MAVERIC

Engineering Trusted AI in Banking

The Maveric Quadrant Framework for Moving AI from Isolated Intelligence to Governed Action

Maveric Systems · November 2026



> In KPMG's Q2 2026 AI Pulse survey of US banking leaders, 39% of banks reported that they had already deployed AI agents. Only 3% said accountability for AI-informed decisions sits with a centralized AI governance or risk committee.

Agents are reaching production faster than the structures that answer for them. For US regional and super-regional banks between $10 billion and $150 billion in assets, the gap is sharper still: the model-risk guidance most relevant to them now explicitly leaves generative and agentic AI out of scope.

Executive summary

AI in banking has moved from analytics that informed a human decision to agents that plan, retrieve data, call tools and act across a workflow. That shift changes what a bank has to trust. Accuracy of a single model output is no longer enough; the bank must show that a whole chain of decisions and actions stayed within purpose, authority and policy.

Existing controls for model risk, data, security, third parties and compliance remain necessary, yet AI crosses those domains faster than conventional governance joins them. The revised US model-risk guidance (SR 26-2) does not reach generative or agentic AI, and the Financial Stability Board's October 2026 sound practices call for risk-based triggers that bring a person into the loop.

The **Maveric Quadrant Framework** closes that gap at the level of the AI-enabled application. Its four quadrants, Agentic Approach, Contextual Data Catalog, Orchestration and Safe AI by Design, each govern one source of failure and each produce evidence. Banks adopt it one consequential journey at a time.



Why agentic AI changes the unit of trust in banking

Banking decisions have moved, over time, from analytics that supported human judgment to decision recommendations. Agentic AI goes a step further: it decomposes an objective, selects data and tools, delegates tasks and responds to intermediate results, progressing a workflow on its own.

Risk has grown at several levels with that shift. A conversational assistant that summarizes a policy creates one type of risk. An agent that interprets the policy, retrieves customer information, suggests treatment, triggers a case, drafts a communication and updates a system of record creates a connected chain of operational, conduct, model, data and compliance risk.

Adoption is advancing with some caution. The latest published Bank of England and FCA survey of UK financial services, from November 2024, found that 75% of responding firms were using AI and a further 10% planned to within three years. Only 2% of reported use cases were fully autonomous, while 55% already involved some automated decision-making. Near-term plans concentrated on internal-process optimization, customer support, financial-crime controls and regulatory reporting: the places where an output becomes embedded in a consequential workflow. A 2026 successor survey extends coverage to generative and agentic AI, with results expected by the end of 2026.

The unit of trust has therefore changed. Banks can no longer evaluate only whether a model is accurate. They must evaluate whether the application behaves within purpose across a sequence of decisions and actions.

Where existing bank controls break down

Most banks already have mature controls for information security, data management, model risk, third-party risk, operational resilience, privacy and compliance. Those controls exist; the difficulty is that AI crosses these control domains faster than conventional governance joins them.

A model may pass validation while retrieving data whose lineage or permitted use is unclear. An agent may hold valid credentials yet use them for a task outside the customer's intent. A workflow may generate accurate individual outputs while producing an unacceptable cumulative outcome. A third-party foundation model may change without an equivalent change being visible in the surrounding banking process. That last risk matters most to regional banks, where much AI arrives inside core and vendor platforms: in the UK survey, a third of all AI use cases were third-party implementations, and the OCC's Spring 2026 Semiannual Risk Perspective says it seeks to support community banks that leverage third-party technology as it reviews supervisory expectations for AI.

Does the revised US model-risk guidance cover generative and agentic AI?

No. On 17 April 2026, the Federal Reserve, OCC and FDIC issued revised model-risk guidance (SR 26-2), also published as OCC Bulletin 2026-13 and FDIC FIL-15-2026. It supersedes SR 11-7 and states that "generative AI and agentic AI models are novel and rapidly evolving. As such, they are not within the scope of this guidance." In a May 2026 speech, Federal Reserve Governor Michelle W. Bowman said the revised guidance "applies narrowly to traditional models and basic AI applications." The agencies have indicated that generative and agentic AI will be addressed separately through future dialogue and requests for information.

The guidance is also, in its own words, "expected to be most relevant to banking organizations with over $30 billion in total assets." That places many US regional banks inside the guidance for traditional models and outside any equivalent US guidance for the agents they are now deploying.

Internationally, the Financial Stability Board published its final sound practices for responsible AI adoption on 15 October 2026. As reported by Reuters, the final text strengthened its language on agentic AI and removed an earlier suggestion that firms treat agents as synthetic employees. Agents, the FSB says, do not reason like humans and may interpret pre-defined goals in unexpected ways. In higher-risk use cases, more intensive human oversight is critical, with clear, risk-based triggers for when an agent must seek approval.

Banks consequently need an engineering layer that connects established risk disciplines at the level where AI applications actually operate.

Regulation is moving closer to design

Regulatory expectations increasingly reach into the way AI systems are designed, documented and operated. For banks with EU customers or operations, the EU AI Act classifies AI used to evaluate the creditworthiness or credit score of natural persons as high-risk, with limited exceptions, and requires a documented risk-management process that continues throughout the system lifecycle. NIST's AI Risk Management Framework similarly treats trustworthiness as a combination of validity, reliability, safety, security, resilience, accountability, transparency, explainability, privacy and managed bias, organized through the functions Govern, Map, Measure and Manage.

