FUTURE-PROOFING FOR THE A.I. ERA

FROM OPERATIONAL GAPS TO AGENTIC READINESS

JULY 2026

QUINLAN
&ASSOCIATES

TABLE OF CONTENTS

EXECUTIVE SUMMARY

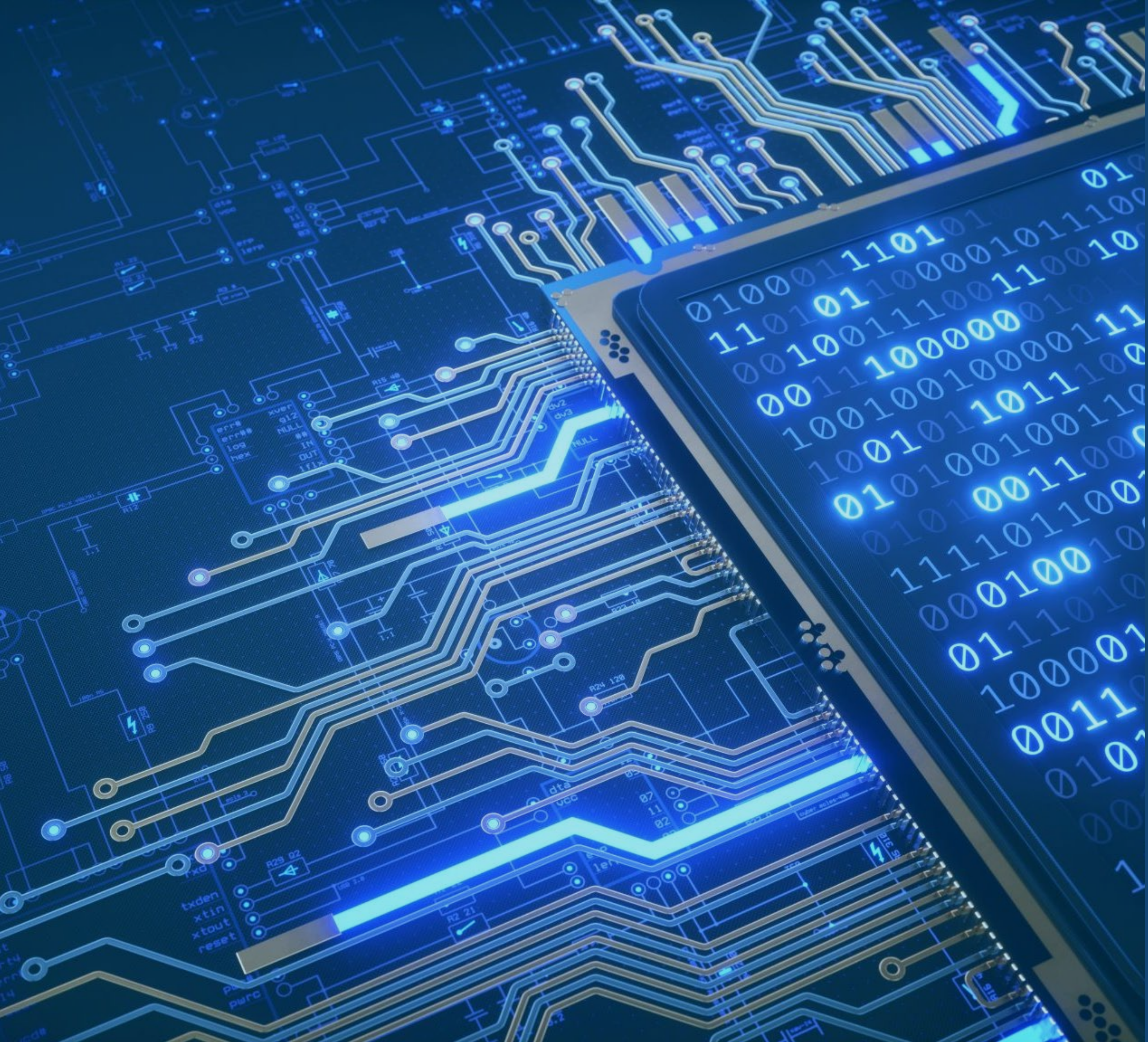
Artificial intelligence ("A.I.") is entering a defining phase of enterprise adoption across financial services. While financial institutions are increasingly deploying A.I. across their organisations, adoption remains largely concentrated in assistive use cases, such as chatbots and copilots. More advanced agentic A.I. deployments, where systems can plan, act, and self-correct toward defined objectives, are being actively explored, but have yet to achieve widespread production adoption.

This shift is being driven by a combination of strategic ambition, competitive pressure, and rapidly expanding investment. The global financial services industry invested approximately USD 90 billion in A.I.-related initiatives in 2025, with spending projected to grow at a 27% CAGR to reach USD 291 billion by 2030. However, despite significant interest and capital commitment, value realisation remains uneven. Nearly half of A.I. initiatives are cancelled, while one-third continue to underperform, reflecting persistent barriers around use case prioritisation, workflow integration, data readiness, control design, governance, and operating model maturity. If these challenges are left unaddressed, financial institutions are expected face operational, regulatory, and reputational risks, particularly as A.I. applications move closer to core business processes, sensitive data environments, and customer-facing decision flows. As a result, A.I. should not be viewed merely as a tool to automate existing processes, but as an opportunity to design fundamentally better, faster, and more intelligent ways of working.

To realise value at scale, financial institutions should **SET** three interconnected priorities that collectively form the roadmap for enterprise A.I. adoption:

- **Strategise:** Establish a clear link between A.I. adoption and enterprise strategy, balancing outside-in awareness of what is technologically possible with an inside-out assessment of where A.I. can address strategic priorities, operational frictions, and workflow bottlenecks. This requires institutions to identify high-value use cases, assess expected impact and feasibility, select fit-for-purpose model architectures, and embed KPIs across the A.I. value chain to measure adoption, productivity, quality, risk reduction, and business outcomes.
- **Engineer:** Build robust A.I. control foundations that reliably translate model capability into production-grade outcomes. As institutions move beyond prompt and context engineering toward more advanced harness engineering, they will need reusable prompts, skills libraries, context engines, orchestration layers, and workflow controls that codify human expertise into rules, logic, examples, validation mechanisms, and human checkpoints. These components are especially important for agentic A.I., where the quality of outputs depends not only on the model itself, but also on the surrounding system that guides, grounds, tests, and refines its actions.
- **Transform:** Redesign the enterprise A.I. operating model to support scalable, controlled, and accountable deployment. Many financial institutions are already experiencing fragmented A.I. implementations that create internal friction rather than sustained business value. As A.I. roadmaps evolve toward higher-autonomy systems, legacy governance models will become increasingly insufficient. Institutions will need clearer structures, assessment processes, accountability frameworks, and role definitions, recognising that employees are increasingly becoming managers of A.I. outputs, while leaders must orchestrate a broader transformation across people, processes, technology, and risk.

While the trajectory of A.I. adoption is becoming clearer, the pathway to scalable value realisation remains highly institution-specific. Financial institutions that treat A.I. as a strategic transformation agenda, rather than a collection of disconnected technology pilots, will be better positioned to convert investment into measurable impact. The next phase of advantage will depend not only on access to powerful models, but on the ability to embed A.I. into workflows, codify institutional expertise, govern autonomous systems, and build an operating model capable of supporting responsible enterprise-scale adoption.



SECTION 1

THE RAPID ACCELERATION OF A.I.

