Akholi

Outsourcing 2026: Agentic AI and Tier 1 Financial Institutions

The Delivery-Side Record from 150 Engagements

July 2026

First Edition

Phil Hatch
p.hatch@akholi.com

Outsourcing 2026: Agentic AI and Tier 1 Financial Institutions

The Delivery-Side Record from 150 Engagements

Phillip J. Hatch

Document Information

Publisher:	Akholi	Study period:	Q1-Q2 2026
Edition:	First Edition (1.0.1)	ISBN:	
Published:	July 2026	DOI:	10.6084/m9.figshare.33005630
Author:	Phil Hatch	ORCID:	0000-0003-1709-8860
LinkedIn:	linkedin.com/in/phillipatch/		

Copyright and Rights

© 2026 Akholi. All rights reserved.

This report may be quoted and cited with attribution to the author and Akholi. Reproduction, redistribution, or adaptation of the whole or of substantial portions requires written permission. Requests go to p.hatch@akholi.com.

How to Cite This Report

Hatch, Phillip J. 2026. Outsourcing 2026: Agentic AI and Tier 1 Financial Institutions: The Delivery-Side Record from 150 Engagements. Akholi.

Contact

Media and general inquiries: info@akholi.com
Permissions, citation, and author correspondence: p.hatch@akholi.com
Web: akholi.com

Table of Contents

Table of Figures

Definitions

This report holds a fixed vocabulary.

1. The study set. The 150 agentic AI outsourcing engagements at Tier 1 financial institutions that this report draws on.
2. Active and canceled engagements. 94 engagements were still running on the study date, and 56 had been canceled.
3. Lifecycle stage. Every engagement sits in planning, proof of concept, or production. 30 engagements had reached production, and the remaining 120 were divided between planning and proof-of-concept.
4. The 10-point scale. Respondents rated 100 engagement-condition questions on a scale from 1 to 10, where 1 is the worst condition a respondent can report, and 10 is the best.
5. Distressed. Any condition rated 3 or below.
6. Overall condition. An engagement's mean rating across all 100 conditions. Ranked by this measure, the study set divides into a weakest, middle, and strongest third, each comprising 50 engagements.
7. Cancellation likelihood. The delivery team's own estimate of the likelihood that the client will cancel an active engagement is rated 1 to 10 and presented as a percentage.
8. High risk. An active engagement with a team-assessed cancellation likelihood of 70% or higher. 19 engagements qualify. Together, these 19 and the 56 canceled engagements account for half of the study set.
9. Moderate risk. An active engagement with a cancellation likelihood of 50% to just below 70%. Adding this band to the high-risk and canceled engagements lifts the canceled-or-at-risk share to three-quarters of the study set.
10. Customer satisfaction. The instrument's satisfaction metric, with a study mean of 43.85%.
11. Return. A measured return on engagement, reported for 17 of the 30 production engagements, averaging 24.71%.

Executive Summary

Between April and June 2026, 850 delivery staff at major outsourcing firms gave this study structured ratings and written testimony. Respondents work on 150 agentic AI outsourcing engagements, serving Tier 1 financial institutions. Their engagements span 10 types of financial institutions and 4 primary outsourcing models. Respondents build the agents. They operate the processes and assemble the client reporting. Their account is vendor-side testimony and uncorroborated by client records.

The study surfaced 4 primary findings:

1. The industry is new. Every participant is inventing the practice in live engagements. No respondents reported more than 1 year of experience with agentic AI.
2. Failure is material now. Cancellation is 37.33% for a study set with a median engagement length of 5 months. Client preparation was rated the weakest among all measured factors. Vendor capability rated second weakest. Co-management of the engagement becomes the primary success factor once in production.
3. The people who assemble client-facing reporting rated it as more likely to be manipulated than accurate. Respondents put the likelihood of manipulation at 6.26 of 10.
4. Oversight of the agents rests on a junior and dissatisfied workforce that is likely to leave. Respondents rated their job satisfaction at 4.19 out of 10. Most indicated they intend to leave their current job. What holds these engagements together is undocumented knowledge, and departures remove it.

The record supports 6 executive actions:

1. Gate every agentic contract behind an institutional readiness standard.
2. Build co-management discipline before contract signature, and improve it throughout the engagement.
3. Audit the design and execution of vendor reporting rather than its arithmetic.
4. Restructure commercial terms around outcomes and exit.
5. Treat the human oversight layer as a control function.
6. Listen directly to the vendor's delivery team.

Every action is within the institution's own authority. Engagements that met these conditions were successful. Those that did not meet them failed at higher rates and had lower customer satisfaction.

Part 1: The Market

Adoption is broad, young, and already carrying material losses. The 150 engagements studied span all primary institution types and outsourcing models. Their outcomes differ based on the conditions the buying institution controls, not on the type of financial institution or the outsourcing model.

1.1 Adoption

Agentic AI outsourcing adoption is industry-wide. 10 primary financial institution types appear, from asset management with 24 engagements to private equity with 9. All 4 primary outsourcing models were also represented, with customer experience management (CXM) the largest at 28.00% of the study set. No institution type or outsourcing model dominates the study set by design.

1.2 Maturity

The engagements are young.

1. The mean engagement age is 5.69 months, and the median is 5 months.
2. 68.67% of the study set is 6 months old or newer.
3. Only 10 engagements have passed the 1-year mark.
4. The longest engagement in the study is 22 months.

Every figure in this report is an early reading of a young market.

Conversion to live operation remains limited. Only 30 engagements have reached production, and those 30 run a total of 181 agents. The remaining 120 engagements divide evenly between the planning and proof-of-concept stages.

Production work consistently concentrates on high-volume processes whose outputs can be checked against a defined standard.

1.3 Outcomes and the Forward Pipeline

Lifecycle stage	n	Cancel rate	At high risk	ROI n	Mean ROI
Planning	60	48.33%	5	-	-
Proof-of-concept	60	28.33%	10	-	-
Production	30	33.33%	4	17	24.71%
Study set	150	37.33%	19	17	24.71%

Table 1: Study Engagement Breakdown by Phase

Cancellation has claimed 56 of the 150 engagements, a rate of 37.33%. Attrition falls unevenly across the lifecycle.

1. Planning-stage engagements cancel at 48.33%.
2. Proof-of-concept engagements cancel at 28.33%.
3. Production engagements cancel at 33.33%.

The heaviest losses come before the first milestone. At that stage, the concept meets the client's real data and systems for the first time.

Among the 30 engagements that reached production, 17 measured returns. Returns average 24.71%, with a range of 6.00% to 47.00%. All 17 ROI reporting engagements remained active at the time of the study.

Customer satisfaction is split to the extremes rather than clustering around the mean (43.85%). On the 100-point scale:

1. 56 engagements score below 40.
2. 59 score at 60 or above.
3. Only 35 fall within the middle band.

1.4 Performance Across Institution Types

Institution type	n	Cancel rate	Satisfaction, active engagements	Reporting a return
Asset management	24	54.17%	69.45%	4
Investment banking	19	47.37%	56.40%	2
Credit card	12	41.67%	61.57%	1
Insurance	14	35.71%	66.11%	0
Wealth management and private banking	23	34.78%	64.00%	2
Retail and consumer banking	12	33.33%	70.38%	4
Commercial and business banking	13	30.77%	63.89%	0
Markets and trading	11	27.27%	65.38%	0
Payments and merchant acquiring	13	23.08%	63.00%	2
Private equity and private capital	9	22.22%	60.00%	2

Table 2: Performance by Institution Type

Cancellation rates vary widely across the 10 institution types, from 22.22% at private equity firms to 54.17% at asset management firms. None of that variation survives testing at these sample sizes. With 9 to 24 engagements per type, differences of this size arise by chance. Satisfaction behaves similarly.

Problem composition holds nearly constant across institution types. Client preparation issues account for 38.00% to 45.00% of the prioritized problems across all financial institution types. The failure pattern is associated with the work itself rather than with the type of institution buying it.

1.5 Corporate-Functional Versus Line-of-Business Operations

Corporate Functional Vs. Line of Business Operational	n	Cancel rate	Customer satisfaction, active	Reporting a return
Corporate functional	74	35.14%	64.35%	6
Line of business operational	76	39.47%	63.83%	11

Table 3: Performance by Operational Scope

Operational scope shows a timing difference with no material performance difference. Institutions first placed line-of-business operations into agentic outsourcing. A corporate-functional wave followed roughly 2 months behind and has had less time to mature.

The 2 waves are nearly equal in size. The study collected 74 corporate-functional engagements and 76 line-of-business engagements. Line-of-business work averages 6.63 months of age, compared with 4.73 for corporate-functional work. Maturity alone separates them.

Cancellation rates are 35.14% for corporate-functional work and 39.47% for line-of-business work. That difference is within sampling noise. Customer satisfaction among active engagements is flat at 64.35% and 63.83% between the 2.

The ROI consequences of the timing gap are mechanical. Line-of-business work holds 21 of the 30 production engagements and 11 of the 17 reporting returns. Corporate-functional work holds 9 and 6. The data indicate the split is a queue position. Corporate-functional engagements have had fewer months to reach production.

Conditions that determine the outcome do not differ between the 2. Mean scores across all 5 problem areas differ by 0.34 points between corporate functional and line-of-business operational outsourcing engagements.

1.6 Outcomes by Outsourcing Model

Outsourcing model	n	Cancel rate	Customer satisfaction, active	Reporting a return
Customer experience management (CXM)	42	35.71%	64.44%	5
Information technology outsourcing (ITO)	40	37.50%	64.36%	5
Business process outsourcing (BPO)	37	37.84%	66.30%	6
Knowledge process outsourcing (KPO)	31	38.71%	60.58%	1

Table 4: Performance by Outsourcing Model

Outsourcing model selection is not a risk lever. The 4 models differ by 3.00 points on cancellation and by 5.73 on customer satisfaction in active engagements. Pricing model and deal size behave the same way. Neither the 5 pricing forms nor the annual contract value bands separate outcomes.

