

FOR MANAGEMENT AND DECISION-MAKERS

# Agentic AI in Financial Institutions

Agents running complex business work in banks and insurers  
— with human intent, direction and control built in

## IN THIS EDITION

- Why agentic AI is now a top management priority
- What leading institutions and vendors have built, by function
- An operating model for human intent, direction, and control
- Reference architecture and a concrete technology stack
- Worked use cases: reinsurance accounting and KYC/AML, with team, effort and cost
- Estimating templates, first steps without central infrastructure, regulation, risks and roadmap

How to read this paper: "financial institutions" or "institutions" means banks, insurers and reinsurers; where a point applies to only one of them, the text says so. Abbreviations are spelled out at first use and listed in the glossary. Numbers in square brackets refer to the sources at the end. Effort and cost figures are informal estimates for orientation, not quotes.

## At a glance

### THE SHORT VERSION

Artificial intelligence (AI) agents, software that plans and carries out multi-step tasks using tools, can now carry out much of the preparation behind complex work in financial institutions: checking and booking reinsurance accounts, refreshing know-your-customer (KYC) files, investigating alerts, drafting risk and investment reports. The value is large, but in banks, insurers and reinsurers every material decision belongs to a named person. The answer is an operating model and architecture in which **people set the intent, direct the work at defined checkpoints and take every commit decision, while agents prepare and the platform records the evidence.**

### Seven key messages

- 01 **Agents have moved into business processes.** Banks such as JPMorgan, BNY, Goldman Sachs and Santander, and insurers such as Allianz, AIG and AXA, run agents in operations, accounting, compliance, claims and underwriting. Risk, investment reporting and finance are following (Section 2).
- 02 **Control is the design question, not autonomy.** Every task follows five control points: intent, plan, checkpoints, commit, review. Autonomy is set per task type and raised only on evidence (Section 3).
- 03 **The agent proposes, the system commits.** A commit boundary separates agent work from anything that changes the institution's records; write actions need a human approval token (Section 4).
- 04 **Centralise control, federate knowledge.** Identity, gateways, policy, audit and evaluation are built once; mandates, checklists, domain tools and test cases belong to the business (Section 5).
- 05 **The technology is available.** All three large cloud providers (hyperscalers) and the main case-management, finance and insurance software vendors now ship agents with identity, policy and tracing (Section 6).
- 06 **Two lighthouses take about eight months and €3.7 million.** Reinsurance technical accounting (about €1.75 million) and KYC and anti-money laundering (AML) for a bank or life insurer (about €2.1 million); people and integration dominate cost, model usage is small (Sections 7–9).
- 07 **Institutions can start before the platform exists.** With an approved assistant, business teams can build the checklists, golden cases (test cases with known correct answers) and process maps every agent needs; a first lighthouse project can build a minimum control plane, the shared controls for identity, access, logging and evaluation, designed for hand-over (Section 10).

### What management needs to decide

- **Principles:** human intent, direction and control as binding design rules for all agent work.
- **Organisation:** which components are central and which belong to domain teams, and how vendor-embedded agents are governed.
- **Investment:** funding for two lighthouse pilots with go/no-go gates based on measured accuracy, time saved and adoption.
- **Ownership:** named mandate owners and product owners in the business.
- **Providers:** approved models, clouds and vendors under the EU Digital Operational Resilience Act (DORA), with exit plans.

## 01 Why this matters now

Agents have moved from answering questions to carrying out multi-step business work. In financial institutions, the limiting factor is no longer model capability but the ability to direct and control that work.

Much of the skilled capacity of banks, insurers and reinsurers goes into case work with a common pattern. People gather documents and data from several systems, apply rules and policy, judge the exceptions, write it up and get it approved. A reinsurance statement of account, a KYC periodic review, an AML alert, a quarterly investment report, a Solvency II (the EU prudential regime for insurers) or other regulatory return, a reconciliation break and a risk report commentary all follow this pattern. The judgement is the valuable part; most of the hours go into finding, checking and documenting.

### What has changed

- **Models can run multi-step work with tools.** Current large language models (LLMs), the AI models behind tools such as ChatGPT, Claude and Gemini, plan, call systems, check intermediate results and recover from errors. Citi's Chief Technology Officer said in 2025 that early models were poor at invoking tools but that current ones are good at it [4].
- **Peers have moved agents into business processes.** JPMorgan runs agents on back-office workflows; BNY operates around 140 "digital employees" with human managers; Goldman Sachs built agents for trade accounting, onboarding and compliance; Allianz runs a seven-agent claims process in which payout decisions are never automated [1][2][3][13].
- **Platforms now ship the control machinery.** Cloud providers offer agent runtimes with built-in agent identity, policy and tracing, and enterprise software vendors embed agents in the case management, finance, policy administration and customer relationship management (CRM) systems institutions already run (Section 6).
- **Open standards reduce lock-in.** The Model Context Protocol (MCP) for connecting agents to tools and the Agent2Agent protocol (A2A) for agent-to-agent communication are both governed by the Linux Foundation [25].
- **Supervisors are engaging.** The European Central Bank (ECB) has made banks' AI strategies and governance, explicitly including generative AI, a supervisory priority for 2026–28 [59]. The European Insurance and Occupational Pensions Authority (EIOPA) published an Opinion on AI governance and risk management for insurers in August 2025 and expects agentic AI to have a greater impact over the medium term [70][71].

### Why intent, direction and control is the design question

In financial institutions, accountability is personal and documented. Signing authorities, four-eyes principles and the governance requirements of banking and insurance supervision assign every material decision to a named person or committee. An agent cannot hold an underwriting or credit authority, sign off a KYC file or approve a booking. The design question is therefore not how autonomous an agent can be, but how a named person's intent stays in charge of the work and how the institution proves afterwards what happened. That requires three things:

- **Intent is explicit.** Every agent task starts from a mandate set by an accountable person or a governed process: goal, scope, constraints, materiality and the evidence expected.
- **Direction stays with people.** People see the plan, can intervene at defined checkpoints, and decide on everything that commits the institution.
- **Control is built in, not added on.** Entitlements, the commit boundary, audit and evaluation are part of the platform, so every task produces its own evidence.

## Where the value is

| Lever                        | What changes                                                                           | Evidence to measure                                                         |
|------------------------------|----------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>Capacity in case work</b> | Agents assemble files, draft analyses and documentation; specialists review and decide | Hours per case; backlog                                                     |
| <b>Cycle time</b>            | Fewer hand-offs and waiting times between collection, analysis and approval            | End-to-end lead time; days to close; service-level agreement (SLA) breaches |
| <b>Consistency</b>           | The same checklist and evidence standard on every case                                 | Quality assurance (QA) finding rate; rework rate                            |
| <b>Control evidence</b>      | Every source, step and approval recorded automatically                                 | Audit findings; time to answer supervisory requests                         |

### A WORD OF CAUTION

Gartner expects more than 40% of agentic AI projects to be cancelled by the end of 2027, citing rising costs, unclear value and weak risk controls [24]. JPMorgan's Chief Analytics Officer has described a "value gap" between what AI can do in theory and what it delivers inside thousands of existing systems [1]. Projects that succeed choose work with measurable volume, redesign the process around human checkpoints, and treat data, integration and control as the main engineering effort.

02

## What leading institutions and vendors have built

The public record shows a consistent sequence: firmwide model platforms first, then agents in process-heavy operations, then agents in judgement-heavy case work with human sign-off. Banks are furthest in finance operations and financial crime; insurers in claims and underwriting. Risk, investment reporting and finance close are at pilot or vendor stage in both sectors, which is where early movers can still build an advantage.

### Banks

| Institution           | What has been built                                                                                                                                                                    | Lesson                                                                         |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| <b>JPMorgan Chase</b> | LLM Suite for around 250,000 employees; agents now run back-office workflows [1]                                                                                                       | A model-agnostic platform with growing system connections is the foundation    |
| <b>BNY</b>            | Around 140 "digital employees" in reconciliation, payment repair and client inquiries, each with a human manager; payment validation down from 5–6 minutes to under 30 seconds [2][11] | Agents are managed like staff, with owners, review boards and compliance gates |
| <b>Goldman Sachs</b>  | Agents for trade and transaction accounting, client onboarding and compliance, built with Anthropic engineers embedded for six months [3]                                              | Rule-based, document-heavy accounting work is feasible under oversight         |
| <b>Santander</b>      | More than 280 agents in production across credit, fraud, KYC and operations; at Openbank about 100,000 AML alerts a year, investigation time from hours to minutes [9]                 | Financial crime is the first domain to scale                                   |

| Institution       | What has been built                                                                                                                                                                   | Lesson                                                            |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|
| HSBC              | Partnership with Google Cloud for agentic AI in financial-crime detection, expecting about twice as fast responses, and for encoding regulatory procedures (announced June 2026) [10] | Human judgement and accountability stated as the core design rule |
| Citi, Commerzbank | Agentic features piloted with 5,000 staff (Citi); group-wide Gemini Enterprise and Microsoft 365 Copilot (Commerzbank) [4] [7]                                                        | Many institutions are still building the assistant layer first    |

## Insurers and reinsurers

| Institution            | What has been built                                                                                                                                                                                                                   | Lesson                                                       |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|
| Allianz                | Project Nemo: seven agents handle small food-spoilage claims in Australia, processing time down 80%; payout decisions are never automated. Group partnership with Anthropic for agents with traceability and human oversight [13][14] | A narrow, well-defined case type is the right first scope    |
| AIG                    | Multi-agent underwriting with Anthropic and Palantir, including a "critic" agent; in one middle-market portfolio about 55% less time to quote [15]                                                                                    | An orchestrator decides when a human must step in            |
| AXA                    | Group-wide, model-agnostic Global AI Hub with governance, cost control and human oversight of high-stakes decisions; five entities live by September 2026 [16]                                                                        | Platform first, then use cases across entities               |
| Swiss Re               | Redesign of core processes; one pricing workflow cut from 25 steps and 14 applications to fewer than five applications [17]                                                                                                           | Value comes from redesigning the workflow, not adding a tool |
| Zurich, HDI, ERGO, W&W | Over 200 prototypes at Zurich, including agents for financial reporting and actuarial analysis; German insurers report agents with humans in the loop moving from pilot to scale [18][21]                                             | Finance and actuarial agents are at prototype stage          |

