁴

Measure the Buyer's Full Cost Without Counting It Twice

The ledger should reflect the buyer's investment perspective: provider charges, buyer-retained work, and separately purchased services. Following activities through both firms exposes hidden effort and dependencies, but provider wages, model consumption, and infrastructure already recovered through its invoice must not be added again. A separate joint-economics analysis may examine provider internal cost; it should not be mixed into the buyer's ROI calculation.

Within that rule, capture review, corrections, reruns, exceptions, decisions, data preparation, model and tool consumption, orchestration, integrations, access, monitoring, training, controls, and fallback. Include transition, parallel running, termination or migration charges, and commitments that remain payable after capacity is released. Distinguish one-time incremental investment from ongoing operating cost and reconcile every item to a contract, invoice, usage record, or defensible allocation.

EY's September 2026 analysis identifies costs beyond tokens, including subscriptions, infrastructure, governance, organizational change, and failure recovery. Use those categories to guide reconciliation to contracts, invoices, usage, and retained effort. Upstream supplier economics may inform future price exposure; they should not become a second addition to costs already recovered through the buyer's invoice.⁵

Complementary investment also needs explicit treatment. Brynjolfsson, Rock, and Syverson's MIT research brief explains how new processes, organizational learning, and human capital can consume resources before a technology produces measured gains. Its productivity model

⁴ Hatch, *Agentic AI Outsourcing*, Part 1, "Measured Return in Production," engagement-manager ROI-tracker account; Part 2, "Changes to the Business Case and Return Definition."

⁵ Shawn Smith et al., "Agentic AI ROI: Can AI Pay for Itself?," EY, September 21, 2026, "The AI Token Is Just the Entry Fee" and "What Costs Belong in an Enterprise AI Business Case?," https://www.ey.com/en_us/insights/ai/agentic-ai-return-on-investment.

does not predict this engagement's return or recovery date. The implication for finance is to include those resources in the investment case and test their expected contribution, rather than omit them or assume eventual payback.⁶

Separate Operating Improvement, Captured Benefit, and Investment Return

Functional operating unit cost divides operating cost by accepted outcomes on the agreed basis. Released capacity measures resources no longer needed for the same workload. If staffing, licenses, minimum commitments, and invoices remain unchanged, lower resource use may create useful capacity without reducing expenditure. Keep it as capacity until there is evidence of a financial conversion.

Brynjolfsson and colleagues' published study of 5,172 customer-support workers illustrates the distinction. AI assistance increased issues resolved per hour by 15 percent on average in one software company and its subcontractors, with people responsible for conversations. This is evidence of an operational gain in that setting, not autonomous-outsourcing ROI or a 15 percent cash saving for another buyer.⁷

Captured benefit requires its own evidence. Cash savings need reduced outlay; avoided spending needs a documented demand or hiring requirement that no longer needs funding. Redeployment requires an identified destination and productive use. Incremental revenue should enter the case as attributable contribution after delivery costs. FinOps distinguishes avoidance, savings, and cost reduction; the same hour must not be counted as reduced spending, avoided hiring, and additional output.⁸

Finance should select the horizon, perspective, and convention before calculating ROI. For the simple cash-based return used here, cumulative net benefit equals comparable legacy operating cost that would otherwise have been incurred, plus verified incremental contribution, less new operating cost and incremental investment. ROI divides that net benefit by the same incremental investment. State the tax, timing, residual-value, and financing assumptions. An undefined or zero investment base does not support this ratio; use the cash-flow and value evidence directly.

⁶ Erik Brynjolfsson et al., *The Productivity J-Curve: How Intangibles Complement General Purpose Technologies*, MIT Initiative on the Digital Economy Research Brief 2019, vol. 2, 1–2, https://ide.mit.edu/sites/default/files/publications/2019-04JCurvebrief.final2_.pdf.

⁷ Erik Brynjolfsson et al., "Generative AI at Work," *The Quarterly Journal of Economics* 140, no. 2 (2025): 889–942, secs. I and III.B, <https://doi.org/10.1093/qje/qjae044>.

⁸ FinOps Foundation, "Unit Economics."

A Lower Functional Unit Cost Can Still Produce a Negative Return

Figure 1 uses a hypothetical two-year comparison, not study observations. Each year contains 100,000 matched accepted outcomes at unchanged scope, mix, quality, control, and timing. Legacy operating cost is \$1 million annually; the new service costs \$850,000 annually, including human work, tools, controls, and fallback. Assume the \$150,000 annual reduction is realized cash saving. A separate \$400,000 investment paid at the outset includes all one-time transition and parallel-running costs; there is no additional contribution or residual value.

