



AGENTIC AI — FUNCTIONAL USE CASES IN FINANCIAL SERVICES

Where the value is — the deployment record, the evidence tiers, and the portfolio

Paper II in the Wissen agentic-AI series. A continuation of "Agentic AI in Financial Services: the infrastructure that unlocks the value — and contains the risk."

Prepared By

Natarajan K Venkatasubramaniam

Executive Director, Wissen Technology
August 2026

DELIVERING 2X IMPACT

www.wissen.com

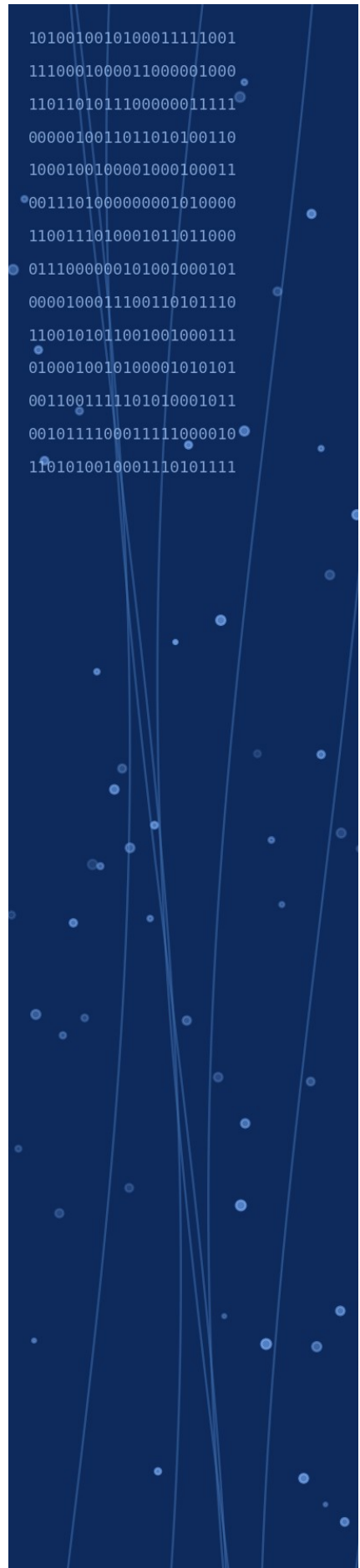




Table of Contents

About this series.....	3
Who this paper is for.....	3
Executive summary.....	4
1. What this paper adds.....	5
2. The Wissen Naturalness Test.....	5
2.1 The test, applied: three candidates, three verdicts.....	6
3. Reading the deployment record.....	7
3.1 The base rates.....	7
3.2 The evidence tiers.....	8
3.3 Where the deployments actually sit on the staircase.....	8
4. Investment banking.....	9
4.1 Where the hours are.....	9
4.2 The use-case inventory.....	9
5. Broker-dealers.....	11
5.1 Where the hours are.....	11
5.2 The use-case inventory.....	12
6. Asset managers.....	13
6.1 Where the hours are.....	13
6.2 The use-case inventory.....	13
7. Custodians and fund servicers.....	15
7.1 The segment that proves the thesis.....	15
7.2 The use-case inventory.....	16
8. Private wealth and wealth management.....	17
8.1 Where the hours are.....	17
8.2 The use-case inventory.....	17
9. The cross-segment value map.....	18
10. The retreat file: what fails, on the record.....	19
11. Constructing the portfolio.....	20
11.1 The selection method.....	20
11.2 Wave 1 playbooks: scope, controls, baseline, benchmark, exit.....	21
11.3 The measurement dictionary.....	22
11.4 What never enters the portfolio.....	23
11.5 Build, buy, or configure.....	23
12. Conclusion — and five questions for your portfolio.....	24
Appendix A — The full inventory, scored.....	24
A.1 Investment banking.....	24
A.2 Broker-dealers.....	25
A.3 Asset managers.....	26
A.4 Custodians & fund servicers.....	26
A.5 Private wealth & wealth management.....	27
References.....	28
About Wissen.....	31

About this series

This is the second in the series of agentic-AI papers produced by Wissen Technology. The first in the series, entitled “**Agentic AI in Financial Services: the infrastructure that unlocks the value, and contains the risk**” (Wissen, July 2026), will be referred to throughout this paper as **Paper I**. That paper answered the how. It provided the non-obvious answers to the twelve challenges that must be met in order to transition from an agent pilot to a full-scale production solution. It included a reference architecture based on the deterministic spine concept, with control and assurance planes. It included a full description of a reporting break remediation agent developed in end-to-end fashion. It provided a five-level autonomy staircase and a first-90-day execution path.

This paper answers the where, the who and the what left open by Paper I. In the where, we provide a catalog of the public deployment record across five segments (investment banking, broker-dealers, asset management, custody and fund services, wealth management). In the who, we provide a ranking of each use case based on the difficulty of its evidence. And in the what, we include a full section on failures, since the industry has already paid for these lessons and there is no point in paying for them again.

There is no required reading order between the two papers: you can start with either. This is the market proof of where to deploy the architecture presented in Paper I. All cross references to Paper I are noted as (Paper I, §n). We tried not to repeat any material from there beyond a single referencing sentence.

Who this paper is for

Experience teaches us that choosing an AI portfolio for a regulated company is rarely the purview of a single decision-maker. There are six of them and every one of them usually walks into the room with a different question. This paper is about answering all of them.

If you lead...	The question you are accountable for	Where the paper answers it
Business / P&L	Where is the proven value in my segment, and what is the external benchmark my initiative will be held against?	The segment chapters (Sections 4–8), the measured-outcomes wall (Figure 3), and the Wave 1 playbooks (§11.2).
Operations	Which of my queues — breaks, alerts, questionnaires, documents — have named-firm precedents, and what did those firms measure?	The use-case inventories per segment, the triage pattern (§5), and the custody case study (§7.1).
Compliance / Risk	Which use cases have regulators validated, which have they cautioned against, and where is the client-facing boundary today?	The regulatory codification and retreat file (Section 10), the Tier 3 exclusions (§11.4), and the segment regulatory anchors (§5.1).
Technology / AI	What do we build first on Paper I’s architecture, and what should we buy instead of building?	Portfolio construction (§11.1), build-vs-buy (§11.5), and the per-use-case playbooks (§11.2).
Finance	Which initiatives can be baselined and budgeted, and what does “proof” look like at review time?	The Naturalness Test filter 4 (Section 2), the measurement dictionary (§11.3), and the five portfolio questions (Section 12).
Distribution / Sales	How does AI competence become a client- and allocator-facing asset rather than a cost line?	The allocator DDQ evidence (§6.1), the RFP/DDQ playbook (§11.2), and the measurement-as-moat argument (§3.1).



Executive summary

1. Value concentrates. Diversified strategies fail. Study the deployment history in the open literature and the realized value keeps concentrating in the same six patterns: knowledge retrieval, document intelligence, exception triaging, supervised drafting, interaction capturing and software engineering. Same goes for the independent outcome evidence. Among the use cases reported by banks that report any measurable outcomes, roughly two-thirds concentrate in just two domains: front office and technology. If the firm gets good at implementing six patterns, it will have covered most of the reliable value that this generation of AI can offer. Scattered pilots will cost the firm only demonstration of its capabilities.

2. Let the evidence drive the portfolio choice and forget the ambition. Each use case is graded into three tiers: Tier 1, publicly measured outcome of the use case in a named firm; Tier 2, pattern is proven but the evidence is sparse and/or comes from a vendor; Tier 3, documented failure modes. The Tier 1 list has grown big enough that no consulting estimate should be necessary to justify your first wave of implementation. 75% reduction in contract review process at BNY. 33% analyst-estimated reduction in surveillance investigation time at Nasdaq. Reduction of due-diligence questionnaire (DDQ) processing time from three days to four hours. Up to four hours reduction per advisor meeting at Bank of America.

3. Autonomy frontier lies where the errors can be cheaply made, and Paper I's architectural framework makes such prediction. Map the named production deployments on Paper I's autonomy staircase and you will see that everything above Level 2 is implemented within an internal, reversible and error-checkable domain: programming agents at Goldman Sachs and Citi, digital employees with human managers at BNY, digital worker squads at SS&C, overnight research agents at BlackRock. There is no autonomous AI used in any regulator-facing process at any regulated company today. The single instance of a client-facing boundary test, Schwab's assistant, went live in June 2026 and hasn't published any results. Reflect on that use case for a minute: the firms that have climbed highest are those that have built the most management infrastructure for their agents.

4. The list of failures is complete and rhymes. Klarna renounced publicly its AI-first service business approach. Commonwealth Bank of Australia reinstated 45 positions and labeled the substitution of its voice-bots as an "error". Both Citi and Bank of America, among the two companies with the richest troves of client data globally, evaluated client facing generative AI and decided not to pursue it. The common denominator is pretty obvious: substitution does not work, while augmentation multiplies. Section 10 presents the retreats as data, not anecdote.

5. Measurement is the moat. More than seventy percent of publicly announced bank AI use cases provide no metric at all for their outcome. It is a unique kind of gift: in the market where most of the players cannot provide proof of effectiveness of their AI, those who could have a distinct competitive advantage both with respect to customers, allocators, and regulators. This is why the portfolio approach described in Section 11 considers the baseline as a precondition for funding.

This paper by the numbers.

137 discrete source citations across 70 references	5 industry segments covered end to end	65 use cases scored (Appendix A)
6 patterns carrying the measured value	2 documented public reversals	0 documented reversals of human-gated deployments

1. What this paper adds

There are two conversations in almost every boardroom that we sit in. Paper I provided the answer to one of them.

The first is the **autonomy** conversation: why does our AI need human approval of everything? Paper I answered that question structurally: autonomy is earned level-by-level against assurance gates (Paper I, §6). What Paper I did not do is show the market evidence, i.e., who earns the level and what he/she gets for it. This paper provides the evidence.

The second is the **coverage** conversation. A leadership team puts a functional capability map on the table and asks where AI is applicable. The answer “almost everywhere, trivially; in a few places, transformatively” satisfies nobody without the ranking of places. We watched what happens with those companies that bypass the ranking and implement AI on every box on the map. We call the approach **AI theater**: the demonstration that will impress a steering committee but will disappear on implementation. A language model brings value proportionally to three things: to the share of language-shape work, to its hours spent, and to the possibility of substituting guaranteed answer with the reviewed draft. The three quantities vary two orders of magnitude across the capability map. The ranking is what this paper does.

On methodology. Over the last eighteen months, we have monitored the published record of AI adoption into the five categories of this paper. The evidence base is 135 distinct sources within the 69 references at the end, and these include firm press releases and management statements, regulators’ reporting and enforcement cases, industry survey results with disclosed sample sizes, paid-vendor claims, and the still small but increasing number of firm-reported impact measurements. All facts in this paper come with a citation. Numbers from vendor press releases are treated as vendor claims, statements from executives are taken as such and not audited as results, and anything we cannot verify goes into the garbage bin.

2. The Wissen Naturalness Test

There are naturally occurring use cases, where the work must be done this way and where AI removes the existing friction. There are forced use cases, where AI is used simply because there is an obligation to use it. Five filters, applied in sequence, and all five must be satisfied for a use case to qualify for the portfolio.

Figure 1 — The Wissen Naturalness Test: five filters, applied in sequence; most capability map items fall at filter 2.