## Use cases by function

| Function                      | Banks                                                                                                                     | Insurers and reinsurers                                                                                                                                                                                      | Evidence and maturity                                                                                                                                                                                                           |
|-------------------------------|---------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Risk</b>                   | Risk report commentary; value-at-risk (VaR) and exposure move explanation; credit memo drafting; early-warning monitoring | Solvency II and IFRS 17 (the accounting standard for insurance contracts) report production and commentary; Own Risk and Solvency Assessment (ORSA) drafting; preparing exposure data for catastrophe models | Pilot and vendor stage. A European bancassurer cut Solvency II and IFRS 17 report development time by 90% with Milliman [20]; Moody's agents for credit memos and insurance risk data [5][26][40]                               |
| <b>Investment reporting</b>   | Fund and client reporting; performance and attribution commentary; portfolio narratives                                   | Quarterly investment reports; Solvency II asset look-through data; investment accounting reconciliations                                                                                                     | Early. Generative commentary in BlackRock Aladdin Wealth [37]; AI for cash matching and reconciliation breaks at Clearwater [36]; no named agent deployment published yet                                                       |
| <b>Compliance</b>             | KYC periodic review; AML alert investigation; sanctions; trade and communications surveillance; regulatory change         | AML for life and investment business; sanctions screening; regulatory change; conduct and complaints                                                                                                         | Most mature. Santander, HSBC, FIS with Amalgamated Bank, Deutsche Bank trade surveillance; Icatu Seguros reports 63% productivity gain from an AML agent [6][9][10][12][19][38]                                                 |
| <b>Accounting and finance</b> | Trade and transaction accounting; reconciliations; close; regulatory reporting tie-out                                    | Reinsurance technical accounting; bordereaux (periodic premium and claims listings) and cash matching; IFRS 17 close; accruals                                                                               | Growing. Goldman Sachs and BNY in production [3][11]; vendor agents for reconciliation (BlackLine, up to 92% less preparation time), accruals (SAP), report tie-out (Workiva) and bordereaux matching (Brisic) [33][34][35][39] |

## Vendors with agents for complex financial work

| Vendor                                | Offering                                                                                                                       | Reported results (vendor claims)                                                                            |
|---------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| <b>Moody's</b>                        | Agentic Solutions for credit memos, monitoring and KYC; MCP server and data agents for insurance risk analytics [26][40]       | Unnamed European banking group: first-draft credit memos in under 15 minutes instead of about two weeks [5] |
| <b>nCino, S&amp;P Global</b>          | Agents for commercial credit reviews and credit memos [27][28]                                                                 | nCino: 60–70% less review effort                                                                            |
| <b>Anthropic</b>                      | Claude for Financial Services: data connectors and agent templates including a KYC screener [29][48]                           | Named customers include BNY, Mizuho and FIS                                                                 |
| <b>Fenergo, FIS, WorkFusion, Hawk</b> | Agents for client lifecycle, periodic review and AML investigations [30][31][32][38]                                           | Fenergo: up to 45% faster periodic reviews; WorkFusion: one bank auto-closes about 60% of one alert type    |
| <b>BlackLine, SAP, Workiva</b>        | Agents for reconciliation preparation, accruals and journal entries, and report tie-out, with accountant sign-off [33][34][35] | BlackLine: up to 92% less manual preparation time                                                           |

| Vendor                                | Offering                                                                    | Reported results (vendor claims) |
|---------------------------------------|-----------------------------------------------------------------------------|----------------------------------|
| <b>Brisic</b>                         | Bordereaux and cash-matching agent for insurers and reinsurers [39]         | 97% automated match rate claimed |
| <b>Duck Creek, Guidewire, Sapiens</b> | Agents in insurance core platforms, mainly for underwriting and claims [41] | No metrics published             |

Two cautions apply. First, these are mostly vendor-reported figures from unnamed clients. Second, no named European insurer or reinsurer has yet published agent results in ORSA, IFRS 17 close or reinsurance technical accounting, and most bank results concern first drafts or alert pre-work, not final decisions. ING's mortgage automation in Germany, with a target of about 30 minutes instead of about six days and a human final decision, shows the same preparation-plus-decision pattern without the agent label [8].

### What research and practice tell us

- **Free-form agents fail on enterprise data.** On realistic enterprise databases, LLMs writing their own database queries in Structured Query Language (SQL) dropped from around 87% to around 10% accuracy [22]. Agents need governed tools and defined metrics, not raw access.
- **The deterministic boundary is the reference pattern.** BCG's 2026 reference architecture for agentic banking places a probabilistic AI core inside a deterministic boundary: the agent proposes, the system commits [23].
- **Process orchestration and agents are converging.** Case-management, finance and insurance-platform vendors now run agents as steps inside governed processes with human tasks around them (Section 6).
- **The bottleneck is the business, not the build.** Experience from data and process programmes suggests that defining checklists, golden test cases and decision rights takes longer than the engineering.

## 03 Human intent, direction and control

Agent work in a financial institution should be organised like delegated work to a capable but unaccredited team member: a named person sets the mandate, the agent prepares, and only people or deterministic systems commit the institution. This section defines that operating model. The architecture in Section 4 implements it.

### Five control points along every task

1. **Intent.** A mandate is set by an accountable person or by a governed process trigger (a review falling due, an alert, a client request). It states the goal, the scope, the constraints, the materiality thresholds and the evidence the result must contain. Mandates are templates owned by the business, not free text.
2. **Plan.** The agent proposes a plan: which sources, which checks, which tools. For routine work the plan is fixed by the template; for open investigations a person approves or edits it.
3. **Checkpoints.** The process defines points where a person must look before work continues, for example when a red flag is found, when data conflict, or when cost or time exceeds a budget. The agent stops and asks; it does not decide around a checkpoint.
4. **Commit.** Anything that changes the institution's position, records or obligations passes through a commit gate: an underwriting or credit decision, a booking, a KYC file closure, a suspicious activity report, a regulatory submission, a client message. The agent proposes; a person with the authority approves; a deterministic system executes.

5. **Review.** After the fact, quality assurance samples completed cases, compares agent work with expert work, and feeds findings into the evaluation set. The second and third lines can replay any case from its trace.

## Autonomy levels

Autonomy is set per task type, not per agent or per platform. It is raised only on evidence from the review step.

| Level                       | What the agent may do                                                                      | Human role                                             | Typical examples                                                                            |
|-----------------------------|--------------------------------------------------------------------------------------------|--------------------------------------------------------|---------------------------------------------------------------------------------------------|
| <b>L0 Assist</b>            | Answer, summarise, draft on request within the user's own session                          | User reads everything                                  | Employee assistant; research                                                                |
| <b>L1 Prepare</b>           | Run a defined task end to end and deliver a draft file with evidence; read-only on systems | Reviewer checks and owns the result                    | Reinsurance statement check; KYC file refresh                                               |
| <b>L2 Propose</b>           | As L1, plus propose a decision or action with reasoning; write to a work queue only        | Approver accepts, edits or rejects each proposal       | AML alert disposition proposal; booking proposal                                            |
| <b>L3 Act with sampling</b> | Execute low-risk actions within hard limits; humans sample after the fact                  | Supervisor sets limits, reviews samples and exceptions | Closing clear false-positive name-screening hits; requesting missing documents from clients |
| <b>L4 Act autonomously</b>  | Execute within a delegated authority without routine review                                | Owner monitors outcomes                                | Not recommended for material decisions today                                                |

Most complex business work in financial institutions belongs at L1 or L2 for the first two years. L3 is realistic for high-volume, low-risk steps with a measured error rate. Accountable decisions stay with people.

## Human roles

| Role                 | Responsibility                                                                                                        | Typical holder                                                                       |
|----------------------|-----------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| <b>Mandate owner</b> | Owens the task template, the checklist, the autonomy level and the success measures                                   | Head of the business process, e.g. head of technical accounting or of KYC operations |
| <b>Case owner</b>    | Sets or accepts the individual task, directs at checkpoints, owns the result                                          | Technical accountant, KYC analyst, investigator                                      |
| <b>Approver</b>      | Takes the commit decision under existing authority rules                                                              | Senior accountant, money laundering reporting officer (MLRO) delegate, team lead     |
| <b>Agent owner</b>   | Responsible for the agent's instructions, tools, evaluation and incidents, like a line manager for a digital employee | Product owner in the business, supported by the platform team                        |
| <b>Assurance</b>     | Validates, samples, audits; can replay any case                                                                       | Model risk, compliance, internal audit                                               |

BNY uses the same principle at scale: its digital employees have human managers who own their performance [2].

## Rules that make the model hold

- **The agent's authority never exceeds the user's.** Agents act on behalf of a named user or a named service identity, and entitlements are enforced in the tool and data layers, not by instructions to the model.
- **Proposals and commits are separate.** No tool available to an agent can commit a material change without a human approval token from the work queue.
- **Every task leaves a case file.** Mandate, plan, sources, tool calls, intermediate results, drafts, human interventions and the final decision are stored together and kept for the record-keeping period of the underlying process.
- **The agent that drafts may not approve.** Four-eyes applies to agents too; a second agent can check, but only a person approves.
- **Stop is always possible.** Every running task can be paused or cancelled by its case owner, and every agent type can be switched off centrally.

## 04 Reference architecture

The architecture has six layers, a human work surface and a cross-cutting control plane. Its defining feature is a **commit boundary**: above it, agents interpret, plan, gather and draft; below it, everything that changes the institution's records runs through deterministic, versioned systems after a human approval.

## Agents prepare above the commit boundary; people approve, systems commit

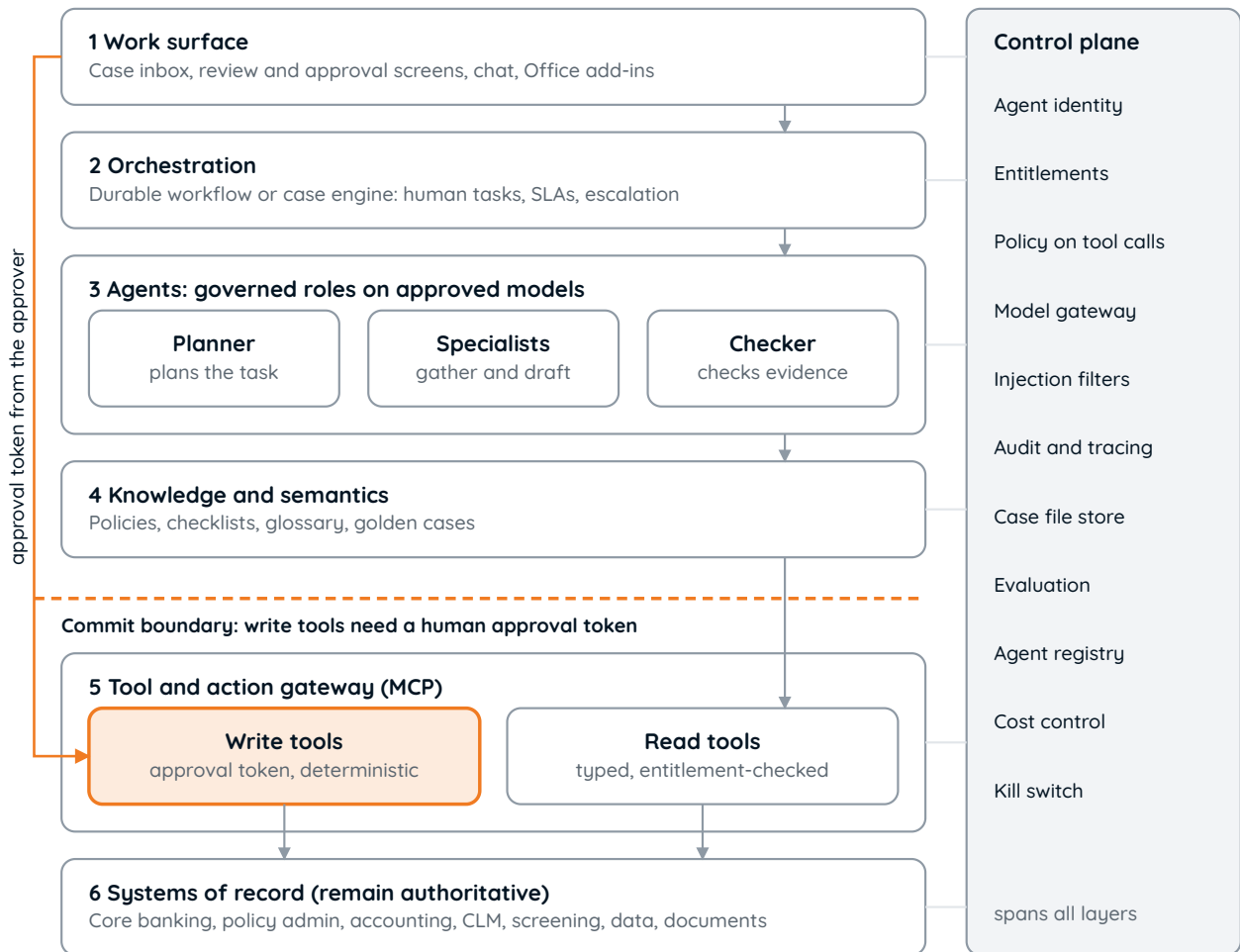


Figure 1. Reference architecture: six layers, commit boundary, control plane.