QUINLAN & ASSOCIATES

STATE OF A.I. ADOPTION IN FINANCIAL SERVICES

While financial institutions are increasingly deploying A.I., adoption remains heavily concentrated in assistive use cases, including chatbots and copilots; agentic A.I. deployments cases are being proactively explored but are yet to achieve widespread adoption

Types of A.I. Applications

By Level of Complexity & Adoption

	Chatbot	Retrieval Assistant	Copilot	A.I. Agent	Agentic Operating System
Description	A rule-based or A.I.-powered conversational interface that handles inbound customer queries through text or voice	An A.I. tool that ingests large volumes of proprietary documents, data, and surfaces relevant information on demand	An A.I. assistant embedded into the business applications professionals use to better perform daily tasks with enterprise context	An A.I. system that can interpret a defined goal, plan the required steps, and execute tasks on behalf within predefined constraints	A coordinated network of specialised A.I. agents handling complex, multi-step workflows end-to-end across functions
Adoption Status	**Deployed Nearly Everywhere** The world's 10 largest banks by assets have all deployed A.I. chatbots	**Becoming Table Stakes** Idea generation, data processing, and market analysis are top 3 adoption areas for AMs[1]	**Most Institutions are Here** Microsoft's introduction of Copilot in Nov 2023 marked the beginning of accessible Copilots	**High Intent, Limited Deployment** 42% of finance professionals use or assess agentic A.I. capabilities	**Aspirational** Following single-agent pilots, leading firms are exploring the adoption of multi-agent system
Early Adopter[2]	ally BANK. — **2015** Ally Assist	STATE STREET — **2017** Quantextual Idea Lab	Swiss Re — **2024** Copilot rollout	Allianz — **2025** Project Nemo	BNY — **2026** Multi-agent System
Other Examples[2]	Schroders — Schroders Go TIGER BROKERS — TigerGPT GEICO — Kate	J.P.Morgan — Contract Intelligence Vanguard — Client Summaries AVIVA — Medical Summary Tool	DBS — CSO Assistant Aladdin by BlackRock — Auto Commentary 中信证券 CITIC SECURITIES — Copilot	BANK OF SINGAPORE — SOWA Goldman Sachs — Devin	AIG — Agent Orchestration Layer

[1]Asset managers; [2]Examples are non-exhaustive

Source: Public disclosures, AITechtonic, FinText, Microsoft, NVIDIA, OpenAI, Reinsurance News, Quinlan & Associates analysis

A.I. DEPLOYMENT AREAS

Financial institutions see A.I. generating business value across various functions, with the front office attracting the most interest across banks, asset & wealth managers, and insurance firms

Current and Expected Deployment Areas

By Sector, 2025

BANKING

Where Banks See A.I. Generating Most Business Value

	Business Function	Weighted Avg. Ranking Score
Front Office	Customer Service	22%
	Sales and Marketing	15%
	Portfolio Management	6%
Middle Office	Cybersecurity	15%
	Regulatory	9%
Back Office	Business Operations	11%
	Software Engineering	9%
	Workforce / Collaboration Tools	8%
	IT Operations	6%

0% 8% 16% 24%

CAPITAL MARKETS

Where Asset & Wealth Managers ("AM / WMs") Use A.I.

	Business Function	% of AM / WMs Using A.I.
Front Office	Customer Analysis	59%
	Self-service Portals	54%
	Conversational Support	48%
	Personalised Experiences	48%
	Product Development	46%
Middle Office	Reg. & Tax Monitoring	57%
	Data Security & Privacy	52%
	Risk Mgmt. & Fraud Prevention	49%
	Data Mgmt. & Analysis	38%
	Portfolio Support	35%
Back Office	Writing & Editing Code	46%
	Business Processes	42%
	Custody Services	39%
	Statements & Recordkeeping	37%
	Employee Training	36%

0% 20% 40% 60%

INSURANCE

Where Supervisors See Insurers Adopting GenA.I.

	Business Function	% Observing Use Cases[1]
Front Office	Claims Handling	63%
	External Chatbots	60%
	Complaints Handling	57%
	Agency / Broker Management	40%
Middle Office	Fraud Detection	60%
	Underwriting	47%
	Pricing	41%
	Risk Management	26%
Back Office	Process Optimisation	57%
	Informative Retrieval	57%
	Compliance	50%
	Capital Planning	22%
	Asset & Liability Management	22%

0% 25% 50% 75%

[1]Average of Personal and Commercial Lines

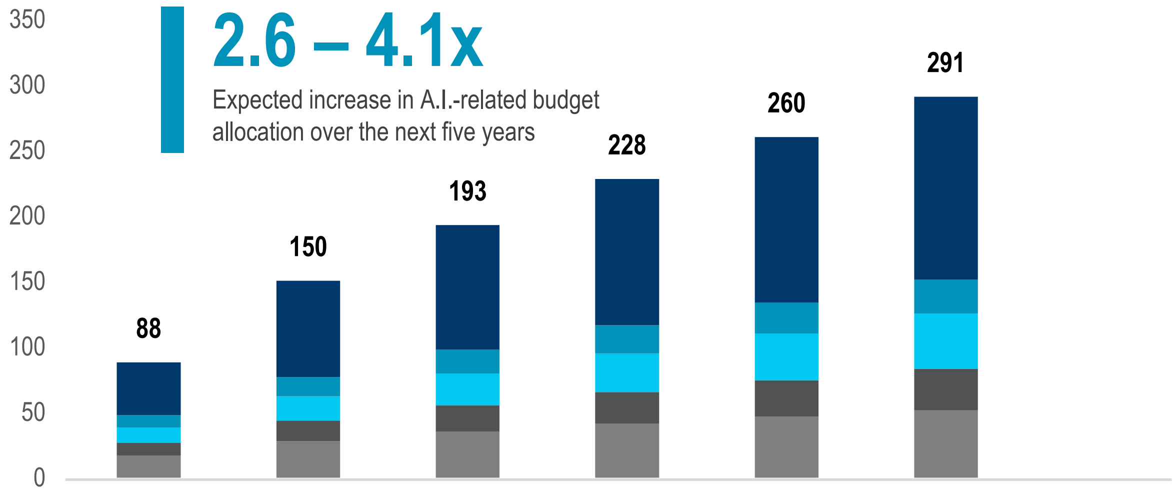
Source: Grant Thornton, Infosys, International Association of Insurance Supervisors, Quinlan & Associates analysis

GLOBAL A.I. SPENDING ACROSS THE FINANCIAL SERVICES INDUSTRY

The global financial services industry invested ~USD 90 billion in A.I.-related initiatives in 2025, with spending projected to grow at a 27% CAGR to reach USD 291 billion by 2030

Global A.I. Spending Size

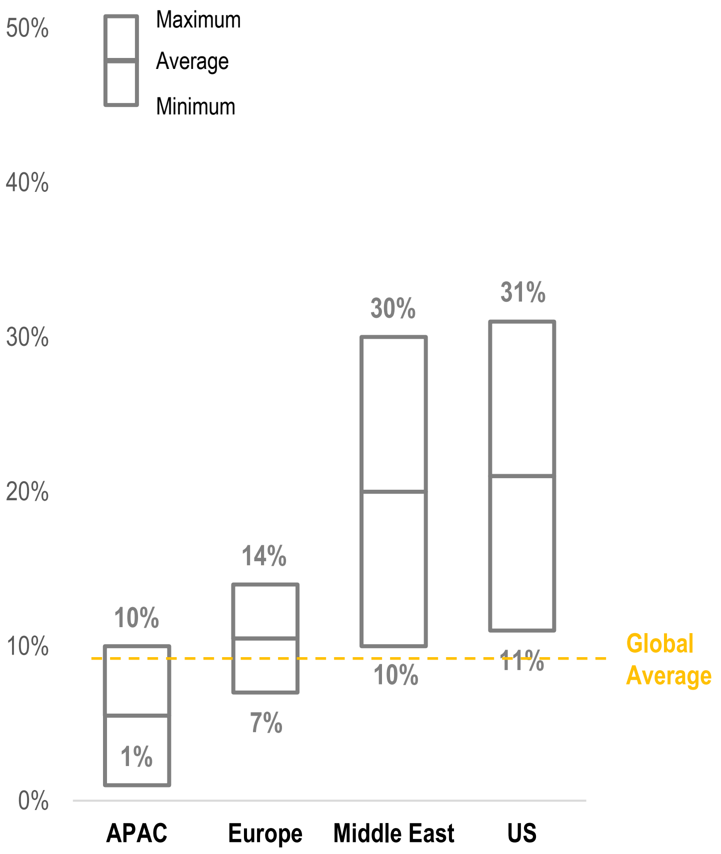
Financial Services, USD Billion, 2025-30E



	2025	2026E	2027E	2028E	2029E	2030E	CAGR
Banking[1]	40	73	95	112	126	140	28%
Sell-side Markets[2]	9	15	18	21	24	26	23%
Buy-side Markets[3]	11	18	24	30	36	42	30%
Market Infra & Payments[4]	10	16	20	24	28	32	27%
Insurance	17	28	35	41	46	51	25%
Total A.I. Spending	**88**	**150**	**193**	**228**	**260**	**291**	**27%**

A.I. Spending Allocation

% Tech Budget Spent on A.I, 2025



[1]Retail, commercial, and corporate banks; [2]Investment banks and brokerages; [3]Asset managers, wealth managers, private banks, and high frequency trading firms; [4]Securities exchanges, digital assets exchanges, and payment providers