A use case must pass all five filters, in order, to enter the portfolio — most capability-map entries fail at filter 2



Filter 1: Language-native work. The nature of the work must be reading, extracting, comparing, drafting, and summarizing. When the work is numerical computation or execution of deterministic rules (calculations of NAV, margins, net capital, settlement instructions), conventional software already wins hands down, and there is no reason to introduce a probabilistic model in the way. It works both ways. AI explanations of why a number has changed are natural. Artificial intelligence to produce the number is malpractice, and Paper I's deterministic spine is intended precisely to prevent this (Paper I, §§2.1-2.2).

Filter 2: A large, costly hours pool. Value equals hours times loaded cost times automatable fraction. A use case involving the daily work of forty people outweighs an elegant use case involving the quarterly work of four people. The vast majority of any capability map's long tail vanishes here, for a prosaic reason: the AI works; the value does not.

Filter 3: Human review errors survive. Draft-to-review workflows admit of imperfection. Nothing auto-committed to a client, a market, or a regulator does. All reported failures of AI in this industry are filtered through this criterion (Section 10).

Filter 4: There is a measurable before. Hours spent in development, in responding to alerts, in answering questionnaires, in producing meeting minutes, in "not in good order" ratios. Without this "before" to measure, nobody will believe the "after," and the initiative will have no basis when budget time comes around.

Filter 5: Sophisticated buyer revealed preference. When JPMorgan, Goldman Sachs, Morgan Stanley, BNY, and BlackRock all converge on the same strategy independently, that speaks volumes about what is to come. Throughout this essay, we have privileged firm deployments and measurements over announcements, and announcements over model predictions.

2.1 The test, applied: three candidates, three verdicts

For tests to be useful, they must kill projects. Here's one, applied to three common contenders.

Candidate A: monthly FOCUS variance narratives.¹ Passes filter 1; explaining is language work. Fail filter 2; the analyst-hours are few every month, so AI would work, but not generate value greater than governance costs.

Verdict: cut. *Reasoning: hours pool too small to care, along with its close cousins: treasury commentary, HR form QC, and facilities workflows.*

Candidate B: an outward-facing client service chatbot based on generative capabilities. Passes filters 1 and 2, fails 3: outputs get committed to clients with no reviews, and Section 10 retreats file illustrates the cost of that. Candidate B fails filter 5 too: the sophisticated buyers who used it changed their minds, and those who never did so deliberately decided not to. **Verdict: out,** for the first two years, across all segments. *Rationale: every single documented example of AI reversal in this industry falls into precisely this category; the compliant near-neighbor, an assistive bot operated behind the curtain by a human being, captures almost all of the value (§11.4).*

Candidate C: DDQ/RFP response automation for the distribution desk. DDQs and RFPS are due diligence questionnaires that institutional investors send before investing in something; the response is a standing, deadline-driven process for any distribution desk. This candidate passes all five filters. The task is inherently language-native (filter 1). Hours pool is large, recurring, and growing, as more and more allocators ask questions around AI governance (filter 2). Workflow is a draft-to-review by definition (filter 3). Turnaround time is easily measurable, there is a known benchmark (filter 4): HedgeServ reduced DDQ turnaround from three days to four hours.² Finally, the buyer market has validated the category (filter 5). **Verdict: Wave 1.** *Rationale: the only*

¹FOCUS is the monthly financial and operational report a broker-dealer files with FINRA. The candidate here is an AI that drafts the narrative explaining month-over-month movements in that report.

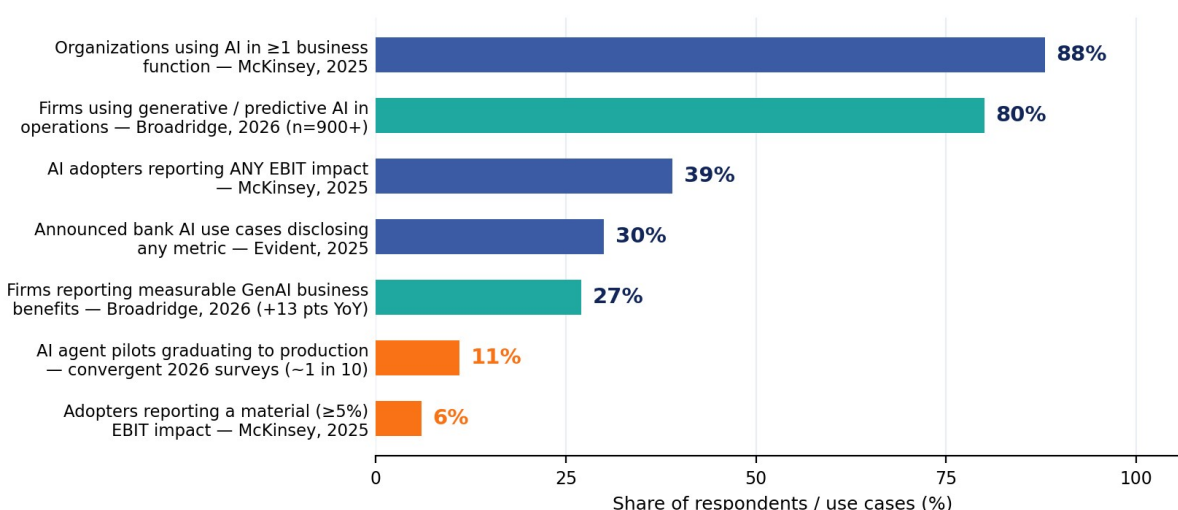
candidate of the three with a proper alignment of evidence, hours, and controls; its playbook is spelled out in §11.2.

3. Reading the deployment record

3.1 The base rates

The gap between the modeled AI value creation potential in this industry and the realized value creation potential is a few orders of magnitude, and an honest portfolio starts with this assumption. Consultancy firm valuations place the generative AI value creation potential in the banking sector in the tens of billions per year. The highest figures confirmed by banks are JPMorgan’s roughly \$1.5-2 billion of annual realized value against comparable AI expenditure and DBS’s self-reported S\$750 million-S\$1 billion. Even McKinsey’s own 2025 global survey showed 88% of organizations having some form of AI deployment, but only 39% of organizations reporting *any* AI effect on earnings, and 6% reporting a material one. Evident’s 2025 Outcomes Report³ revealed that out of 173 AI use cases reported by 47 of the 50 tracked banks, only 30% disclosed the impact metrics, and out of those only 68% were front office and IT/security-related.

Figure 2 — Adoption-Impact Funnel, 2025-2026: Near-universal adoption, scarcity of measured impact. Sources: McKinsey State of AI (2025); Evident (2025); Broadridge Digital Transformation Study (Feb 2026, n=900+); convergent 2026 agent-pilot surveys



Even the latest bank disclosures do nothing but strengthen the message. In its Q2 2026 earnings call, JPMorgan disclosed that it has roughly 1,000 AI use cases in production, and about 150,000 of its 300,000+ employees use the internal LLM every week. At the same time, Jamie Dimon warned that “you don’t uniquely benefit from AI,” since the competitors are using the same tools. Second-quarter 2026 of BNY is a case of structural proof: 7% fewer headcount compared to previous year, 13% increase in revenue to a record \$5.7 billion, with CFO calling it “a natural output, not a target.” The realized value is starting to show in the P&L, and it is showing first at the banks that built the strongest control environments.

The 2026 data shifts the funnel without tightening it. Broadridge’s sixth annual survey (February 2026, over 900 technology and operations leaders) finds 80% of companies already deploying either generative or predictive AI in their operations, compared to 31% just a year earlier, but still only 27% see tangible business impact. And the

²HedgeServ is a large independent fund administrator; it runs fund accounting and investor services for hedge funds and private-markets managers, so answering investor DDQs is core daily work there. The figure comes from a case study published by its software vendor, Loopio. And an administrator is not an asset manager: treat the benchmark as directional for a distribution desk, not a like-for-like peer.

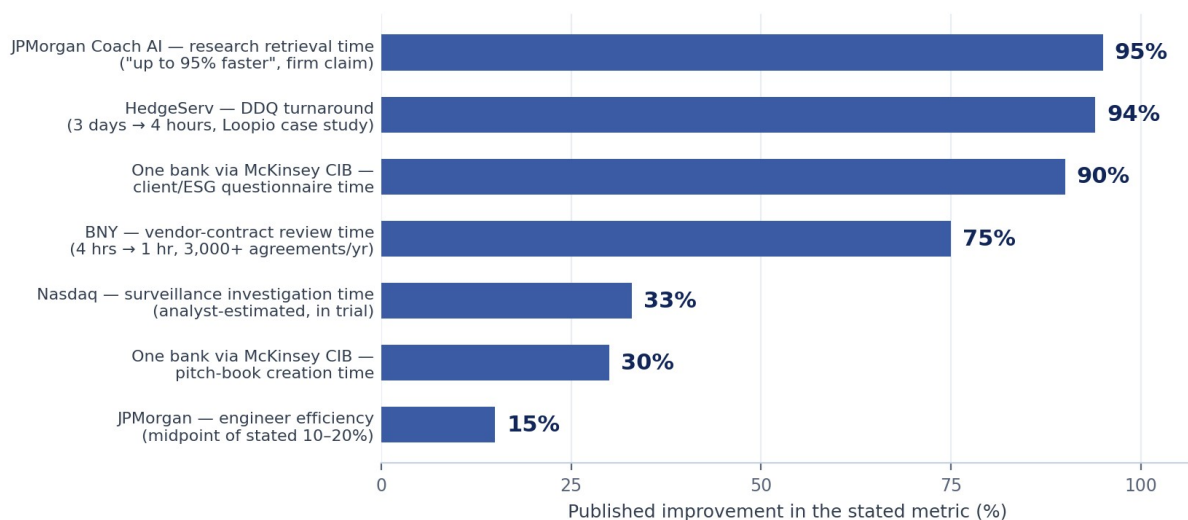
³Evident is an independent benchmarking firm that tracks AI adoption across the 50 largest banks. Its Outcomes Report is the most systematic public accounting of announced bank AI use cases that we know of.

2026 individual-agent surveys converge on around one in ten agent pilots being in production; this is the gap which Paper I starts with. A disputed finding by MIT's NANDA initiative was published in 2025⁴: ~95% of enterprise Gen-AI pilots generate no measurable P&L improvement. What matters less here is the headline, the framework underneath which remained undisputed: external vendor tools outperformed internal projects by around twofold margins, budgets tended to chase visible front-office pilots, while the higher ROI back-office work around document processing was underfunded, and the successful categories were boring things such as document automation, code generation, and call summarization. RAND has found quite similar independently. AI initiatives fail at twice the rate compared to regular IT projects for very similar reasons, predating the language models: the problem is misunderstood, the data is bad, or the technology is chased. The failure, in other words, is primarily a portfolio design error, and Naturalness Test is designed to avoid that error.

3.2 The evidence tiers

All the use cases in Sections 4-8 are labeled with two identifiers, and this section defines them; the inventory tables and Appendix A will employ them without repeating.⁵ The first identifier is the **evidence tier**. T1 denotes that some named company published a concrete outcome of the project. T2 means that the pattern is proven in production but the publicly available data is scarce or provided by vendors. T3 is a documented failure mode and we explicitly rule out all T3s from any first-two-years roadmap. A hybrid designation such as "T1-2" means that the evidence falls in between two descriptions: usually an outcome is measured and reported, but the numbers come from the vendor rather than the firm itself. The second identifier is the **autonomy level** of the staircase on Paper I (Paper I, §6): Level 1 is assisted use case, Level 2 is tool-using, Level 3 is supervised-autonomous, Level 4 is multi-agent, Level 5 is compounding. A range such as "Level 1-3" implies that a use case is deployed by various firms at various rungs of the staircase, or evolves through rungs.