Reads flow freely through governed tools; only the write path crosses the commit boundary, and only with an approval token issued to a named person.

| Layer                            | What it does                                                                                                                         | Who typically owns it                         |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|
| <b>1 Work surface</b>            | Where people set intent and direct work: case inbox, review and approval screens, chat or Teams, Office add-ins                      | Business application or platform team         |
| <b>2 Orchestration</b>           | Runs the process: a durable workflow or case engine with human tasks, timers, SLAs and escalation; agents are steps inside it        | Process owner with the platform team          |
| <b>3 Agents</b>                  | Defined roles (planner, specialist, checker), each a governed set of instructions, tools and permissions on top of approved models   | Business domain (agent owner)                 |
| <b>4 Knowledge and semantics</b> | Policies, procedures and checklists for retrieval; business glossary and ontology; metric and rule definitions; golden cases         | Business domain                               |
| <b>5 Tool and action gateway</b> | The only route from agents to systems: typed, versioned tools exposed via MCP; read tools open, write tools gated by approval tokens | Platform team (gateway); domain teams (tools) |

| Layer               | What it does                                                                                                                                                                               | Who typically owns it  |
|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| 6 Systems of record | Core banking, policy administration, accounting, client lifecycle management (CLM) for KYC, screening, data platforms, document stores, external data providers; they remain authoritative | Existing system owners |
| Control plane       | Agent identity and entitlements, policy, model gateway, audit and tracing, evaluation, registry, cost control, kill switch                                                                 | Central platform team  |

Design principles

- **Agents are steps in a process, not the process.** A governed workflow decides what happens next, who is involved and when to escalate. Agents do the knowledge work inside the steps. This keeps processes auditable and lets agents be replaced.
- **The agent proposes, the system commits.** Write actions are separate tools that require an approval token issued by the work queue to a named approver.
- **Delegated identity.** Agents act with the user’s entitlements (“on behalf of”) or a narrowly scoped service identity. Entitlements are enforced at the tool and data layer.
- **Facts from tools, words from the model.** Figures, ratings, limits and screening results come from systems; the model explains and drafts. Every figure in a draft is matched against a tool result before the draft is shown.
- **Business knowledge is a first-class asset.** Checklists, policies, definitions and golden cases are versioned like code and owned by the business. They outlast any model or vendor.
- **Open interfaces.** MCP for tools, A2A for agent-to-agent calls, OpenTelemetry (an open standard for recording what software does) for traces. These protect investments when models or platforms change.

The evidence contract

Every agent output that reaches a reviewer meets a fixed contract, checked automatically:

- **Sources** for every material statement: document, page or record, retrieval time and version.
- **Number check:** every figure matches a tool result exactly, or the output is blocked.
- **Checklist coverage:** each item of the mandate checklist is marked done, not applicable with reason, or open.
- **Open points and confidence:** what the agent could not establish, and where it disagrees with the source data.
- **Trace:** model, prompt and tool versions, all tool calls and human interventions, stored with the case.

05

## What to provide centrally and what projects build

The rule of thumb: **centralise what concerns control, security and reuse; leave to the business what concerns domain knowledge and process.** The control plane and gateways are built once; mandates, checklists, tools for domain systems and evaluation sets belong to the business areas.

## Responsibility split by component

| Component                                                                                  | Central platform                                    | Domain project                              | Notes                                                                               |
|--------------------------------------------------------------------------------------------|-----------------------------------------------------|---------------------------------------------|-------------------------------------------------------------------------------------|
| Model access and model gateway                                                             | Builds and runs                                     | Uses                                        | Approved models, EU regions, budgets, data-loss controls, model switch in one place |
| Agent identity and entitlements                                                            | Builds and runs                                     | Defines roles and scopes                    | On-behalf-of flows; agent registry with owners                                      |
| Policy and guardrails                                                                      | Builds framework, firm-wide rules                   | Adds domain rules                           | Prompt-injection and data-leak filters centrally; business rules in the domain      |
| Tool gateway (MCP)                                                                         | Builds and runs                                     | Registers its tools                         | Allow-list, approval-token enforcement, rate limits                                 |
| Tools for shared systems (documents, email, CRM, data platform)                            | Builds                                              | Uses                                        | Built once, used by every domain                                                    |
| Tools for domain systems (policy administration, technical accounting, screening, ratings) | Standards, review                                   | Builds and owns                             | Owned by the system's domain team                                                   |
| Orchestration engine and human task framework                                              | Provides                                            | Models its processes                        | One or two approved engines firm-wide                                               |
| Work surface                                                                               | Provides components (inbox, approval, trace viewer) | Configures screens                          | Embedded in existing case tools where possible                                      |
| Agent framework and templates                                                              | Provides, maintains reference patterns              | Builds agents                               | Planner-specialist-checker as a reusable pattern                                    |
| Knowledge base and retrieval infrastructure                                                | Provides                                            | Curates content                             | Content owners in the business                                                      |
| Audit, tracing and case file store                                                         | Builds and runs                                     | Uses                                        | Immutable, retention per process                                                    |
| Evaluation harness                                                                         | Provides                                            | Owns golden cases and acceptance thresholds | Regression on every model, prompt or tool change                                    |
| Cost control                                                                               | Provides budgets and reports                        | Owns its budget                             | Cost per case as a standard metric                                                  |
| Risk classification and validation                                                         | Defines method (with model risk)                    | Completes per use case                      | One control story for supervisors                                                   |

## Four ways to organise it

| Option                                                 | How it works                                                                                                                                              | Fits when                                                     | Watch out for                                                              |
|--------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|----------------------------------------------------------------------------|
| <b>A Central agent factory</b>                         | One central team builds and runs all agents; business supplies requirements and experts                                                                   | Early stage; few, high-visibility use cases; scarce AI skills | Central team becomes a bottleneck; weak business ownership                 |
| <b>B Platform plus federated domains (recommended)</b> | Central team runs control plane and gateways; domain teams build agents, tools and evaluation sets on it under common standards                           | Several business areas with their own IT capacity             | Needs clear standards, a review gate and inner-source contribution         |
| <b>C Vendor-embedded agents</b>                        | Agents come with business applications (e.g. policy administration, finance close, KYC/CLM, case management); central function governs and registers them | Strong vendor estate; standard processes                      | Fragmented control story; lock-in; agents cannot easily cross applications |
| <b>D Lighthouse first, platform later</b>              | One domain builds a complete use case including minimum shared components, designed for hand-over to a central team                                       | No central platform yet; urgent business case                 | Risk of a silo if hand-over is not planned and funded from the start       |

In practice, most institutions will run **B with elements of C**: a central control plane that also registers and monitors vendor-embedded agents, and domain teams that build where the institution's own knowledge differentiates. Option D is the usual starting point when the central platform is not ready (Section 10). An "inner source" rule applies throughout: shared components built by a project are contributed back, with the central team as maintainer.

## Standards a central function should set from day one

Even before a central platform exists, a small central function can set the standards that make later consolidation cheap:

- Approved models, regions and providers, with DORA third-party assessments.
- Tool interface standard (MCP) and naming, versioning and approval-token conventions for write tools.
- Trace format (OpenTelemetry) and minimum audit fields per case.
- Agent registry entry: owner, purpose, autonomy level, tools, data classes, evaluation results.
- Risk classification method for agent use cases, agreed with model risk, compliance and data protection.

## 06 Technology landscape

Every major cloud provider, Amazon Web Services (AWS), Microsoft Azure and Google Cloud, now offers a complete agent control plane, so an institution's main technology choice is which estate to anchor on, not whether the components exist. The table names concrete products as of October 2026; they are examples, not endorsements, and availability in EU regions must be confirmed at procurement.

## Hyperscaler agent platforms

| Capability                          | AWS                                                                                         | Microsoft                                                                        | Google Cloud                                                                           |
|-------------------------------------|---------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| <b>Agent runtime</b>                | Amazon Bedrock AgentCore Runtime (general availability (GA) October 2025) [42]              | Microsoft Foundry Agent Service, hosted agents [46]                              | Gemini Enterprise Agent Platform, Agent Runtime (formerly Vertex AI Agent Engine) [47] |
| <b>Agent framework</b>              | Strands Agents; any framework                                                               | Microsoft Agent Framework 1.0 (successor to AutoGen and Semantic Kernel) [46]    | Agent Development Kit (ADK)                                                            |
| <b>Tool gateway</b>                 | AgentCore Gateway (application programming interfaces (APIs) and MCP servers as tools) [42] | Azure API Management AI gateway (AI Gateway tier in preview since August 2026)   | Agent Gateway with Model Armor; Apigee MCP support [47]                                |
| <b>Agent identity</b>               | AgentCore Identity                                                                          | Microsoft Entra Agent ID [45]                                                    | Agent Identity [47]                                                                    |
| <b>Policy on tool calls</b>         | AgentCore Policy, Cedar rules checked outside agent code (GA March 2026) [43]               | Agent 365 (registry, least-privilege access, Conditional Access for agents) [44] | Agent Gateway policies                                                                 |
| <b>Evaluation and observability</b> | AgentCore Observability and Evaluations (GA March 2026)                                     | Foundry evaluation and tracing; Purview, Defender                                | Agent simulation and evaluation; anomaly detection                                     |
| <b>End-user front end</b>           | Amazon Quick Suite or custom                                                                | Microsoft 365 Copilot, Copilot Studio                                            | Gemini Enterprise                                                                      |

Models from Anthropic (Claude) are available on all three clouds; OpenAI models are available through Microsoft; Google's Gemini models through Google Cloud [48]. Using two providers through one model gateway reduces DORA concentration risk.