1.7 Commercial Term Length

Contract term	n	Cancel rate	Customer satisfaction, active
12 months	29	24.14%	63.59%
24 months	48	56.25%	63.71%
36 months	48	35.42%	65.10%
48 months	13	23.08%	60.10%
60 months	12	16.67%	66.90%

Table 5: Performance by Contract Term Length

Engagements on 24-month terms cancel at 56.25%, compared with 28.00% for all other terms pooled. The difference survives correction for multiple comparisons, at a corrected p of 0.010. The available checks do not explain it away. 24-month engagements are younger than average. Their composition resembles the rest of the study set. Satisfaction among survivors is flat, ranging from 60.10% to 66.90% across the 5 terms.

The correlation between the age of the engagement and the cancellation is not in the data. One reading is that a 2-year term is long enough to judge an engagement and short enough to abandon it. 1-year deals are treated as disposable pilots, and 4- and 5-year deals incur switching costs that keep them in place.

1.8 Engagement Headcount

Vendor headcount across the study set peaked at 1,502 and is now 1,253. Planned headcount a year out is 856. Respondents forecast a 31.68% decline for the full study set. Canceled engagements are winding down, driving almost all of it. The 94 active engagements plan growth of 11.90%, from 765 today to 856. Within the study set, a shrinking team signals an engagement approaching cancellation. 2 boundaries apply.

1. Respondents staff the agentic delivery itself. The record says nothing about employment inside the client institutions or the outsourcing vendor itself.
2. Canceled engagements still reported teams largely in place or in early transition.

1.9 Compliance, Oversight, Regulatory, and Legal (CORAL)

The study identified material CORAL concerns. Notably:

1. Individual Contributors frequently lacked any awareness of CORAL topics.

2. Respondents regularly noted significant human effort needed to address preparation and agentic AI challenges. Decisions are made, and a junior team takes action without a basic understanding of CORAL. Most of these human efforts are carried out without the bank's involvement in decision-making or basic awareness.

While the risk is material, the data indicated that CORAL concerns would not have materially altered the performance or final disposition of any engagement included in the study. However, given respondents' lack of awareness of this topic and the young mean age of agentic AI outsourcing, it is highly likely that this topic is a greater concern than the data indicate.

This report will not further highlight the CORAL issues found in the study.

Part 2: Why Agentic AI Outsourcing Engagements Fail

Problem area	Ratings captured	Mean score (1-10)	Distressed
Bank-side (client) preparation	15,966	3.52	51.48%
Vendor-side preparation and management	15,943	3.66	48.22%
Joint bank-vendor co-management	15,968	3.65	49.26%
AI tools	15,911	4.58	32.14%
3rd-party ecosystem (tools, data, and services)	15,947	4.62	31.74%

Table 6: 5 Failure Categories

Failure in the engagements studied is concentrated in 5 problem areas. The weight of these areas shifts as an engagement matures. Client and vendor problems dominate as the most problematic in the early stages. Joint co-management leads as engagements reach production. No participant in the study reported an established best practice or proven playbook for client preparation, vendor delivery, or joint co-management.

The technology itself is immature and still ranks behind institutional failures across every engagement phase. The 3rd-party ecosystem of tools, data, and services ranks last in this study.

The 5 areas move together statistically. An engagement weak in one tends to be weak in the others. The data cannot cleanly assign an outcome to any single area. Outcomes are attributed to the overall condition of the engagement. The concentration of the evidence ranks the 5 areas.

2.1 Bank-Side Preparation

Client-readiness group	Group mean rating (1-10)	Rated distressed
Governance and responsiveness	3.51	51.83%
Process readiness	3.51	52.15%
People and skills	3.52	51.36%
Legacy estate	3.52	52.06%
Data readiness	3.55	50.11%

Table 7: Client-Readiness Groups

Client institution conditions were rated the weakest of all the measures in this study. Across all 100 rated conditions, the 5 weakest question groups are those describing the client organization. More than half of all client-side ratings are categorized as distressed. 5 gaps carry the area, in rated order from weakest to strongest:

1. Governance and responsiveness rated 3.51. It measures governance, oversight, and the speed of the client's response. Governance and responsiveness challenges led to damaging issue selections on 87 of the 150 engagements, with 51.83% rated as distressed.

2. Process readiness rated 3.51. It measures whether the processes handed to agents are documented and mapped at the depth automation requires. Documentation gaps led to damaging issue selections on 94 of the 150 engagements, with 52.15% rated as distressed.

3. People and skills rated 3.52. The condition covers the availability of client staff who understand the processes, the data, and agentic AI. Respondents selected it in 58 engagements, with 51.36% rated as distressed.

4. Legacy estate rated 3.52. It measures the accessibility of core systems built decades before the advent of agentic integration. Respondents selected it in 96 engagements, with 52.06% of those engagements rated as distressed.

5. Data readiness rated 3.55. It covers the quality, normalization, and governance of the data agents run on. Respondents selected it in 93 engagements, with 50.11% rated as distressed.

These ratings cluster within a tenth of a point. Client preparation is a uniform condition across everything the engagement depends on. No single remediation addresses it.

Lowest-rated client institution conditions	Mean (1-10)	Rated distressed
Ownership and accountability for agentic AI	3.47	52.00%
Stability of the business case, KPIs, and ROI definition	3.48	52.90%
Agentic AI knowledge and skills of client staff	3.48	52.71%
Willingness to redesign processes for agentic AI	3.48	53.36%
Timely access to functional, legacy-system, and process experts	3.49	53.85%
Capacity to govern and oversee the risks of AI agents	3.49	52.32%
Process mapping and maturity at the depth agentic AI requires	3.49	54.09%
Data quality, normalization, and consistency across source systems	3.50	51.05%
Data accessibility	3.51	50.38%
Legacy infrastructure accessibility, stability, and compatibility	3.51	52.18%

Table 8: Lowest-Rated Client Institution Conditions

Qualified AI Ownership Inside the Client Institution

> *"Tell us who owns this stuff before we go live."*
>
> *Individual Contributor, BPO, Canceled Engagement*

> *"... no one in charge. The people they sent us don't know AI and can't get things done inside the bank."*
>
> *Individual Contributor, BPO, Canceled Engagement*

Clarity over who owns AI within the client institution was rated 3.47, the lowest rating in the study. The stability of the client's business case and ROI definition was rated 3.48, the second-lowest rating. Both differ in kind from the operational gaps above them. Each measures whether the client has decided who is responsible and to what end. An institution that has not settled ownership cannot direct or co-manage an engagement. During working engagements, a clear client-side owner was in place.

Agentic AI knowledge among client staff was rated 3.48, and the capacity to govern AI agent risks was rated 3.49. For canceled engagements, the ownership condition fell to 2.32 from 3.64 among engagements that reached production.

Process Maturity, Mapping, and Optimization for Agentic AI

"Never scope an agent onto a process nobody wrote down."

Engagement Manager, CXM, Canceled Engagement

"They had 50 different ways to do the same job. Every person we talked to told us to do it a different way."

Individual Contributor, BPO, Canceled Engagement

"Printing paper and waiting for the manager to read doesn't work with agents."

Individual Contributor, BPO, Canceled Engagement

"We spent months with the bank mapping processes.... No one in the bank consistently followed the plan. Every due diligence packet we got was different. AI was self-training based on these variations. We lost control of the quality of the output."

Engagement Manager, BPO, Canceled Engagement

Process mapping and documentation rated 3.49, with 54.09% of ratings distressed. Standardization sat 2 hundredths higher at 3.51. Consistency of process execution across business units followed at 3.56. Willingness to redesign processes for agentic AI was the weakest of the 4 at 3.48.

Canceled engagements rated process mapping at 2.38. Engagements that reached production rated it 3.78, a spread of 1.40 points.

Respondents in 47 engagements noted that processes were never re-engineered for agentic AI. An agent executes the process it was given, exactly as mapped, at machine speed.

Where the map is missing, the vendor's team reconstructs the process based on observations before any agent can run. Respondents frequently noted that the bank did not validate vendor-team-mapped or optimized processes. Changes were often made without the bank’s awareness.

Client Data Cleanliness, Consistency, and Accessibility

"Data-sharing limits slow agent access to source systems."

Individual Contributor, ITO, Active Engagement

"Threat intel feeds arrive in inconsistent formats. Our system fails each time."

Individual Contributor, ITO, Active Engagement

"Data quality gaps still force manual patching before the agent can run."

Individual Contributor, BPO, Active Engagement

"Their data was bad. Garbage in, garbage out."

Individual Contributor, CXM, Canceled Engagement

Data quality and AI-readiness led to damaging issue selections in 93 of the 150 engagements. Data governance and model-risk readiness followed at 87. Data-sharing strain and residency limits appeared in 45. Confidentiality challenges appeared in 38.

Normalization and format consistency across source systems rated 3.50, with 51.05% of ratings distressed. Timely accessibility was rated 3.51. The quality of the data itself was rated 3.59.

Engagements that reached production rated their data conditions between 3.62 and 3.84, while canceled engagements rated between 2.33 and 2.45.

As with bank-side process mapping, maturity, and agentic AI optimization, respondents noted that they are resolving data issues manually. Respondents often do so without the bank's input or awareness during the data-modification decision-making process.