Figure 3 – The measured-outcomes wall: firm-published improvements, mid-2024 to mid-2026. Every bar is a disclosure of a named institution, not a model.



3.3 Where the deployments actually sit on the staircase

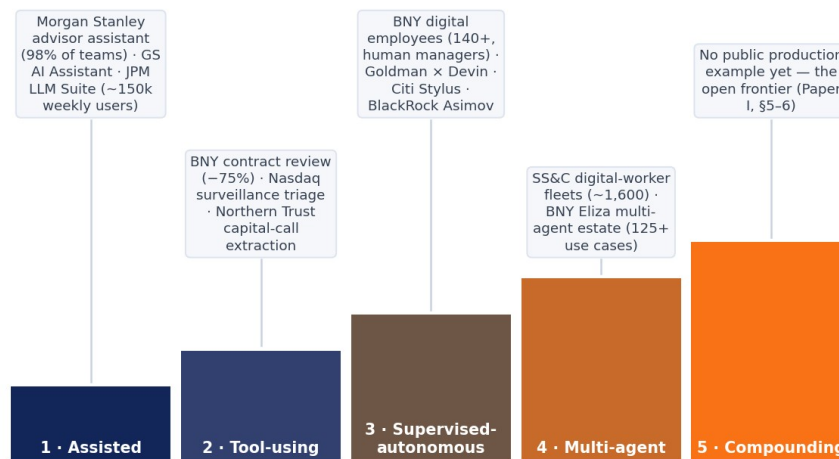
Populate the autonomy staircase from Paper I with the deployments of mid-2026 and we get what seems to be the most strategically useful picture in this paper. Levels 1 and 2 abound with measured wins. Level 3 is real but remains strictly internal: coding, research, operations. Level 4 exists at scale only within custody and fund

⁴NANDA is a research initiative at MIT. The 95% headline drew fair criticism: the failure definition counted learning pilots and vendor evaluations as failures, and the interview base was thin, which is why we lean on the report's structural findings rather than the number.

⁵"Level" is the autonomy level on Paper I's five-rung staircase and "Tier" is the evidence tier; both are defined in §3.2, including the hybrid values (a "Tier 1-2" is measured but partly vendor-sourced; a level range means firms run it at different rungs). Evidence anchors name the firm and the figure; qualifiers like "capability claim" or "analyst-estimated" mark the numbers that are not measured production results.

servicing. Level 5, in which new agents compose sanctioned capabilities from the governed graph, does not have any publicly demonstrated production deployment in financial services, the frontier is wide open. Paper I's Section 5 outlines the machinery that such a firm would need to build.

Figure 4 – Named production deployments plotted against Paper I's autonomy staircase. Everything above Level 2 is internal, reversible, and error-checkable.



The autonomy staircase from Paper I (its Figure 8), populated with the named production deployments of mid-2026

First, an implication warrants highlighting before proceeding to the segments. The top of the ladder was not attained thanks to trust in models. It was accomplished through adding more management to each agent: identity, bounded scope, named accountable individual, performance review. BNY's digital workers get personal system accounts and about a hundred human managers are reviewing their performance. This is the control plane of Paper I (Paper I, §3). The market has, effectively, already conducted the experiment recommended in Paper I, and has published its findings.

4. Investment banking

4.1 Where the hours are

Cost base of an investment bank is dominated by two pools of interest for the Naturalness Test. The first is front-office compensation: Goldman Sachs' cost of compensation for 2024 amounted to about \$16.7 billion against revenue of \$53.5 billion, and average for top-14 banks amount to about \$11.3 million net revenue per front-office employee, according to Deloitte. The second is technology, which consumes 13–26% of headcount. JPMorgan has more than 40,000 engineers, and Goldman has 12,000 developers among 46,000 employees. Within the front office, junior banker's time is the most condensed automatable resource. First-year analysts work 95–100 hours weekly, and much of it is spent on pitches, comparables, diligence, and documentation preparation. This is Filter 1 work, almost by definition.

4.2 The use-case inventory

This is the first of five segment inventories, so a brief explanation of the reading of these should be provided. The same format will be used throughout Sections 5–8.

Use case	Level	Tier	Evidence anchor
Firmwide knowledge assistant over	1	1	GS AI Assistant firmwide (2025); JPM LLM Suite —



Use case	Level	Tier	Evidence anchor
proprietary content			~150,000 weekly users and ~1,000 use cases in production (Q2 2026 call); Citi Assist / Stylus, ~175k staff
Research retrieval & synthesis for sales and trading	1–2	1	Morgan Stanley AskResearchGPT (~3x prior-tool engagement); JPM Coach AI — retrieval up to 95% faster; 20% AWM gross-sales lift partly attributed
Deal & offering document drafting (S-1s, memos, pitch books)	2	1	Goldman CEO: AI drafts “95% of an S-1 in minutes”; Rogo (a finance-specific AI drafting and analysis platform) at Jefferies, Lazard and Baird — Baird reports 95% engagement across Global IB, ~10,000 workflows/week, top users saving 5+ hrs/day (vendor-hosted, firm-attributed); ~30% pitch-book time cut at one bank (McKinsey CIB study)
Legal & vendor contract intelligence	2	1	JPMorgan COiN: ~360,000 hours/yr of lawyer and loan-officer work automated since 2017; BNY contract review –75%
Master-agreement lifecycle (ISDA / CSA / GMRA): term extraction & effective-dated amendment tracking	2	2	ISDA Create (Linklaters' CreateIQ) captures negotiated terms as structured data, integrated with S&P Global's Counterparty Manager (2023); the IBOR-transition repapering wave proved document AI at portfolio scale
Meeting & call capture into CRM	2	1	MS Debrief — one adviser reports ~30 min/meeting; BofA Meeting Journey — up to 4 hrs/meeting (capability claim)
Software engineering acceleration	1–3	1	JPM 10–20% engineer efficiency; Citi ~100k developer-hours freed weekly; MS DevGen.AI 280k hours; Goldman piloting Devin, Cognition's autonomous coding agent
M&A / client screening at machine scale	2	2	UBS: 300,000 companies scanned in ~20 seconds
ESG / client questionnaire completion	2	1–2	~90% questionnaire-time cut at one bank (McKinsey CIB study)
Research production & distribution leverage	2	2	Deutsche Bank DB Lumina; UBS analyst avatars (video output 1,000 → 5,000/yr target)
Post-trade commentary & desk-supervision narratives	2	2	Summarization is FINRA's top-observed use case; keep AI off the order path
Client-facing autonomous advice or execution agents	5	3	No production precedent; regulatory posture adverse (Section 10)

There are two observations worth making here. **The first one is that junior banker's workbench has turned from a science project into a procurement decision.** Rogo claims Jefferies and Lazard as its customers. OpenAI hires more than 100 investment bankers in 2025 to develop its drafting and modeling capabilities for financials. JPMorgan is open in discussing compressing junior-to-senior ratio from 6:1 toward 4:1, and Goldman's October 2025 OneGS 3.0 memo links limited headcount expansion directly to AI efficiency. The difference now is not in access to the capability, but in the quality of control around it: versioned prompts, citation discipline, supervisory review that could withstand an SEC comment letter.

Second, **engineering is the quiet compounder**. The hardest numbers in this whole industry are developer numbers, and gains compound into every other initiative. Reporting pipelines come out faster, data platform cleaning becomes easier, AI program progress accelerates. We find it odd how many AI strategies fail to address the engineering function at all; it is debating the second largest pools while ignoring the largest one. The most recent datapoint raises the bar even further. On its July 2026 earnings call BNY announced that about 40% of software developed by its engineers use AI tools, within an engineering budget where AI spending is “de minimis.”

There is one particular case of contract intelligence that calls for a brief treatment here: **master-agreement lifecycle management**.⁶ Unlike other contractual agreements, ISDA, CSA, and GMRA documents are amendable continuously over the lifetime of their contractual relationship that could stretch through many decades, with each amendment governing the relationship only from the effective date onwards. Instead of asking “What does the agreement say?”, the business needs to know “Which terms were governing this trade on this date?”, and doing so involves knowing the effective dated terms and not just the last version of the agreement. The division of labor aligns very neatly with the Naturalness Test: AI extracts the negotiated and amended terms, the effective date of the amendment, and the deltas from the previous version; while a deterministic, date-versioned terms store knows the timeline and answers the “as of” question. The margin call, close-out calculation, and collateral eligibility queries read from this store and never from the model. The industry has already proved both halves: ISDA Create captures negotiated terms as structured data and feeds them to S&P Global’s Counterparty Manager, while the IBOR repapering rush showed document AI processing thousands of old agreements at once. What is missing is the effective dated terms store in between, and that is the domain of databases, not AI, and much cheaper to acquire than the latter.

What to do first. Start with research retrieval and meeting capture, as they are revenue-adjacent, Tier 1 projects with low integration complexity. Run the engineering program in parallel as a controlled initiative. Pilot the deal document drafting within a single product group and with document-level baselines, before letting it spread.

5. Broker-dealers

5.1 Where the hours are

The opportunity of applying AI to control and operation function at a broker-dealer emerges from the confluence of document volumes, alert volumes, and regulatory consequences there. Cost pools are well quantified. Global financial crime compliance exceeds \$206 billion a year and \$61 billion a year in the US and Canada alone; and it is labor driven; 99% of firms report increasing costs (LexisNexis). The penalty side is quantified as well. Finra’s Consolidated Audit Trail enforcement already includes Instinet (\$3.8 million), Goldman Sachs (\$1.4 million), Citadel Securities, and IMC.⁷ A 2% settlement fail rate is projected to cost the industry up to \$3 billion a year.⁸ And the regulator is uncharacteristically direct about what it expects generative AI to do. Finra’s 2026 Annual Regulatory Oversight Report identifies summary and information extraction as the dominating member-firm use case; Finra itself applies AI to hundreds of billions of daily market events and uses generative AI to create summaries out of its 10,000+ annual investor complaints.

⁶ISDA Master Agreements govern OTC derivatives relationships (with the CSA covering collateral terms) and the GMRA (ICMA’s Global Master Repurchase Agreement) governs repo. Both sides of the street sign them: dealers, asset managers, and their counterparties alike.

⁷The characterization of the Instinet settlement as FINRA’s first significant CAT reporting enforcement action is law firm Katten’s, in its client alert on the case.

⁸The \$3 billion figure originates in a 2016 TABB Group study and is cited by DTCC in its settlement-fails material; treat it as an industry estimate, not a measured annual figure.