## Process, case and business-application platforms with agents

Complex business work usually already runs in a case or process system. These vendors now run agents as governed steps inside it:

| Category                              | Products (examples)                                                                                                                                                                                                                                                                                                                         | Role in the architecture                                              |
|---------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| <b>Process and case orchestration</b> | Pega Infinity 26 (governed orchestration of Pega and third-party agents, MCP, pricing per resolved case; GA July 2026) [49]; Camunda 8.9 (agents inside Business Process Model and Notation (BPMN) processes, MCP and A2A connectors, human-task listeners; April 2026) [50]; ServiceNow AI Agents; UiPath Maestro; IBM watsonx Orchestrate | Orchestration layer and work surface; human tasks and audit           |
| <b>Durable code-first workflows</b>   | Temporal (durable execution with agent software development kit (SDK) integrations) [58]; LangGraph                                                                                                                                                                                                                                         | Orchestration for engineering-led teams                               |
| <b>AI governance across vendors</b>   | ServiceNow AI Control Tower [51]; IBM Agentic Control Plane [52]; Microsoft Agent 365 [44]                                                                                                                                                                                                                                                  | Inventory and oversight of all agents, including vendor-embedded ones |
| <b>Finance and accounting</b>         | BlackLine Verity agents for reconciliation preparation [33]; SAP Joule agents for accruals, journal entries and close [35]; Workiva tie-out and disclosure agents with an MCP gateway [34]; Oracle Fusion agentic applications                                                                                                              | Domain agents inside the finance close                                |

| Category                              | Products (examples)                                                                                                                                                                      | Role in the architecture                                 |
|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|
| <b>Insurance core and reinsurance</b> | Duck Creek, Guidewire and Sapiens agent frameworks [41]; Brisc bordereaux and cash-matching agent [39]; SAP FS-RI and other reinsurance administration systems as systems of record      | Domain agents and tools for insurance operations         |
| <b>Risk and investment data</b>       | Moody's Agentic Solutions and insurance risk MCP server [26][40]; S&P Global [28]; Clearwater Analytics [36]; BlackRock Aladdin [37]                                                     | Data and domain agents for risk and investment reporting |
| <b>KYC, AML and screening</b>         | Fenergo FinCrime OS [30]; FIS Financial Crimes AI Agent [38]; WorkFusion (UiPath) [31]; Hawk [32]; Quantexa (entity resolution); LSEG World-Check, Dow Jones, Moody's for screening data | Domain agents, data and entity resolution                |
| <b>Financial data via MCP</b>         | S&P Global, LSEG, FactSet, Moody's, PitchBook, Dun & Bradstreet connectors [48]                                                                                                          | External data as governed tools                          |

## Gateways, standards and observability

| Capability                    | Products (examples)                                                                                                                                                                                                                                                                                            |
|-------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Model and MCP gateway</b>  | Kong AI Gateway 2.x with AI Registry and Token Vault (September 2026) [54]; LiteLLM (open source, MCP and A2A gateway) [55]; cloud-native gateways above                                                                                                                                                       |
| <b>Standards</b>              | MCP, current specification 2026-07-28 with hardened authorisation based on the OAuth standard and trace context from the World Wide Web Consortium (W3C) [53]; A2A 1.0 (April 2026) [25]; both under the Linux Foundation, whose Agentic AI Foundation counts banks such as Wells Fargo among its members [25] |
| <b>Tracing and evaluation</b> | Langfuse (open source, self-hostable; now part of ClickHouse) [56]; LangSmith with Fleet approval inbox [57]; Arize; OpenTelemetry into the institution's security information and event management (SIEM) system                                                                                              |
| <b>Document intelligence</b>  | Azure AI Document Intelligence, Google Document AI, Amazon Textract; Moody's QUIQspread for financial spreading                                                                                                                                                                                                |

## Build, buy or embed?

- **Buy the control plane** from the institution's strategic cloud and, where needed, one gateway vendor. Building identity, policy and tracing in-house no longer pays.
- **Embed where the process lives.** If the process runs in Pega, ServiceNow, SAP, Guidewire, Fenergo or nCino, start with that vendor's agent capability, connected to the institution's model gateway and registered in the central inventory.
- **Build where knowledge differentiates:** mandates, checklists, policy retrieval, domain tools, evaluation sets and the cross-system agents that no single vendor covers.

## 07 Use case A: reinsurance technical accounting

Agent-prepared processing of statements of account can cut technical accountants' time per statement substantially while every booking stays approved by a person. The illustrative pilot below takes about seven months, a team of about 12 full-time equivalents (FTE), roughly 325 person-weeks and about €1.75 million for a mid-sized European reinsurer.

Illustrative scenario. Effort, durations and costs are informal estimates for orientation, not quotes or benchmarks.

### The business task

A reinsurer receives statements of account and bordereaux from cedents (the insurers that buy reinsurance) and brokers every quarter, as PDF, Excel and, increasingly, structured messages. Each statement has to be read, assigned to the right treaty and period, checked against the contract terms, reconciled with cash and with earlier estimates, booked, and passed on to IFRS 17 and Solvency II reporting. Differences lead to queries to the broker or cedent. Most of a technical accountant's time goes into reading, re-keying, checking and chasing. The same pattern applies to a primary insurer's inward and outward reinsurance and to delegated-authority bordereaux.

**Pilot scope:** a reinsurer with about 3,000 treaties, about 20,000 statements and bordereaux a year and about 35 technical accountants. The pilot covers proportional and non-proportional property and casualty treaties in two regions. Out of scope: claims handling, treaty underwriting, commutations, and any change to IFRS 17 or Solvency II measurement.

### How the work is divided

Agents prepare the statement, the accountant approves, only approved entries post

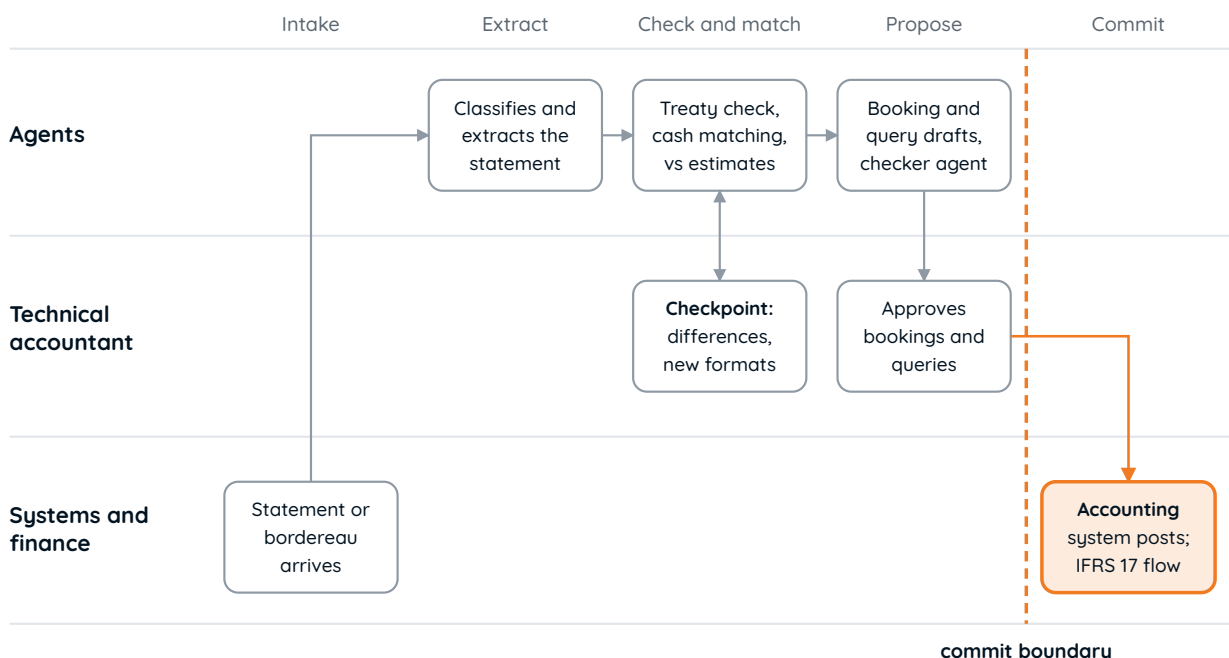


Figure 2. Use case A: statement of account flow, illustrative.

The technical accountant directs the work at the checkpoint and approves every booking and query; only approved entries cross the commit boundary into the accounting system.

| Step                        | Agent role                                                                                                         | Tools                                                           | Human control                                          | Autonomy                                           |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|--------------------------------------------------------|----------------------------------------------------|
| <b>1 Intake</b>             | Intake agent classifies incoming statements and bordereaux; identifies cedent, broker, treaty and period           | Mailbox and broker feeds; document store; document intelligence | Unidentified documents go to the accountant            | L1                                                 |
| <b>2 Extraction</b>         | Extraction agent reads figures into a standard technical-account structure                                         | Document intelligence; bordereaux templates                     | Low-confidence fields shown for confirmation           | L1                                                 |
| <b>3 Treaty check</b>       | Treaty agent compares figures with contract terms: shares, commission, brokerage, premiums, reinstatements, limits | Contract administration system; deterministic rule checks       | Breaches of tolerance trigger a checkpoint             | L1                                                 |
| <b>4 Cash and estimates</b> | Matching agent matches balances with cash received and compares with estimates and prior periods                   | Bank statements; matching engine; technical accounting system   | Unmatched items go to the accountant                   | L2; later L3 for matches within tolerance, sampled |
| <b>5 Booking proposal</b>   | Booking agent prepares entries per treaty and period                                                               | Technical accounting system (write, via approval)               | Accountant approves, edits or rejects                  | L2                                                 |
| <b>6 Queries</b>            | Query agent drafts queries to brokers or cedents with the differences and evidence                                 | Email or broker platform (write, via approval)                  | Accountant releases                                    | L2                                                 |
| <b>7 Check</b>              | Checker agent verifies figures, sources and checklist coverage                                                     | Number check; accounting rules                                  | Findings shown before approval                         | -                                                  |
| <b>8 Commit</b>             | None                                                                                                               | Technical accounting system posts; IFRS 17 data flow            | Accountant approves; finance controls the period close | -                                                  |

The agent never posts and never computes IFRS 17 or Solvency II figures. Contract terms come from the contract system, calculations from deterministic rules and engines, and postings from the accounting system after approval. Vendors report that agents can prepare reconciliations and bordereaux matching at high automation rates [33][39]; the pilot has to show this on the institution's own statements.