Respondents Are Absorbing the Preparation Gap

> *"Our monthly report shows uptime and ticket counts without a field for how many book-building outputs the syndicate desk quietly recalculated by hand, hiding the real hand-holding this account still needs."*
>
> *Engagement Manager, KPO, Active Engagement*

> *"Our dashboard reported ticket resolution as agent-driven once the agent posts any response, even when a human had to redo the work behind the scenes."*
>
> *Engagement Manager, BPO, Canceled Engagement*

> *"I spend half my time fixing data issues that the bank should have fixed before they sent it to us."*
>
> *Individual Contributor, CXM, Active Engagement*

> *"Legacy systems are forcing manual workarounds every week. They need constant babysitting."*
>
> *Engagement Manager, CXM, Active Engagement*

92.00% of Individual Contributors reported manually resolving issues with client-provided materials. Respondents are resolving the client's preparation gap through 3 methods of unrecorded work:

1. Manual repair of data, legacy access, and other inputs before any agent operates.
2. Mapping or enhancing processes to the levels required by agentic AI.
3. Defining agentic AI output quality standards and resolving output errors.

A junior team without material agentic AI or financial industry experience is making decisions that directly affect output. Further, most of the vendor’s delivery team lacks a basic understanding of CORAL as it applies to the financial industry.

Respondents noted that the client is frequently neither involved nor aware. Those decisions are often undocumented and are passed down within the vendor team as informal knowledge.

Part of the workforce described this backstop as work it values. The testimony names 4 as points of professional pride.

1. Catching agent errors before the client sees the output.
2. Repairing input data before an agent runs.

3. Adjusting the client's process to fit agentic delivery, without the client's awareness.
4. Resolving agentic AI output before submission to the bank.

Every prior generation of outsourcing tolerated a lack of client preparation, and human delivery teams absorbed it invisibly. Agentic delivery was sold on removing that human layer. Within these 150 engagements, the human layer takes on a new role. It absorbs the client's lack of preparation at a speed and volume it was never sized for.

2.2 Vendor Challenges

Lowest-rated vendor conditions	Mean (1-10)	Rated distressed
Depth of knowledge of agentic AI	3.58	51.55%
Effectiveness of the employer's top leadership team	3.60	48.11%
Overall qualification to deliver elite agentic AI outsourcing	3.61	47.95%
Ability to manage agentic engagements for Tier 1 banks	3.63	48.91%
Experience running agentic AI in live production beyond pilots	3.64	48.24%
Depth of knowledge of the financial industry	3.64	48.88%
Transition from FTE-based to agentic commercial models	3.65	48.57%
Playbook from concept to governed production	3.65	48.57%
Quality and investment in agentic AI internal training	3.66	48.56%
Internal management redesigned for agentic delivery	3.66	48.56%

Table 9: Lowest-Rated Vendor Conditions

Issues with the vendor's own capabilities led to 1,262 damaging-issue selections among all 850 respondents and affected 148 of the 150 engagements. Only client unreadiness drew more, at 1,632 selections. Most of the identified issues are directly related to the vendor's lack of experience with agentic AI.

Entirely New to the Vendor

The agentic AI outsourcing industry is less than 2 years old and staffed primarily by people with less than 1 year of experience. Most of the challenges respondents noted, specific to vendor delivery capabilities, are rooted within this context. Outsourcing firms are selling agentic delivery to Tier 1 financial institutions with little to no additional agentic AI experience or understanding than the banks themselves.

Vendor agentic AI experience separates engagement outcomes. Engagements with vendors that had delivered agentic AI services for 2 years were canceled at a rate of 19.23%. Engagements with vendors below that mark were canceled at a rate of over 40%, a difference significant at $p = 0.04$.

Recorded customer satisfaction moves similarly. Outsourcing vendors with 2 years of agentic experience achieved an average customer satisfaction score of 50.77%. Vendors with 1 or fewer years of experience generated below 43%.

Defining the New Commercial Model Under Load

The FTE-based billing model is leaving this industry faster than its replacement is arriving. Agentic delivery removes the person from the unit of work. Billing must therefore move from seats supplied to output delivered. The move demands a commercial offering with more structure and more precision than seat pricing ever required. Vendors in this study have not completed the transformation of their core business model and are struggling to do so under load.

Senior respondents, the 183 at Engagement Manager rank and above, rated their employer's remaking of the commercial model for agentic delivery at 4.04. Respondents rated their employer’s ability to price agentic work profitably when billable hours no longer measure the work at 4.02.

Respondents noted that the change in the commercial model has created financial pressure within the outsourcing vendor. Every study indicator within the commercial model segment ranked between 37.70% and 40.56% as distressed.

Weak pricing models directly signal failure and suggest the vendor may be in the early stages of understanding the fundamental changes AI is having on the industry. Pricing capability correlates with cancellation ($r = -0.52$) and recorded customer satisfaction ($r = 0.62$); $p < 0.001$ for both.

Leadership and Management Without Agentic Experience

"Im a fresher and know more about AI than anyone in our executive team."

Individual Contributor, KPO, Active Engagement

"My management chain struggles to understand basic AI realities. I bounce between meetings with executives that either think AI is overpromised or vastly under-respected."

Engagement Manager, ITO, Active Engagement

"...they tell us to do stuff that doesn't make sense. I don't think managers know enough to make decisions."

Individual Contributor, CXM, Active Engagement

"Executives talk like our marketing materials. In sidebar discussions, I question if they understand AI realities at all. ...nervous about the company direction..."

Portfolio Director, BPO, Active Engagement

Vendor leadership teams do not yet hold the experience their firms are selling.

Senior respondents rated the overall effectiveness of their employer's top leadership team at 4.09, compared with 3.60 across all 850 respondents in Table 9. 39.89% of the senior ratings on that condition fall in the distressed band. Leadership effectiveness tracks outcomes at the engagement level, with correlations of $r = -0.51$ with cancellation and $r = 0.61$ with recorded customer satisfaction ($p < 0.001$ for both). Engagements where the senior leadership reporting was 3 or below were canceled in 64.29% (36 of 56), with a recorded satisfaction of 25.75%. Engagements where it sits at 6 or higher were canceled at 8.57% (3 of 35), with a satisfaction rate of 67.60%.

Leadership's understanding of agentic AI further widens the gap. Engagements where the leadership's knowledge of agentic AI scored 3 or below canceled at 89.66% (26 of the 29) and recorded a customer satisfaction rate of 13.66%.

In contrast, engagements with a leadership agentic AI rating of 6 or higher recorded no cancellations across all 41 engagements and achieved a mean customer satisfaction rate of 71.68%.

The pattern extends to mid-tier management as well. All 7 direct-manager measures separate canceled from active engagements more sharply than any of the 20 employer-capability conditions. The 5 manager competence measures correlate with cancellation between $r = -0.62$ and -0.67 and with recorded

satisfaction from r = 0.74 to 0.78, p < 0.001 throughout. The record does not settle the direction of cause, and failing engagements may sour their teams on the manager.

The managers running these engagements reported their direct lack of experience with agentic AI.

1. 53.33% of the study's 150 Engagement Managers reported 0 agentic AI experience, with a mean of 0.47 years. No Engagement Manager reported more than 1 year of agentic AI experience.
2. 64.52% of the 31 Portfolio Directors reported 0 experience with agentic AI, with a mean of 0.33 years.

Manager inexperience carries a measurable price. Engagements whose Engagement Manager reported 0 agentic years canceled at 45.00% (36 of 80 engagements), with a recorded satisfaction of 38.44%. Engagements whose Engagement Manager reported the single year canceled at 28.57% on satisfaction of 50.04% (20 of 70 engagements), a difference significant at p = 0.04.

The Missing Vendor-Side Playbook

Vendors lack a mature agentic AI playbook to govern internal vendor operations and the full customer delivery lifecycle.

The internal rulebook gap is material. Respondents rated the extent to which internal management processes have been redesigned for agentic delivery at 3.66, with 48.56% distressed. That reading describes management running an agentic business through a pre-agentic internal process. No firm-level rules exist to inherit.

More concerning, there is a significant gap in a playbook that walks any customer through the entire outsourcing lifecycle. Respondents rated their employer's ability to manage agentic engagements end-to-end at 3.69.

Respondents rated the maturity of their employer's playbook for taking agentic use cases from concept to governed production at 3.65, with 48.57% of ratings in the distressed range. On canceled engagements, the reading fell to 2.40. Engagements in production stood at 3.94. The absence of a mature playbook led to damaging issue selections in 111 of the 150 engagements. Playbook maturity correlates with cancellation at r = -0.59 and with recorded customer satisfaction at r = 0.71 (p < 0.001 for both).

The data indicate that vendors are not maturing their customer-facing playbooks fast enough. Among the 94 active engagements, playbook maturity is flat against engagement age. Engagements up to 3 months old rated it 4.47. Engagements at 4 to 6 months rated 4.34. Engagements older than 7 months rated 4.45.

Findings are problematic. Respondents noted that without an operationally mature client-facing playbook, they invent the rules themselves. Continuing a theme found throughout this study, junior

respondents are making delivery decisions without formal guidance from their employer, without direct customer involvement in decision-making, and often outside the visibility of either party's management team.

The Vendor's Own Platforms Predate Agents

> *"We sold an AI solution that uses a hosted econometric modeling platform we built many years ago... failing with every agent interaction."*
>
> *Engagement Manager, KPO, Active Engagement*

> *"The sales team pitched our dev kit we built as being optimized for AI. No one in sales checked if it was. We panicked for a couple of weeks as we validated AI compatibility after contracts were signed."*
>
> *Portfolio Director, ITO, Active Engagement*

> *"Our quality control system for customer support tickets was built for batch processing at the end of the day. No one thought about how agents would interact with it before we went live. We fixed it but we had to keep our team in the office over the weekend."*
>
> *Individual Contributor, CXM, Active Engagement*

> *"We are the only team that is using agents with our SF plugin. We can't change the plugin now, or every other customer may have problems. We are fighting about how to fix this."*
>
> *Portfolio Director, BPO, Active Engagement*

Over the last decade, outsourcing vendors have expanded hybrid outsourcing, or the bundling of cloud, SaaS, and others with the human delivery layer. Respondents described these platforms as problematic within an agentic AI context.