Supervisors also benchmark data discipline against BCBS 239. The principles formally apply to systemically important banks, yet they set out the standard examiners and boards use for risk data: governance, data architecture, aggregation, accuracy, completeness, timeliness and useful reporting. More than a decade after the principles were issued, the Basel Committee still found that substantial work remained across banks and emphasized board responsibility, ownership and accountability for data quality.

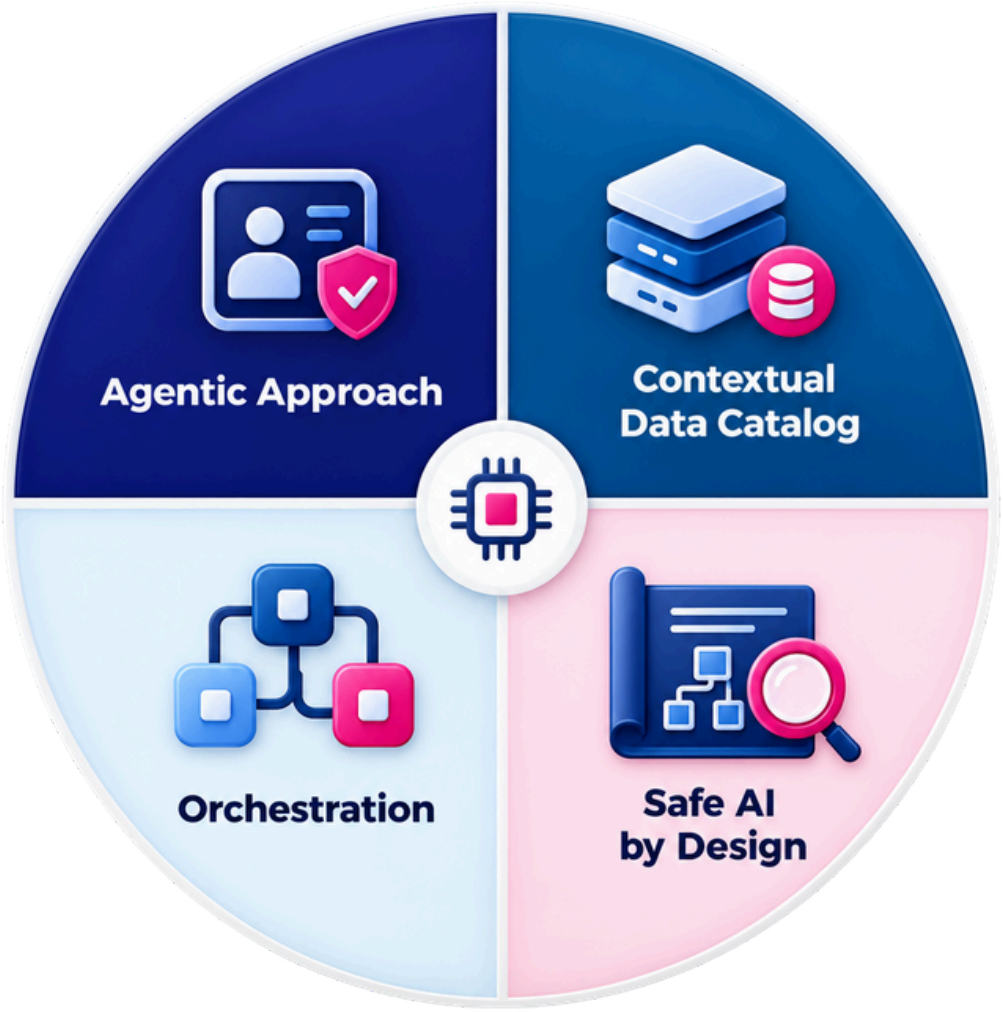
> These expectations point to the same conclusion: trust cannot be attached as a policy statement above an AI landscape. It must be expressed in the architecture and evidenced during execution.