5.2 The use-case inventory

Use case	Level	Tier	Evidence anchor
Surveillance alert triage & investigation narratives	2	1	Nasdaq: 33% investigation-time reduction (analyst-estimated, in-trial); platform AI serving ~50 exchanges and 20 regulators; Nasdaq Verafin agentic workforce adopted by 650+ FIs — vendor-claims up to 90% sanctions-alert workload reduction; NICE SURVEIL-X; Behavox Quantum live at BNY
AML alert triage & SAR drafting support	2	1	HSBC × Google Cloud: 2–4x true positives, ~60% fewer alerts, against the \$206B cost pool
Reg-reporting reject triage & repair drafting (CAT / TRACE)	2–3	2	CAT fines as penalty-side quantification; FINRA Rule 6893 accuracy and repair-deadline duties; Paper I works this exact agent end to end (Paper I, §4)
Enforcement & regulatory-change intelligence	2	2	Commercial category proven: Corlytics; CUBE (acquired Thomson Reuters Regulatory Intelligence, 2025); AscentAI
Membership & periodic due-diligence lifecycle with year-over-year comparison	2	1–2	FINRA Rule 4517 annual verification; OCC / CME / NSCC member reporting duties; HedgeServ benchmark — DDQ 3 days → 4 hours
Onboarding & KYC document AI; periodic-review diff	2	1–2	63% of custodians already use Gen AI in onboarding/KYC (Citi survey, n=537); McKinsey agentic-KYC direction
Complaint intake classification (4530) & response drafting	2	2	FINRA itself Gen-AI-summarizes complaints at scale — the regulator validated the pattern on its own volume
Reconciliation & settlement-fail triage with seg-risk flagging	2	2	Established reconciliation-software vendors (SmartStream, Duco, Gresham) now ship this; 47% of custodians testing AI reconciliation; 15c3-3 makes stock-record breaks regulatory
Corporate-actions capture & election-deadline risk	2	2	ISO 20022 transition live (Fedwire 2025; Swift MT sunset Nov 2025); SmartStream Smart Agents “proven across Tier 1 pilots”
Best-execution committee packs & 605/606 exception narratives	2	2	Deterministic data, AI narrates; summarization is FINRA's top-observed use case
Client-facing generative brokerage chatbot	5	3	Documented failure mode (Section 10)

Triage is one engineering problem presented in three uniforms: surveillance alerts, AML alerts, and post-trade exceptions are one and the same underneath. A deterministic system produces the items; AI clusters them, describes the root cause in English, and recommends an action together with evidence citations; the human acts on it. Those who grasp this, build the pattern once, on Paper I's spine, and reuse it for compliance and operations. Those who do not find themselves having to buy it three times from three vendors.

The reporting remediation white space is documented. Only 24% of firms have generative AI deployed in post-trade processes compared to 53% in others (Citi Securities Services Evolution 2025, 537 respondents); fines are real, the pattern is established in adjacent domains, and productivity numbers cited in conversations

are model numbers provided by vendors. That means the first firm publishing their actual measured baseline numbers will have a unique claim to credibility with its regulator and its board. Paper I's worked example (Paper I, §4) is the engineering blueprint for this exact agent. What this paper contributes to it is the market context: nobody has claimed this metric yet.

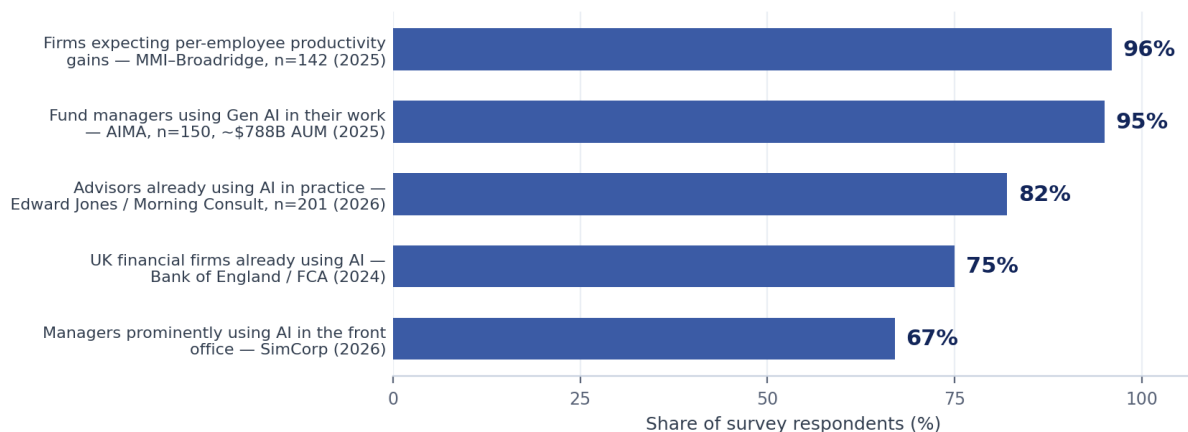
Where the quality improvement process should begin. Surveillance or AML triage first, where the largest measurable quality gains intersect the largest cost center. Remediation of reporting breakdowns is quick whenever the firm's own repair backlog is documented. Enforcement intelligence makes a good starting point because it runs entirely off of live, public data with no confidentiality concerns whatsoever.

6. Asset managers

6.1 Where the hours are

Asset management comes into play in the context of fee pressure and an unusually steep adoption curve. In a September 2025 survey conducted by AIMA, the Alternative Investment Management Association, of 150 fund managers (around \$788 billion of AUM), 95% reported use of generative AI in their processes. But the two numbers that truly count are the ones on the other side of the equation: **29% of surveyed investors included generative-AI questions in their due-diligence questionnaire, and 60% reported increased likelihood of investing in a fund with an AI R&D budget.**⁹ In other words, AI proficiency became a criterion for investment. The general surveys are telling the same story. A survey (2025) conducted by Money Management Institute–Broadridge found 96% of firms expecting per employee productivity increase, while SimCorp reports that more than two-thirds of investment managers are extensively using AI in front-office operations. BCG's 2026 asset-management report models agency-oriented workflows increasing investment operations capacity by 55-65% and reducing costs by roughly 40%. This is a model rather than a measurement, but the custody evidence from Section 7 makes it hard to dismiss.

Figure 5 — Industry surveys' view on adoption. Sources and samples are provided as labels. Edward Jones figure reflects an industry survey rather than individual firm adoption.



6.2 The use-case inventory

Use case	Level	Tier	Evidence anchor
RFP / DDQ response automation for distribution	2	1	Fund administrator HedgeServ, via RFP-response software Loopio: DDQ 3 days → 4 hours, RFP first drafts 7 days → 2 days; Ontra (private-markets legal

⁹The investor-side questions in AIMA's study were answered by a separate sample of just 18 institutional investors, so treat these two percentages as directional rather than precise.



Use case	Level	Tier	Evidence anchor
			automation), 1M+ documents processed; demand rising — 29% of allocators embed AI questions
Research synthesis & earnings-call digestion at scale	2–3	1–2	BlackRock Asimov — agentic overnight research analyst (2025); AllianceBernstein ABAI gateway, 10–50% time savings cited on operational processes
Private-markets deal-memo drafting	2	2	Schroders Capital GAiiA in production since June 2024
Alternative-asset document processing (capital calls, distributions, K-1s)	2	1–2	Northern Trust in production since 2021; Canoe Intelligence — 44,000+ funds, vendor-claimed 94% retrieval-time cut; Bloomberg agreed to acquire Canoe (announced Jul 2026, pending approvals); BlackRock paid ~\$3.2B for Preqin, the private-markets data provider, to own the data layer
Thematic & idea screening	2	2	BlackRock Thematic Robot: theme → stock basket in ~1 hour versus weeks
Client & advisor content with automatic disclosures	2	2	Vanguard client-ready summaries for its 150,000-advisor network (launched in beta); Fidelity Saifr × Microsoft compliant-content agent
Portfolio commentary drafting	2	2	Aladdin Wealth AI commentary tool; Morgan Stanley first implementer (2025)
Distribution intelligence & meeting prep	2	2	Franklin Templeton Intelligence Hub with Microsoft (2026); Vanguard Expert Insights (2026)
ML-driven strategy research / signal generation	3	2	Bridgewater AIA Labs \$2B fund; Man Group agentic quant signals — errors caught by PMs, not clients
NAV oversight & anomaly detection	2	2	Temenos NAVigator; Milestone pControl live at Waystone and Artemis — classic ML; NAV computation stays deterministic
Autonomous client-reporting or trading agents	5	3	No production precedent in traditional asset management; excluded

Tier 1's dark horse is **distribution** desk. RFP and DDQ handling is a function where all of the work, growth potential, and evidential value come together. It expands as investors ask their own AI governance questions. The benchmark is stark: three days reduced to four hours. Its supporting infrastructure, an approved content library with citations and review gates, is a small but representative case for Paper I's action catalog hypothesis. There is even some compounding effect available here. When a manager has automated its DDQ desk and can handle an investor's AI governance questions straight from its own audit log, it has converted its cost project into capital-raising tool.

Private markets are being redesigned around document AI in terms of industry structure. Blackrock acquired Preqin in a deal worth \$3.2 billion. In July 2026 Bloomberg entered into agreement to buy Canoe. Goldman Sachs led growth funding for 73 Strings, an AI valuation and monitoring platform for private markets. Three acquirers, three bets on one thing: whoever organizes documents of private market ends up controlling its workflow. If you manage private investments portfolio, capital call and distribution extraction have become table stakes.



Where we would start. Start with the automation of RFP/DDQ, which is measured to have the fastest payback and is visible to allocators. Follow with the extraction of alternative documents. Limit the research synthesis to the earnings season to limit the scope. And tackle content generation with the compliance gate incorporated into the process from the get-go.

7. Custodians and fund servicers

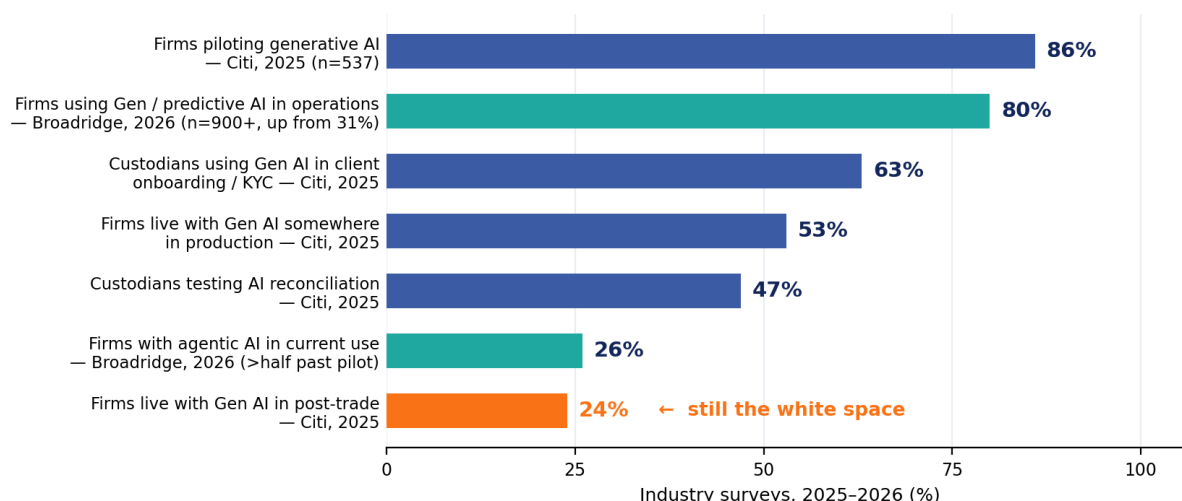
7.1 The segment that proves the thesis

It turns out that custody has quietly emerged as the leader in applying agentic AI, and it's not hard to see why. It does high volume, internal, document-driven, and error-checkable work. This happens to be the sweet spot that Paper I's staircase allows crossing safely beyond Level 2. Custody also has the most advanced publicly available case study of financial services. It would serve operational leaders of other segments well to study it.