Respondents rated how well their employer's bundled legacy platforms are optimized for agentic delivery at 3.70, with 48.36% of ratings in the distressed range. On canceled engagements, the reading fell to 2.52. Respondents selected vendor-built platforms that are not agent-ready as the most damaging issue on 26 engagements.

No Agentic Talent Exists for the Vendor to Hire

> *"Our team is full of freshers. We need senior workers with some agentic AI experience. We've had open reqs for months, and can't find anyone. I will have to tell the customer we will miss targets if we can't find them by the end of the month."*
>
> *Portfolio Director, BPO, Active Engagement*

> *"... was my first job out of university. They sent me to training. I knew more about AI than the trainer."*
>
> *Individual Contributor, BPO, Canceled Engagement*

Respondents rated their employer's supply of qualified agentic AI talent against realized demand at 3.69, with 47.63% of ratings in the distressed range. The rating does not describe a hiring backlog. There is no pool of experienced practitioners for any vendor to recruit. A firm cannot hire a depth of agentic delivery experience the field has not yet produced.

The shortage reached the engagements directly. Respondents selected talent shortage as one of the 5 most damaging issues in 69 of the 150 engagements. Selections concentrated in failure. 57.14% of canceled engagements had a talent-shortage selection, compared with 39.36% of active ones. Canceled engagements rated the talent supply at 2.46, compared to 3.93 in production.

No vendor in the study had solved the agentic AI talent shortage.

With no talent to hire, the vendor's only remaining capacity development is through aggressive training programs. Respondents rated vendor-supplied training as problematic.

Respondents rated how well employer training programs build practical agentic AI skills at 3.69 and the ability to develop financial industry knowledge at 3.72. On canceled engagements, the 2 readings fell to 2.50 and 2.57.

Of the 800 respondents who named a primary source of agentic AI knowledge, 125 named their employer's training program (15.62% share). Vendor-provided training ranked last among the 6 sources measured. Vendor and partner training led at 140. Online courses and on-the-job learning followed at 136 each, university coursework at 132, and self-teaching at 131.

Training coverage is a condition that an institution can include in an engagement and monitor quarterly, and Action 5 in Part 5 of this report places it in the contract terms.

2.3 Co-Management

Lowest-rated co-management conditions	Mean (1-10)	Rated distressed
Readiness of the joint management model for agent autonomy	3.60	50.37%
Quality measures capture the errors that matter	3.61	51.14%
Clarity on what agents may do without human approval	3.61	50.25%
Speed to jointly stop an agent behaving unexpectedly	3.62	50.50%
Agent updates are visible to the other firm before they take effect	3.62	49.07%
The governance model fits the work performed by agents	3.63	49.19%
Single shared view of the risks agents introduce	3.64	48.30%
Human review points placed where errors matter most	3.64	49.31%
Genuine scrutiny of agent output at review points	3.65	49.50%
Joint issue resolution matches agent speed	3.65	49.44%

Table 10: Lowest-Rated Co-Management Conditions

Joint daily management of an engagement whose work is executed by autonomous systems has no precedent in outsourcing practice. The methods and maturity of joint customer-vendor delivery throughout the outsourcing lifecycle have yet to be developed within an agentic AI context.

Traditional outsourcing governance was paced to people. Review meetings, issue resolution, and change boards assume work at human speed. Agentic delivery moves faster than any of those processes, and it arrives in volumes they cannot clear. Neither the client institutions nor the vendors in this study have built replacements. The market is too young to have produced best practices to adopt.

Among production-stage engagements, 3 conditions rated lowest.

1. Cross-firm visibility of agent actions was rated 3.25, the lowest among the production engagement conditions.
2. Joint human review capacity rated 3.26. A review built for human throughput cannot clear machine throughput.
3. Joint change control rated 3.3. Agents change behavior between formal releases. A model updates. A prompt changes. Upstream data shifts. Respondents frequently make material changes to ensure KPIs are met. A change board built for quarterly release cycles governs none of it, and respondents reported changes reaching production without any joint review.

No co-management playbook exists to adopt. Engagements that survive are those in which the client and vendor build one together before the contract is signed. The highest-rated engagements keep refining that playbook throughout the engagement.

Co-Management Becomes Critical as the Engagement Matures

"We were so busy fixing pre-prod issues that we never thought about how to do the actual work."

Engagement Manager, KPO, Active Engagement

"Scaling the agent fleet outrunning our operating model."

Individual Contributor, KPO, Active Engagement

*"...never figured out how to work together. It wasn't a problem until we turned the lights on.... S***-show now that we are live ."*

Individual Contributor, CXM, Active Engagement

"I have been in the industry a long time. The last six months have been rough. I talk with the customer more about basic operational issues now than I ever did in any of the projects I was involved with in the past."

Portfolio Director, BPO, Active Engagement

Early-stage engagements rated client-readiness conditions as their weakest category. Among production engagements and all canceled engagements, the 15 weakest-rated of the 100 conditions are all co-management conditions. Client and vendor problems are worked down as an engagement matures, and the co-management problem becomes the most impactful category.

The 10 lowest-rated co-management conditions range from 3.60 to 3.65 across the study, a spread of 0.5. When the sample was limited to production engagements, the 3 weakest co-management conditions were rated between 3.25 and 3.27.

Decision Rights and Accountability

"We never figured out the balance between letting our agents fly or keeping them in check. We kept tuning it throughout the project. We never got it right. We never found the right balance."

Engagement Manager, KPO, Canceled Engagement

"We have a constant fight with the bank and our management team. We can let AI run more without people to meet the cost and output goals we have. The bank won't let us. I am a shuttlecock in the middle."

Engagement Manager, BPO, Active Engagement

The weakest condition among canceled engagements is clarity on what agents may do without human approval, with a rating of 2.16. Decision rights over agent autonomy are the first question a joint agentic operation must answer. No canceled engagement had settled it.

Severity rises as the rating moves down the management chain toward the daily work. Individual Contributors rated co-management at 3.52 against 4.98 among Portfolio Directors, a gap of 1.46 points on the same engagements.

Cross-Firm Visibility of Agent Actions

"They kept updating their agents and didn't tell us. We found out when our systems crashed."

Portfolio Director, BPO, Canceled Engagement

"We were told to delay giving the customer access to documentation."

Individual Contributor, BPO, Active Engagement

Agent visibility across the firm boundary was rated 3.25 among production engagements. Asked to name the single most damaging issue in their engagement, respondents chose the agent black box across the boundary 54 times. Visibility or awareness of the other party's agent updates is rated at 3.29.

Canceled engagements rated the same condition at 2.3. Genuine scrutiny of agent output rated 2.16 on canceled engagements, and the quality measures both firms used to validate the other party's agents

rated 2.21. Production engagements rated cross-firm visibility at 3.9, 1.6 points above canceled engagements.

Human Review at Machine Output Speed

> *"Human review was the bottleneck."*
>
> *Engagement Manager, BPO, Canceled Engagement*

> *"The validation queue backs up at month-end. Two weeks into the month, and we still haven't finished last month's QA."*
>
> *Individual Contributor, KPO, Active Engagement*

Joint human review capacity was rated 3.26 across production engagements and led to damaging issue selections in 63 engagements. Placement of review points where errors matter most was rated 3.30. Data indicated that agentic AI production volume was problematic. Genuine scrutiny of work product at agentic AI output scale rated 3.4.

The study indicates that much of the human review process operates within a pre-AI context. Given non-deterministic AI outputs, humans need to review the work product. However, the review process must be optimized and scaled specifically for agentic AI.

Joint Change Control at Machine Speed

"Could not keep pace." *Engagement Manager, BPO, Canceled Engagement*	*"We change things every day to keep it working."* *Individual Contributor, CXM, Active Engagement*
"Change control still lags the speed agents can actually deliver at." *Engagement Manager, ITO, Active Engagement*	*"We are creating a new reality every day. AI behaviour changes without explanation, data feeds are not consistent. Our team has to make updates to data, the process, and the AI rules every day. We inventory them, and I talk to the bank about them every day. Most of the time, they defer decisions. Change control doesn't happen."* *Portfolio Director, BPO, Active Engagement*

Respondents noted that they must constantly effect changes in response to variations in inputs, customer demands, 3rd-party systems, and the non-deterministic output of AI. The study demonstrates that the volume of changes the teams make exceeds the capability of the customer-vendor change control process. Demands to achieve KPIs are high.

Joint change control rated 3.3 among production engagements. The speed mismatch beneath it was measured directly. Joint change control processes kept pace with the speed and volume of change, with ratings of 3.4 in production and 2.26 for canceled work.

Canceled engagements rated joint governance of human- or agent-made changes at 2.19, compared to 3.32 in production.

2.4 The Technology Is Immature and Ranks Behind the Institutions

Lowest-rated technology conditions	Mean (1-10)	Rated as distressed
Production readiness of the tools for bank-grade work	4.54	32.04%
Safe handling of sensitive data across outputs and logs	4.54	32.38%
End-to-end reliability on multi-step tasks	4.55	32.48%
Accuracy of citations and references against real sources	4.55	33.50%
Agents took only intended and authorized actions	4.56	32.07%
Support for reliable pre-release testing	4.56	32.70%
Resistance to hostile instructions in processed content	4.56	32.14%
Integrity of persistent memory against contamination	4.57	31.89%
Ability to explain and provide evidence outputs at audit depth	4.57	32.75%
Stability of behavior when nothing changed	4.57	33.00%

Table 11: Lowest-Rated Technology Conditions

Inside these engagements, agentic AI technology does not meet the standards the financial industry demands. A third of all technology ratings sit in the distressed band. Output-quality defects, including fabricated and inaccurate content, appear throughout the written testimony as a routine operating burden. For institutions whose tolerance for error in regulated processes approaches 0, a technology layer rating of 4.54-4.66 falls well short of the bank-grade standard.