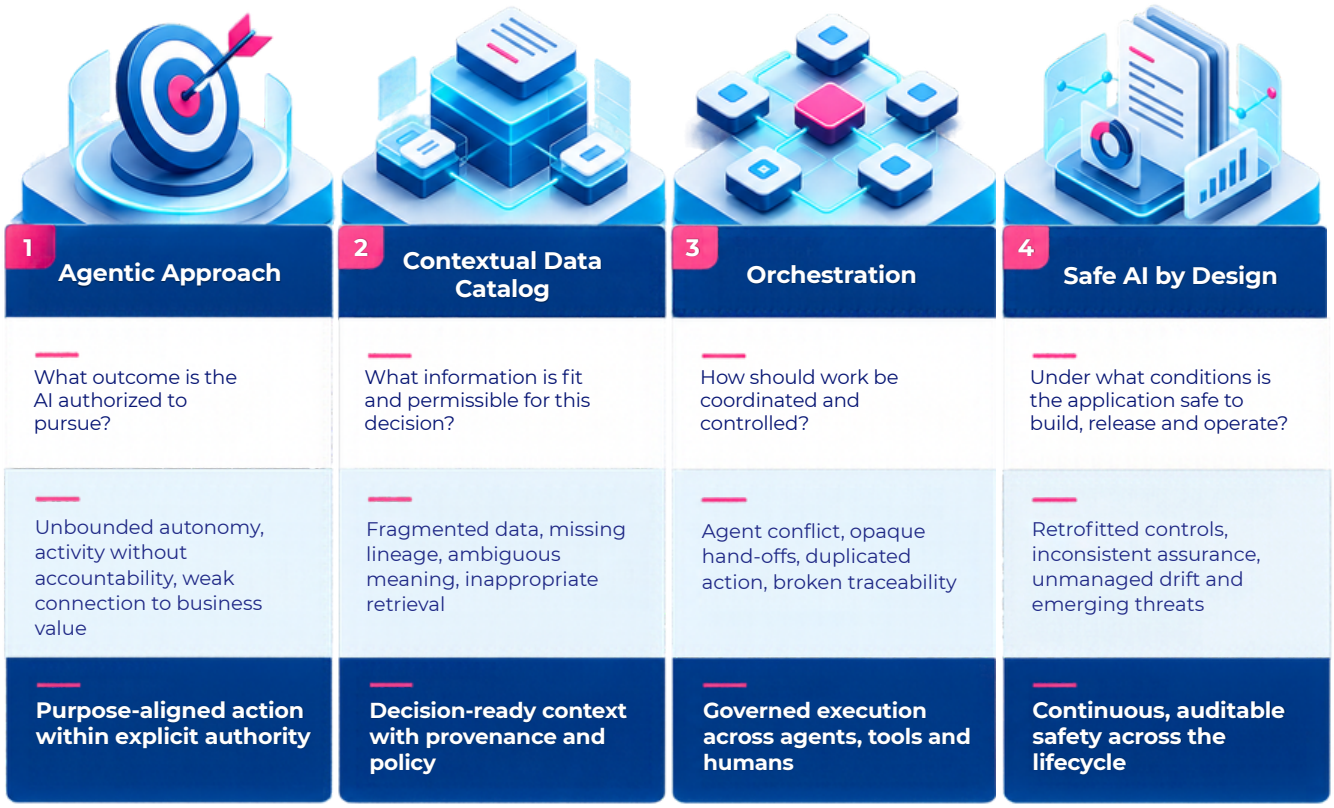


The Maveric Quadrant Framework for trusted AI in banking

A framework for trusted AI in banking has to connect intent, context, execution and assurance. The Maveric Quadrant Framework does this through four quadrants: an agentic approach, a contextual data catalog, orchestration, and safe-by-design principles embedded much earlier in the development cycle.

Two principles sit above the quadrants. AI is designed into the core of the solution from the start, rather than bolted onto an existing process. And the work starts with a well-defined outcome and a business value that can be measured. If either principle is missing, the quadrants become a taxonomy rather than an operating system. With both in place, the four quadrants below are non-negotiable if a use case is to reach production and deliver that value consistently.





Each quadrant has a distinct role, and together they form a system of trust.

The Agentic Approach converts business intent into bounded goals, roles and measurable outcomes.

The Contextual Data Catalog provides the governed information and decision context required to pursue those goals.

Orchestration controls the sequence through which agents, models, tools and people turn context into action.

Safe AI by Design establishes and verifies the conditions under which every component and interaction remains acceptable.

The framework is deliberately application-centric. It operationalizes a bank's existing enterprise governance, security, model-risk and data-management frameworks inside AI-enabled banking applications, and leaves those frameworks in place.