On paper, **Eliza** from BNY Mellon is the most advanced deployment of enterprise AI in the industry. More than 20,000 BNY employees actively develop agents there. More than 125 use cases operate in production using frontier models from multiple providers, and its contract-review assistant reduced legal review from four to one hour, saving 75%, on 3,000+ annual vendor contracts. The most instructive aspect about this is not technical but organizational: BNY operates **140+ digital employees with their own system logins, roughly two dozen skills each, and around 100 human managers who perform their performance** reviews, on payment-instruction validation, data remediation, and code security tasks. This is Paper I's control plane designed as an org structure. Agent identity (Paper I's "know your agent"), limited scope, clear accountability, reviews, all done at production scale, and the CEO is on record that this program is not about reducing headcount. Other companies follow the same path. **SS&C** runs 1,600+ digital workers in its transfer agency and fund administration business lines using the Blue Prism platform it acquired for roughly \$1.6 billion. **State Street** put a natural language layer on top of its Alpha platform and in its 2025 annual report says it is deploying agentic AI on multi-step processes. Swift completed federated-learning fraud pilots with 13 banks producing a model roughly twice as good at detecting known fraud compared to any of the participants' single organizations. It is the first credible template for cross-institutional AI.

The white space appears in the survey conducted by the industry itself. Securities Services Evolution 2025 (537 respondents) showed 86% of the firms piloting generative AI but **only 24%** in operation in post-trade, while 63% of custodians are already using AI for onboarding and KYC and 47% are testing AI reconciliation. Post-trade operations remains the largest under-penetrated document and exception pool in the industry. The 2026 data confirms that this gap is persisting despite the rapid adoption of AI in other places: Broadridge's February 2026 survey finds 80% of the firms deploying generative or predictive AI in operations and 26% with agentic AI deployment, more than half of them past the pilot stage, while no corresponding increase has been recorded for post-trade production use.

Figure 6 – Post-trade white space, 2025-2026: Adoption continues to grow (teal = Broadridge 2026), but post-trade production remains the laggard. Citi has not yet published 2026 version of its survey.



7.2 The use-case inventory

Use case	Level	Tier	Evidence anchor
Contract & agreement review	2	1	BNY: -75% review time across 3,000+ agreements/yr
Digital-worker operations (payment validation, data repair, code security)	3-4	1	BNY 140+ digital employees with human managers; SS&C ~1,600 digital workers; Forrester TEI on Blue Prism: 330% three-year ROI
Client onboarding / KYC document processing	2	1-2	63% of custodians already using Gen AI here (Citi survey)
Alt-asset document intake (capital calls, distributions, statements)	2	1-2	Northern Trust production since 2021; Canoe scale validation
Natural-language data access over the servicing platform	1-2	2	State Street “ask your assets anything” on Alpha; 400 AI engineers
Reconciliation & exception management	2	2	47% of custodians testing; SmartStream / Duco / Gresham category
Corporate-actions capture & election processing	2	2	SmartStream Smart Agents (2026); no custodian-published accuracy numbers yet — the first mover owns the metric
Cross-border payment fraud detection (federated)	2	2	Swift-led pilots, 13 banks: ~2x detection effectiveness (synthetic-data experiments) ¹⁰
Transfer agency & shareholder services automation	3	2	SS&C GIDS among the heaviest digital-worker deployments
Autonomous settlement or payment release	5	3	No precedent; blast radius disqualifying

The place to begin. Contract analysis and alternate document processing as our first proofs because they are document-centric and have precedents for measurement. Reconciliation and corporate actions focused on the

¹⁰Participants included BNY, ANZ and Intesa Sanpaolo, with Google Cloud among the technology partners. The two-times result was measured in pilot experiments on synthetic transactions, not yet on live flows.



24% post-trade gap. Our recommendation for year two is to use the BNY identity and manager approach, not to develop one. The build sequence is the 90-day Paper I approach (Paper I, §6).

8. Private wealth and wealth management

8.1 Where the hours are

Scarcity of resource in wealth management lies in advisor's time; it comes as no surprise that AI evidence in this sector is the most actionable of the five: meetings per minute, plans per document, days per account opening. In fact, in this particular field the distinction between assist and replace is the most clearly defined, simply because the relationship *is* the product.

8.2 The use-case inventory

Use case	Level	Tier	Evidence anchor
Advisor knowledge copilot	1	1	Morgan Stanley: 98% advisor-team adoption over ~100k documents; UBS Red over 60,000 documents; JPM Coach AI — 95% faster retrieval, 20% AWM gross-sales lift partly attributed; LPL × Anthropic — Claude tools rolling out to 30,000+ advisors (Feb 2026); Ameriprise \$1B AI-centered tech spend; Raymond James “Raimond”; Edward Jones—commissioned industry survey: 82% of advisors already use AI (n=201, industry-wide)
Meeting prep, capture & follow-up	2	1	BofA Meeting Journey (2026): up to 4 hours saved per meeting across millions of meetings (capability claim); MS Debrief — one adviser reports ~30 min; Raymond James × Zoom AI; funding wave in advisor meeting-AI vendors — Jump \$80M, Zocks \$45M Series B
Estate & trust document extraction	2	2	Wealth.com \$65M Series B (“Ester” doc AI); Vanilla \$35M
Tax-return-driven planning automation	2	1–2	Holistiplan: 39% share in the T3 advisor-software survey; 2,100-firm XYPN partnership — advisors demonstrably pay for document AI
Account-opening quality — reducing “not in good order” (NIGO) paperwork	2	2	Docupace vendor claim: ~90% errors; no named-firm AI figure — instrument your own
Next-best-client / organic-growth prediction	2	2	RBC Wealth Management × TIFIN AG (an AI client-growth vendor); LPL advisor-growth tools
Wealth-ops service agenting (requests, document chase)	3	2	Raymond James proprietary AI operations agent + Client 360
Retail-scale deterministic assistants	2	1	BofA Erica: 3.2B+ interactions, an ~11,000-FTE equivalent — deliberately non-generative; Wells Fargo Fargo: 1B+ interactions in under three years (Mar 2026), no PII to the LLM
Client-facing generative agents	5	3	Klarna and CBA reversals; Citi “risky for customer support”; Schwab's client-facing assistant went live June 2026 — no results published yet

Meeting intelligence provides the fastest ROI on wealth, and the figures speak for themselves. Four hours per meeting, max: that's what Bank of America claims about all those millions of meetings per year, and no financial organization has committed so much advisor time before. Something else we believe is undervalued: the compliance ROI. The systematic capture, permissioned and organized record of every client meeting is a compliance resource, too.

The two most successful “chatbots” in the business are non-generative in nature. Erica's 3.2 billion interactions and Fargo's more than a billion interactions were accomplished through deterministic, highly routed NLP, and personal data was separated from the model. The companies that conduct the most customer conversations took control instead of going with generation, and this is the most concrete data point that can tell you about the location of the client-facing line today. Schwab is putting the line to the test right now. Schwab's first investor-focused AI assistant went live in June 2026, and voice agents and portfolio insight features are being introduced throughout the year. We will not say anything else until some data is shared regarding adoption and quality.

Where to begin. Starting with intelligence. The intelligent copilot where there is already a proprietary library of content. Extracting tax and estate documentation for planning-driven firms. Also, account opening quality based on the firm's own NIGO benchmark and not a vendor's marketing number.

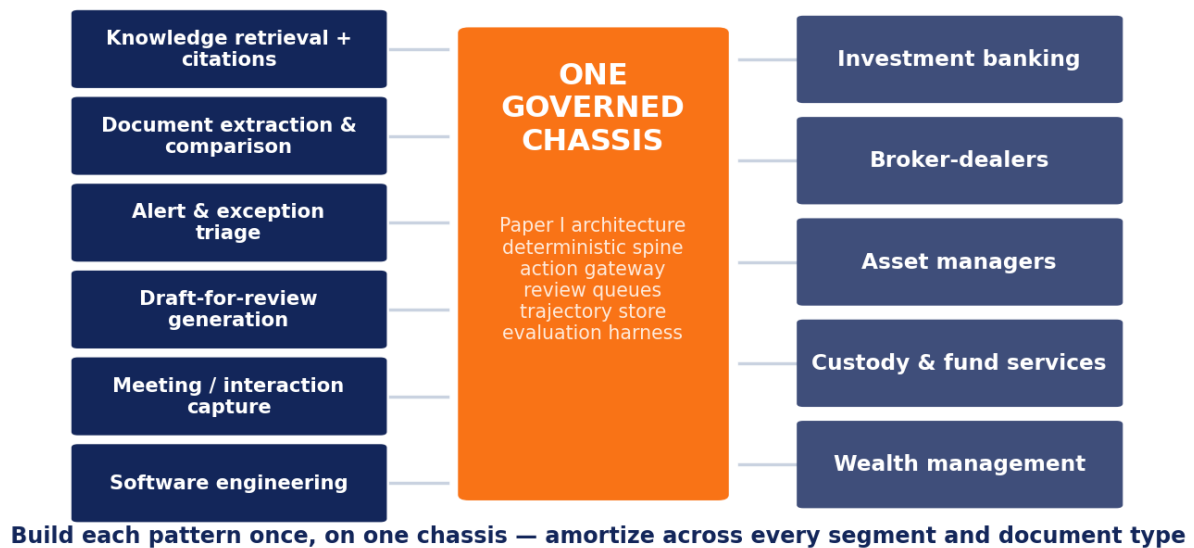
9. The cross-segment value map

Six patterns drive almost all of the value measured in every segment, and one anti-pattern crops up in every segment as well. Here are three key takeaways from the map. First, while **the patterns themselves are portable, the documents are not**. The same extraction engine will do for the S-1, the DDQ, the capital call, and the trust document, which is why having one platform per pattern makes much more sense than having one tool per department. That is the economic rationale underlying Paper I's action catalog (Paper I, §2.4, §5). Second, the **Tier 1 density is greatest where the content is proprietary and reviewable** (research corpora, contracts, meetings) and least where the outputs are destined for clients. Third, **the Level 3–4 governance of tomorrow is being built out in custody** today. Its governance model will migrate to the rest of the segments from custody, not the other way around.

Figure 7 — Six patterns × five segments, scored by evidence tier. The last row is the only Tier 3 in the industry.