The challenges with the core technology are widely documented beyond this study. 2 quotes provide general examples of the frustration respondents reported.

Non-Repeating AI Output and Hallucinations

"AI invented a husband for a widow on her loan application. I looked at the processing logs. I don't know why it thought this was important, or how it created this new person."

Portfolio Director, BPO, Active Engagement

"Complete guesswork what the agent will produce."

Individual Contributor, BPO, Active Engagement

"We did a POC with the bank to see if we could handle regulatory filings. Using the same sample documents, the AI never generated the same filing reports. The bank canceled the pilot and started looking at every other AI project."

Engagement Manager, CXM, Canceled Engagement

"When the agent could not locate a match, it generated a plausible-sounding term instead of flagging a gap. Our dashboard did not distinguish sourced terms from generated ones until the customer manually spot-checked months after the problem started."

Engagement Manager, CXM, Canceled Engagement

Agentic AI output is non-deterministic. The same agent, given identical input and configuration, returns different outputs across runs. Respondents rated the tools' consistency under identical inputs at 4.58.

No vendor in the study carried a proven method for evaluating non-repeating output. Employer quality assurance capability for such output rated 3.68, with 47.48% of ratings distressed. QA and evaluation gaps led to damaging issue selections in 111 of the 150 engagements, matching the breadth of the missing delivery playbook. An engagement without a playbook improvises its delivery, and an engagement without an evaluation method cannot prove what the improvisation produced.

Teams re-run outputs and compare versions by hand. Respondents maintain non-AI systems in parallel with the agents they operate to validate the agents' output. They check samples to ensure the product has been certified. None of that labor appears in any client-facing measure. Part of the human workload documented in 2.2 is exactly this work. The effort is the cost of operating a non-deterministic system under a deterministic quality model. The data indicate the vendor absorbs it silently.

2.5 The 3rd-Party Ecosystem Ranks Last and Carries Real Exposure

Lowest-rated 3rd-party conditions	Mean (1-10)	Rated distressed
Deliberate handoff of the quiet human fixes for 3rd-party quirks	4.58	33.54%
Change notice reaching the people running the agents	4.58	31.95%
Managing dependence on a small number of providers	4.59	32.04%
Rate limits and quotas under agent volume	4.60	32.92%
Licenses covering work done by agents rather than named users	4.61	31.91%
Knowing where humans quietly fixed 3rd-party issues	4.61	33.08%
Agent dependence on gated external rails	4.61	30.86%
Fallback readiness for a critical 3rd-party failure	4.61	32.04%

Table 12: Lowest-Rated 3rd-Party Conditions

Respondents rated the 3rd-party ecosystem the least severe of the 5 problem areas. Failures inside it remain material and reach every engagement.

Change and Service Disruption Notifications

> *"The bank uses a vendor for KYC. I don't know if that vendor tells the bank if their system is down or not. We never hear about it until our agent stops."*
>
> *Portfolio Director, BPO, Active Engagement*

> *"...The API changed the number of decimals in the feed. We were never told that the change was coming, and our system crashed. I don't know if the bank knew. The bank had a service disruption for almost an hour."*
>
> *Individual Contributor, BPO, Active Engagement*

Change arrives from third parties unannounced. A model version retires, or an API shifts underneath running agents. Upstream data structures change on a cadence the providers alone control, and they owe the engagement nothing.

Change notifications and service interruptions are not reaching the delivery team. Notice quality rated 4.6. The paths that would carry notice to the agents' operators rated the same.

Programmatic Access

"No API, no way in." *Individual Contributor, CXM, Canceled Engagement*	*"I built a script to scrape website for data..."* *Individual Contributor, ITO, Active Engagement*

Respondents rated the availability of effective machine interfaces at 4.63, with 29.80% of ratings distressed. A machine interface replaces a screen built for people. Where APIs exist, their stability is rated 4.66. The quality of the data arriving through them was rated 4.64. Gated external rails include payment networks and regulatory submission systems. Access to those rails is rated 4.61.

Production engagements rated 3rd-party machine interface availability at 5.04. Canceled engagements rated it 3.14, among the widest spreads in the 3rd-party area. The exposure sits in the ecosystem itself. Systems built for human operators now run under agents that cannot use them. Survival and programmatic access move together in this record, whichever drives the other.

2.6 The 5 Problem Areas Across the Lifecycle

An engagement opens against a client institution that has not prepared adequately for agentic AI and has not settled ownership. Delivery then falls to a vendor with at most 2 years of experience doing the work and depends on a workforce without any meaningful agentic AI experience. An engagement that survives those 2 conditions long enough to mature meets the third. No one in this study has built an effective joint co-management discipline. Underneath sits an immature technology, and around it an ecosystem that has not yet evolved for agentic use.

Failure in this study set is institutional in origin rather than technical. Those conditions sit within the buying institution's authority to change. Part 5 defines resolution actions.

Part 3: The Integrity of Performance Information

An institution buying agentic AI outsourcing monitors it through reporting assembled by the vendor's staff. Part 3 examines that channel through the only witnesses available to the study: The people who build the reports. Their account describes a reporting layer with a systemic problem. The people who assembled the dashboards rated them as more likely to be manipulated than accurate. Findings soften at each step up the management chain that interacts with the client. Everything in this part is the vendor workforce describing its own output, with no client verification and no audit.

"We used data from the last cycle that met performance goals. Data is exported into an Excel file. We used that data to load Power BI. If the bank questioned the dashboard, we sent them the Excel file we created."

Individual Contributor, KPO, Canceled Engagement

"We tuned the agent's confidence threshold right before each client review so more breaks got auto-closed as 'resolved' that week. We reset it after the meeting."

Individual Contributor, BPO, Canceled Engagement

"On the release-governance side, failed or rolled-back deploys get logged in a separate incident tracker that never feeds into the published release success rate. The dashboard shows a clean record that we all know is fake."

Individual Contributor, ITO, Active Engagement

"I think we limited the client's audit access to a rolling 30-day dashboard view. Any quarter-over-quarter recalculation of a reported metric became invisible once the earlier snapshot aged out."

Individual Contributor, KPO, Active Engagement

"Our dashboard counted a question as closed when the bot replied, even when the answer was wrong. The employee had to re-escalate through email that never touched the ticketing system."

Individual Contributor, CXM, Canceled Engagement

"We base the dashboards on real data, but AI still creates statistics that are not correct. We try to fix these before the customer can see but the dashboards are real time. If they look at the dashboard during their work hours, I don't know if what they see is good."

Individual Contributor, KPO, Active Engagement

© Akholi 2026

3.1 What the Report Producers Say

Asked to rate the likelihood that their employer manipulates the dashboards and reports the client receives, respondents rated it 6.26 out of 10. 63.61% answered 6 or higher. Respondents rated the accuracy of client-facing reporting at 4.81. Team-average manipulation ratings reach 7 or higher on 56 engagements, and 15 of those remain active today.

Rank softens the reading without materially changing it. Individual Contributors rated the likelihood of report manipulation at 6.39. Engagement Managers, the rank that typically owns the client-facing reports, rated it 5.81. Portfolio Directors rated it 5.59. On average, every rank rated manipulation as more likely than not.

Reported as More Common Than 5 Years Ago

Asked whether manipulation is more or less common than 5 years ago, respondents averaged 6.18, compared with a no-change mark of 5.

Most of this workforce is too new to make a 5-year comparison. Findings hold among those who can. Respondents with 5 or more years of career history answered 5.64, above the 'no change' option.

Morale Moves the Manipulation Reading

The workforce reporting manipulation is also a sign of workforce distress—respondents who are content and intend to stay rated the manipulation at 4.39. Likely leavers rated it 7.41. The correlation between the manipulation rating and job dissatisfaction is $r = 0.62$ ($p < 0.001$).

An unhappy workforce can still testify accurately. The mechanism descriptions in 3.2 are specific and consistent.

3.2 Reports Shaped by Design

Mechanism	Share of descriptions	Survives an accuracy audit
Proxy-metric or favorable-comparator selection	27.89%	Yes
Exclusion from the sample or denominator	18.60%	Yes
Permissive completion definitions	18.18%	Yes
Human-corrected work counted as automated success	17.77%	Yes
Checkbox and point-in-time self-certification	10.33%	Partial
Reclassification or blame shift	8.88%	Partial
Aggregation dilution	8.06%	Partial
Timing and smoothing	5.58%	Partial
Hidden queues and side channels	5.37%	Partial to No
Audit trail suppression	5.37%	No
Threshold gaming	1.86%	Partial
Baseline or target gaming	1.65%	Partial
Outright fabrication	1.45%	No
Narrative framing	1.24%	Yes

Table 13: Reporting Mechanisms Coded from 484 Descriptions

484 respondents across 127 engagements described how client-facing dashboards and reports are manipulated. Their descriptions were independently coded against 14 mechanisms. 4 dominate, and each produces arithmetically correct figures that would likely survive a traditional audit.

1. 27.89% described proxy-metric and favorable-comparator selection. The report carries a metric chosen to flatter. A volume count appears where a correctness rate would be embarrassed. Accuracy is often measured against the agent's own prior output rather than the client's actuals.

2. 18.60% described exclusion from the sample or denominator. Failed and awkward cases route to side queues, review buckets, and categories that never enter the published rate. In many cases, poor-performing metrics are excluded from all reporting.

3. 18.18% described permissive completion definitions. The definition of a finished unit of work becomes more lenient as the workload goal is approached.

4. 17.77% described human-corrected work as automated success. Staff correct agent output by hand, and the dashboard counts the corrected result as the agent's. The human absorption documented in Part 2 is reported to the client as the automation succeeding.