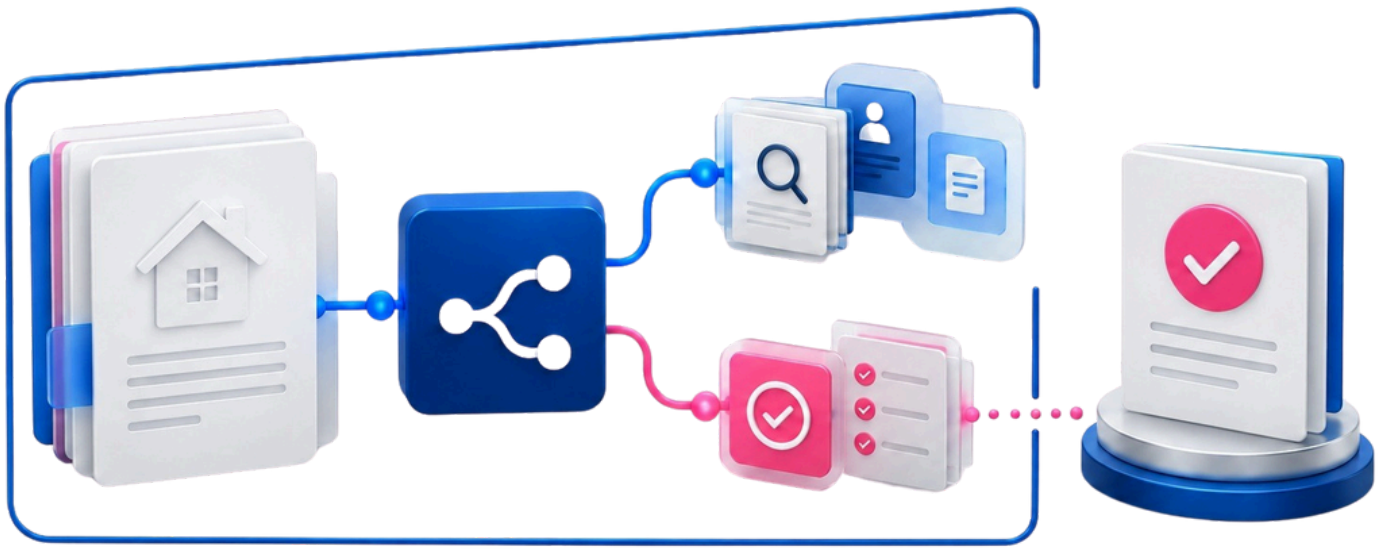
Quadrant 1: Agentic Approach

Why banking needs an agentic approach

> The agentic approach translates a banking objective into a controlled operating contract that defines what an agent may pursue, with what authority, and under whose accountability.

Many banking processes are extended journeys involving interpretation, investigation, exception handling, coordination and follow-through, rather than single decisions. A financial-crime alert may require transaction analysis, customer-history retrieval, typology matching, evidence collection, case prioritization and narrative preparation. A service complaint may cross policy interpretation, account investigation, hardship assessment, remediation calculation and communication.

Conventional automation works well when the path is stable and exceptions are limited. Agentic systems are valuable where the objective is clear but the route must adapt to context. Specialized agents can plan, execute and revise a sequence while drawing on different tools and knowledge domains. This makes them potentially powerful for complex banking work, and it also makes uncontrolled agentic systems unacceptable. The framework therefore begins with an agent's mandate before considering its capability.



The challenges it solves

Activity without outcome alignment.

AI initiatives frequently measure technical performance (accuracy, latency, token consumption or task completion) without proving an improvement in the banking outcome. An agent may process more alerts but increase false escalations; shorten handling time but reduce first-contact resolution; improve approval speed but worsen exception losses.

Authority without clear boundaries.

Agents require permissions to retrieve information and invoke tools. Excessive permissions enlarge the impact of error, manipulation or compromise. OWASP identifies excessive autonomy, tool abuse, goal hijacking, memory poisoning and high-impact action abuse among material agent risks, recommending least privilege, isolated execution and human approval for high-risk actions.

One agent doing everything.

A single agent that retrieves, decides and acts is hard to inspect and easy to over-privilege. Narrow roles keep authority small and review possible.

Local success with system-level failure.

An agent can complete itsassigned step correctly while undermining the broader customer or risk outcome. An eligibility agent may correctly identify an offer, for example, yet miss a hardship flag held elsewhere in the process.

No accountable owner.

When objectives are loosely written and actions are distributed across agents, responsibility can become ambiguous. Banking trust requires every autonomous action to resolve to an approved business purpose, a system owner and, where necessary, an accountable human decision-maker.

How the agentic approach works

Five design elements establish the operating contract.

Outcome definition: Every agent begins with an approved outcome, such as reducing false-positive investigation effort while maintaining detection effectiveness, improving complaint resolution without weakening conduct controls, or shortening lending turnaround while preserving policy adherence and fairness. The outcome includes value, risk and customer measures so optimization cannot silently trade one against another.

Role decomposition: Work is divided among narrow, inspectable roles: planner, retriever, analyst, policy checker, executor and validator. Specialization limits authority and makes each contribution easier to evaluate. No agent should both originate and unilaterally approve a high-impact action.

Bounded agency: Each role receives explicit permissions, tools, data domains, monetary or operational limits, prohibited actions, escalation conditions and time or resource budgets. Agent identity must be distinct and verifiable. NIST's work on agent identity and the FSB's October 2026 sound practices point the same way: do not share human credentials with an agent, and do not let a goal be reinterpreted without a trigger back to a person.

Decision rights: Actions are classified by materiality and reversibility. Low-risk, reversible steps may execute automatically. Material recommendations may require human review. Irreversible or high-impact actions, such as declining credit, moving funds, filing a regulatory report or changing a customer's status, require stronger validation and approval. Human oversight becomes risk-sensitive instead of being added indiscriminately to every step.

Outcome feedback: The system observes downstream results and feeds them back into agent evaluation. Trust is assessed through sustained business, customer, risk and compliance performance; task completion alone is insufficient evidence.

The result is accountable autonomy: agents can adapt how work is performed, while being unable to redefine why it is performed, enlarge their own authority or conceal the path to an outcome.

Evidence of trust

A bank should be able to inspect an agent register containing the agent owner, purpose, version, approved tasks, identity, entitlements, tools, data access, decision limits, escalation rules and current status. Execution records should connect each action to the originating objective and authorization. Performance monitoring should show both operational results and boundary events: rejected actions, permission violations, escalations, human overrides and goal-drift alerts.

With this evidence, the agentic approach becomes a controllable workforce of execution specialists instead of an abstract capability.

Quadrant 2: Contextual Data Catalog

What is a contextual data catalog in banking?

> A contextual data catalog is a governed semantic and policy layer that tells an AI application which data is fit, permissible and authoritative for a specific banking decision, and proves it.