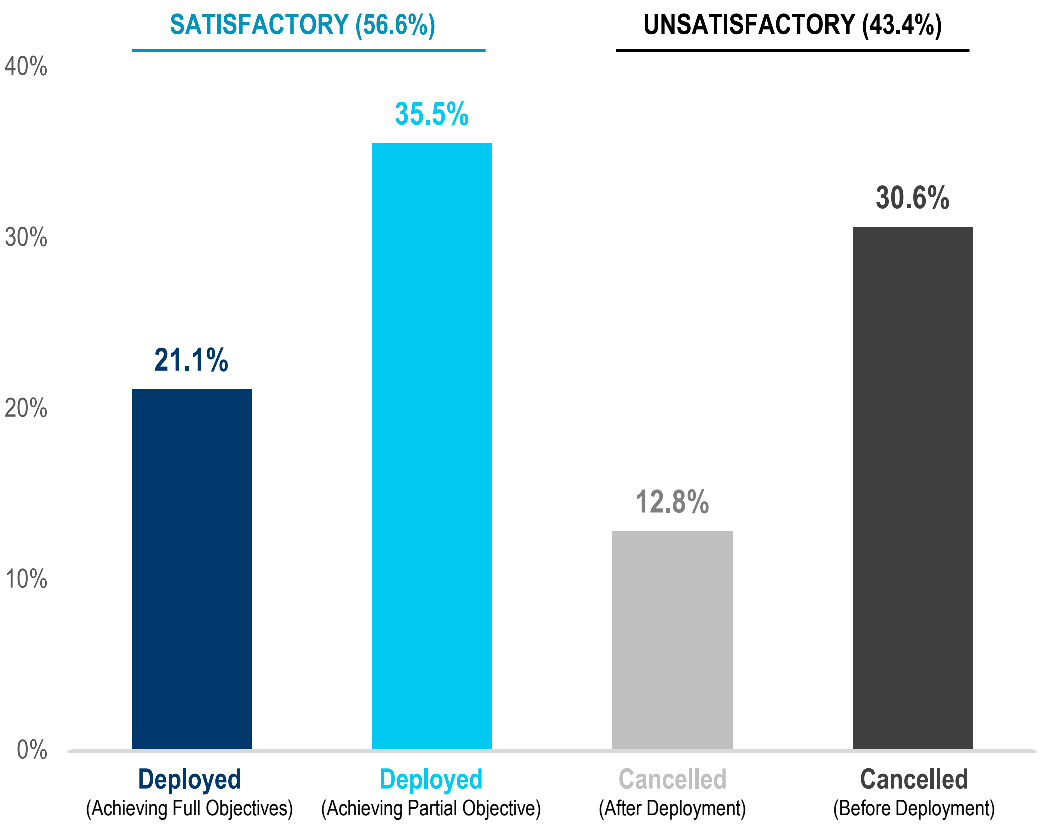
Source: Coalition Greenwich, Nvidia, IDC, public disclosures (e.g., annual statements, financial positions, etc.), Quinlan & Associates estimation

A.I. INITIATIVE CHALLENGES

Despite sizable interest and capital commitment, nearly half of A.I. initiatives are cancelled, while one-third continue to underperform, reflecting the widespread adoption barriers that persist across financial institutions.

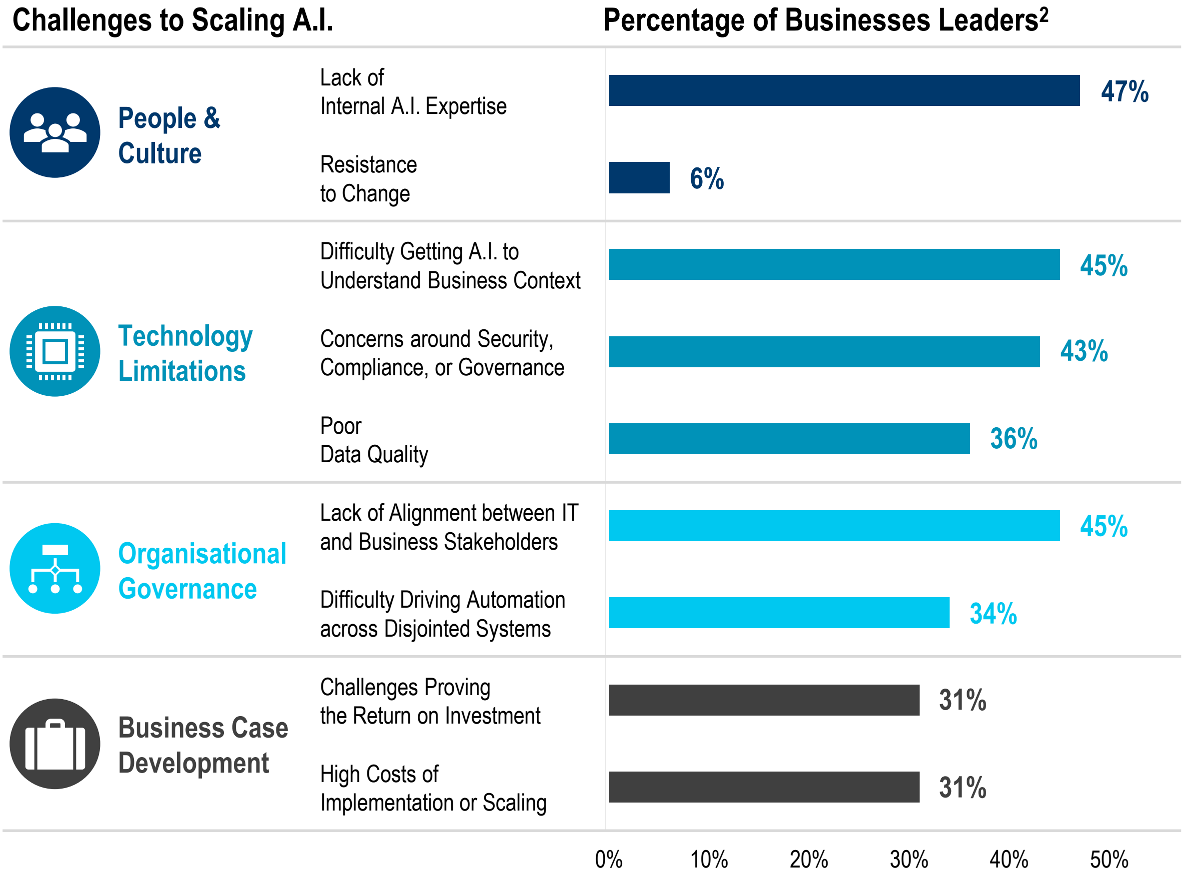
The Reality of A.I. Adoption

% of A.I. Initiatives[1], Dec 2025



Key Challenges

Interview Results, Jun & Jul 2025



[1]Out of 25,289 A.I. initiatives developed by 400 surveyed banks; [2]1,649 business leaders worldwide across 14 sectors, including 15% in banking and 4% in insurance

Source: Celonis, Infosys, Quinlan & Associates analysis

CONSEQUENCES OF NOT GETTING IT RIGHT

If challenges are left unaddressed, financial institutions risk eroding the value of their A.I. investments through economic waste, unreliable outputs, and compromised system integrity, exposing them to operational, legal, and reputational risks

Key Consequences
Use of A.I. in Other Industries

1 ECONOMIC WASTE

Runaway Cost from Context Rot
Without disciplined skill and workflow design, A.I. systems reprocess bloated context at every stage and consume far more tokens

Lost Staff Time from Manual Rework
Vague skill definitions and undocumented failure modes renders outputs unusable, forcing employees to re-run the same task or redo it manually

J.P.Morgan

Erosion of returns from ungoverned A.I. Usage
J.P. Morgan's Chief Data and Analytics Officer noted that some employees are incurring A.I. token expenses exceeding their salaries, raising skepticism about the true efficacy and return on A.I.