A conventional accuracy audit recomputes the reported numbers and confirms them. All 4 dominant mechanisms pass it. The defect sits in the choice of metric and the definition of done. It also sits in what enters the sample. The arithmetic itself is clean. Catching it requires an audit focused on the exact mechanisms used to generate client-facing reporting.

Mechanism of Shaping by the Outsourcing Model

Outsourcing model	Most described	Second	Third
Customer experience management (n = 133)	Permissive completion definitions, 24.06%	Human-corrected as automated success, 24.06%	Proxy-metric or favorable-comparator selection, 24.06%
Information technology outsourcing (n = 143)	Proxy-metric or favorable-comparator selection, 29.37%	Exclusion from the sample, 18.18%	Permissive completion definitions, 13.99%
Business process outsourcing (n = 122)	Proxy-metric or favorable-comparator selection, 26.23%	Human-corrected as automated success, 19.67%	Exclusion from the sample, 19.67%
Knowledge process outsourcing (n = 86)	Proxy-metric or favorable-comparator selection, 33.72%	Human-corrected as automated success, 19.77%	Checkbox and point-in-time self-certification, 15.12%

Table 14: The 3 Most-Described Mechanisms by Outsourcing Model

Shares are of each model's describers. Third place tied in 2 models. Reclassification also reached 13.99% in information technology outsourcing, and permissive completion reached 15.12% in knowledge process outsourcing.

3.3 Reporting Softens Between the Delivery Team and the Client

Reading	**Individual Contributor**	**Engagement Manager**	**Portfolio Director**
Rated their own knowledge of the client's success measures as adequate, among those able to answer	35.63%	38.67%	50.00%
Gave no numeric read on client satisfaction	36.58%	0%	0%
Rating mildness relative to Individual Contributors, all 100 conditions	reference	+0.43	+1.26
Manipulation likelihood rating, mean	6.39	5.81	5.59

Table 15: Line of Sight and Reading by Rank

Across all 100 rated conditions, Engagement Managers rated their engagements 0.43 points better than their own Individual Contributors did. Portfolio Directors rated them 1.26 points better. The ranks that brief client leadership sit furthest from the daily delivery and carry the most favorable view. Reporting travels upward through layers that each see less and rate better than the layer below.

The Line-of-Sight Gap Beneath the Softening

Much of the delivery organization is unaware of client-defined KPIs or customer satisfaction. 36.58% of Individual Contributors provided no numeric read on whether the client is satisfied. All Engagement Managers and Portfolio Directors provided data on both customer KPIs and customer satisfaction.

Where clarity in customer KPIs exists at the delivery layer, reporting trends are toward being accurate. Self-assessments closely track the client's recorded outcomes as reported by respondents. The remainder lack the sight to report against, whatever their intent. A reporting layer that cannot see the client's definition of success drifts toward the definitions that flatter the vendor.

Part 4: The Delivery Workforce

Every control framework applied to agentic outsourcing assumes a competent and stable human layer that develops, operates, and supervises the agents. The people who staff that layer described themselves in this study. Their testimony carries 4 facts.

1. Respondents lack agentic AI experience.
2. They report low job satisfaction.
3. A majority intended to leave within a year.
4. Their preferred destination is the bank's global capability centers (GCC) more often than not.

4.1 Experience and Credentials

Role	Career years	Agentic years	First agentic engagement	Self-rated expertise	Holds AI certification
Individual Contributor	1.87	0.40	51.70%	4.75	40.63%
Engagement Manager	9.97	0.47	28.00%	5.26	43.33%
Portfolio Director	15.06	0.33	29.03%	6.10	48.39%

Table 16: Experience and Credentials by Role

Meaningful agentic tenure does not exist in this workforce, at any rank. Among the 850 respondents, 467 reported no agentic experience, and 326 reported a single year of experience. The maximum anywhere in the study is 1 year.

Junior in the Trade, New to the Technology

> *"I don't know if it is a Gen Z issue or if it is because they have never had a job before. They may be smart and have some AI skills, but I can't get them to do basics like responding to a client email or showing up to meetings on time. I spend more time teaching them how to be an employee than I do about the actual work."*
>
> *Engagement Manager, KPO, Active Engagement*

> *"I am tired of managers who have been in outsourcing for decades telling us what to do. They don't understand AI. I gave up trying to explain to them why what they did before doesn't work now."*
>
> *Individual Contributor, BPO, Active Engagement*

The workforce is junior in the trade. 77.04% reported fewer than 5 career years, and the median career length is 2.

The extreme sits with the Individual Contributors. 51.70% were on their first agentic engagement, and 59.74% reported having no prior agentic AI experience. 9.93% of Individual Contributors reported having no career experience before the specific engagement.

Management holds deep delivery seniority and no agentic seniority. Engagement Managers averaged 10 years of career experience and 0.5 years of agentic work. Portfolio Directors averaged 15.06 years of career experience and even less experience with AI. Signing off on agentic delivery falls to people who learned the outsourcing trade in an era that ended before the technology arrived.

Agentic AI is still new. There is no solution to the shortage of workers with collective wisdom gained through years of experience. However, the risk to financial institutions is material. Those who build, operate, manage, and govern agentic AI projects for the banks are learning on the job.

Certifications and Credentials

No credential separates confidence from capability. Across the workforce, 58.71% hold no agentic AI certification, and the only AI certification that does appear is one of 5 generic platform-fundamental courses. Management holds full credentials in the previous outsourcing era, including Six Sigma, ITIL, and project management.

4.2 Morale and Departure

Role	Job satisfaction, mean (1-10)	Likely to leave within a year
Individual Contributor	4.09	55.52%
Engagement Manager	4.45	50.00%
Portfolio Director	5.00	41.94%

Table 17: Morale and Departure Intent by Role

Job satisfaction averaged 4.19, and the most common rating was 1. 46.79% rated satisfaction at 3 or below.

Departure intent mirrors it. 53.88% rated themselves as likely to leave within a year, and satisfaction and departure intent correlated at $r = -0.91$ ($p < 0.001$). 43.29% rated satisfaction at 4 or below and departure likelihood at 7 or higher. Only 16.47% of respondents indicated they were both content and staying.

Feeling Wasted Drives the Departure

I wrote a proposal for my managers on how we "could improve our AI practice. They didn't read it. They told me I need to have more experience before I can understand how the business works. I might not have been working as long as them but I have more AI experience than my entire management team combined."

Individual Contributor, BPO, Active Engagement

"I am just meat... I try to explain the problem. Nobody wants to listen to me."

Individual Contributor, CXM, Active Engagement

While direct management, job stress, and other indicators surfaced, the strongest correlate of departure intent is whether staff believe their employer uses them effectively, with a correlation of $r = -0.78$, $p < 0.001$. Respondents noted frustration that their employer is not using them to help define the agentic AI offering strategy and to optimize the vendor playbooks now being created.

Management Read the Distress as Milder

Management read the team's distress as milder than the people reporting it. Portfolio Directors assessed their teams' satisfaction 1.69 points above what the delivery layer reported for itself. Portfolio directors understated the team's intent to depart by 1.26 points.

4.3 Where the Work and the People Are Going

Reading	Individual Contributor	Engagement Manager	Portfolio Director
Strategy expectation			
The client is scaling its agentic AI outsourcing engagements	20.22%	20.00%	17.24%
The client is shifting more work to its own global capability centers	51.25%	41.33%	23.33%
Career direction			
Prefers to work directly for a bank in its global capability center	42.79%	40.67%	20.00%
Working for a bank would be preferable over outsourcing vendors	46.77%	36.67%	13.33%
Wants to continue working in the financial industry	19.90%	32.67%	48.28%

Table 18: Strategy Expectations and Career Direction by Role

Respondents' strongest strategy expectation is that the client will move the work into its own global capability center (GCC). 44.28% rated the move as likely, compared with 19.78% who expected the client to scale its outsourcing within the respondent's employer.

Expectation intensifies inside failing work. Respondents on canceled or high-risk engagements rated the GCC move at 7.06 of 10, with 64.91% rating it 7 or higher. On canceled engagements alone, the mean reaches 7.31.

41.55% of respondents rated themselves as likely to seek employment within a bank's GCC. The pull concentrates on the population leaving. Likely leavers rated GCC preference at 7.45 out of 10, while likely stayers rated it at 2.77.

4.4 What Departure Costs the Institution

Section 2.1 (Client Preparation) documents the human backstop. Manual repair and undocumented decisions hold these engagements together outside any reported measure. The cost of that dependence lies in what is left when the people leave. An engagement's operating knowledge has no backup copy, and the error-catching that keeps defects from the client departs with the staff who performed it.

Continuity condition, rated 1-10	Mean	Rated distressed
The engagement can still be performed by hand, and the manual workarounds agents now depend on	4.66	30.82%
The engagement knows where humans quietly fixed 3rd-party and data issues	4.61	33.08%
The quiet human fixes were deliberately handed over rather than left tacit	4.58	33.54%
Share of the workforce likely to leave within a year		53.88%
Workforce readings on canceled engagements, relative to active		2 to 3 points lower

Table 19: The Backstop and Its Exposure

All 3 continuity conditions ranged from 4.58 to 4.66, with 30.82% to 33.54% of ratings distressed. The engagements do not reliably know where their own human fixes live, and they have not deliberately handed those fixes to anything.

Against those readings sits a 53.88% departure intent. The gradient makes the exposure concrete. Every workforce indicator is 2 to 3 points lower for canceled engagements than for active ones. The engagements most dependent on undocumented human knowledge are those whose staff sit closest to departure.

What Leaves With the team

“The knowledge of how to fix things walks out the door faster than we capture it."

Portfolio Director, BPO, Active Engagement

"SME knowledge-transfer completion was tracked by the number of sessions held rather than whether the underwriting logic was actually captured, so the metric looked complete while real institutional knowledge kept leaving with departing staff."