AI quality depends on data quality, but banking decisions require more than technically clean data. They require data that is meaningful and permissible in context. A transaction amount may be accurate but stale for a real-time fraud decision. A customer attribute may be available but prohibited for a particular use. A policy document may be authoritative in one jurisdiction and obsolete in another. A hardship indicator may be essential for service treatment but inappropriate for marketing. The same data element can be valid, restricted, irrelevant or misleading depending on who is using it, for what decision, at which point in a process and under which regulation.

A conventional inventory is insufficient for this reason, and it is where data governance in banking has to extend into AI. A Contextual Data Catalog must connect metadata with banking meaning and decision policy.

The urgency shows in industry data. In the US, 63% of banks in KPMG's Q2 2026 survey cited data readiness and access as the biggest challenge to deploying AI agents. In the UK, four of the top five perceived current AI risks concerned data: privacy and protection, quality, security, and bias or representativeness. BCBS 239 also expects risk data to be accurate, complete, timely, adaptable and available across relevant business, legal-entity, asset, industry and regional dimensions.



The challenges it solves

Fragmented semantics.
Banking terms such as product or category names may have different definitions across products, entities and systems. An AI application can retrieve technically correct but semantically incompatible information.

Unknown provenance.
Without lineage, an application cannot show where a data element originated, how it was transformed, which quality checks were applied or whether an upstream change altered its meaning.

Retrieval without entitlement.
Generative systems can surface information through prompts or retrieved documents even when the user or agent should not access it. Permission must follow the data into the AI context instead of stopping at the source system.

Context loss.
Models often receive values without the policy, temporal, jurisdictional or customer context necessary to interpret them. Missing context creates plausible but unsafe conclusions.

Data-policy drift.
Consent, retention,regulatory classification, product terms and authoritative sources can change. Static controls do not reliably govern dynamic retrieval.

How the contextual data catalog works

The Contextual Data Catalog sits between enterprise information and AI action and combines six forms of context.



Business context.

Common banking terms, product hierarchies, customer relationships, process stages, risk taxonomies and KPI definitions establish what the data means.



Technical context.

Source, schema, format, transformation, lineage, owner, refresh rate and quality history establish where the data came from and how it was prepared.



Decision context.

The catalog records which data is authoritative for a use case, what supplementary evidence is required, acceptable freshness, known limitations and whether a value may support, recommend or determine an action.



Regulatory context.

Jurisdiction, legal basis, consent, retention, residency, sensitivity, protected characteristics and permitted-purpose rules establish whether and how the data may be used.



Operational context.

Process status, recent events, channel, device, prior decisions, open cases and exception history help the application interpret the customer's current situation.



AI context.

Embedding source, retrieval policy, prompt eligibility, training-use permission, model compatibility, grounding confidence and evaluation history establish how information may enter an AI system.

At runtime, the catalog should act as well as describe. A context service can resolve the requesting agent's identity, customer or account scope, purpose, process stage and jurisdiction; select approved sources; enforce field- and document-level policy; attach provenance and freshness; and return a decision-ready context package. Sensitive or low-confidence information can be masked, excluded or routed for review.

Data lineage is central to this operation. The Basel Committee's January 2026 implementation newsletter on BCBS 239 describes lineage as "the traceability of data from its origin to its final use," important for confirming data quality, and notes that robust data management becomes more important because the quality of AI-driven outputs depends on high-quality data.

From catalog to evidence

For every consequential AI-supported outcome, the bank should be able to answer:

- Which data and documents were used?
- Which versions and sources were authoritative at that moment?
- What transformations and quality checks occurred?
- What was excluded, masked or rejected, and why?
- Which consent, entitlement, residency and purpose rules applied?
- How fresh was the data, and what limitations were known?

> **Answering these questions creates contextual integrity: the assurance that an AI action was shaped by the right information, interpreted in the right banking context and used under the right authority.**

Quadrant 3: Orchestration

Why intelligence needs a control plane

Orchestration is the control plane that turns distributed intelligence into governed execution.

An AI-enabled banking application may contain multiple agents, predictive models, foundation models, rule engines, APIs, workflow platforms, data services and human work queues. Trust depends on how the components interact, so approving each one separately is insufficient.

Orchestration determines which agent acts, which model or tool it may use, what context it receives, when a human decision is required, how conflicts are resolved and what happens when confidence, policy or system conditions deteriorate.

It goes beyond workflow automation. A conventional workflow follows a predefined path. AI orchestration must control paths that may be selected dynamically while preserving the bank's non-negotiable rules.



The challenges it solves



Opaque hand-offs.

In a multi-agent process, the output of one component becomes the input to another. Without a shared execution record, the bank cannot reconstruct how an initial instruction became a final action.



Conflicting agents.

A service agent may optimize customer resolution while a fraud agent recommends restriction and a fair-treatment check flags the customer for special handling. The system needs explicit precedence, conflict-resolution and escalation policies.



Duplicated or runaway execution.

Retries, recursive planning or asynchronous events can create duplicate payments, repeated communications, excess API calls or uncontrolled cost. Resource budgets and circuit breakers are necessary controls for agentic operation.



Implicit trust between agents.

A sub-agent may assume that instructions from an orchestrator are valid. OWASP warns that compromise of a central orchestration layer can direct downstream agents at scale and recommends signed workflows, independent scope validation, strong authentication and network segmentation.



Human oversight at the wrong point.