## Reference setup

The example assumes a reinsurer on Microsoft Azure with Entra ID and Microsoft 365, SAP S/4HANA with a reinsurance module such as SAP FS-RI (or a comparable reinsurance administration system), and Camunda as the firm's process engine. Equivalent choices on AWS or Google Cloud are listed in Section 6.

| Module                        | Reference choice (example)                                                                                                                                                 | Alternatives                                               |
|-------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|
| <b>Models</b>                 | Claude (larger model for analysis and drafting, small model for classification and checks) via Microsoft Foundry in an EU region; second provider through the same gateway | Azure OpenAI; Gemini or Claude on Google Cloud; Bedrock EU |
| <b>Model and tool gateway</b> | Azure API Management AI gateway; or Kong AI Gateway                                                                                                                        | LiteLLM                                                    |

| Module                      | Reference choice (example)                                                                                                    | Alternatives                                                           |
|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| Agent runtime and framework | Foundry Agent Service with Microsoft Agent Framework or Claude Agent SDK                                                      | LangGraph on Azure Kubernetes Service (AKS)                            |
| Orchestration               | Camunda 8 process with human tasks; agents as service tasks                                                                   | Pega; SAP Build Process Automation                                     |
| Agent identity              | Entra Agent ID; on behalf of the accountant                                                                                   | Service principals with scoped roles                                   |
| Documents and bordereaux    | Azure AI Document Intelligence with standard bordereaux templates                                                             | Google Document AI; Brisc as a specialist buy option [39]              |
| Matching and reconciliation | BlackLine with Verity agents, or the reinsurance system's own matching [33]                                                   | In-house matching engine                                               |
| Systems of record           | SAP FS-RI and S/4HANA via APIs, with SAP Joule finance agents where licensed [35]; contract administration; IFRS 17 subledger | Sapiens or other reinsurance administration software; in-house systems |
| Knowledge                   | Azure AI Search over accounting manuals, treaty wordings and query templates                                                  | pgvector                                                               |
| Work surface                | Review and approval screens in Camunda Tasklist or SAP Fiori; Excel add-in for statement review                               | Custom web front end                                                   |
| Tracing and evaluation      | OpenTelemetry to self-hosted Langfuse and the SIEM; golden-case suite in the continuous integration (CI) pipeline             | Foundry evaluation; LangSmith                                          |
| Inventory                   | Agent 365 or ServiceNow AI Control Tower                                                                                      | Central registry in the platform                                       |

## Team, effort and cost

| Workstream                       | Roles                                                                                 | FTE | Weeks | Person-weeks |
|----------------------------------|---------------------------------------------------------------------------------------|-----|-------|--------------|
| 1 Document intake and extraction | Data engineers, document AI engineer                                                  | 2.5 | 20    | 50           |
| 2 Domain tools and integrations  | Backend and SAP engineers (contract system, technical accounting, cash, IFRS 17 data) | 2.5 | 20    | 50           |
| 3 Agents and rule templates      | AI engineers                                                                          | 2.0 | 22    | 44           |
| 4 Orchestration and work surface | Process engineer, front-end engineer                                                  | 1.5 | 20    | 30           |
| 5 Minimum control plane          | Platform and security engineers                                                       | 1.5 | 20    | 30           |
| 6 Evaluation and validation      | Quality engineer; internal control (part-time)                                        | 1.5 | 20    | 30           |
| 7 Product and delivery           | Product owner from technical accounting, delivery lead                                | 2.0 | 28    | 56           |

| Workstream                          | Roles                                                              | FTE | Weeks     | Person-weeks |
|-------------------------------------|--------------------------------------------------------------------|-----|-----------|--------------|
| 8 Accounting subject-matter experts | Senior technical accountants, treaty specialists, finance controls | 1.5 | 24        | 36           |
| <b>Total</b>                        | about 12 FTE on average, 15 at peak                                |     | <b>28</b> | <b>326</b>   |

- **Budget:** about €1.75 million including 20% contingency (range €1.4–2.2 million), at a blended rate of about €1,000 per day for engineering; subject-matter expert (SME) time counted separately. If the institution already runs a central agent platform, workstream 5 shrinks to configuration and the budget falls by roughly €0.15 million.
- **Model usage:** a statement run of about 300,000 input and 15,000 output tokens costs roughly \$1 at list prices in the region of \$3 and \$15 per million tokens for a mid-sized model; 20,000 statements a year come to about \$22,000. People and integration dominate the cost.
- **Run cost after the pilot:** platform and cloud infrastructure of an estimated €8–20k per month, plus licences for reconciliation or specialist software where bought.

### Success measures for the go/no-go decision

| Measure                                                          | Target at end of pilot                |
|------------------------------------------------------------------|---------------------------------------|
| Accountant hours per statement                                   | at least 40% lower than baseline      |
| Postings without human approval                                  | 0                                     |
| Figures in proposals not matching the statement or a tool result | 0 released                            |
| Booking proposals accepted without change                        | at least 80%                          |
| Differences missed by the agent but found in QA                  | no worse than analyst-only processing |
| Days from statement receipt to booking                           | at least 50% lower than baseline      |

## 08 Use case B: KYC periodic review and AML alert investigation

Financial-crime operations are the strongest candidate for scaled agent work in banks and in life insurers: high volumes, documented procedures and a large share of routine cases. The illustrative pilot below takes about eight months, roughly 400 person-weeks and about €2.1 million; unlike use case A, it reaches autonomy level L3 for a narrow set of clear cases. Santander already handles about 100,000 AML alerts a year at Openbank with AI, and HSBC and FIS are building agentic financial-crime investigation [9][10][38].

Illustrative scenario. Effort, durations and costs are informal estimates for orientation, not quotes or benchmarks.

### The business task

Two processes share most of their data and tools:

- **Periodic KYC review.** Every client is reviewed on a risk-based cycle: identity and ownership data refreshed, beneficial owners verified against registers, screening re-run, the risk score updated and the file documented.
- **AML alert investigation.** Transaction-monitoring (TM) alerts are investigated at level 1: the transactions are put in context of the client profile, counterparties are checked, and the alert is closed with a rationale or escalated. Escalated cases may lead to a suspicious activity report (SAR) to the Financial Intelligence Unit (FIU).

**Pilot scope:** a mid-sized bank with about 30,000 periodic reviews and about 40,000 alerts a year and about 120 people in KYC and AML operations. The pilot covers low- and medium-risk corporate clients and two transaction-monitoring scenarios. A life insurer, which is an obliged entity under the EU Anti-Money Laundering Regulation (AMLR) for its life and investment-related business, can use the same design with policy administration in place of core banking [73]. Out of scope: high-risk clients, onboarding decisions, sanctions true-match handling, SAR filing and any exit decision.

## How the work is divided

Agents investigate, analysts decide; only clear cases close at L3, sampled daily

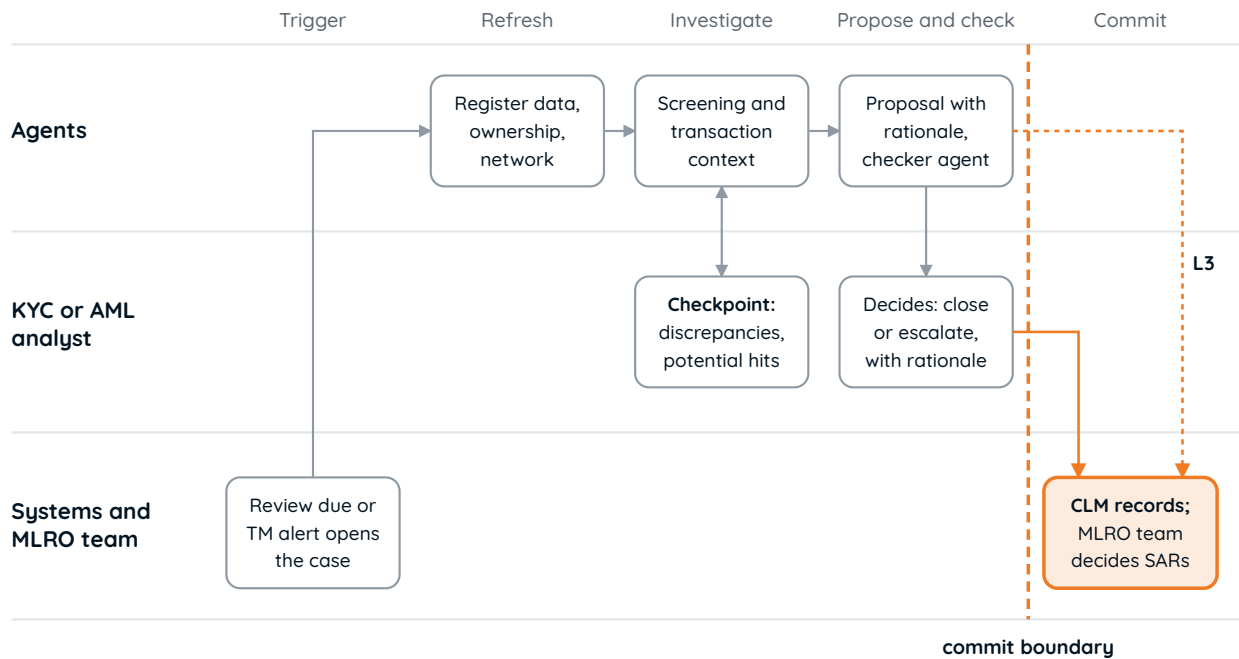


Figure 3. Use case B: KYC review and AML alert flow, illustrative.

The solid path is the default: the analyst decides every case. The dotted L3 path opens only after the pilot, for alerts that meet strict no-risk criteria, with a daily human sample; SAR decisions always stay with the MLRO team.

| Step                           | Agent role                                                                  | Tools                                                                                                                                                                              | Human control                             | Autonomy                            |
|--------------------------------|-----------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|-------------------------------------|
| <b>1 Trigger</b>               | Case engine opens a review or alert case                                    | CLM or case system; TM system                                                                                                                                                      | Team lead assigns or auto-routes          | –                                   |
| <b>2 Data refresh</b>          | KYC agent compares the file with registers and client data                  | German transparency and commercial registers (Transparenzregister, Handelsregister); Global Legal Entity Identifier Foundation (GLEIF); core banking or policy administration; CLM | Discrepancies go to the analyst           | L1                                  |
| <b>3 Ownership and network</b> | Network agent resolves entities and ownership chains                        | Entity resolution (e.g. Quantexa); register data                                                                                                                                   | Complex structures go to the analyst      | L1                                  |
| <b>4 Screening</b>             | Screening agent re-runs and pre-assesses name-screening hits                | Screening service (e.g. LSEG World-Check, Dow Jones)                                                                                                                               | Potential true matches always to a person | L2, or L3 for clear false positives |
| <b>5 Transaction context</b>   | Investigation agent analyses alerted transactions against profile and peers | TM system; payments data; counterparty data                                                                                                                                        | Unusual patterns trigger a checkpoint     | L1                                  |

| Step                            | Agent role                                                                          | Tools                                                                    | Human control                                          | Autonomy |
|---------------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------|--------------------------------------------------------|----------|
| <b>6 Documents and outreach</b> | Outreach agent drafts a request for missing documents or explanations               | Client portal or email (write, via approval)                             | Analyst releases; L3 later for standard requests       | L2 → L3  |
| <b>7 Proposal</b>               | Disposition agent proposes close or escalate, with a written rationale and evidence | Procedure retrieval; risk scoring engine                                 | Analyst decides                                        | L2       |
| <b>8 Check</b>                  | Checker agent verifies completeness, sources and consistency with the procedure     | Checklist rules; number check                                            | Findings to the analyst                                | –        |
| <b>9 Commit</b>                 | None                                                                                | CLM or case system closes the case; risk score set by the scoring engine | Analyst closes or escalates; MLRO team decides on SARs | –        |

**Sampled auto-closure.** After the pilot has shown its accuracy, alerts that meet strict, predefined criteria for no risk indicators can be closed by the agent at L3, with a fixed share sampled by people every day. The draft guidelines of the EU Anti-Money Laundering Authority (AMLA) on ongoing monitoring take the same line: automated closure only where no suspicious activity or risk indicators are found, with explainable, tested and documented monitoring [61]. Escalation and SAR decisions always stay with people.