	Investment banking	Broker-dealers	Asset managers	Custody & fund services	Wealth management
Knowledge retrieval + citations	T1	T2	T1-2	T2	T1
Document extraction & comparison	T1	T2	T1	T1	T1-2
Alert & exception triage	-	T1	T2	T2	T2
Draft-for-review generation	T1-2	T2	T2	T2	T1
Meeting / interaction capture	T1	-	T2	-	T1
Software engineering	T1	T1	T1	T1	T1
Client-facing autonomous agents	T3	T3	T3	T3	T3
	Tier 1 — firm-measured, published	Tier 2 — proven pattern, weak public measurement	Tier 1-2 — measured, partly vendor	Tier 3 — documented failure mode	

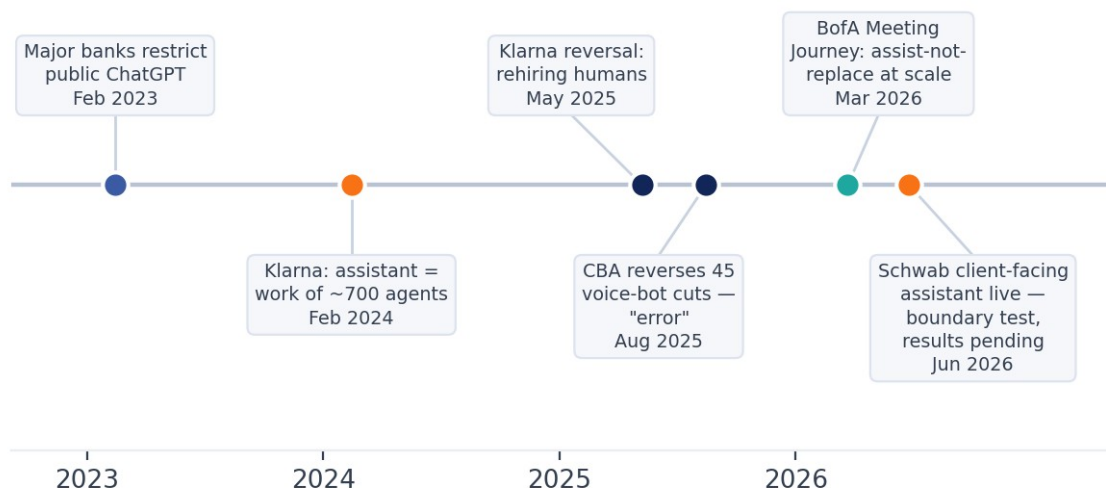
Figure 8 — Six patterns, one chassis, five segments: the platform economics driving the map. The chassis is Paper I's reference architecture.



10. The retreat file: what fails, on the record

There have been enough public reversals now that failure itself can serve as a data point, not just a rumor. Four stories tell the tale.

Figure 9 — The reversal timeline: expansion narratives (orange), reversals and refusals (navy), and the boundary test in progress.



Klarna (2024–2025). The most oft-cited case study of AI-first customer service showed their AI doing 66% of all customer-service chats, the work of about 700 human agents, leading to \$40 million in savings. But after 15 months, the CEO admitted the cost-first strategy had resulted in “lower quality” service, and Klarna hired back humans even as it retained the AI to do what it did best. Lesson: The ceiling for automation in customer contact is real, and breaching it makes a cost gain a brand cost.

Commonwealth Bank of Australia (2025). CBA eliminated 45 jobs in its call centers based on the voice bot. Call volume went up. The bank reversed the job cuts and called it a “mistake,” the most candid admission by any

major bank that an AI replacement attempt had failed. Lesson: Substitution decisions based on pilot scale evidence cannot survive production scale variance.

The quiet refusals. Citi executives have publicly stated that generative AI poses too much risk for direct customer service. And the owner of the most successful automation tool in retail banking, Bank of America, designed Erica to be deterministic. The institutions with the largest amount of customer data set the boundary on purpose, and it deserves more respect.

Regulatory codification. The FINRA report from 2026 embraces the extraction approach and mentions the agentic risks (scope, auditability, and leakage) in the same breath, and its effective practices would read like a control specification: prompt and output logging, human-in-the-loop reviews, vendor management, Rule 3110 applying in full. The SEC's abandonment of its predictive analytics proposal in 2025 leaves us with technology-neutral conduct rules, putting the burden on each firm's governance (i.e., Paper I architecture). The additions in 2026 all move in the same direction. The US Treasury's framework on AI risk management in financial services (February 2026; 230 control objectives), IOSCO's final Supervisory Toolkit (May 2026), and the second cohort of FCA's AI live testing¹¹ are all institutionalizing supervised AI rather than restricting it.

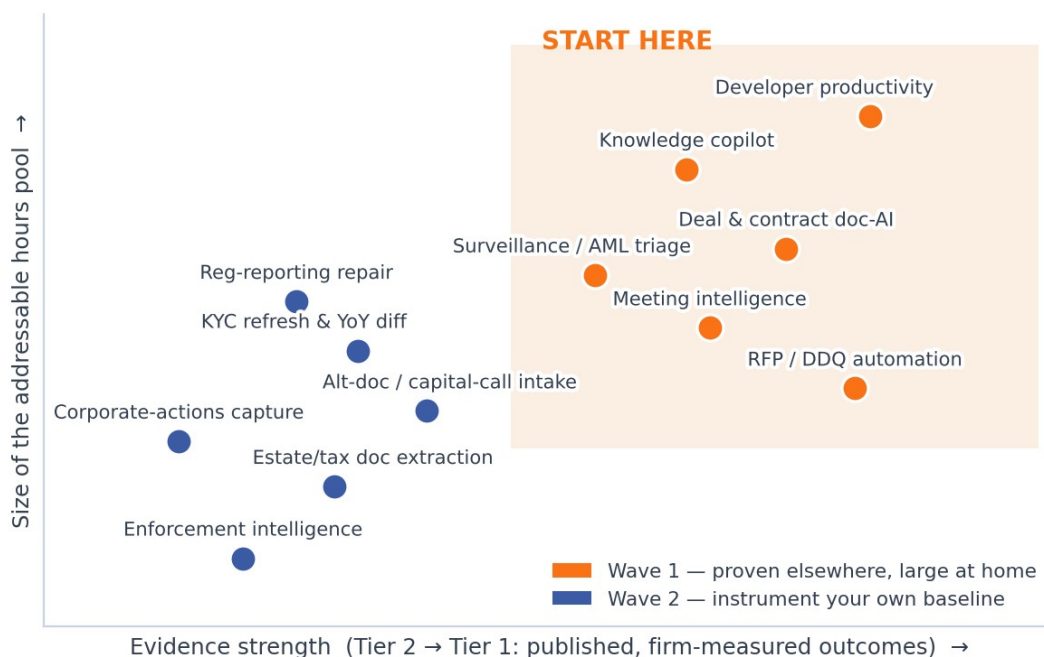
What the failure file contains: *Not* one documented reversal of an internal, human-gated deployment. All failures come from beyond the Naturalness Filter 3. External failure on one side, internal compounding on the other. Among all things in this series, that asymmetry might be the most persuasive empirical argument.

11. Constructing the portfolio

11.1 The selection method

Sections 4-8 gave the menu. This is the selection method. Rate each candidate on two dimensions: the quality of the **evidence** (its tier, emphasizing firm-level measurement) and **the size of the hours pool, Filter 2, that is at your institution**. Top right corner is Wave 1. All else will have to wait, or die.

Figure 10 — Portfolio construction: evidence strength x hours pool. Wave 1 (orange) is proven elsewhere and large at home; Wave 2 (navy) is when your own baseline becomes the differentiator.



¹¹Cohort two, announced April 2026, includes Barclays, Lloyds/Scottish Widows, UBS, Experian and GoCardless, with use cases spanning AML/KYC and agentic payments; the FCA has committed to a "good and poor practice" report later in 2026.



To convert the quadrant from a visual into a decision-making framework, rate each candidate 1-5 on the criteria below, weight and rank according to weighted score. The weights are designed to reflect this paper's logic, with external evidence and internal hours pools being critical. No row is reserved for strategic appeal, and that is a deliberate omission.

Criterion	Weight	How to score it
External evidence strength	25%	5 = named-firm published outcome (Tier 1); 3 = production precedent, vendor-measured; 1 = modeled only
Addressable hours pool	25%	Measured during scoping: FTE-hours per month in the target queue × loaded cost. 5 = top decile of your candidate list
Error tolerance with review	15%	5 = pure draft-for-review; 1 = output would reach a client, market or regulator without a gate — automatic disqualification
Baseline measurability	15%	5 = metric already tracked (alert-handling time, NIGO rate, turnaround); 1 = metric would need to be invented
Data & integration readiness	10%	5 = documents/queues already digital and accessible; 1 = data program required first (Paper I, §2.11)
Reuse of prior build	10%	5 = composes actions and controls an earlier agent already proved (Paper I, §5); 1 = greenfield everything

Wave 1: proven elsewhere, large at home. Copilot for knowledge, meetings intelligence, deal and contract document AI, RFP/ DDQ automation, surveillance and AML triage, developer productivity. Each has its firm-level, published benchmark against which it can be assessed. The logic is the same across all of these: first capture the baseline, then apply it to Paper I's spine, and publish the internal benchmark. In other words, it is Paper I's first-90-days sequence (Paper I, §6), targeted at use cases that have the highest level of external evidence.

Wave 2: instrument your own baseline. Reporting-break remediation, alt-document intake, KYC refresh, corporate actions capture, estate and tax extraction, enforcement intelligence. Of these, two are conditional: alt-document intake and onboarding/ KYC processing qualify as Wave 1 use cases for custodians and alternatives-heavy portfolios, as peer adoption is already published (see Appendix A.4). For the remaining use cases, the patterns are proven but the numbers are not public or come from vendors. We would consider this a good sign, as the first institution to publish its results in this space has a reference point. Coincidentally, Paper I's worked example belongs to Wave 2.

11.2 Wave 1 playbooks: scope, controls, baseline, benchmark, exit

Each use case from Wave 1 collapses into a set of five decisions, all of which determine whether the initiative goes forward: What specifically is to be built first, the control set within which it has to operate (all Paper I cross-references), the baseline to capture prior to applying any model to the process, the external benchmark against which the initiative is to be held, and the 90-day exit criterion that determines whether it lives.

Use case	First scope (build exactly this)	Baseline to capture	External benchmark	90-day exit criterion
Knowledge copilot	One corpus (research, policies, or procedures), retrieval-only with mandatory citations and	Time-to-answer for the 20 most common questions; % answered without	JPM Coach AI: retrieval up to 95% faster; MS: 98% team adoption	≥60% weekly active usage in the pilot team; zero uncited answers in sampled



Use case	First scope (build exactly this)	Baseline to capture	External benchmark	90-day exit criterion
	entitlement filters; no open generation	escalation		audit
Meeting intelligence	One desk or advisor pod; consented transcript → summary, actions, CRM draft, flags; no auto-send	Minutes per meeting on notes + CRM entry; CRM completeness rate	BofA: up to 4 hrs/meeting (capability claim); MS Debrief: ~30 min (adviser-reported)	≥50% of pilot meetings processed; measured minutes saved ≥30/meeting; all drafts human-approved
Deal & contract doc-AI	One document class (e.g., vendor agreements or NDAs): extraction + clause-risk flags to reviewer	Review hours per document; backlog age	BNY: 4 hrs → 1 hr across 3,000+ agreements/yr	≥50% review-time reduction on the pilot class at equal-or-better issue-catch rate
RFP / DDQ automation	Approved content library + drafted answers with coverage labels (library / adapted / no-source); reviewer gate	Turnaround days per DDQ; % of questions answered from library	HedgeServ: DDQ 3 days → 4 hrs; RFP drafts 7 → 2 days	Turnaround cut ≥50% on two live questionnaires; every shipped answer citation-backed
Surveillance / AML triage	One alert type; AI recommends disposition + drafts narrative; analyst disposes; no auto-close	Minutes per alert; alerts per analyst-day; escalation precision	Nasdaq: ~33% investigation time (analyst-est.); HSBC: 2–4x true positives, ~60% alerts	≥25% triage-time reduction with zero missed true positives on a replayed golden set
Developer productivity	Coding assistants to one platform team + measurement harness (PR cycle time, review load)	PR cycle time; story throughput; % AI-assisted commits	JPM: 10–20% efficiency; Citi: ~100k hrs/week	Measured cycle-time gain ≥10% on the instrumented team; expansion plan costed

There are three implementation rules that apply to every row. The first rule: The baseline is captured **before** the model is switched on, by the operations team, in the operations team's own metric. Anything else doesn't cut it. The second rule: Each playbook ends at Paper I's review-queue and trajectory store controls (Paper I, §2.3, §3). In our experience, the playbook which “temporarily” misses out on those controls never comes back. The third rule: The exit criterion is binding both ways. Hit it and you get funding for further expansion; miss it and the initiative is terminated, in writing. Nothing beats this for giving the entire portfolio credibility with a CFO.