2 UNRELIABLE OUTPUTS

Hallucination and Factual Errors
Without skills in cross-verifying facts, users may accept A.I.-generated fabrications, leading to compliance, legal, or technical issues

Poor Judgement and Reasoning
Outputs may fail to reflect sound judgment, domain expertise, or organisational intent due to flawed decision logic / workflow

Virgin MONEY

Poor experience due to misinterpretation
A U.K. bank was forced to remove its A.I. chatbot after it mistakenly flagged the bank's own name, "Virgin", within the customer's question as inappropriate language, refusing to process a legitimate service ask

3 INTEGRITY COMPROMISE

Data Breach
Users frequently leak proprietary information into the A.I. open context if restrictions and data boundaries are not clearly defined in the prompt

External Manipulation
Without strict isolation layers, an A.I. can process malicious instructions (prompt injections) hidden within external sites, compromising systems

cb Community Bank

Sensitive customer data exposed via A.I. tool
A US bank employee uploaded customer data, including Social Security numbers, into an unsanctioned A.I. application, triggering severe regulatory and operational consequences

Source: Forbes, Fortune, Semafor,Quinlan & Associates analysis

NEED FOR FUNDAMENTAL REWIRNG

Importantly, A.I. should not be viewed merely as a tool to automate existing processes but as an opportunity to design fundamentally better, faster, and more intelligent ways of working

START (REQUEST)
The client submits a specific business request

1 ROUTE
The request is captured, logged, and assigned to the appropriate team.

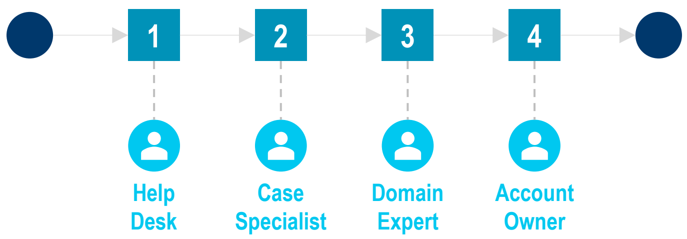
2 PROCESS
Specialist addresses the case by putting together data and applying judgement

3 VALIDATE
The proposed output is checked for completeness, accuracy, and eligibility

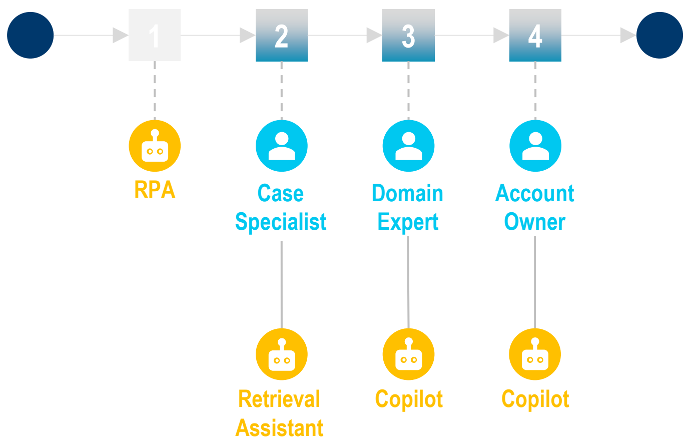
4 APPROVE
An accountable decision-maker provides final authorisation and sign-off

FINISH (DELIVERY)
The client receives the outcome

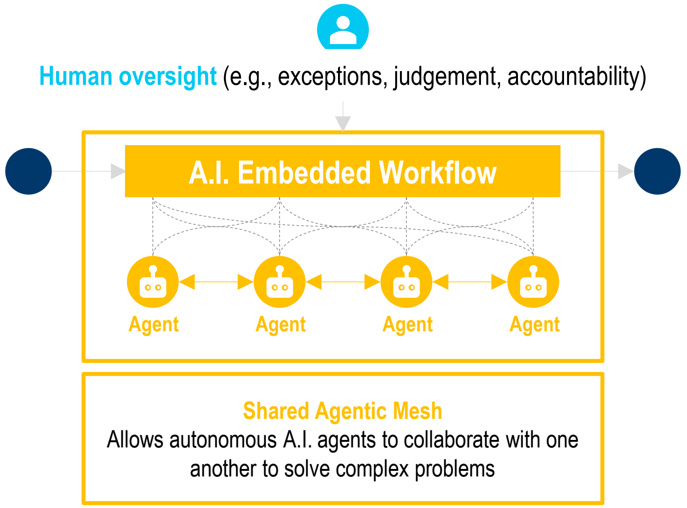
NO AUTOMATION
HUMAN AT EVERY STAGE



PROCESS ENHANCEMENT
WORKFLOW AUTOMOATION



PROCESS RE-INVENTION
A.I.-EMBEDDED WORKFLOW REVAMP



QUINLAN &ASSOCIATES

BUILDING BLOCKS

To overcome the challenges associated with A.I. initiatives and realise value at scale, financial institutions must focus on three interconnected elements that collectively form the roadmap for A.I. adoption

Building Blocks
Description

STRATEGISE
Use Case Portfolio & Business Case · Model Selection · KPIs

Define Objectives and Direct Value
The test of whether A.I. is driven by strategy or by enthusiasm

Is our A.I. strategy and plan surrounding the use cases and models applied to the right problems, and do we have a clear line of sight from use cases to business outcomes?

- Translate business objectives into a prioritised portfolio of use cases
- Select the right models and approaches for each use case, not just the most capable
- Define KPIs that measure outcomes, not activity, and assign ownership of each

ENGINEER
Prompt & Context Engineering
Human-in-the-loop
Evaluation, Monitoring & Drift Detection

Execute Reliably
The gap between A.I. that works in a pilot vs. survives in production

Is A.I. operating consistently and at the quality level the institution and its customers require?

- Engineer prompts and context to produce consistent, high-quality outputs
- Define and enforce system-level controls and human-in-the-loop checkpoints
- Monitor for drift, quality degradation, and safety failures in real time

TRANSFORM

GOVERNANCE	PEOPLE & CULTURE
• Data Governance and Ownership	• Leadership Development
• Decision-making Mechanism	• Training and Awareness
• Terms of Reference, Accountability, and Standards	• KPI Tracking and Monitoring
• Access Control and Oversight Mechanism	• Incentive Design
	• Coordination Structure

Enable Everything with Standards
The distance between A.I. adoption that lands in the organisation vs. merely tolerated by a few

Do we have the appropriate controls and capability to adopt A.I. at scale and in a way that sustains value over time?

- Define and maintain policy
- Assign clear accountability
- Build A.I. literacy across the organisation, from leaders to front-line staff
- Establish ways of working that embed A.I. into day-to-day processes, not as an add-on

Capabilities underpin the entire stack, **driving output quality** in order to **achieve the desired outcomes** defined at the top of the pyramid

Source: Quinlan & Associates analysis

SECTION 2

STRATEGIC IMPERATIVES FOR SCALING A.I.

QUINLAN &ASSOCIATES

DEFINING THE A.I. STRATEGY

A.I. value is maximised when organisations balance outside-in awareness of what is possible with a rigorous inside-out approach that embeds A.I. into core workflows and strategy-led priorities

A.I. Strategy
Double-edge Perspectives

To ensure that **market-proven A.I. capabilities** are selectively embedded into workflows and operating models to **generate actual value…**

BE INFORMED BY THE MARKET…NOT DRIVEN BY IT

Look both within and outside the financial sector, defining what is technically and commercially possible by leveraging what is already proven in the market

→ ***Short time-to-market based on best practices***

START WITH STRATEGY…THEN REDESIGN WORKFLOW

Ensure that A.I. adoption is tied to the overarching strategy first and foremost, and review existing workflows to spot operational friction, which are strong automation targets, applying A.I. as a process engineering lever

→ ***Tailored opportunities to pain points and value creation***

Strategic Considerations
Description

Key Considerations	Description	Goals
1 STRATEGY FIRST APPROACH	A.I. should directly serve the overarching business strategy, with every use case explicitly traced to a defined objective (e.g., growth, risk reduction, efficiency)	*Define Business Objectives*
2 WORKFLOW-LED DESIGN	Existing end-to-end workflows should be deconstructed to identify frictions and where A.I. can fit in, rather than starting from available A.I. tools / capabilities	*Transfer Domain Expertise*
3 USE CASE PRIORITISATION	Subsequently, A.I. use cases should be prioritised based on expected impact and implementation feasibility, rather than pursuing broad experimentation	*Place Strategic Bets*
4 MODEL SELECTION	Appropriate LLM (i.e. the engine) should be selected by applying a structured scoring methodology across, with the weighting tailored to each business case	*Determine Fit-for-purpose Technology*
5 MEASUREMENT FRAMEWORK	Organisations should embed relevant KPIs across the value chain of A.I. deployment to ensure that adoption impacts are properly monitored	*Track Adoption Impact*

Source: Quinlan & Associates analysis

STRATEGY-FIRST APPROACH

Financial institutions should first validate if and where A.I. supports the overarching strategy to establish directional intent before mapping promising use cases where A.I. adoption can add value