Individual Contributor, BPO, Active Engagement

*"Capture process documentation continuously. Don't wait for the exit to write it down. We got f******"*

Individual Contributor, BPO, Canceled Engagement

"The team that did the work left to work for the bank. I had no understanding of how much they did beyond their defined job to keep things running. A few days after they left, the FX agent failed. I had to call the bank and ask them to send the team back to us for a few days to fix it."

Portfolio Director, BPO, Active Engagement

Undocumented workarounds and the map of past fixes leave with each departure. For an institution, the conclusion is structural and requires direct mitigation. Failure to address team departures or document their methods will lead to a new team forced to reinvent what the previous team refined. Given the volume of undocumented and behind-the-scenes efforts by the respondents, as regularly called out in this report, the loss of team members will affect the performance of the outsourcing engagement.

Part 5: Conclusions and the Executive Agenda

The primary finding of this study is that the agentic AI outsourcing industry is in its infancy. As with past industry transformations, the outsourcing industry has yet to fully understand the context needed to achieve high success rates and strong project returns.

Failures that led to low customer satisfaction and high engagement cancellation rates were consistent in the data provided by respondents.

1. The level of customer preparation, governance, and oversight was the top callout by respondents.
2. Outsourcing vendors lack experience and an established playbook for providing agentic AI outsourcing services to any customer, not just banks. Further, outsourcing vendors are attempting to define this new playbook with a team that lacks any material experience with the technology itself.
3. As the customer and vendor resolve institutional and technical challenges, co-management becomes the primary success factor in production-stage engagements. The industry is attempting to apply best practices gained over the past 20 years to functions such as issue and risk management, decision-making, and change control. All were optimized for human-centric delivery. These functions no longer work with the speed and volume that agentic AI delivery models produce. All engagements in the study that achieved an ROI invested in co-management planning before the engagement started. Further, all successful engagements made the co-management process improvement a senior executive priority throughout the engagement.

The bank's core technologies and the broader 3rd-party ecosystem of cloud, SaaS, services, and data providers warrant intentional investigation and remediation. However, the most impactful and likely largest single workload is in bank-side preparation, governance, and oversight. The study demonstrates that banks must immediately:

1. Thoroughly document, optimize for AI, and ensure consistency of all processes that will be affected by agentic AI.
2. Address data quality, consistency (normalization), and accessibility challenges that are routinely affecting agentic AI.
3. Optimize governance, oversight, risk, and compliance for the new era of agentic AI. If one has not yet been identified and communicated to outsourcing vendors, the bank must define a senior executive as the official owner for all agentic AI work.

4. Fix legacy systems access, allowing agentic AI to consistently access the tools and data it needs at the required speed.

5.1 6 Immediate Executive Actions

Action 1

Gate Every Agentic Contract Behind an Institutional Readiness Standard

If this role does not yet exist, assign a senior executive the role of the AI Tsar. Within this person's charter, gate every agentic contract behind a written institutional readiness standard, met in full before signature. Review every existing engagement against the same standard on a regular schedule.

The standard covers 5 elements.

1. Data quality and accessibility.
2. Process documentation, consistency, and optimization to the depth required by agentic AI.
3. Legacy-system access.
4. Expert availability.

Action 2

Build Co-Manage Discipline Before Signatures, and Improve It Throughout

Include an intentional, senior-executive-led co-management definition and optimization in every request for proposal. Build the co-management discipline jointly with the vendor before signing the contract. Throughout the engagement, intentionally work with the vendor to improve co-management. Failure by the outsourcing vendor to give co-management material weight must be grounds for termination.

Key functions to include :

1. Risk management, issue resolution, decision-making, change control, and more, optimized for agentic AI delivery.
2. Regulatory, legal, and compliance.
3. Active co-management of every engagement, including transparent weekly reviews between named engagement managers.

Action 3

Audit the Design of Vendor Reporting Rather Than Its Arithmetic

Commission a 3rd-party design audit of vendor reporting. Repeat these 3rd-party audits annually. This audit must answer 4 questions.

1. Who defines each metric and its comparator?
2. What counts as complete?
3. What enters the sample and what routes around it?
4. Is human-corrected work reported as automated success?

If any issues are found in the audit, embed a bank employee directly into the outsourcing vendor's team until behavior changes.

Beyond the 3rd-party audits, change how the bank reviews the vendor's performance reporting. Add an intentional question set that asks the vendor to demonstrate how all dashboards and reports are generated. Ensure the bank has an internal technical and functional expert involved in reviewing the mechanical process the vendor used to create each report. Given the rate of change in agentic AI delivery, this conversation must be appended to every performance report review. Regularly, have the bank's internal team audit the vendor's reporting mechanisms.

Action 4

Restructure Commercial Terms Around Outcomes and Exit

Price the work on measured and validated outcomes. Define the measured outcome in the contract with the design-audit protections of Action 3. Write the exit while the institution still holds negotiating power. Staged gates with evidence requirements.

1. Data and knowledge escrow.
2. Transition obligations.
3. Termination terms that make abandoning a failing engagement a decision rather than a write-off.

Action 5

Treat the Human Oversight and Development Layer as a Control Function

Every assurance the institution receives about supervised agentic delivery rests on people averaging under 1 year of experience with the technology they supervise.

Treat the human oversight and development layer as the institution treats any control function.

4 terms belong in the contract.

1. Qualification standards for the people supervising agents.
2. Retention strategy and continuity for key persons.
3. Visibility into turnover on the institution's engagements.
4. Verified human capital development programs.

Action 6

Listen Directly to the Vendor's Delivery Team

The study uncovered an incredible depth of knowledge among the delivery team that is undocumented and largely out of the customer's sight. More than any dashboard or report, developing a relationship and listening to the vendor's team will provide the bank with the best view of reality.

Key components to include:

1. Begin immediate rotations of bank team members to the vendor's delivery site. Similarly, rotate members of the vendor's delivery team to the bank for short-stay immersion sessions. Get to know each team member. Develop the relationship necessary to elicit important information. Doing so will also reduce unnecessary turnover of high-value team members.
2. Hold regular joint retrospectives that include Individual Contributors. Ensure that all vendor team members can speak freely.
3. Be involved in some form of the exit interview process with departing delivery staff. Ensure the bank gains a better understanding of why the person is leaving (if voluntary), what the bank could improve, and how the overall engagement could be improved.

The Actions Sit Inside the Institution's Authority

None of the 6 actions requires a technology decision or the permission of a regulator. Each is a governance choice within the authority the institution already holds.

The record already prices inaction. Cancellation stands at 37.33% realized, with half the study set canceled or at high risk. The conditions the 6 actions govern do. Standards for this industry are being set now by whoever moves first. That work belongs on the board's agenda this quarter.

Appendices

About Phil Hatch

Phil Hatch is a 30-year senior executive and authority on global outsourcing. Hatch built and led ManpowerGroup's outsourcing business unit. He advises Tier 1 banks, outsourcing vendors, Fortune 500 companies, and governments on making large-scale outsourcing work and preparing global labor markets for job growth. His focus now is the hardest version of that problem, transitioning the outsourcing industry to prepare for artificial intelligence.

He is the author of The State of Agentic AI Outsourcing at Tier 1 Financial Institutions, a study based on 150 engagements and 850 people who deliver them. The report continues a line of research he began 2 decades ago into why outsourcing succeeds or fails.

As chief executive of Ventoro, later acquired by ManpowerGroup, he built the workforce research function for a major investment bank and helped it open its offshore captive center in Hyderabad. Hatch has been a board advisor for numerous major outsourcing firms, including Infosys and Wipro, on industry direction, customer management, risk, and the labor economics behind their client engagements. For more than a decade, he has assessed the AI readiness of Fortune 100 companies, including leading an enterprise AI rollout for a global technology firm where the binding constraint proved to be data rather than technology.

Alongside this work, Hatch advises governments and international organizations, among them the World Bank, the ILO, UNDP, and the African Development Bank, on workforce and economic development. Most recently, the US State Department asked Hatch to return to Ukraine as a senior advisor on economic and labor policy reform.

Akholi

Akholi is a research-led advisory firm for agentic-era outsourcing, service engagements, and global capability centers. Its authority rests on primary evidence rather than opinion.

In the first half of 2026, 850 of the people who deliver agentic AI outsourcing described 150 live engagements at Tier 1 financial institutions. Of those engagements, 37.33% had already been canceled. The record supports one governing claim. In the engagements studied, outcomes track the conditions the buying institution builds rather than the deals it signs. Failure there is institutional in origin rather than technical. The determinants sit within the institution's own authority to change.

Akholi helps boards and executives act on that finding. The Report sets out the evidence. The Book codifies the operating discipline it demands. The Academy teaches that discipline. The Practice applies it to a live engagement.

The firm's work is led by Phil Hatch, a thirty-year Fortune 500 executive authority on global outsourcing who built and led ManpowerGroup's outsourcing business unit and has advised Tier 1 banks, the World Bank, the International Labour Organization, and the US State Department.

The Book

An operational playbook for preparing and jointly managing agentic AI service engagements. It sets out the discipline across the full lifecycle: five phases, five decision gates, and one measurement floor, from preparation and procurement through transition, steady-state co-management, and exit. It reads as a practitioner's desk reference, with the tools and templates hosted on Akholi.com. It releases in the fourth quarter of 2026.

The Practice

Fixed-scope, time-boxed clinics that apply the discipline to a live situation.

- Outsourcing Performance Tuning. A live engagement drifting toward cancellation. Three months.
- Outsourcing Preparation. Readiness measured and closed before signing. Three months.
- Global Capability Center Planning. Whether and how to stand up, expand, or repatriate captive work. Six months.
- Global Capability Center Performance Tuning. An underperforming captive center. Three months.

- Agentic-AI Governance, Oversight and Risk. Standing up or repairing the control function. Six months.
- Crisis Response. Acute failure, led by a senior executive, within twenty-four hours.