## Reference setup

This example deliberately uses a different estate from use case A: an institution on AWS, Okta or Entra identity, and Pega as the case-management platform for financial-crime operations.

| Module                                          | Reference choice (example)                                                                                            | Alternatives                                                  |
|-------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|
| <b>Models</b>                                   | Claude on Amazon Bedrock in the Frankfurt region; small model for classification and screening pre-assessment         | Other Bedrock models; Azure or Google equivalents             |
| <b>Agent runtime, gateway, identity, policy</b> | Amazon Bedrock AgentCore: Runtime, Gateway, Identity, Policy (Cedar rules on every tool call), Observability [42][43] | Kong AI Gateway with self-hosted runtime                      |
| <b>Agent framework</b>                          | Strands Agents or Claude Agent SDK                                                                                    | LangGraph                                                     |
| <b>Case orchestration and work surface</b>      | Pega Infinity: cases, human tasks, SLAs, Pega-governed agent steps [49]                                               | Fenergo CLM with FinCrime OS agents [30]; Camunda; ServiceNow |
| <b>Specialist AML agents (buy option)</b>       | WorkFusion or Hawk for alert pre-investigation, connected through the same gateway [31][32]                           | In-house agents as above                                      |
| <b>Entity resolution</b>                        | Quantexa                                                                                                              | In-house graph on Neptune or Neo4j                            |

| Module                           | Reference choice (example)                                                                         | Alternatives                |
|----------------------------------|----------------------------------------------------------------------------------------------------|-----------------------------|
| Screening and registers          | LSEG World-Check or Dow Jones; register data providers; GLEIF                                      | Moody's KYC data            |
| Documents                        | Amazon Textract plus a document classification agent                                               | Azure or Google document AI |
| Knowledge                        | Retrieval over AML procedures, typologies and FIU guidance (Bedrock Knowledge Bases or OpenSearch) | pgvector                    |
| Tracing, evaluation, QA sampling | AgentCore Observability and Evaluations; Langfuse; daily QA sample in Pega                         | LangSmith; Arize            |

## Team, effort and cost

| Workstream                            | Roles                                                           | FTE | Weeks     | Person-weeks |
|---------------------------------------|-----------------------------------------------------------------|-----|-----------|--------------|
| 1 Data and entity resolution          | Data engineers                                                  | 2.5 | 22        | 55           |
| 2 Domain tools                        | Backend engineers (CLM, TM, screening, core banking, registers) | 3.0 | 22        | 66           |
| 3 Agents and rationales               | AI engineers                                                    | 2.0 | 24        | 48           |
| 4 Case orchestration and work surface | Pega developers                                                 | 2.0 | 22        | 44           |
| 5 Minimum control plane               | Platform and security engineers                                 | 1.5 | 20        | 30           |
| 6 Evaluation, QA sampling, validation | Quality engineers; model risk (part-time)                       | 2.0 | 22        | 44           |
| 7 Product and delivery                | Product owner from financial-crime operations, delivery lead    | 2.0 | 32        | 64           |
| 8 KYC, AML and compliance experts     | Senior investigators, compliance, MLRO office                   | 2.0 | 28        | 56           |
| <b>Total</b>                          | about 13 FTE on average, 17 at peak                             |     | <b>32</b> | <b>407</b>   |

- **Budget:** about €2.1 million including 20% contingency (range €1.8–2.7 million) at the same rate assumptions as use case A; buying a specialist AML agent shifts part of workstreams 3 and 6 into licence cost.
- **Model usage:** about \$2 per periodic review and about \$1 per alert at the list prices used above; at full volume roughly \$100,000 a year.
- **Benefit logic:** the business case rests on handling time per case and on backlog reduction. Vendor figures of 45% faster periodic reviews and 60% auto-closure of one alert type indicate the range, but must be confirmed on the institution's own data [30][31].

## Success measures for the go/no-go decision

| Measure                                         | Target at end of pilot                                    |
|-------------------------------------------------|-----------------------------------------------------------|
| Known true positives in the test set escalated  | 100%                                                      |
| Handling time per periodic review and per alert | at least 40% lower than baseline                          |
| QA findings on agent-prepared cases             | no worse than analyst-only cases                          |
| Error rate of sampled L3 closures               | below the threshold agreed with compliance before go-live |
| Rationales accepted without material edits      | at least 80%                                              |
| Backlog of overdue reviews                      | falling month on month                                    |

## 09 Estimating effort and timelines

A first agent use case for complex business work typically needs 250–450 person-weeks and six to eight months to reach a go/no-go decision. The estimate is driven far more by integrations, data and business availability than by the agents themselves. The template below can be applied to any candidate use case.

### Sizing template

| Building block                                                                                | Unit effort (person-weeks)                   | Comment                                           |
|-----------------------------------------------------------------------------------------------|----------------------------------------------|---------------------------------------------------|
| Minimum control plane (gateway, identity, tracing, registry)                                  | 25–35 once; 0–5 if a central platform exists | The largest saving from a central platform        |
| Read tool to an existing system with an API                                                   | 4–8 per system                               | Includes entitlement checks, tests, documentation |
| Write tool with approval token                                                                | 8–12 per action                              | Includes commit gate and reconciliation           |
| System without an API (screen, file or end-user computing (EUC) source such as a spreadsheet) | 10–20 per system                             | Or treat as a prerequisite project                |
| Agent role (instructions, tools, tests)                                                       | 6–10 per role                                | Planner, specialists, checker                     |
| Process modelling and human tasks                                                             | 1–3 per process step                         | In the institution's case or workflow engine      |
| Work surface                                                                                  | 10–30                                        | Less if embedded in an existing case tool         |
| Document intelligence                                                                         | 10–25                                        | Depends on document variety and language          |
| Evaluation set and harness                                                                    | 20–40, plus SME time                         | 100–300 golden cases with expected results        |

| Building block                  | Unit effort (person-weeks)                       | Comment                                      |
|---------------------------------|--------------------------------------------------|----------------------------------------------|
| Product, delivery and change    | about 20% of the total                           | Product owner from the business is essential |
| Business subject-matter experts | 15–25% of engineering effort, counted separately | Usually the bottleneck                       |

Costs follow from person-weeks at the institution's blended rate (about €5,000 per person-week at €1,000 per day), plus 20% contingency. Model usage is usually below 2% of the pilot budget; external data licences and the run organisation are the larger recurring items.

## What moves the estimate

| Driver                              | Effect                                                                   | How to reduce the uncertainty                                            |
|-------------------------------------|--------------------------------------------------------------------------|--------------------------------------------------------------------------|
| Integration readiness               | Systems without APIs can double tool effort                              | Inventory interfaces in the first month                                  |
| Data and document quality           | Poor scans, inconsistent master data and missing history slow every step | Sample 50 real cases before fixing the plan                              |
| Write actions and autonomy level    | Each write action adds a commit gate, reconciliation and controls        | Start at L1/L2; add write tools one at a time                            |
| Business definitions and checklists | Agreeing what "complete" means takes longer than building it             | Named mandate owner with authority                                       |
| Platform readiness                  | Without central identity, gateway and tracing, the project builds them   | Agree with the Chief Information Officer (CIO) what is provided and when |
| Regulatory classification           | High-risk or model-risk classification adds validation effort            | Classify in the first month with compliance and model risk               |
| Vendor choice                       | Buying domain agents shifts effort into integration and licence cost     | Run a short proof on the institution's own cases before committing       |

## A typical pilot plan

### A first use case reaches its go/no-go decision in about 28 weeks

Indicative plan in weeks; each gate widens the user group only if accuracy targets are met

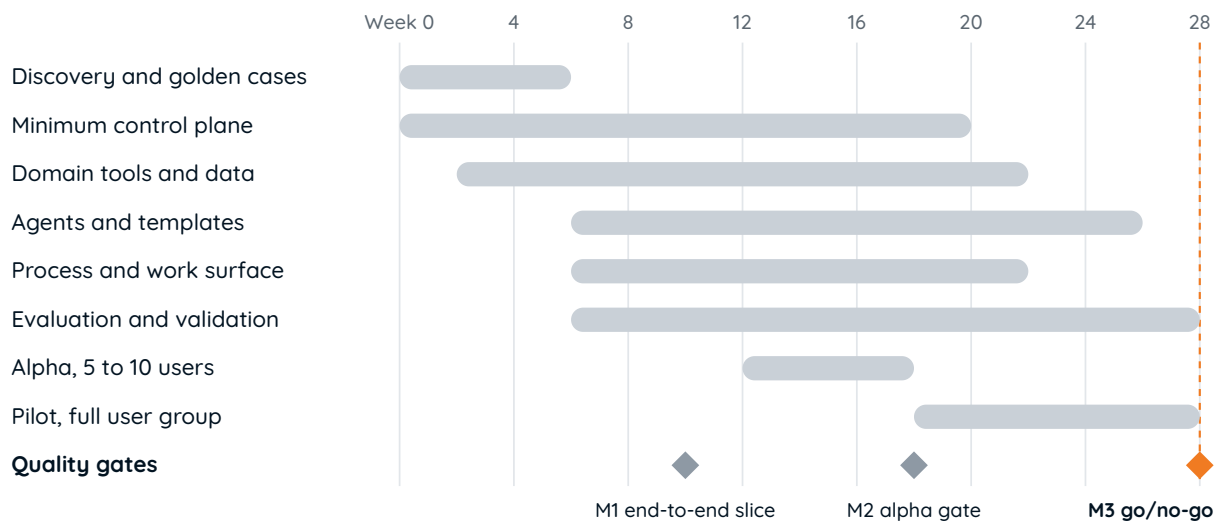


Figure 4. Indicative pilot plan: informal estimate.

Discovery and the golden case set start in week 0 because every later gate is measured against them; use case B needs about four weeks more than shown.

## 10 First steps when central infrastructure is missing or partial

Institutions do not need a finished platform to start; they need to start in a way that a platform can later absorb. Each starting position has a useful first step that produces real value and the inputs the full solution needs, without creating a silo.