Strategic Alignment

Existing Strategy and A.I. Capabilities, Illustrative

Type of Institutions	BANKING INSTITUTIONS	ASSET / WEALTH MANAGERS	SECURITIES BROKERAGES	INSURANCE FIRMS
Overarching Strategy	**Simplify the Organisation** Streamline the bank's structure and operations by existing non-core markets and services	**Adapt to the Wealth Transfer** Ensure smooth asset transition and reinvent current workflows for the next generation with new values	**Tap into Retail Clients** Further monetise from the retail clients by reinventing the do-it-yourself approach to investment	**Expand Protection Services** Introduce innovative products to protect clients and their assets from a range of emerging risks

Extent of strategic alignment across potential A.I. applications by function

	Identify Emerging Risks	Develop Business Cases	Design Product & Pricing	Offload Aggregate Risks	Launch & Market Products	Process Future Claims
Description	Pinpoint market gaps and test if the new risks are insurable	Build formal business cases to establish the target market	Set the premium pricing and write the policy guidelines	Manage the insurer's exposure and obtain regulatory approvals	Train agents and brokers for marketing and process claims	Set up concrete workflows to review and pay out claims
Relevant A.I. Capabilities *(Example)*	**Horizon Scanning** A.I. can continuously scrape unstructured data (e.g., litigation, research) for signals of new risks	**Persona Testing** A.I. can simulate client personas and stress-test if and how the product may be exploited	**Policy Drafting** A.I. can draft precise definitions and exclusions before comparing them with existing policies	**Pitch Creation** A.I. can create collateral, such as impact projects, that are presented to reinsurers and regulators	**Sales Enablement** A.I. chatbots and employee copilots can inform customers and agents more intuitively	**Pre-processing** A.I. can automate document processing and provide preliminary opinions on each claim
Alignment Assessment	Very Aligned	Somewhat Aligned	Very Aligned	Somewhat Aligned	Very Aligned	Very Aligned

Source: Quinlan & Associates analysis

WORKFLOW-LED DESIGN (1/2) – APPLICABLE TO ALL DEPARTMENTS

Across the relevant functional areas identified, operational frictions and bottlenecks are identified across the end-to-end workflow, establishing a structured foundation for where and how A.I. can be applied, freeing up resources toward higher-value activities

Use Case Development

A.I. Deployment

	Description	Implications of Inaction
DEFINE	**STANDARD OPERATING PROCEDURE / WORKFLOW** Map the end-to-end workflow across business lines, including inputs, outputs, decision points, approvals, controls, and hand-offs	Automating fragmented, undocumented processes may result in greater inefficiencies and weakened controls
ASSESS	**MANUAL GRIND AND PAIN POINT** Identify workflow bottlenecks and manual-intensive activities by assessing frequency, effort, turnaround time, rework, error rates, and control sensitivity	Deploying A.I. based on assumptions or hype rather than measurable pain points and business needs may not generate value
MATCH	**TECHNOLOGY AND RISK SUITABILITY** Evaluate whether each activity is suitable for automation, augmentation, or continued human execution, taking into account regulatory requirements and model risk / impact	Applying A.I. to unsuitable use cases where risks exceed potential value / benefits (e.g., compliance, operational risks, etc.)
REDESIGN	**A.I.-INTEGRATED WORKFLOW** Revisit and update the end-to-end workflow to incorporate A.I., defining new roles, controls, escalation paths, and oversight as needed	Resources intended to be 'freed up' ended up being spent more on correcting A.I.'s mistakes due to an unclear, outdated process

Manual vs. High-Value Tasks

Illustrative, Front Office and Back Office

FRONT OFFICE
(A Private Banker)

× Manual Tasks		✓ Higher-Value Tasks
× Client Meeting Notes	Free up Resources for →	✓ Needs Discovery
× Portfolio Summaries		✓ Tailored Advisory Provision
× Product Information Retrieval		✓ Client Relationship Management
× Internal Documentation		✓ Commercial Prioritisation
× Routine Enquiry Response		✓ Estate Planning Coordination

BACK OFFICE
(A HR Recruiter)

× Manual Tasks		✓ Higher-Value Tasks
× CV Screening Support	Free up Resources for →	✓ Candidate Interview
× Candidate Outreach Draft		✓ Workforce Planning / Advice
× Interview Scheduling		✓ Hiring Judgement
× Feedback Summaries		✓ Talent Event Participation
× Candidate Sourcing		✓ Candidate Relationship Building

× Manual Tasks Led / Augmented by A.I. ✓ Higher-Value Human-led Tasks

Source: Quinlan & Associates analysis

WORKFLOW-LED DESIGN (2/2) – IMPORTANCE OF DOMAIN EXPERTISE

As A.I. applications are deployed from simple workflow automation to advanced A.I.-embedded workflow revamp, the quality of transformation depends on how well human expertise is codified into rules, context, controls, and validation logic

Complementary Relationship

Human Domain Expertise & A.I. Execution Capabilities

HUMAN
(Domain Expertise)

ARTIFICIAL INTELLIGENCE
(Execution Capabilities)

1. **Documentation:** Human domain expertise is captured from documents, procedures, examples, business rules, and operational experience
   - Business intent
   - Real-life workflow and example
   - Judgement criteria

→ ***Raw expertise base that explains what matters and why***

2. **Codification:** A.I. helps structure, clarify, test, and convert domain expertise into repeatable workflow logic
   - Codified standard operating procedure
   - Dynamic workflow development
   - Decision rules & Control thresholds

→ ***Codified logic that can be consistently applied and governed***

3. **Automation:** A.I. applies the codified expertise to automate workflow steps within defined permissions and boundaries.
   - Retrieval
   - Drafting
   - Analysis and routing

→ ***Faster and more consistent execution of repeatable manual work***

4. **Refinement:** Human experts review outputs, correct errors, handle exceptions, and refine the workflow based on real-world performance
   - Output review
   - Exception handling
   - Rule refinement

→ ***Improved expertise base and stronger next-cycle automation***

Source: Quinlan & Associates analysis

USE CASE PRIORITISATION

Once specific use cases are identified, financial institutions should assess them across expected impact and feasibility to determine which ones are to be piloted first

Use Case Prioritisation

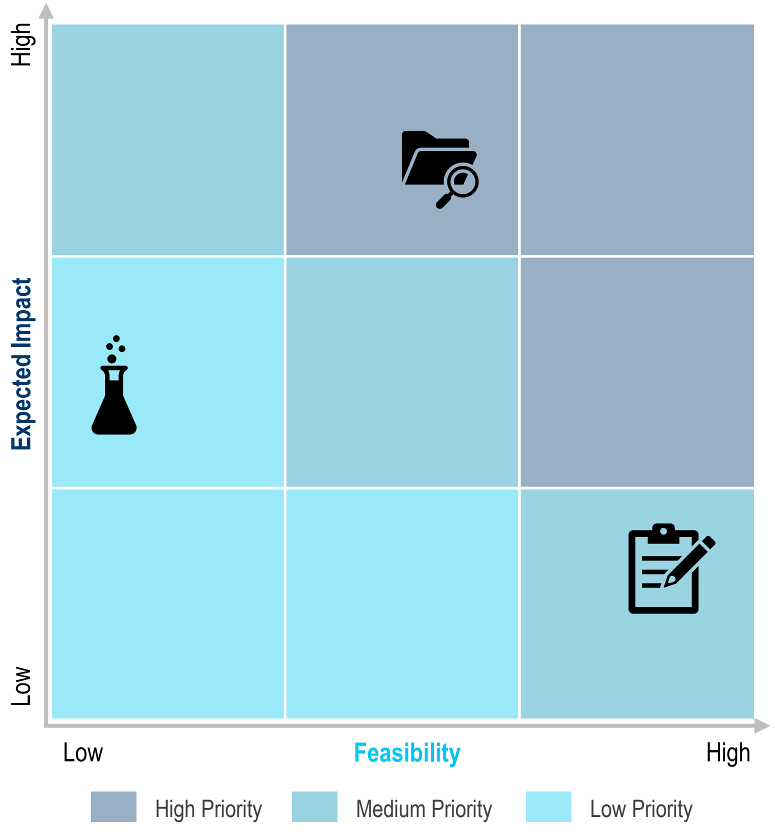
Expected Impact vs. Feasibility, Illustrative