The Academy

Executive education that transfers the discipline to the people who run engagements. Each class runs onsite for a cohort of fifteen.

- Outsourcing Lifecycle Management. The foundational course across the full five-phase lifecycle, and the prerequisite for the master classes. Two days.
- Governance, Oversight and Risk Master Class. The control function in depth. One day.
- Performance Optimization Master Class. Reading and improving a live engagement's true performance. One day.

Study Information

1. Design, collection, and independence

The study recorded structured ratings and written testimony from 850 vendor delivery staff who work on the 150 agentic AI outsourcing engagements at Tier 1 financial institutions. Collection ran throughout the first half of 2026. Respondents hold the 4 vendor delivery roles named throughout the report:

1. Individual Contributors, who execute the work.
2. Engagement Managers, who run a single client engagement.
3. Portfolio Directors, who oversee several Engagement Managers.
4. Customer Management Executives, who own a client relationship.

All 10 primary institution types and all 4 primary outsourcing models are represented.

Independence was built into the collection rather than added after it. A 3rd-party research firm administered the survey using its own instrument and retained the raw responses. The author never saw any response in which identifying details might have surfaced. That firm reviewed every response and certified that nothing identifying a bank, a vendor, a bank customer, or a proprietary method had been collected. The questions were written so that no respondent could name an employer, a client, or a client customer. Both firms behind every engagement stay anonymous by construction. The anonymity was deliberate. Delivery staff describe their own employer more fully when neither firm can be identified, and the candor of the testimony in Parts 3 and 5 reflects it.

Recruitment and screening followed a fixed order:

1. A specialist provider built the candidate pool from professional networks, published profiles, and vetted panels, then added referrals from early respondents. Participation was compensated.
2. Recruitment reached more than 900 completed responses.
3. Screening removed anyone whose answers showed no working knowledge of banking, the outsourcing industry, or agentic AI. The screen ran before any analysis, so it could not be tuned toward a preferred result.
4. The retained sample was 850. A validation round followed, during which 122 respondents completed short telephone interviews to confirm their roles and responses. No material discrepancy appeared.

Two design choices bound what the study can claim:

1. Sampling was purposive rather than proportional. Smaller institution types, among them investment banking and wealth management, were oversampled so that each type had enough observations to be described. Engagement counts, therefore, say nothing about market size or share.
2. The account is vendor-side self-report. It is systematically instrumented and internally validated, but not corroborated by client records.

The data also hold unresolved internal contradictions, reported as they stand rather than smoothed by assumption. The workforce rates its own confidence against the grain of its experience, and the delivery layer reports both the highest likelihood of manipulation and the least sight of the client measures it is judged against. A market this young, drawn from a small qualified population, produces such patterns, and the significant findings survive them.

2. Statistical conventions and precision

The instrument carried 200 questions, grouped as follows:

1. Respondent demographics and sentiment.
2. Engagement facts.
3. Client and vendor profiles.
4. 100 universal engagement-condition ratings, set in 5 areas of 20.
5. A forced ranking of the 5 issues most limiting each engagement.
6. Open written testimony.

Rating scales range from 1 to 10, with 1 being the worst. A rating of 3 or below is distressed. The answers "I do not know" and "I prefer not to answer" were valid everywhere, and item nonresponse is itself reported as a finding where it occurs. The text quotes ratings to 1 decimal place and shares to whole numbers, and the exhibits keep full precision.

The study carries 2 units of inference, and the report holds them apart:

1. Engagement-level facts, among them age, stage, cancellation, satisfaction, term, and headcount, take one value per engagement and rest on n = 150.
2. Respondent-level readings, among them ratings, sentiment, and testimony, rest on the 850 individual responses. These cluster within engagements, since colleagues on one engagement answer more alike than strangers, with intraclass correlations (ICC) between 0.48 and 0.84 across the rated conditions. Clustering cuts the information in the 850 responses to an effective sample of roughly 170 to 280, and every respondent-level interval below carries that design effect rather than the raw count.

The 95 percent confidence intervals are these:

1. Realized cancellation of 37.33 percent runs from 30.0 to 45.3 percent, on n = 150.
2. Mean engagement satisfaction of 43.85 percent carries about plus or minus 4.5 points, and the active-engagement mean of 64.1 percent about plus or minus 2.3 points.
3. Workforce job satisfaction of 4.19 carries about plus or minus 0.4 of a point.
4. Departure intent of 53.88 percent runs from about 47 to 61 percent.
5. Manipulation rated 6 or higher, at 63.61 percent among those able to answer, runs from about 58 to 70 percent.

As a working rule, whole-study engagement facts are reliable to within about 8 points, respondent-level shares to within about 7 points, and subgroup readings widen as the cell shrinks.

A difference is called a finding only where it survives testing at conventional thresholds, with correction for the number of comparisons. Three cross-cutting results pass:

1. Overall engagement condition is the strongest correlate of cancellation, at r = -0.72. Ranked into thirds by condition, the weakest, middle, and strongest thirds cancel at 88, 24, and 0 percent.
2. Engagements on 24-month terms cancel at 56.25 percent, against 28.00 percent for all other terms pooled, at a corrected p = 0.010.
3. The age gap between corporate-functional and line-of-business engagements holds at p = 0.003.

The differences the report declines to call findings fail those tests. Cancellation and satisfaction do not differ significantly across institution types, at p = 0.59 or worse, nor across outsourcing models, pricing models, or contract-value bands. Where a cell is small, the report says so, and the 2 Customer Management Executives are reported with their counts and support, not percentages.

The mechanism is shared in Part 3, and the post-mortem themes in Part 2 come from coding the open testimony. 484 manipulation descriptions and 759 advice responses were each coded against a fixed set of categories, with multi-coding allowed. Coded shares move by about 3 points under reasonable changes to the coding rules, and the claims rest on the order and clustering of the stable categories rather than on any single share.

3. Limitations

The report carries 7 limitations, stated together:

1. The account is vendor-side self-report, uncorroborated by client records.

2. The portfolio is young, at a median age of 5 months, and was observed once rather than tracked over time.
3. Purposive sampling bars any inference about market composition.
4. Segment cells are too small to separate outcomes by institution type or model.
5. Respondent clustering reduces the effective sample for workforce readings.
6. Forward risk figures are the delivery teams' own assessments, uncalibrated to realized outcomes.
7. Text-coded shares carry coder sensitivity of about 3 points.

None of these is unusual for a first systematic study of a market this young, and none is hidden in the body. Each finding cites its basis where it is stated.

The report is published for research purposes and does not constitute legal, regulatory, or investment advice.

Artificial Intelligence Disclosure

The author prepared this report with the assistance of an artificial intelligence system, Claude from Anthropic. The system worked as a directed tool on defined tasks, and the author reviewed its output. Its role covered 3 areas:

1. It assisted in the statistical analysis of the collected data.
2. It helped structure the report and refine its prose.
3. It checked the dataset and the manuscript to confirm that no bank, no bank client, and no client customer could be identified.

Grammarly was used to address spelling and punctuation throughout the report. Quotations from respondents appear as they were submitted.

The observations, findings, and conclusions in this report are the author's own. The artificial intelligence does not generate them. Stylistic patterns sometimes read as markers of machine writing. Where they appear here, they are artifacts of Claude's work in refining the text. The analysis and the judgments remain the author's.

Acronyms and Abbreviations

The terms below are the acronyms and abbreviations used in the report, with their expansions and the sense in which the report uses each. Every definition is drawn from the report.

Acronym	Definition
AI	Artificial intelligence. The subject of the report: agentic AI outsourcing, in which vendors build and operate autonomous AI agents for financial institutions.
API	Application programming interface. The connection between software systems; cited by a respondent whose agent failed after an API changed.
BI	Business intelligence. Reporting and analytics software, such as Power BI, referenced in the delivery workflow.
BPO	Business process outsourcing. One of the four primary outsourcing models in the study set (37 engagements).
CORAL	Compliance, Oversight, Regulatory, and Legal. The report's heading for the compliance dimension of an engagement (Section 1.9).
CXM	Customer experience management. The largest of the four outsourcing models in the study set (42 engagements, 28.00% of the set).
DOI	Digital Object Identifier. The persistent identifier assigned to the published study.
FTE	Full-time equivalent. A unit of staffing; the report tracks the shift from FTE-based delivery to agentic delivery.
FX	Foreign exchange. Referenced by a respondent describing an agent failure in a foreign-exchange process.
GCC	Global capability center. A captive, in-house delivery center; 44.28% of respondents expect the client to move the work into its own global capability center.
ICC	Intraclass correlation. A statistical measure of how closely respondents on the same engagement agree, used in the analysis.
ILO	International Labour Organization. One of the international bodies the report's author has advised.
ISBN	International Standard Book Number. The standard identifier for the published report.
ITIL	Information Technology Infrastructure Library. A service-management framework, cited alongside Six Sigma as prior-era practice.
ITO	Information technology outsourcing. One of the four primary outsourcing models in the study set (40 engagements).
KPO	Knowledge process outsourcing. One of the four primary outsourcing models in the study set (31 engagements).
KYC	Know your customer. A financial-compliance process referenced by respondents.
ORCID	Open Researcher and Contributor ID. The persistent researcher identifier for the report's author.
POC	Proof of concept. One of the three lifecycle stages of an engagement: planning, proof of concept, and production.
QA	Quality assurance. The checking of delivered work, referenced in respondent testimony.
ROI	Return on investment. A measured return on engagement, reported for 17 of the 30 production engagements and averaging 24.71%.

Acronym	Definition
SME	Subject matter expert. The holder of specialized process knowledge, referenced in the context of knowledge transfer.
UNDP	United Nations Development Programme. One of the international bodies the report's author has advised.
US	United States. Appears in reference to the US State Department.