11.3 The measurement dictionary

When firms cannot demonstrate the functionality of their AI, the problem usually does not reside in the modeling. Much more likely, there was no agreement on the metric prior to the deployment. Six patterns, six metrics. Nail them down during scoping rather than review.

Pattern	Canonical metric	Measurement method
Knowledge retrieval	Time-to-answer; deflection rate	Timed task panel pre/post on the same question set; sampled citation audit for groundedness
Document extraction	Minutes per document; field-level accuracy	Golden set of ≥20 hand-labeled documents; field precision/recall re-run on every prompt or model change
Exception triage	Minutes per item; recommendation	Replayed historical queue with known outcomes; zero-

Pattern	Canonical metric	Measurement method
	agreement rate; missed-true-positive count	miss constraint on true positives as a hard gate
Draft-for-review	Draft acceptance rate; edit distance; turnaround time	Reviewer decisions logged in the queue (accept / edit / reject); before-after turnaround on live volume
Interaction capture	Minutes saved per meeting; CRM completeness; flag recall	Advisor time survey + CRM field-population rates; seeded compliance moments in test transcripts
Software engineering	PR cycle time; review load; throughput	Repo analytics on the instrumented team vs. a control team, same sprint calendar

11.4 What never enters the portfolio

Three exclusion categories apply for the first two years, in every segment: client-facing generative agents, AI computation of regulated numbers, and autonomous submission to an external party. While none of these is driven solely by caution, all three are carried by the retreat file (Section 10). Also, each has a compliant variation that captures most of the value: agent-assist, not client-facing agents; an AI-drafted explanation of a number produced by deterministic code; a machine-generated, human-authorized submission through Paper I's double-controls gate (Paper I, §2.4).

11.5 Build, buy, or configure

The strongest structural insight to emerge from the failures literature concerns allocation, not technology: External vendor tools made it to production roughly twice as often as internally built solutions (MIT NANDA). Paper I's pattern-level guidance is consistent with this. Purchase where commoditization has rendered the category commodity and where the data is yours. Build where the differentiation lies in your corpus, your controls, and your integrations.

Pattern	Default posture	Rationale and category examples
Meeting intelligence	Buy, configure	Mature vendor category (Jump, Zocks, Zeplyn; platform-native options); differentiation is CRM integration and compliance flags, both configurable
Surveillance / AML triage	Buy the detection, build the triage narrative layer	Detection vendors are entrenched (NICE, Nasdaq, Behavox); the narrative/triage layer must sit on your queues and your policy language
RFP / DDQ automation	Buy the workflow, own the content library	Loopio/Responsive-class tools are proven; the moat is your approved, cited content corpus — never outsource its governance
Knowledge copilot	Build on platform primitives	The corpus, entitlements and citation discipline are the product; renting the UI is fine, renting the retrieval governance is not
Document extraction	Configure foundation-model extraction; buy niche parsers only where volume justifies	General extraction quality has commoditized; specialty parsers (e.g., alt-doc feeds) earn their keep only at scale
Reg-reporting repair	Build on Paper I's spine	Deeply coupled to your systems of record, entitlements and submission gates; no vendor owns your trajectory store



12. Conclusion — and five questions for your portfolio

Paper I ended with five questions regarding the engineering of your agent estate. Paper II ends with five questions around the economics of your portfolio. Treat them as a production readiness check for the value side:

- Of your current initiatives that involve AI, how many have an external, published benchmark they are being measured against, and how many are supported only by an estimate?
- Of your three largest initiatives, were the baselines captured pre-deployment, and would your CFO sign off on the delta?
- What percentage of your spend on AI goes into the two categories that comprise industry measurement outcomes, engineering and front office enablement, and how much is spent elsewhere?
- If you were to tell an allocator, client or regulator tomorrow that none of your AI outputs reached their desks without some human involvement in the process, could you prove it from logs?
- Which Wave 2 lane is your institution going to lead to the publication of a measured result in first, and what value would you gain from owning this reference point?

Here's how the reality of AI in financial services looks in mid-2026, based on all publicly available information. Money in measurable terms is in the area of assistance: retrieval, extraction, triage, drafting, capture, concentrated in front-office enablement, engineering, and document-intensive processes. The autonomy frontier is real and working but remains confined to internal domains where mistakes are cheap to catch. Documented mistakes happen exactly in those use cases where companies gave themselves autonomy without controls from Paper I. Firms that will survive the next few years will not believe the biggest numbers. They will choose natural use cases, hold to the published benchmarks, measure the performance against their baseline, and let the autonomy come as a result of control they already had in place. They will not need any predictions, ours or others'. They will have their own numbers.

Appendix A — The full inventory, scored

The Appendix presents all the use cases we've found in five segments, each one evaluated according to the rubric of §11.1 (weighted, 1–5) and provided with a verdict. We have it here to ensure that every workflow owner knows whether their use case is included or not and if not, why that is the case. The scores are our cross-industry defaults; the firm-specific scoping will re-score the two 25% criteria, since the evidence transfers unchanged while the pools of hours stay company-specific. Verdicts: **W1** = Wave 1; **W2** = Wave 2; **Watch** = revisit as soon as the evidence and regulation change; **Excl.** = excluded, Tier 3 or an outright failure of filtering. Verdicts are conditional to the segment by design. The same use case can be Wave 1 in the segment where there is published local evidence and Wave 2 in all other segments, and that is clear from the conditional verdicts below.

A.1 Investment banking

Use case	Tier	Level	Score	Verdict
Firmwide knowledge assistant	1	1	4.6	W1
Research retrieval & synthesis (sales/trading)	1	1–2	4.7	W1
Deal & offering document drafting	1	2	4.5	W1
Legal & vendor contract intelligence	1	2	4.4	W1
Master-agreement lifecycle (ISDA / CSA / GMRA),	2	2	3.6	W2 — the dated terms

Use case	Tier	Level	Score	Verdict
effective-dated terms				store is deterministic; AI extracts
Meeting & call capture into CRM	1	2	4.3	W1
Software engineering acceleration	1	1–3	4.8	W1
ESG / client questionnaire completion	1–2	2	4.0	W1
M&A / client screening at machine scale	2	2	3.4	W2
Research production & distribution leverage (avatars, formatting)	2	2	3.0	W2 — research-franchise firms only
Post-trade commentary & desk-supervision narratives	2	2	2.9	W2
Syndicate / roadshow logistics workflows	2	2	2.6	Watch
Pitch-personalization at scale (CRM-driven)	2	2	2.8	Watch
Client-facing autonomous advice / execution agents	3	5	—	Excl. — retreat file

A.2 Broker-dealers

Use case	Tier	Level	Score	Verdict
Surveillance alert triage & narratives	1	2	4.5	W1
AML alert classification & SAR preparation support	1	2	4.4	W1
Rejection of reg-reporting submissions classification & correction (CAT/TRACE)	2	2–3	3.9	W2 — Paper I's worked agent
Membership and due-diligence management, year-over-year comparison	1–2	2	3.8	W2
On-boarding and KYC document review by AI; difference in periodic review	1–2	2	3.9	W2
Complaints classification (4530) & response preparation	2	2	3.4	W2
Reconciliation/settlement failure triage (segmentation risk alerting)	2	2	3.7	W2
Corporate actions data capture & deadline risk	2	2	3.5	W2
Enforcement & regulatory-change intelligence	2	2	3.2	W2
Best execution committees & 605/606 stories	2	2	3.0	W2
Registration life-cycle quality control (U4/U5 compliance)	2	2	2.4	Watch — hours pool small at most firms
Blue-sheet / regulatory-inquiry response assembly	2	2	2.7	Watch

Use case	Tier	Level	Score	Verdict
Client-facing generative brokerage chatbot	3	5	—	Excl. — retreat file
AI computation of net capital / reserve numbers	3	—	—	Excl. — fails Filter 1

A.3 Asset managers

Use case	Tier	Level	Score	Verdict
Request for proposal/DDQ response	1	2	4.5	W1
Research synthesis & earnings-call digestion	1–2	2–3	4.2	W1
Alt-assets documentation processing (calls on capital, K-1s)	1–2	2	4.0	W1/W2 — W1 for alt-books
Private-markets deal-memo drafting	2	2	3.5	W2
Client and advisor documentation with auto-disclosures	2	2	3.4	W2
Portfolio commentary drafting	2	2	3.3	W2
Distribution intelligence & meeting prep	2	2	3.4	W2
Thematic & idea screening	2	2	3.0	W2
Fund marketing-compliance review	2	2	3.1	W2
NAV oversight & anomaly detection (classic ML)	2	2	3.0	W2 — not Gen AI; label honestly
ML-driven strategy research / signal generation	2	3	2.9	Watch — hedge-fund domain
Index / benchmark operations QC	2	2	2.6	Watch
Proxy-voting research summarization	2	2	2.5	Watch
Autonomous client reporting or trading agents	3	5	—	Excl.

A.4 Custodians & fund servicers

Use case	Tier	Level	Score	Verdict
Contract & agreement review	1	2	4.5	W1
Client onboarding / KYC documentation processing	1–2	2	4.1	W1 for custodians (63% of peers live); W2 elsewhere
Alt-asset document intake	1–2	2	4.0	W1 for custody / alternatives volume
Digital worker ops (verification, data fixes)	1	3–4	4.2	W1 pattern; year-2 destination via Paper I gates

Use case	Tier	Level	Score	Verdict
Reconciliation & exception management	2	2	3.8	W2 – address the post-trade gap of 24%
Corporate actions capture & elections processing	2	2	3.6	W2 – metric belongs to the first mover
Natural-language data access over servicing platforms	2	1–2	3.3	W2
Cross-border payments fraud detection (federated)	2	2	3.2	W2 — consortium-dependent
Transfer agency & shareholder services automation	2	3	3.4	W2
Client reports & billing narratives	2	2	2.9	Watch
NAV contingency / backup-NAV	2	2	2.8	Watch — classic ML
Autonomous settlement or payment release	3	5	—	Excl. — blast radius

A.5 Private wealth & wealth management

Use case	Tier	Level	Score	Verdict
Advisor knowledge copilot	1	1	4.7	W1
Meeting prep, capture & follow-up	1	2	4.6	W1
Tax-return-driven planning automation	1–2	2	3.9	W2 — W1 for planning-led firms
Estate & trust document extraction	2	2	3.6	W2
Account-opening quality (NIGO reduction)	2	2	3.5	W2
Next-best-client / organic-growth prediction	2	2	3.2	W2 — classic ML
Wealth-ops service agenting (requests, document chase)	2	3	3.3	W2
Client-content personalization with auto-disclosures	2	2	3.2	W2
Retail-scale deterministic assistants	1	2	3.8	W2 — control-first counterpoint (Erica, Fargo)
Suitability / appropriateness narrative support	2	2	2.8	Watch
Client-facing generative agents	3	5	—	Excl. — Schwab boundary test underway; revisit when results publish

References

Vendor-published figures (Canoe, Docupace, Loopio, Forrester TEI) are marked as vendor claims wherever used. The facts ascribed to the executives are statements, not measured results. All URLs verified in July 2026. The paper is not legal or investment advice. Paper I of this series — “Agentic AI in Financial Services: the infrastructure that unlocks the value, and contains the risk” (Wissen Technology, July 2026) — contains its own reference list; citations below are for Paper II exclusively.