Use Case Examples

Key Criteria		Product Info. Retrieval	Meeting Minutes	A.I. Sandbox
Expected Impact (50%)				
Profitability Growth (20%)	How easily can the initiative contribute to profitability?	✓	✗	✗
Productivity Growth (20%)	How much productivity boost can the initiative enable?	✓	✓	–
Innovative Culture (5%)	Does the initiative encourage employees to innovate?	✗	✗	✓
Company Morale (5%)	Would the initiative contribute to team-building?	–	–	✓
Feasibility (50%)				
Infrastructure & Data (15%)	Are incumbent technologies sufficient for the solutions?	–	✓	–
Budgeting Ease (15%)	How much budget would the deployment require?	–	✓	✗
Employee Readiness (10%)	Can staff navigate easily or do they need training?	–	✓	✗
Deployment Time (10%)	How long would it take from ideation to adoption?	–	✓	✗
Overall Prioritisation (100%)		High	Medium	Low

Favourable Dependent Less Favourable

PRIORITISATION MATRIX



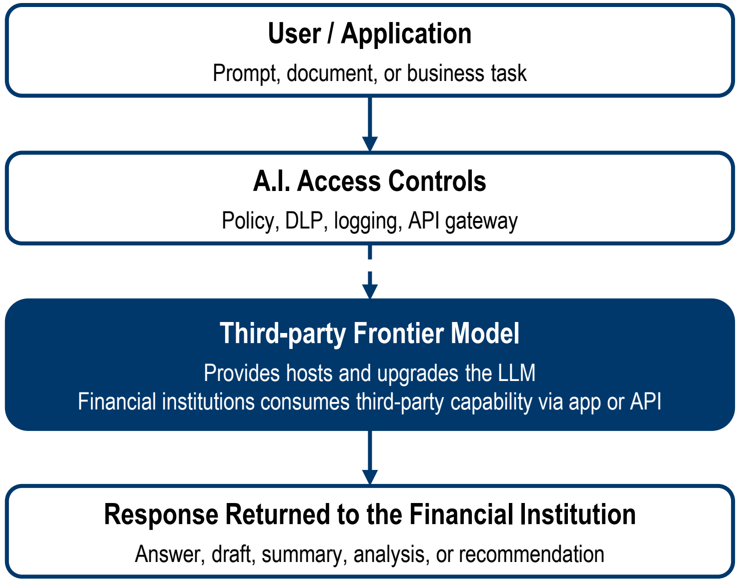
Source: Quinlan & Associates analysis

MODEL SELECTION (1/2) – DEPLOYMENT APPROACH

To support the business case, financial institutions may consider deploying frontier models, open-source or open-weight models, or a multi-model architecture supported by an LLM orchestration layer

Deployment Approach

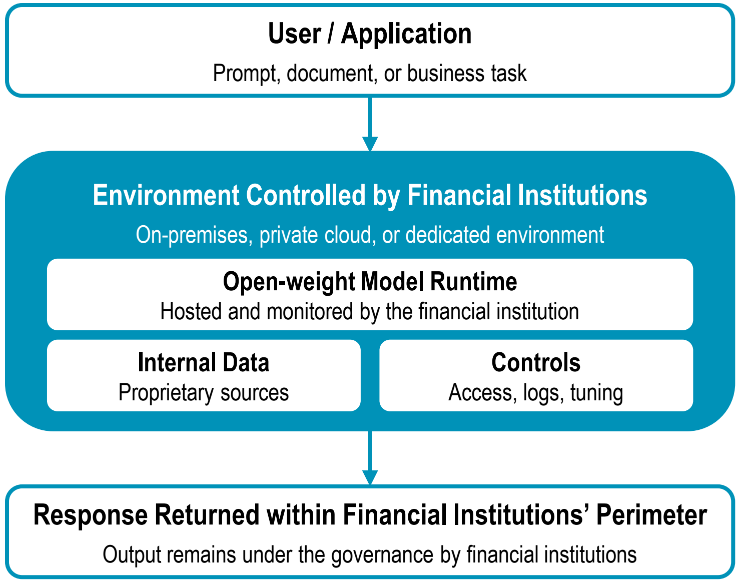
Frontier vs. Open Source vs. Hybrid



THIRD-PARTY LLM DEPLOYMENT

(Frontier Model)

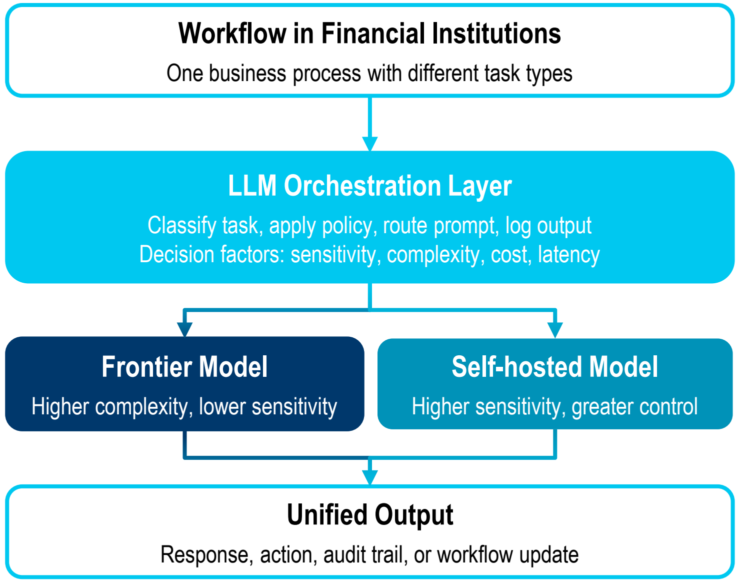
Third-party managed LLMs accessed through web, application, and/or API channels, allowing financial institutions to access leading model capabilities without directly operating the underlying model or infrastructure



CONTROLLED DEPLOYMENT

(Self-hosted Open Source / Weighted Model)

LLM is controlled by the financial institutions and deployed inside on-premises or private cloud environments, enabling tighter control over hosting, data handling, access permissions, finetuning, and infrastructure integration



MULTI-MODEL DEPLOYMENT

(Hybrid Orchestration)

An orchestration approach that routes different tasks or workflow steps to the most suitable model environment based on sensitivity, performance, cost, governance, and operational requirements

Source: Microsoft, Google, Quinlan & Associates analysis

MODEL SELECTION (2/2) – METHODOLOGY

When selecting an appropriate LLM, financial institutions can apply a structured scoring methodology across key decision criteria, with the weighting of each criterion is adjusted according to the needs and sensitivity levels of different business units

	Methodology Key Criteria and Assigned Weights (%)	Frontier LLM (Closed Source)	Self-hosted LLM (Open Weight/Source)	Front-office (Weight)	Back-office (Weight)
BUSINESS	**Data Privacy:** How sensitive are the data inputs and conversations processed by the LLM?	■□□□□	■■■■■	20%	5%
	Domain Finetuning: To what extent should the LLM be fine-tuned for the financial institution's business-specific requirements?	■□□□□	■■■■■	20%	5%
	Counterparty Risks: How critical is it for the financial institution to control the model lifecycle and steer clear of vendor lock-in?	■■□□□	■■■■■	10%	5%
MODEL	**Capability:** How important is long-context understanding and advanced reasoning for achieving quality outcomes?	■■■■■	■■■□□	20%	20%
	Latency: Is there a hard requirement for response time, throughput, uptime, and fallback?	■■■■■	■□□□□	10%	5%
IMPLEMENTATION	**Time-to-Market:** What is the financial institution's internal timeline for the LLM to be deployed into production environments?	■■■■■	■■□□□	5%	20%
	Total Cost of Ownership: What is the financial institution's appetite to shoulder the upfront and ongoing cost (e.g., API calls)?	■■■□□	■■■□□	5%	30%
	Operational Readiness: Does the financial institution have internal capabilities (e.g., infrastructure and talents, etc.) to support usage?	■■■■■	■■□□□	5%	10%
SUITABLE MODEL				**Self-hosted LLM**	**Frontier LLM**

Source: Quinlan & Associates analysis

MEASUREMENT FRAMEWORK

With a clear use case roadmap and model selection, financial institutions should embed relevant KPIs across the value chain of A.I. deployment and ensure that these impacts are measured for usage incentivisation and powerful stakeholder communications