## Starting positions and first steps

| What is in place                                                                                  | Useful first step                                                                                                                                                              | Value now                                           | Prepares for                                                               |
|---------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|----------------------------------------------------------------------------|
| Nothing approved yet                                                                              | Approve one enterprise assistant (e.g. Microsoft 365 Copilot, Gemini Enterprise, Claude Enterprise, ChatGPT Enterprise) for internal, non-client data; set minimum usage rules | Drafting, research, procedure Q&A                   | Skills, governance basics, demand overview                                 |
| Approved enterprise assistant only                                                                | Business teams build reusable projects or skills: checklists, memo templates, procedure retrieval; humans bring the documents                                                  | Time saved on drafting steps, measured from day one | Mandate templates, golden cases, prompts that move into the platform later |
| Cloud AI access, no agent platform                                                                | A lighthouse use case with a minimum viable control plane (see below), built for hand-over                                                                                     | First end-to-end use case at L1                     | The central platform's first components                                    |
| Business application with built-in agents (e.g. Guidewire, SAP, Pega, ServiceNow, Fenargo, nCino) | Switch on the vendor's agents inside the existing perimeter for one process; connect them to the institution's model policy and register them                                  | Fast value where the process already lives          | Inventory and control story for embedded agents                            |
| Partial platform (e.g. gateway and models, no agent identity or tracing)                          | Add tracing and an agent registry first, then delegated identity; run the first use case on what exists                                                                        | Use case live while the platform grows              | Platform completion driven by real demand                                  |

## Preparation that needs only an approved assistant

The most time-consuming inputs to any agent use case are business work and can start today, without client data:

| Task                                   | What the assistant helps with                                                                        | Who checks                              |
|----------------------------------------|------------------------------------------------------------------------------------------------------|-----------------------------------------|
| Write mandate templates and checklists | Turns procedures and accounting or KYC manuals into structured checklists with evidence requirements | Process owner                           |
| Build the golden case set              | Turns anonymised past cases and QA findings into test cases with expected results                    | Senior analysts, QA                     |
| Map the process and decision rights    | Drafts the process with checkpoints, approvers and autonomy level per step                           | Process owner, compliance               |
| Inventory interfaces and data sources  | Summarises system documentation; lists APIs, files and EUC tools per step                            | IT system owners                        |
| Draft the risk classification          | Applies the institution's AI and model risk method to the use case                                   | Model risk, compliance, data protection |

## The minimum viable control plane

If a lighthouse project has to build its own foundations, it should build exactly these, to central standards, and hand them over:

- **Model gateway** in front of all model calls (e.g. LiteLLM, Kong AI Gateway, or the cloud’s own gateway), with budgets and logging.
- **Tool gateway with MCP**, read-only by default, write tools only with approval tokens.
- **Delegated identity**: agents never hold their own database passwords; they act with the user’s token or a scoped service identity.
- **Tracing** with OpenTelemetry into a self-hosted store (e.g. Langfuse) and the SIEM.
- **Agent registry**, even if it is a governed table at first: owner, purpose, autonomy level, tools, data classes, evaluation results.
- **Evaluation harness** in the CI pipeline with the golden case set.

This typically costs 25–35 person-weeks (Section 9). Funded as a firm asset with a named future owner, it becomes the core of the central platform rather than a silo.

Rules that keep early steps from becoming technical debt

- **Build as if central**: use the standards a central function would set (Section 5), even if no central function exists yet.
- **Keep business assets portable**: mandates, checklists, golden cases and prompts are stored in the institution’s repositories, not only inside a vendor tool.
- **No shadow processes**: assistant-based work produces drafts inside the existing case tool; it never replaces the documented process or its approvals.
- **Contract for exit**: DORA exit plans and data-return clauses for every model and agent vendor from the first contract.
- **Agree the hand-over at the start**: who will own each shared component after the pilot, and how it is funded.

11

## Regulation and governance

There is no separate rulebook for AI agents in financial services; existing banking and insurance rules apply in full and assign accountability to people. Supervisors expect clear ownership, second-line review before material use cases go live, explainability, human oversight and DORA-compliant exit options [65][70].

| Framework                                          | What it means for agent work                                                                                                                                                                                                                                                                 | Status as of October 2026                                                                                                                                                                                                                                                     |
|----------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| EU AI Act                                          | High-risk uses (Annex III) include credit scoring of natural persons and risk assessment and pricing for natural persons in life and health insurance. Both use cases in this paper fall outside these categories. AI literacy and transparency duties apply to all uses                     | Digital Omnibus, Regulation (EU) 2026/1744, in force since 27 July 2026, moves stand-alone Annex III obligations to 2 December 2027 [62]. Germany’s Federal Financial Supervisory Authority (BaFin) will supervise financial firms’ AI under the German implementing law [69] |
| General Data Protection Regulation (GDPR), Art. 22 | Decisions with legal effect on natural persons may not rest solely on automated processing; the judgment of the Court of Justice of the European Union (CJEU) on the German credit bureau SCHUFA (C-634/21) shows that a heavily relied-on automated score can count as such a decision [63] | Relevant for any agent work on retail, private-client or policyholder cases                                                                                                                                                                                                   |

| Framework                                                                                                                | What it means for agent work                                                                                                                                                                                                                                                                               | Status as of October 2026                                                                                                                                                                                                                                                  |
|--------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>DORA</b>                                                                                                              | Model, cloud and agent vendors are information and communication technology (ICT) third-party providers for banks, insurers and reinsurers: assessment, contracts, exit plans, incident reporting                                                                                                          | Applies since January 2025; BaFin's December 2025 guidance places AI within DORA ICT risk management [64]. The ECB asked significant banks for action plans on AI-enabled cyber threats, including controls for AI agents as non-human identities, by 31 October 2026 [74] |
| <b>Governance rules in banking and insurance law (Capital Requirements Directive (CRD), Solvency II)</b>                 | Sound governance, four-eyes principle, outsourcing rules and documentation remain unchanged; agents fit as preparation, decisions stay with authorised people                                                                                                                                              | The Solvency II review, Directive (EU) 2025/2, applies from 30 January 2027 and changes reporting deadlines and frequencies [72]                                                                                                                                           |
| <b>Supervisory guidance on AI</b>                                                                                        | EIOPA's Opinion on AI governance and risk management expects proportionate data governance, record-keeping, fairness, explainability and human oversight [70]. European Banking Authority (EBA) guidelines require traceable inputs, human overrides and expert judgement for automated credit models [66] | EIOPA's 2026 survey: about two thirds of insurers use generative AI, mostly at proof-of-concept stage [71]. The EBA found no structural conflict between the AI Act and banking rules [68]                                                                                 |
| <b>EU AML package (AMLR, AMLA)</b>                                                                                       | Customer due diligence (CDD) and ongoing monitoring must be documented and explainable; automated alert closure only where no risk indicators are found [61]. Applies to banks and to insurers for life and investment-related business [73]                                                               | AMLR applies from 10 July 2027; AMLA direct supervision of selected entities from January 2028 [60]                                                                                                                                                                        |
| <b>Basel Committee risk data principles (BCBS 239) and ECB guide on risk data aggregation and risk reporting (RDARR)</b> | Risk data used by agents in banks must keep accuracy, completeness and lineage                                                                                                                                                                                                                             | ECB focus in 2026–28 [59]                                                                                                                                                                                                                                                  |
| <b>Model risk management</b>                                                                                             | The institution decides how agent components are classified and validated under its own framework                                                                                                                                                                                                          | In the US, supervisory letter SR 26-2 on model risk management (April 2026) replaced SR 11-7 and explicitly excludes generative and agentic AI, leaving it to institutions' own governance [67]                                                                            |

## Controls that make agent work supervisable

| Before the task                                    | During the task                         | At commit                                | After the task                         |
|----------------------------------------------------|-----------------------------------------|------------------------------------------|----------------------------------------|
| Use case risk classification and approval          | Entitlements enforced in tools and data | Human approval by authorised person      | Complete case file and trace           |
| Mandate template with checklist and autonomy level | Policy checks on every tool call        | Approval token required by write tools   | QA sampling and replay                 |
| Evaluation against golden cases                    | Number check and source citation        | Deterministic system executes            | Regression tests on every change       |
| Vendor and DORA assessment                         | Checkpoints and kill switch             | Four-eyes: drafting agent never approves | Second and third line access to traces |

12

## Risks and how to manage them

The largest risks of agent work are a plausible but wrong result reaching a decision-maker, and people approving agent proposals without real review. Both are managed by design, not by instructions to the model.

| Risk                             | How it shows                                                                        | How to manage it                                                                                                      |
|----------------------------------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| Plausible but wrong content      | A well-written memo or rationale contains a wrong figure or a missed fact           | Facts only from tools; number check; source per statement; checker agent; golden-case regression                      |
| Rubber-stamping                  | Approvers accept proposals without reading them as volume rises                     | Show evidence, not just conclusions; random hidden test cases; approval-rate and dwell-time monitoring; QA sampling   |
| Missed true positives            | An agent closes or downgrades a case that should have been escalated                | L2 for all dispositions until recall is proven; strict L3 criteria; daily sampling; 100% recall target on known cases |
| Entitlement breach               | An agent reaches data its user may not see                                          | Delegated identity; entitlements in tools and data; filtered retrieval; full audit                                    |
| Prompt injection                 | Instructions hidden in client documents or emails steer the agent                   | Documents treated as data; read-only tools; write tools behind approval tokens; injection filters in the gateway      |
| Uncontrolled actions             | An agent sends, books or changes something                                          | No commit without human approval token; deterministic execution; kill switch per agent type                           |
| Process drift                    | Agents quietly change how work is done; procedures and reality diverge              | Mandate templates under change control; agent registry; regression tests on every change                              |
| Vendor lock-in and concentration | Switching model or platform becomes expensive; one provider outage stops operations | Business assets in the institution's repositories; MCP and A2A; model gateway; two providers; DORA exit plans         |

| Risk                  | How it shows                                           | How to manage it                                                                              |
|-----------------------|--------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| <b>Skills erosion</b> | Junior staff stop learning the craft behind the drafts | Training paths that include unassisted cases; reviewers rotate; explain-why in every proposal |
| <b>Low adoption</b>   | Tools exist but people keep their old routines         | Product owner from the business; start with daily pain points; measure weekly use             |

13

## Roadmap and decisions requested

An institution can have two agent use cases in controlled production within about eight months and a scalable platform within two years. Each phase ends with a management gate based on measured results.

### Lighthouses first, then platform, then a federated factory

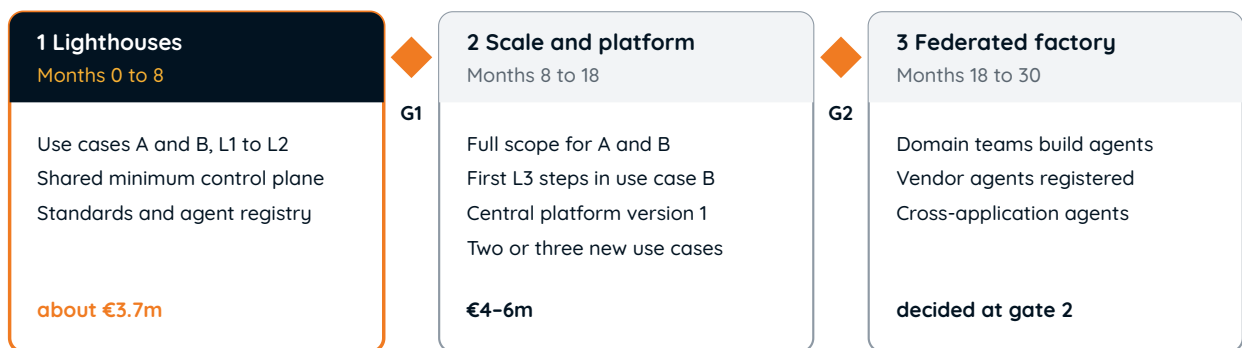


Figure 5. Programme roadmap: indicative.

Gate 1 opens phase 2 only if both pilots meet their success measures (Sections 7 and 8); gate 2 decides the platform's organisation model and the domain budgets.

| Phase                                     | Content                                                                                                                                                                                          | Indicative budget                                                   |
|-------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| <b>1 Lighthouses (months 0-8)</b>         | Use cases A and B at L1/L2; minimum control plane built once and shared; standards and agent registry                                                                                            | about €3.7 million (both pilots, shared control plane counted once) |
| <b>2 Scale and platform (months 8-18)</b> | Full scope for A and B, first L3 steps in B; central platform v1; two or three further use cases, e.g. Solvency II or regulatory reporting commentary, investment reporting, complaints handling | €4-6 million, depending on platform readiness                       |
| <b>3 Federated factory (months 18-30)</b> | Domain teams build on the platform; vendor-embedded agents registered and governed; cross-application agents                                                                                     | Run and change budget per domain, decided at gate 2                 |

Phase 2 and 3 figures are orientation only and depend heavily on the decisions in Section 5.