Figure 1. Hypothetical functional savings and two-year investment ROI

Illustrative measure	Legacy function	Agentic function
Accepted outcomes each year	100,000	100,000
Annual operating cost	\$1,000,000	\$850,000
Operating cost per accepted outcome	\$10.00	\$8.50
One-time incremental investment	Not applicable	\$400,000
Two-year operating cash saving	Not applicable	\$300,000
Two-year net benefit after investment	Not applicable	–\$100,000
Two-year simple investment ROI	Not applicable	–25%
Payback within two years	Not applicable	Not achieved

Hypothetical cash-based example, not study data. Scope, mix, quality, controls and timing held equal. \$400,000 paid initially; \$150,000 savings at each year-end. No extra contribution, residual value, tax or discounting. No cost counted twice.

Source: Author's illustrative assumptions and arithmetic, not study observations or a forecast. Cash-based, before tax, undiscounted; investment paid at the outset and annual savings received at year-end.

Operating cost per accepted outcome falls from \$10 to \$8.50, a 15 percent improvement. Two years yield \$300,000 of operating savings against the \$400,000 investment. Net benefit is negative \$100,000 and simple investment ROI is negative 25 percent. The cash flows are negative \$400,000 initially and positive \$150,000 at each year-end, so payback is not achieved within the horizon. Reporting the unit-cost improvement as ROI would conceal the negative investment result.

For a multiyear decision, use the buyer's justified hurdle rate to show discounted cash flows and net present value alongside payback and the chosen ROI convention. Any positive discount rate makes this illustration's already negative net present value more negative. Test delayed ramp-up, weaker demand, greater review effort, price changes, and lower captured

benefits. Forecasts should remain separate from realized return; future benefits cannot be reported as already earned.

Trace Each Value Constraint to Its Operating or Commercial Cause

Once the comparison and ledger are credible, the audit can establish where expected value is lost. Trace elapsed time from valid demand to accepted outcome, separating agent execution from decisions, approvals, review, and exceptions. A delay becomes a financial loss through an evidenced monetary consequence, such as paid rework, overtime, contractual service cost, or foregone contribution. Unmonetized waiting and constrained capacity remain operating evidence, not a financial amount. The study's response-speed and joint-decision accounts identify mechanisms to investigate, not measured losses.⁹ Figure 2 connects those mechanisms to buyer value capture.

Figure 2. Find the constraint between agent activity and captured buyer value



Diagnostic hypotheses, not measured losses. Necessary human work and controls are not automatically waste.

Source: Author's proposed diagnostic based on the study's mechanisms and cited economics guidance. These are hypotheses, not measured loss amounts or causal attribution.

Start with the constrained part of the service. If fast agent execution feeds a growing review queue, additional agent capacity may worsen the backlog. Establish whether the constraint is decision authority, qualified capacity, source quality, or an unsuitable process or interface. McKinsey's July 2026 economics analysis connects value with correctly completed work,

⁹ Hatch, *Agentic AI Outsourcing*, Part 2, "Speed of Financial Institution Response to Issues Raised by the Engagement" and "Joint Decision Speed Mismatched to Agent Speed."

human supervision or repair, and full operating cost. That perspective directs measurement to the service rather than its cheapest model call.¹⁰

The cost evidence should then distinguish necessary protection from avoidable work. Unusable feeds, access delays, repeated context, uncontrolled retries, unnecessary calls, duplicate checks, and poor exception routing require different remedies. A legally or operationally necessary review remains part of delivery cost even when removing it would improve a nominal efficiency figure. The audit should establish the purpose and financial consequence of the work before classifying it as waste.

Reconcile manual intervention with performance reports and the ledger. Study accounts describe efficiency calculations omitting verification hours and manual completion after an interface failure. They warrant looking for missing effort, not applying a universal overstatement factor. Commercial records must then show whether improved capacity changes the buyer's expenditure. Fixed fees, minimum volumes, unused licenses, duplicated commitments, or provider-retained savings may prevent capture; those are terms to examine, not evidence that a pricing model is inherently improper.¹¹

Assign the Repair, Test the Change, and Remeasure the Benefit

During days 0–30, the CFO and operational sponsor should secure records, approve or reconstruct the baseline, reconcile material claims, and identify unsupported savings commitments with provider delivery and commercial counterparts. The resulting diagnosis should assign each constraint to the party able to change it. Mandatory legal, safety, and material-control corrections require action independently of discretionary ROI improvement. Contain unacceptable exposure immediately; the proposed timetable is neither a regulatory deadline nor a promise of full closure.