Human approval added indiscriminately produces delay and consent fatigue; added too late, it becomes ceremonial. Oversight must occur before high-impact actions and at ambiguity, conflict or policy thresholds.

How orchestration works

The orchestration layer governs execution through seven mechanisms.

Plan registration: The proposed task plan is checked against the originating objective, approved workflow patterns and agent permissions. Material deviations trigger re-planning or human review.

Policy-aware routing: Routing considers more than availability. It evaluates data sensitivity, jurisdiction, model approval, use-case materiality, customer impact, latency and cost before selecting an agent, model or tool.

Identity and delegation: Each request carries the identities of the originating user or system and acting agents, the delegated purpose, permitted scope and expiry. Authority cannot expand as a task passes through an agent chain.

Separation of duties: Planning, execution, validation and approval are allocated to different roles where risk warrants. A lending recommendation, for example, can be independently checked for policy adherence, evidence sufficiency and fairness before it reaches a decision-maker.

State and conflict management: A shared, controlled process state prevents agents from acting on inconsistent versions of reality. Priority rules and escalation paths resolve conflicts between customer, risk, compliance and operational objectives.

Resilience controls: Timeouts, bounded retries, fallbacks, kill switches, queue controls and graceful degradation prevent a local failure from becoming a banking incident. When AI confidence or service availability falls below an approved threshold, the process moves to a deterministic rule, a human queue or a safe stop.

End-to-end observability: The layer records prompts or task instructions where appropriate, retrieved context references, model and agent versions, tool calls, policy checks, intermediate outputs, approvals, exceptions, timestamps and final actions. Sensitive reasoning content should not be indiscriminately retained; evidence must be sufficient for accountability while respecting privacy, security and record-management policy.

> Orchestration thereby creates execution integrity. The bank can prove what happened and also that each hand-off remained within identity, purpose, policy and control boundaries.

Quadrant 4: Safe AI by Design

Why safety has to run across the lifecycle

Safe AI by Design treats safety as a property engineered into every stage of an AI application's life, from first design to retirement.

Banks commonly govern technology through review points: architecture approval, model validation, security testing, compliance sign-off and production readiness. These remain necessary, but agentic applications can change their path at runtime, rely on changing external models and data, and exhibit behavior that was not visible in a static pre-release test.

Safety must therefore operate as a lifecycle property, with release review as one checkpoint among many. ISO/IEC 42001, adopted as a European standard in 2026 without a change to the 2023 requirements, frames AI governance as a management system that is continually improved: responsibility, risk, data governance, transparency, performance monitoring and corrective action. NIST's AI RMF likewise expects testing before deployment and regularly in operation. The 2026 supervisory record adds a sharper point. Because generative and agentic systems sit outside revised US model-risk guidance, a bank cannot assume that an SR 11-7-style validation will be accepted as the control. The control has to be designed for the application, including retrieval, delegation and tools.

Safe AI by Design also has a wider scope than model risk. It has to satisfy banking regulations and AI regulations, and it has to comply with engineering standards as well as banking standards. Security, fairness, explainability, human oversight, monitoring and guardrails are embedded from the start: what an agent may do, when a person enters the loop, and how decisions and data access are logged. The intent is to codify those guardrails once, so they apply consistently across the models an institution uses instead of being reinterpreted for each model.



The challenges it solves

Retrofitted controls

When security, fairness, privacy or explainability is added after application design, controls often monitor the model but miss the surrounding data retrieval, agent delegation and tool execution.

Static assurance

A one-time test does not capture model updates, prompt changes, data drift, new tools, emerging attacks or altered customer populations.

Disconnected risk disciplines

Security may assess prompt injection, model risk may assess performance, compliance may assess obligations and data teams may assess quality, without a unified view of the application's residual risk.

Evidence assembled after the event

If evidence is not produced during design and execution, audit and regulatory response become manual, slow and incomplete.

Guardrails that do not travel

A control written for one model will not hold when the bank adds another. Codified, reusable policy is the only way to stay consistent.

How Safe AI by Design works

Safe AI by Design translates trust principles into engineering requirements, tests, controls and operating evidence across the application lifecycle.

Design

The bank classifies the use case by materiality and customer impact; maps applicable obligations; identifies affected stakeholders; and defines prohibited outcomes, risk appetite, human-oversight points and safe-failure behavior. It further threat-models the full application, including agents, memory, retrieval, APIs and third parties.

Build

Developers apply secure coding, least privilege, approved model and data components, protected secrets, prompt and input controls, output filtering, privacy controls, provenance capture and segregated environments. Safety policy is expressed as reusable code and configuration wherever possible, reducing inconsistent manual interpretation.

Validate

Testing covers functional quality and banking consequences. It includes representative and edge-case scenarios, fairness analysis, explainability appropriate to the audience, privacy leakage, adversarial and prompt-injection testing, tool-abuse scenarios, agent goal drift, resilience, fallback behavior and human-factor testing. Explainable AI in banking has to serve several audiences: for high-risk decisions, the explanation must be sufficient for the affected customer, bank operator, validator and regulator, and technical availability alone falls short. The OCC's Spring 2026 Semiannual Risk Perspective lists lack of explainability first among the unique challenges of genAI and agentic AI.