Measurement Framework

Key Pillars, Examples, and Potential Pitfall

	Applications of KPIs	Performance Indicators	Common Pitfalls to Avoid
PRE-DEPLOYMENT	**USE CASE SELECTION** Are we choosing the right problems to solve?	• **Strategic Alignment Rate** (% of selected use cases tied to a business priority) • **Use Case Value Score** (i.e., impact x feasibility rating)	Chasing novelty over value
PRE-DEPLOYMENT	**PILOT MANAGEMENT** Are resources spent wisely on ongoing pilots?	• **Time-to-kill Period** • **Pilot Success Rate** • **Cost per Pilot**	Failing pilots kept alive due to sunk cost or internal politics / killing viable pilots before outcomes can materialise
POST-DEPLOYMENT	**MODEL QUALITY** Is the model performing accurately and reliably over time?	• **Hallucination Rate** • **Time Efficacy** (i.e., time reduction from the use of A.I.) • **User Satisfaction**	Poor performance on edge cases that matter most may be masked (e.g., rare fraud types)
POST-DEPLOYMENT	**GOVERNANCE AND COMPLIANCE** Are we operating within regulatory and ethical boundaries?	• **Data Lineage** • **Incident Rate** (i.e., no. of material incidents have arisen and the extent of impact) • **Audit Trail Completeness**	A.I. tools that do not adapt as regulatory context and internal compliance evolves
POST-DEPLOYMENT	**ADOPTION IMPACTS** Are employees using A.I. effectively, and does usage translate into value?	• **Active User Rate** • **User Confidence Score** • **Feedback Score**	Measuring inputs instead of outputs, where tracking token usage may trigger 'tokenmaxxing' and burn A.I. budgets

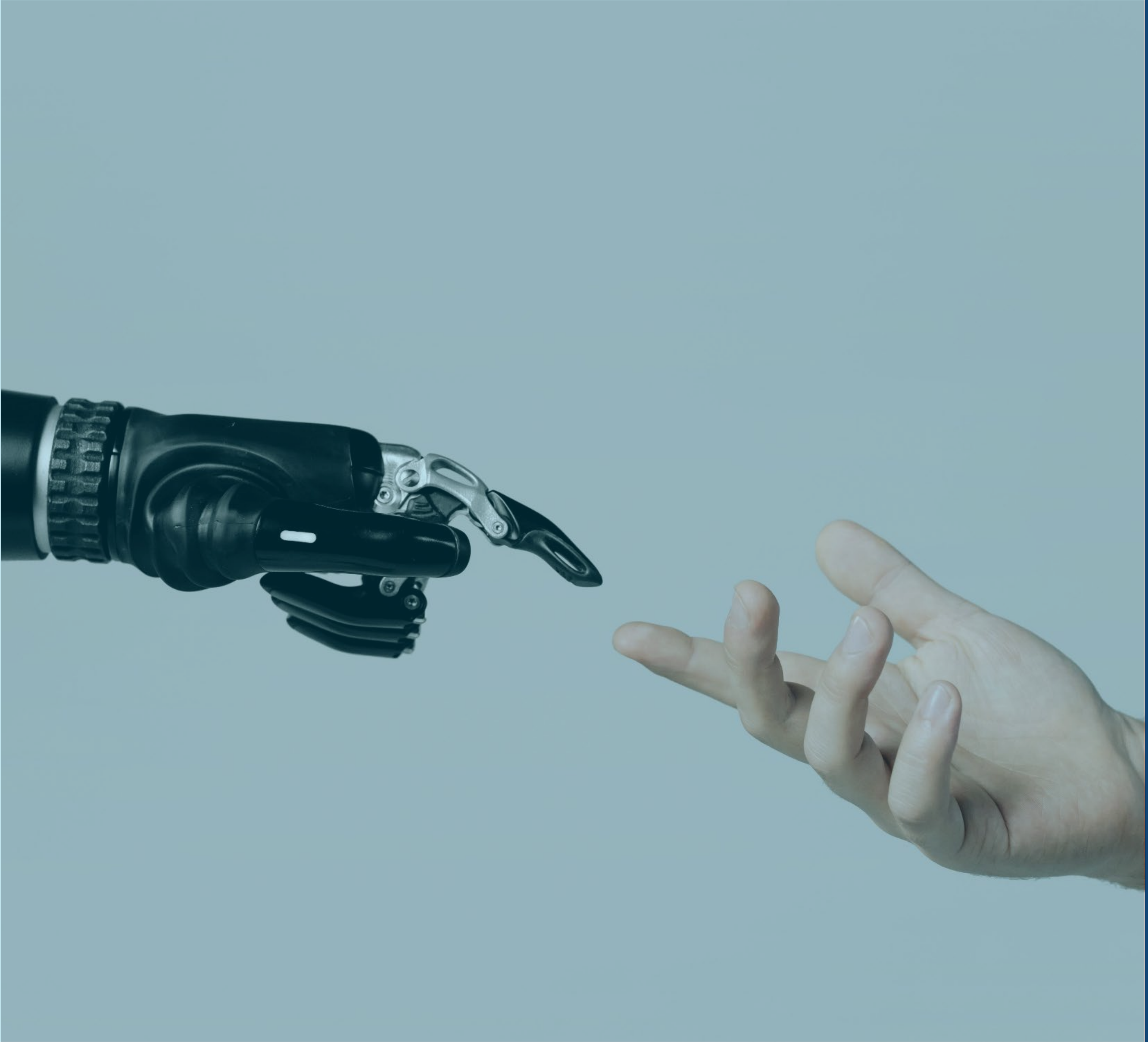
Case Study

DBS

DBS Bank institutionalised A.I. as a core capability across cross-functional teams by embedding indicators into the A.I. scoreboard and leveraging high-profile internal success stories to continuously encourage and accelerate adoption.

A.I. Practices	Impacts
Balanced A.I. Scoreboard DBS developed an internal A.I. scorecard that combines financial and strategic indicators	**Institutionalise A.I. As a Capability**
Value Narrative In DBS's public releases, the impacts of A.I. adoption is reported in highlight figures	**Communicate Impacts Clearly**

Source: DBS, Quinlan & Associates analysis



SECTION 3

FOUNDATIONS OF A.I. ENGINEERING

QUINLAN &ASSOCIATES

FOUNDATIONS OF A.I. ENGINEERING

The success of A.I. deployment depends not only on the model itself, but also on the underlying model-level harness engineering, which must be robust to ensure model capability is reliably translated into production-grade outcomes

A.I. Solution Pillars

Key Components

HARNESS

CONTEXT

PROMPT

Role / Objective | **Steps** | **Examples** | **Constraints**

If reusable

Memory
Implicitly-learnt user behaviour from past conversations

Knowledge Base
Privately curated information given by user(s)

Skills
Reusable prompts approved and managed by user(s)

Tools
Solutions that connect the model to external databases and apps

Agentic Loop
The autonomous self-reinforcing cycle to complete specific task(s)

Dynamic Workflow
The adaptive coordination of autonomous agents

Schema / Rules
The rulebook that defines exactly how agents should behave

PROMPT ENGINEERING (2022)

The natural language instruction or question to trigger a specific response

→ *What does the A.I. model need to do?*

CONTEXT ENGINEERING (2024)

The surrounding context and data to ground the A.I.'s response in relevant facts

→ *What does the model need to know?*

HARNESS ENGINEERING (2026)

The infrastructure around a model to automate, control, and validate its end-to-end execution

→ *How should A.I. perform the task?*

Source: Quinlan & Associates analysis

QUINLAN & ASSOCIATES

THE RISE OF HARNESS ENGINEERING

In the past few years, prompt and context engineering have been the focus areas for many institutions, with the next frontier being harness engineering in order to equip next-generation agentic A.I. solutions that can plan, act, and self-correct

A.I. Solutions Evolution

Applicability of Prompt, Context, and Harness Engineering

✓ Fully Applicable - Applicable to Some Extent × Not Applicable

FIRST GENERATION Prompt Engineering (2022) → **SECOND GENERATION** Context Engineering (2024-25) → **THIRD GENERATION** Harness Engineering (2026) →

CHATBOT
Conversational A.I. interface where the user prompts the model to generate a specific output, such as answering a question, drafting content, and summarising information

RETRIEVAL ASSISTANT
A.I. assistant connected to external knowledge sources and/or internal databases, which retrieves and references relevant data before generating a response

COPILOT
A.I. assistant embedded within end users' workflow and application, supporting them with context-specific tasks within their real working environment

A.I. AGENT
A goal-based software that can interpret an objective, plan steps, call tools, monitor progress, and execute multi-step tasks within end-user-approved parameters

AGENTIC A.I. SYSTEM
System of one or more A.I. agents that pursues defined goals with limited human supervision, often across multiple tools, workflows, data sources, and approval layers