The repair should address the diagnosed cause. Buyer process owners should correct decision authority and response obligations; data and technology owners should repair source and interface defects; provider delivery owners should correct execution, review, and recording. Finance should reconcile the ledger while commercial owners establish how any gain reaches the buyer under the contract. Rank discretionary changes by evidenced loss, consequence, feasibility, dependencies, and repair cost, then select a bounded test that can establish the expected effect.

¹⁰ Lari Hämäläinen et al., “Is That AI Agent Worth It? Agentic Economics and the Modern Operating Model,” McKinsey & Company, July 13, 2026, opening discussion and “Why the Operating Expenditure Bill Keeps Rising: The Six Drivers of Agentic Economics,” <https://www.mckinsey.com/capabilities/quantumblack/our-insights/is-that-ai-agent-worth-it-agentic-economics-and-the-modern-operating-model>.

¹¹ Hatch, *Agentic AI Outsourcing*, Part 1, “Measured Return in Production,” accounts of verification hours and manual completion after an API barrier; Part 5, “Visibility in Reported Performance.”

During days 31–90, test the prioritized changes under the conditions that produced the constraint. Finance should compare accepted outcomes, case mix, quality, timing, full cost, and captured benefits with the governed baseline, preserving the original and explaining adjustments. A completed repair establishes an activity, not a realized financial benefit. Expansion should follow only when operating records and financial evidence support the result, with remaining estimates and any burden shifted to the other firm disclosed.

Approve the Functioning Service and Its Reconciled Investment Case

The investment committee should receive baseline confidence, accepted functional performance, full cost, captured benefit, remaining investment, material obligations, downside exposure, accountable owners, and the next review date. Expansion requires a service and investment case that withstand verification. Redesign is appropriate where an operating constraint can be corrected; renegotiation where terms prevent value capture; restriction or exit where material risk or economics cannot be justified. Continuity obligations remain part of any transition.

Where the baseline is missing, reconstruct it and qualify the claim. Where value is being consumed, repair the evidenced cause and verify the recovery. The board's decision should concern the complete buyer–provider service and the investment required to sustain it. An isolated task-efficiency percentage cannot answer that question.

Evidence and Source Note

The banking study is a nonrepresentative, single cross-section of 850 vendor delivery-team respondents on 150 engagements during January–June 2026, without bank or vendor corroboration. Production return entries are not standardized financial ROI; mechanisms may inform other sectors without transferring prevalence. The author-supplied version 1.8.2, dated November 1, 2026, was consulted as forthcoming material. Public sources were checked through October 23, 2026. The measurement convention, worked example, and repair framework are proposed applications, not validated intervention results.¹²

¹² Hatch, *Agentic AI Outsourcing*, Introduction, “Evidentiary Basis and Its Limits” and “Measured Return and Indicators Distressed.”

Bibliography

- Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. "Generative AI at Work." *The Quarterly Journal of Economics* 140, no. 2 (2025): 889–942. <https://doi.org/10.1093/qje/qjae044>.
- Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. *The Productivity J-Curve: How Intangibles Complement General Purpose Technologies*. MIT Initiative on the Digital Economy Research Brief 2019, vol. 2. 2019. https://ide.mit.edu/sites/default/files/publications/2019-04JCurvebrief.final2_.pdf.
- FinOps Foundation. "Unit Economics." FinOps Framework. Accessed October 23, 2026. <https://www.finops.org/framework/capabilities/unit-economics/>.
- George, Rohith P., Joe Pennell, Brad L. Peterson, and Oliver Yaros. "Contracting for Agentic AI Solutions: Shifting the Model from SaaS to Services." Mayer Brown. February 17, 2026. <https://www.mayerbrown.com/en/insights/publications/2026/02/contracting-for-agentic-ai-solutions-shifting-the-model-from-saas-to-services>.
- Hämäläinen, Lari, Mark Patel, Sven Blumberg, Tanguy Catlin, and Wasim Lala. "Is That AI Agent Worth It? Agentic Economics and the Modern Operating Model." McKinsey & Company. July 13, 2026. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/is-that-ai-agent-worth-it-agentic-economics-and-the-modern-operating-model>.
- 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>.
- Smith, Shawn, Steve Wanner, Hyong Kim, et al. "Agentic AI ROI: Can AI Pay for Itself?" EY. September 21, 2026. https://www.ey.com/en_us/insights/ai/agentic-ai-return-on-investment.