Release

Deployment is permitted only when mandatory evidence exists, residual risks are accepted by the right authority and operational controls are ready. Model, agent, prompt, data-source and policy versions are bound into a release record. Material changes trigger proportionate re-evaluation.

Operate

Continuous monitoring detects performance and data drift, bias, abnormal tool use, permission violations, injection attempts, hallucination or grounding failures, unexpected escalation patterns, third-party changes and divergence from business outcomes. Thresholds trigger investigation, rollback, restricted mode or shutdown.

Release

Incidents, near misses, overrides, complaints, audit findings and outcome deterioration feed into revised policies, evaluations and architecture, so safety matures on the strength of evidence.

Safe AI by Design creates assurance integrity: the ability to demonstrate that trust requirements were considered, implemented, tested and monitored throughout the application's life.

How the four quadrants work together

The framework joins the four quadrants through explicit contracts, so that each contributes its own strength to engineering trust.

The Agentic Approach states the outcome and sets the boundaries of authority. The Contextual Data Catalog returns only information permitted for that purpose and attaches the meaning and provenance required for use. Orchestration ensures that each agent, model, tool and human operates in the approved sequence and that authority is preserved across hand-offs. Safe AI by Design defines the acceptance conditions, tests the system and monitors whether the complete application remains within tolerance.

Consider an AI-enabled retail-lending journey.



A customer submits an application. The agentic contract permits collection, evidence checking and recommendation, but no unilateral approval outside defined low-risk bands.



The Contextual Data Catalog resolves identity, consent, jurisdiction and product; retrieves approved affordability, exposure and policy data; and records source, freshness and permissible use.



Orchestration assigns document validation, fraud screening, affordability analysis and policy checking to specialized agents and services. Conflicting evidence or material exceptions move to a human underwriter.



Safe-AI controls test for discriminatory outcomes, explanation sufficiency, adversarial document manipulation, privacy leakage and model drift. The release and runtime records preserve the evidence required to review the decision.

Remove any quadrant and trust weakens. Agents without boundaries may exceed their decision rights. A system lacking contextual data may reach a defensible-looking answer from the wrong evidence. If orchestration is missing, hand-offs and exceptions become opaque, and controls that were never designed in stay inconsistent while assurance decays after deployment. That interdependence turns trust from a checklist into an operating architecture.

Engineering blueprint for the framework

The framework can be implemented as a set of logical services integrated with the bank's existing platforms, instead of as a monolithic product.

Capability layer	Illustrative components	Primary evidence produced
Outcome and agency	Use-case register, agent registry, outcome contracts, role and entitlement definitions, action limits	Ownership, purpose, authority, KPI and risk linkage
Context and data	Business glossary, data products, lineage, quality rules, policy tags, knowledge sources, runtime context service	Source, meaning, freshness, consent, entitlement and provenance
Execution and orchestration	Planner, policy engine, routing, agent identity, workflow state, approval service, conflict resolution, circuit breakers	Complete action chain, approvals, exceptions and safe-stop events
Assurance and operations	Risk classification, evaluation harness, red teaming, fairness and explainability testing, monitoring, incident response	Test results, residual risk, release decision, drift and control performance

Banks should integrate these capabilities with existing identity and access management, data governance, model inventory, API management, security operations, observability, case management, DevSecOps and governance-risk-compliance systems. The objective is to connect existing control systems around the lifecycle of the AI-enabled application, without creating another isolated AI control stack.



How should a bank adopt the framework?

Start with a consequential journey

A bank should resist the temptation to implement the framework as an enterprise taxonomy exercise. The strongest starting point is a bounded but consequential banking journey, such as complaints, financial-crime investigation, lending operations, payment exception handling or servicing, where outcome, data, workflow and risk can be observed together.

The initial use case should be important enough to test the trust architecture but bounded enough to establish clear ownership and evidence. A low-value assistant may not expose the real control problem; a fully autonomous high-impact decision may create avoidable implementation risk.

Progress through four stages

1 Frame



Define the banking outcome, customer and regulatory impact, risk appetite, accountable owner, human decision rights and measurable baseline. Map the existing process and identify where AI is permitted to recommend, decide or act.

2 Engineer



Create agent roles and permissions, curate contextual data products, implement orchestration controls and convert safety requirements into tests and runtime policies. Establish a common event model so every component contributes to one trace.

3 Prove



Run in observation or recommendation mode before increasing autonomy. Compare AI-supported and existing outcomes, test exceptions and adversarial scenarios, verify explanations, rehearse rollback and demonstrate evidence to risk, compliance, audit and operations stakeholders.

4 Scale.



Reuse policy patterns, agent identities, evaluation suites, context services and evidence schemas across applications. Expand autonomy only when outcome and control evidence support it. This prevents scale from becoming a multiplication of bespoke agents and disconnected controls.

How should banks measure trust in AI?

Trust should be measured through evidence across four dimensions. No single “trust score” can replace judgment, but a balanced scorecard can make deterioration visible and governance actionable.