1. FINRA, Regulatory Notice 24-09, “Artificial Intelligence, Including Large Language Models and Generative AI” (June 2024) — finra.org/rules-guidance/notices/24-09
2. FINRA 2026 Annual Regulatory Oversight Report, Section on Generative Artificial Intelligence (Dec 2025)
3. FINRA, R. Cook, “Finra’s Mission Enhanced by AI” (Oct 2025)
4. FINRA Rules 4517 and 6893 — finra.org/rules-guidance/rulebooks
5. Withdrawal by SEC of Proposal for Predictive Data Analytics (S7-12-23 June 20)
6. IOSCO (Mar 2025), “Artificial Intelligence in Capital Markets” — iosco.org; IOSCO (May 2026), “Supervisory Toolkit for Artificial
7. Bank of England/FCA, “Artificial Intelligence in UK Financial Services 2024,” Nov 2024 — bankofengland.co
8. McKinsey, “The State of AI: Global Survey” (2025) — mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
9. McKinsey, “Been there, doing that: How corporate and investment banks are approaching gen AI” (2024) — mckinsey.com
10. Evident AI Index (Oct 2025) & Evident AI Outcomes Report (H1 202
11. MIT Nanda, “The GenAI Divide: State of AI in Business 2025”; criticism: arnon.dk
12. RAND, “Root Causes of Failure of Artificial Intelligence Initiatives” (2024) — rand.org/pubs/research
13. AIMA, Survey of Adoption of Front-Office Gen AI, 150 Managers/ \$788B Assets Under Management (September
14. Citi Securities Services Evolution 2025, n=537 — citibank.com/icg/docs/securities-services/Citi-Securities-Services-Evolution-2025.pdf
15. Survey of MMI–Broadridge members (Nov 2025) — broadridge.com; SimCorp’s AI usage in front office
16. BCG, “Rebuilding Asset Management for an AI-First World” (2026) –
17. Deloitte Insights, generative AI in investment banking (2023) — deloitte.com
18. LexisNexis Risk Solutions, “The True Cost of Financial Crime Compliance” (2023/2
19. Reuters: Coding Assistant Efficiency at JPMorgan is 10-20% (Mar 2025); JPMorgan Coach AI and more than 20% Gross Sales in AWM (May 2
20. eMarketer/Bloomberg: JPMorgan sal plan om oor die bedrag van \$2 miljard in AI te investeer en is opgewonde om die winne te sien; hulle het meer as
21. CNBC: Goldman Sachs GS AI Assistant firmwide (Jan 2025); Goldman Sachs × Devin autonomous-coder pilot (Jul 2025); banks using AI to hire fewer people (Oct 2025)
22. Fortune: D. Solomon on AI drafting 95% of an S-1 (Jan 2025); Goldman OneGS 3.0 and AI-related headcount reductions (Oct 2025)
23. Press Releases from Morgan Stanley: AI @ Morgan Stanley Debrief (2024); AskResearchGPT (2024); OpenAI-Morgan Stanley Case Study (98% adoption)
24. Citi: Stylus Workspaces agentic release (Sept 2025); American Banker on Citi AI prompt training (~175,000 employees); VentureBeat on Citi and customer support risks
25. Bloomberg: JPMorgan COiN (2017); UBS M&A screening tool (2024); BlackRock Asim
26. Semafor, “The Investment Banking AI push is a rising tide” – Rogo at Jefferies & Lazard; OpenAI hiring at banks (May 2026)
27. Press Releases from Nasdaq: Gen AI for market surveillance — analysts estimated a 33% reduction in investigation time in tests (2024); AI throughout the surveillance lifecycle (2025)
28. Gen AI for NICE Actimize SURVEIL-X (2025); Businesswire, BNY Mellon finishes Behavox Quantum rollout (2025)
29. Google Cloud, AML AI with HSBC – 2-4 times more true positives, around 60% fewer alerts (2023)



30. Compliance Week, Instinet pays \$3.8M CAT penalty; Katten, FINRA's first major CAT enforcement; Goldman Sachs \$1.4
31. Broadridge: Operations GPT (2024); "Efficiency Unleashed" post-trade Gen-AI white paper; Broadridge Digital Transformation Survey (2025)
32. BlackRock: Aladdin Copilot product page; Preqin acquisition (~\$3.2B) press releases; Aladdin Wealth AI comments at Morgan Stanley (2025); eFront Copilot
33. Schroders Capital, GAiiA launch (June 2024) — [schroders.com](https://www.schroders.com)
34. Vanguard press releases: Gen-AI client summaries (May 2025); Expert Insights (Apr 2026)
35. Franklin Templeton, Intelligence Hub with Microsoft (Jan 2026) — investors.franklinresources.com
36. InnoLead: AllianceBernstein CAIO profile (ABAI; 10–50% time savings); T. Rowe Price AI ROI strategy
37. H
38. Case Study: Loopio, HedgeServ — DDQ 3 days to 4 hours; Case Study: Ontra, one million documents processed
39. Claims: Canoe Intelligence; Acquisition: WealthManagement.com, Bloomberg acquires Canoe (2026); Press Release: Businesswire, 73 Strings \$55M Series B (2025)
40. Press Release: PR Newswire, Fidelity Saifr x Microsoft compliant-content AI agent (May 2025)
41. Temenos Multifonds NAVigator; Milestone Group pControl NAV oversight
42. Case Study: OpenAI, BNY — Eliza, contract review -75%, 20,000+ builders, 125+ use cases — openai.com/index/bny
43. CNBC (Feb 2026), Fast Company: BNY digital employees, human managers,
44. FTF News, SS&C digital workers (~1,600; Blue Prism ~\$1.6B; Forrester TEI 330% ROI)
45. VentureBeat, State Street Gen AI on Alpha; State Street 2025 Annual Report (agentic AI)
46. Private Banker International, Northern Trust alternative-asset document extraction (2021)
47. Swift, federated-learning fraud pilots with 13 banks, ~2x detection (Sept 2025) — [swift.com](https://www.swift.com)
48. Businesswire, SmartStream Smart Agents for back-office operations (June 2026)
49. Newsroom of Bank of America: Merrill / Private Bank Meeting Journey, up to 4 hours per meeting (Mar 2026); Erica exceeds 3 billion interactions (2025); American Banker, "How Bank of America's Erica does the work of 11,000 people"; CIO Dive on the intentionally non-generative Erica approach
50. VentureBeat (Apr 2025), Wells Fargo Fargo — no-PII-to-the-LLM architecture, then 245M+ interactions; superseded by 1B+ milestone in ref
51. Microsoft Switzerland, UBS & Microsoft AI alliance (including UBS Red) (Feb 2025); Finextra, UBS AI Analyst Avatars (2025)
52. TIFIN AG x RBC Wealth Management; ThinkAdvisor, LPL AI Tools (2025); InvestmentNews, Raymond James AI Operations Agent; Raymond James x Zoom AI Companion (2025)
53. ThinkAdvisor, Schwab client-facing AI agents rolling out from June 2026
54. WealthManagement.com, Jump \$80M Series B; Businesswire, Zocks Series A/B; Businesswire, Wealth.com \$65M Series B (2026); Insight Partners, Vanilla Funding
55. Access Newswire, Holistiplan 39% Tax Planning Category Share; XYPN Partnership (2,100 Firms)
56. Docupace, NIGO reduction infographic (vendor claim)
57. Klarna Press Release (Feb 2024); Forbes and Entrepreneur, Klarna Reversal Coverage (May
58. Goldman Sachs Q4/FY2024 financials and 2024 annual report (compensation ratio); CNN / Forbes, 2021 Goldman Sachs first-year analyst survey (95 to 100 hours per week)
59. JPMorgan Chase Q2 2026 earnings call (Jul 2026): around 1,000 AI applications used internally; around 150,000 weekly internal LLM users – Motley Fool transcript and analysis
60. BNY Q2 2026 financials and earnings call (Jul 2026): staff down 7% year-over-year, record revenues of \$5.7B; roughly 40% of software developed using AI tools – American Banker; Motley Fool transcript
61. Forrester, "How BNY built its digital workforce backwards – and it worked" (May 2026): over 130 digital workers in operation, up from 70; 1/3 of all staff creating agents
62. Charles Schwab press release (May-June 202



63. Wells Fargo newsroom (Mar 2026): Fargo exceeds 1 billion customer interactions within less than three years
64. ThinkAdvisor (Feb 2026), LPL extends Anthropic partnership – Claude technology to 30,000+ advisors; InvestmentNews (Jul 2026), Ameriprise invests \$1B in AI-powered tech and Raymond James “Raimond”; Edward Jones/Morning Consult industry survey (Jul 2026): 82% of advisors (n=201) already utilize AI
65. Nasdaq Verafin (Jun 2026): Adoption of Agentic AI Workforce by more than 650 financial organizations; claims from the vendor – up to 90% reduction in sanctions alerts and 50% reduction in EDD reviews
66. Rogo customer publication: Baird Global Investment Banking – 95% engagement rate, ~10,000 workflows per week, most active users ~5 hours per day (hosted by vendor, attributed to the firm)
67. U.S. Treasury, AI Risk Management Framework for financial services (Feb 2026, 230
68. Broadridge, sixth annual Digital Transformation & Next-Gen Technology Study (Feb 25, 2026; n=900+ technology and operations leaders): 80% using Gen/predictive AI in operations (up from 31%); 27% reporting measurable GenAI benefits (+13 pts YoY); 26% using agentic AI — broadridge.com/press-release/2026/genai-delivering-now-tokenization-is-next
69. Convergent 2026 agent-pilot surveys (Forrester/Anaconda 2026; WRITER/Workplace Intelligence 2026, n=2,400): roughly one in ten AI agent pilots graduate to production — as compiled in Paper I, reference 1
70. ISDA, S&P Global Market Intelligence and Linklaters: CreateiQ / ISDA Create integration with Counterparty Manager for end-to-end negotiation of derivatives documentation with structured legal data (Sept 2023) — isda.org; ISDA Amend for at-scale amendment workflows — isda.org/isda-solutions-infohub/isda-amend



About Wissen

Wissen was founded in 2000 in the United States. Based in Bangalore, India, Wissen Technologies is a well-known information technology company in the market today. With a strong presence in countries like the US, India, the UK, Australia, Mexico, and Canada, we provide end-to-end solutions to various companies across industries including Banking and Financial Services, Telecom, Healthcare, Manufacturing, and Energy. Using our world-class infrastructure and development centers all over the world, we have delivered over \$1 billion of projects for over 25 Fortune 500 companies till date. With our team of highly qualified 5,500+ professionals spread across the world through the Wissen Group, we leverage our expertise and engineering skills to deliver meaningful results to the world's most demanding organizations.

Services We Offer

ADVISE



Regulatory, Compliance
& Risk Advisory



AI Governance
& Enablement



Target Operating Model
& Cost Transparency

BUILD



Application &
Platform Engineering



Data Engineering
& Analytics



AI/ML & Agentic
Automation



Cloud & DevSecOps



Product Design



QA & Test
Automation

RUN



Managed Services
& KPO



Regulatory Reporting
as a Managed Service



Infrastructure & SRE

 contact@wissen.com

WISSEN 
TECHNOLOGY
DELIVERING 2X IMPACT