### The first 90 days

01

**Appoint sponsors and product owners.** A board-level sponsor; product owners from finance and from financial-crime operations with authority over priorities.

- 02 **Agree the control model with the second and third lines.** Autonomy levels, commit gates, evidence contract, risk classification method.
- 03 **Decide the organisation model** (Section 5) and what the central function provides in year one.
- 04 **Confirm providers and hosting.** EU-region model endpoints, two providers, DORA assessments, no training on the institution's data.
- 05 **Start the business assets.** Mandate templates, checklists and golden case sets for both use cases.
- 06 **Assess integration readiness.** Interfaces of contract administration, technical accounting, CLM, screening and TM systems; sample 50 real cases each.
- 07 **Fix success measures** for every gate, so that each decision rests on evidence.

#### DECISIONS REQUESTED

1. Endorse human intent, direction and control, the commit boundary and the evidence contract as binding design principles for all agent work.
2. Choose the organisation model, with the recommendation of a central control plane plus federated domain teams (Option B).
3. Approve the two lighthouse pilots (about €3.7 million together) with go/no-go decisions at month eight.
4. Mandate the central standards of Section 5 for every agent project, including vendor-embedded agents.
5. Nominate mandate owners and product owners in finance and financial-crime operations.
6. Commission the integration and data readiness assessments for both use cases.

# Glossary

| Term                        | Meaning                                                                                                                                |
|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| <b>A2A</b>                  | Agent2Agent protocol, an open standard for communication between agents                                                                |
| <b>Agent</b>                | Software built on a language model that plans and carries out a multi-step task with tools, under a mandate set by a person or process |
| <b>AML / AMLR</b>           | EU Anti-Money Laundering Authority / EU Anti-Money Laundering Regulation                                                               |
| <b>Approval token</b>       | A credential issued by the work queue when an authorised person approves; write tools refuse to act without it                         |
| <b>Bordereau</b>            | A periodic report from a cedent, broker or coverholder listing premiums or claims for a group of risks                                 |
| <b>CDD</b>                  | Customer due diligence                                                                                                                 |
| <b>CLM</b>                  | Client lifecycle management system (onboarding and KYC)                                                                                |
| <b>Commit boundary</b>      | The line between agent preparation and changes to the institution's records, which require human approval and deterministic execution  |
| <b>DORA</b>                 | Digital Operational Resilience Act, EU rules on ICT risk for banks, insurers and other financial entities, applying since January 2025 |
| <b>EIOPA</b>                | European Insurance and Occupational Pensions Authority                                                                                 |
| <b>FIU</b>                  | Financial Intelligence Unit, the authority receiving suspicious activity reports                                                       |
| <b>Golden cases</b>         | Test cases with known correct results, owned by the business, used to evaluate agents                                                  |
| <b>IFRS 17</b>              | International accounting standard for insurance contracts                                                                              |
| <b>KYC</b>                  | Know your customer                                                                                                                     |
| <b>MCP</b>                  | Model Context Protocol, an open standard for connecting agents to tools and data                                                       |
| <b>MLRO</b>                 | Money laundering reporting officer                                                                                                     |
| <b>OpenTelemetry</b>        | Open standard for traces and metrics from software systems                                                                             |
| <b>ORSA</b>                 | Own Risk and Solvency Assessment under Solvency II                                                                                     |
| <b>SAR</b>                  | Suspicious activity report                                                                                                             |
| <b>Solvency II</b>          | EU prudential framework for insurers and reinsurers                                                                                    |
| <b>Statement of account</b> | Periodic statement from a cedent or broker to a reinsurer showing premiums, commissions, claims and the balance for a treaty           |
| <b>Technical accounting</b> | Recording of premiums, commissions and claims under insurance and reinsurance contracts                                                |
| <b>TM</b>                   | Transaction monitoring                                                                                                                 |

# Sources

Links accessed October 2026. Vendor figures are as reported by the vendors and have not been independently verified. Sources [16], [46] and [53] are secondary reports.

## Banks

- [1] CeFPro Connect, [Inside JPMorgan's AI push](#), October 2025.
- [2] Forrester, [BNY built its digital workforce backward — and it's working](#), 2026.
- [3] AI News, [Goldman Sachs tests autonomous AI agents for process-heavy work](#), February 2026.
- [4] Banking Dive, [Citi deploys agentic tools to in-house AI platform](#), September 2025.
- [5] Moody's, [Major European bank transforms credit analysis with Moody's agentic solutions](#), April 2026.
- [6] Amalgamated Bank, [Collaboration with FIS and Anthropic on financial crimes](#), May 2026.
- [7] Commerzbank, [KI-Partnerschaften](#), July 2026.
- [8] CIO.de, [ING will Baukredite mit KI in 30 Minuten prüfen](#), January 2026.
- [9] Santander, [AI-first strategy turned into measurable impact](#), June 2026.
- [10] HSBC, [HSBC and Google Cloud announce AI banking partnership](#), June 2026.
- [11] Microsoft WorkLab, [The making of a Frontier Firm: how AI is redesigning work at BNY](#), May 2026.
- [12] Bloomberg Government, [Deutsche Bank, Goldman look to AI to flag trader misconduct](#), February 2026.

## Insurers

- [13] Allianz, [When the storm clears, so should the claim queue](#), November 2025.
- [14] FinTech Global, [Allianz and Anthropic partner to drive responsible AI](#), January 2026.
- [15] Reinsurance News, [AIG builds multi-agentic solution](#), May 2026.
- [16] Ideal Investisseur, [AXA partners with Publicis Sapient to deploy AI agents across the group](#), September 2026 (secondary).
- [17] Insurance Business, [Swiss Re CEO bets on AI for productivity gains](#), April 2026.
- [18] Zurich, [Agentic AI Hyper Challenge update](#), July 2025.
- [19] Qorus, [AI AML agent at Icatu Seguros](#), March 2026.
- [20] Milliman, [Could AI simplify the production of Solvency II and IFRS 17 reports? \(vendor case study\)](#).
- [21] IT Finanzmagazin, [Die Tech-Predictions der Versicherer](#), February 2026.

## Research and technology

- [22] F. Lei et al., [Spider 2.0](#), ICLR 2025.
- [23] BCG and SCBX, [The agent is the easy part](#), September 2026.
- [24] OODA Loop, [Over 40% of agentic AI projects will be scrapped by 2027, Gartner says](#), June 2025.
- [25] Linux Foundation, [A2A protocol surpasses 150 organisations](#), 2026; [Agentic AI Foundation welcomes 57 new members](#), August 2026.
- [26] Moody's, [Agentic Solutions announcement](#), September 2025.
- [27] nCino, [Analyst Digital Partner](#), April 2026.
- [28] S&P Global, [Credit Memo Builder launch](#), June 2026.
- [29] Anthropic, [Agents for financial services](#), May 2026.
- [30] Fenergo, [FinCrime Operating System with agentic AI](#), May 2025.
- [31] FinTech Global, [WorkFusion's Isaac rethinks the AML investigator's role](#), October 2026.
- [32] FinTech Global, [Hawk's AI agent matches analyst accuracy in AML reviews](#), September 2026.
- [33] BlackLine, [Governed AI for finance generally available](#), July 2026.
- [34] Workiva, [Specialized AI agents and intelligence layer](#), July 2026.
- [35] ERP Today, [SAP accruals agent for month-end close](#), September 2026.
- [36] Business Wire, [Clearwater Analytics brings AI into institutional investment workflows](#), June 2026.
- [37] MarketScreener, [Aladdin Wealth launches AI-enabled commentary tool](#), October 2025.
- [38] FIS, [FIS brings agentic AI to banking with Anthropic, starting with financial crimes](#), May 2026.
- [39] Coverager, [Brisac launches bordereaux CashOps AI agent](#), February 2026.
- [40] Moody's, [Insurance risk analytics for the agentic era](#), July 2026.

- [41] Insurance Canada, [Duck Creek launches native agentic AI platform](#), April 2026.
- [42] AWS, [Amazon Bedrock AgentCore generally available](#), October 2025.
- [43] AWS, [Policy in Amazon Bedrock AgentCore generally available](#), March 2026.
- [44] Microsoft, [Microsoft Agent 365](#).
- [45] Microsoft Learn, [What's new in Microsoft Entra Agent ID](#).
- [46] byteiota, [Foundry Agent Service GA](#), 2026 (secondary).
- [47] Google Cloud, [Google Cloud Next 2026 wrap-up](#), April 2026.
- [48] Anthropic, [Claude for Financial Services](#).
- [49] Business Wire, [Pega Infinity 26](#), July 2026.
- [50] Camunda, [Camunda 8.9](#), April 2026.
- [51] ServiceNow, [AI Control Tower expansion](#), May 2026.
- [52] IBM, [Introducing the agentic control plane](#), 2026.
- [53] MCP servers blog, [MCP specification 2026-07-28](#), July 2026 (secondary).
- [54] Kong, [Product releases](#), September 2026.
- [55] LiteLLM, [Release notes](#).
- [56] Langfuse, [Langfuse is joining ClickHouse](#), 2026.
- [57] LangChain, [Introducing LangSmith Fleet](#), March 2026.
- [58] Temporal, [OpenAI Agents SDK integration](#), March 2026.

## Regulation

- [59] ECB Banking Supervision, [Supervisory priorities 2026–28](#), November 2025.
- [60] FinTech Global, [EU's AMLA puts high-risk banks on notice for 2028](#), July 2026.
- [61] CMS, [AML draft guidelines on ongoing monitoring under Art. 26\(5\) AMLR](#), June 2026.
- [62] Mondaq, [EU Digital Omnibus on AI enters into force](#), July 2026.
- [63] Bird & Bird, [Key takeaways from the SCHUFA case of the CJEU](#), December 2023.
- [64] Rödl & Partner, [BaFin-Orientierungshilfe zu IKT-Risiken beim Einsatz von KI-Systemen](#), 2026.
- [65] ECB Banking Supervision, [Speech by Pedro Machado](#), 24 February 2026.
- [66] EBA, [Guidelines on loan origination and monitoring \(EBA/GL/2020/06\)](#), paras. 54–55.
- [67] Federal Reserve, OCC, FDIC, [SR 26-2 attachment](#), April 2026.
- [68] Regulation Tomorrow, [EBA letter on its AI Act mapping exercise](#), December 2025.
- [69] t3n, [So kontrolliert die BaFin künftig KI-Systeme](#), July 2026.
- [70] EIOPA, [Opinion on AI governance and risk management](#), August 2025.
- [71] EIOPA, [Survey on generative AI among Europe's insurers](#), February 2026.
- [72] EUR-Lex, [Directive \(EU\) 2025/2 amending Solvency II](#), January 2025.
- [73] Springlex, [AML Article 3: obliged entities](#).
- [74] KPMG, [ECB calls for an action plan to combat AI-driven cyberattacks](#), July 2026.