Dimension	Illustrative measures
Outcome trust	KPI improvement, customer outcome, false-positive or false-negative rate, exception loss, resolution quality, benefit realized versus baseline
Data trust	Authoritative-source coverage, lineage completeness, freshness compliance, rejected retrievals, quality-rule pass rate, policy-tag coverage
Execution trust	Trace completeness, unauthorized-action attempts, approval compliance, conflict and escalation rate, duplicate-action prevention, fallback success
Assurance trust	Evaluation pass rate, fairness variance, explanation sufficiency, security findings, drift alerts, override rate, control time-to-remediate

Metrics must be interpreted by use case. A high human-override rate may signal poor agent performance, but it may also show that controls are correctly stopping ambiguous cases. A low escalation rate may indicate efficiency, or it may indicate under-detection. The framework therefore combines quantitative measures with review of cases, customer effects and control behavior.



The trust architecture for AI-first banking

Banking has always depended on delegated authority. Employees, systems, models and partners act within defined mandates, using approved information and producing evidence for oversight. Agentic AI introduces a new class of actor: one that can interpret, plan, collaborate and act at software speed. Suppressing agency would forfeit the value, and relying on model accuracy as a proxy for trust would leave the chain of actions unexamined. Banks need a way to make agency governable.

The Maveric Quadrant Framework provides that structure. It aligns AI with measurable banking outcomes, supplies decision-ready context, governs multi-agent execution and embeds safety into the application lifecycle. Each quadrant addresses a distinct source of failure; together they establish the evidence chain from intent to outcome.

The next 12 to 18 months will test that chain. US agencies have said generative and agentic AI will be addressed separately from SR 26-2, the FSB will monitor developments and update its practices as the technology evolves, and the Bank of England and FCA are due to publish their first survey results covering agentic AI. Banks that can already produce an agent register, a context record and an execution trace will answer those expectations from evidence, while the rest reconstruct it after the fact.

For AI-first banking, trust has to be engineered into every objective, data interaction, hand-off and action, and remain demonstrable each time the application operates. The test is a practical one:

> **If an examiner asked tomorrow for the agent register behind your most consequential AI workflow, including its owner, entitlements, decision limits and last boundary event, could your team produce it within a day?**

Sources:

- *KPMG, Q2 2026 AI Pulse Survey: Banking (fielded 28 April to 25 May 2026; US leaders at organizations with $1B+ revenue).*
- *Federal Reserve, OCC and FDIC, SR 26-2: Revised Guidance on Model Risk Management (17 April 2026).*
- *Schneider Downs, Banking Agencies Revise Model Risk Management Guidance (OCC Bulletin 2026-13, FDIC FIL-15-2026).*
- *Federal Reserve, Michelle W. Bowman, Artificial Intelligence in the Financial System (speech, 1 May 2026).*
- *OCC, Semiannual Risk Perspective, Spring 2026.*
- *Financial Stability Board, Sound Practices for Responsible Adoption of Artificial Intelligence (AI): Final Report (15 October 2026).*
- *Financial Stability Board, FSB finalises sound practices for the responsible adoption of AI (press release, 15 October 2026).*
- *Reuters via The Manila Times, FSB urges more oversight of agentic AI (16 October 2026). Bank of England and FCA, Artificial Intelligence in UK Financial Services, 2024 (21 November 2024).*
- *Bank of England, Financial Stability in Focus: Artificial Intelligence in the Financial System (April 2025).*
- *Burges Salmon, Keeping an Eye on AI in Financial Services: The AI Survey (2026 Bank of England and FCA survey).*
- *European Union, Regulation (EU) 2024/1689 (Artificial Intelligence Act).*
- *NIST, Artificial Intelligence Risk Management Framework (AI RMF 1.0).*
- *NIST AI Resource Center, AI RMF Core.*
- *Basel Committee on Banking Supervision, Basel Framework: Risk Data Aggregation and Risk Reporting Chapter.*
- *Basel Committee on Banking Supervision, Progress in Adopting the Principles for Effective Risk Data Aggregation and Risk Reporting (November 2023).*
- *Basel Committee on Banking Supervision, Press Release: Basel Committee Report on Implementing BCBS 239 (28 November 2023).*
- *Basel Committee on Banking Supervision, Implementation of the Principles for Effective Risk Data Aggregation and Risk Reporting (Newsletter 36, 6 January 2026).*
- *OWASP, AI Agent Security Cheat Sheet. NIST,*
- *Back to the Future: Why Agentic AI Needs a Strong Identity Foundation. OWASP Cornucopia, Agentic AI Card AAIQ: Orchestration Layer Compromise.*
- *OWASP Cornucopia, Agentic AI Companion AAIQ 1.0.*
- *ISO, ISO/IEC 42001:2023: AI Management Systems. https://www.iso.org/standard/42001.*
- *ISO, ISO/IEC 42001 Explained. https://www.iso.org/home/insights-news/resources/iso-42001-explained-what-it-is.html.*

The Maveric Edge

Maveric Systems is a banking-exclusive technology specialist with 25+ years of domain expertise, engineering trust in AI-first banking. We deploy AI at the core through our proprietary AI@Scale framework and AI-powered platforms, guided by principles that engineer trust by design. Our pragmatic, outcome-driven approach, and AI services spanning advisory, CoE establishment and specialist competencies led engagements make us the trusted engineering partner for the AI era.

25+
years of domain mastery

7+
Operations and Technology AI CoEs

12+
AI powered Platforms and Solutions

AI@Scale
Proprietary Framework for AI adoption at Scale

www.maveric-systems.com

Talk to the Experts ⟶