	Chatbot	Retrieval Assistant	Copilot	A.I. Agent	Agentic A.I. System
PROMPT					
Role / Objective	✓	✓	✓	✓	✓
Steps	✓	✓	✓	✓	✓
Examples	✓	✓	✓	✓	✓
Constraints	✓	✓	✓	✓	✓
CONTEXT					
Memory	✓	✓	✓	✓	✓
Knowledge Base	×	✓	✓	✓	✓
Skills	×	×	×	✓	✓
HARNESS					
Tools	-	-	✓	✓	✓
Agentic Loop	×	×	×	✓	✓
Schema	×	×	×	✓	✓
Orchestration	×	×	×	×	✓

Low — LEVEL OF AUTONOMY — High

Source: AWS, Microsoft, Google, IBM, Quinlan & Associates analysis

QUINLAN & ASSOCIATES

PROMPT ENGINEERING

To ensure consistent and reliable outputs across teams and geographies, financial institutions can create pre-defined prompts that incorporate objectives, examples, and workflow expectations for quick access to employees

Key Components

Well-configured Prompt

SET OBJECTIVES

Designate the model's role and give an overview of the task, as well as relevant contexts to ensure the model captures task-specific nuances

INSTRUCT STEPS

Detail each step involved, including any constraints that the model is prohibited from doing while carrying out the task

INCORPORATE EXAMPLES

Give specific examples of expected outputs help the model learn about the task and the complexity, tone, and concision of expected outputs

USE DELIMITERS (e.g., <>, ```, “”, etc.)

Specify where inputs are inserted in the prompt to help model identify user inputs (e.g., company profiles), especially when they are long

STRUCTURE OUTPUTS

Specify the response with a fixed syntax to parse the output into be more readily available formats to be fed into the following processes

1

Credit Memo | Client Briefing | AML Case

TASK: Act as a Senior Commercial Credit Analyst. Analyse the raw applicant profile below, perform a credit risk analysis, and generate a standardised credit memo.

ANALYSIS FRAMEWORK

Before formulating your final assessment, you must think step-by-step: (1) Financial Assessment: Analyse liquidity, leverage, and repayment capacity with accounting ratios; (2) Qualitative Analysis: Assess the borrower's industry stability, management experience, and collateral; (3) Risk vs. Mitigant Balancing: Explicitly identify the top 2-3 credit risks and how they are mitigated; (4) Deduction: Formulate the final credit decision based on the balance of these factors.

EXAMPLE

Applicant Profile: Logistics firm operating 5 years. Requesting $500,000 for fleet expansion; Financials (Annual Revenue: $4M, Net Profit Margin: 8%. Current Debt Service Coverage Ratio (DSCR): 1.65x (Strong)); Risk Note: Major customer represents 40% of revenue (Concentration Risk); Collateral: Real estate valued at $700,000.

Analysis Reasoning: (a) Financials: Financial health is strong; cash flow comfortably covers new debt service (1.65x); (b) Qualitative: Experienced management (5 years), but high customer concentration risk makes revenue volatile; (c) Mitigants: The concentration risk is heavily mitigated by strong hard collateral ($700k real estate) which provides a clear secondary source of repayment.

Output: "executive_summary": "Logistics provider seeking $500k for fleet expansion. Strong cash flow +solid real estate collateral offset high customer concentration risk.", "key_credit_strengths": ["Strong repayment capacity with a historical DSCR of 1.65x.", "High quality, easily liquidable real estate collateral ($700k value)."], "key_risks_and_mitigants": ["Risk: Customer concentration (40% of revenue from one client). Mitigant: Long-term contract in place plus substantial asset backing."], "rationale": "The borrower exhibits stable profitability and strong debt service capability. While the loss of their primary client would hurt, the loan is structurally secure due to conservative loan-to-value on the collateral.", "opinion": "Approve“.

2

CURRENT TASK

Applicant Profile: ```Data Analytics Firm operating 6 years. Requesting $650,000 GPU server cluster for A.I. modelling; Financials (Annual Revenue: $5.2M, Net Profit Margin: 18%. Current Debt Service Coverage Ratio (DSCR): 2.10x (Strong)); Risk Note: High employee turnover in senior data science roles (Key Person/Execution Risk); Collateral: Intellectual property portfolio and enterprise software contracts valued at $850,000.```

Instructions: (i) Follow the Analysis Framework to process the data; (ii) The `opinion` field must strictly be chosen from: `Approve`, `Ambivalent` (requires conditions/further info), or `Reject`; (iii) Output should match this structure: ```json { "executive_summary": "string", "key_credit_strengths": ["string"], "key_risks_and_mitigants": ["string"], "rationale": "string", "opinion": "string"}```

3

Employee Workflow

1 Select Task

Choose the task for which predefined templates have been created with all key elements to be loaded in the chatbox

2 Input Information

Provide case-specific information for the model to carry out the specific task, such as client profile and financial position

3 Generate Output

Request the model to generate outputs following the format and steps as instructed in the template

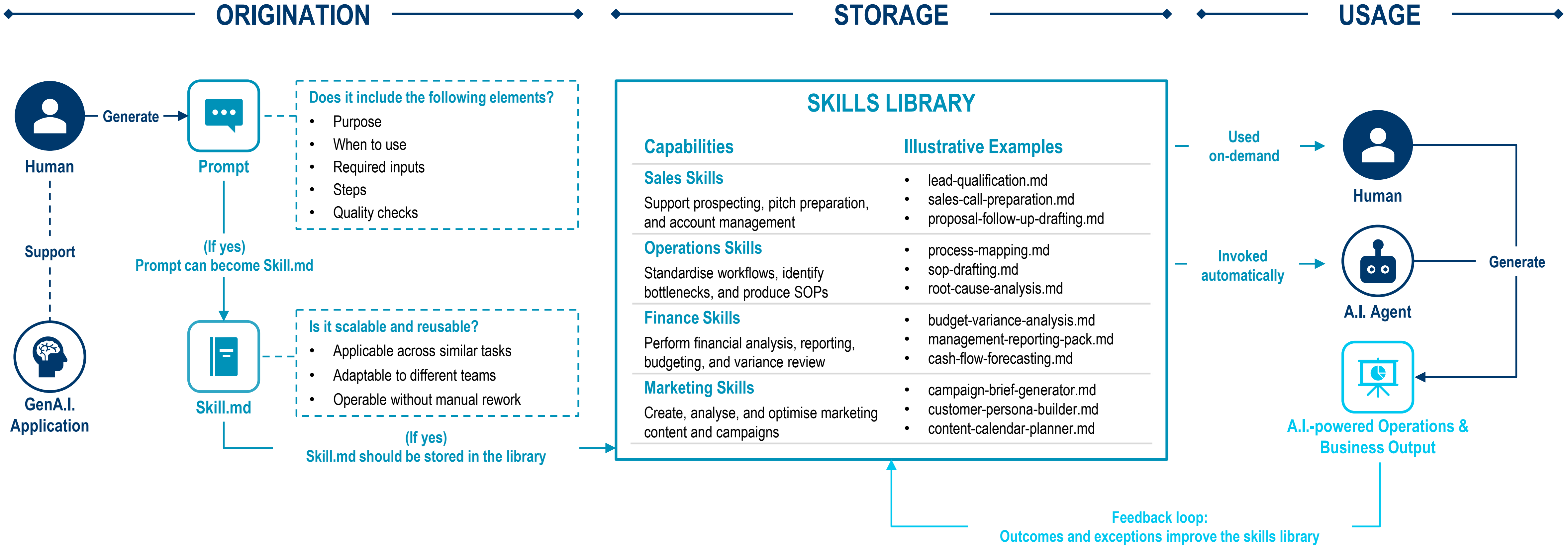
Source: CoreWeave, Comet, Quinlan & Associates analysis

SKILLS (1/2) – OVERVIEW

Skills are reusable and scalable procedural prompts that can be used by humans on demand or automatically invoked by the relevant A.I. agent based on specific configuration

Lifecycle of Skill

Origination, Storage, and Usage



SKILLS LIBRARY

Capabilities	Illustrative Examples
Sales Skills Support prospecting, pitch preparation, and account management	• lead-qualification.md • sales-call-preparation.md • proposal-follow-up-drafting.md
Operations Skills Standardise workflows, identify bottlenecks, and produce SOPs	• process-mapping.md • sop-drafting.md • root-cause-analysis.md
Finance Skills Perform financial analysis, reporting, budgeting, and variance review	• budget-variance-analysis.md • management-reporting-pack.md • cash-flow-forecasting.md
Marketing Skills Create, analyse, and optimise marketing content and campaigns	• campaign-brief-generator.md • customer-persona-builder.md • content-calendar-planner.md

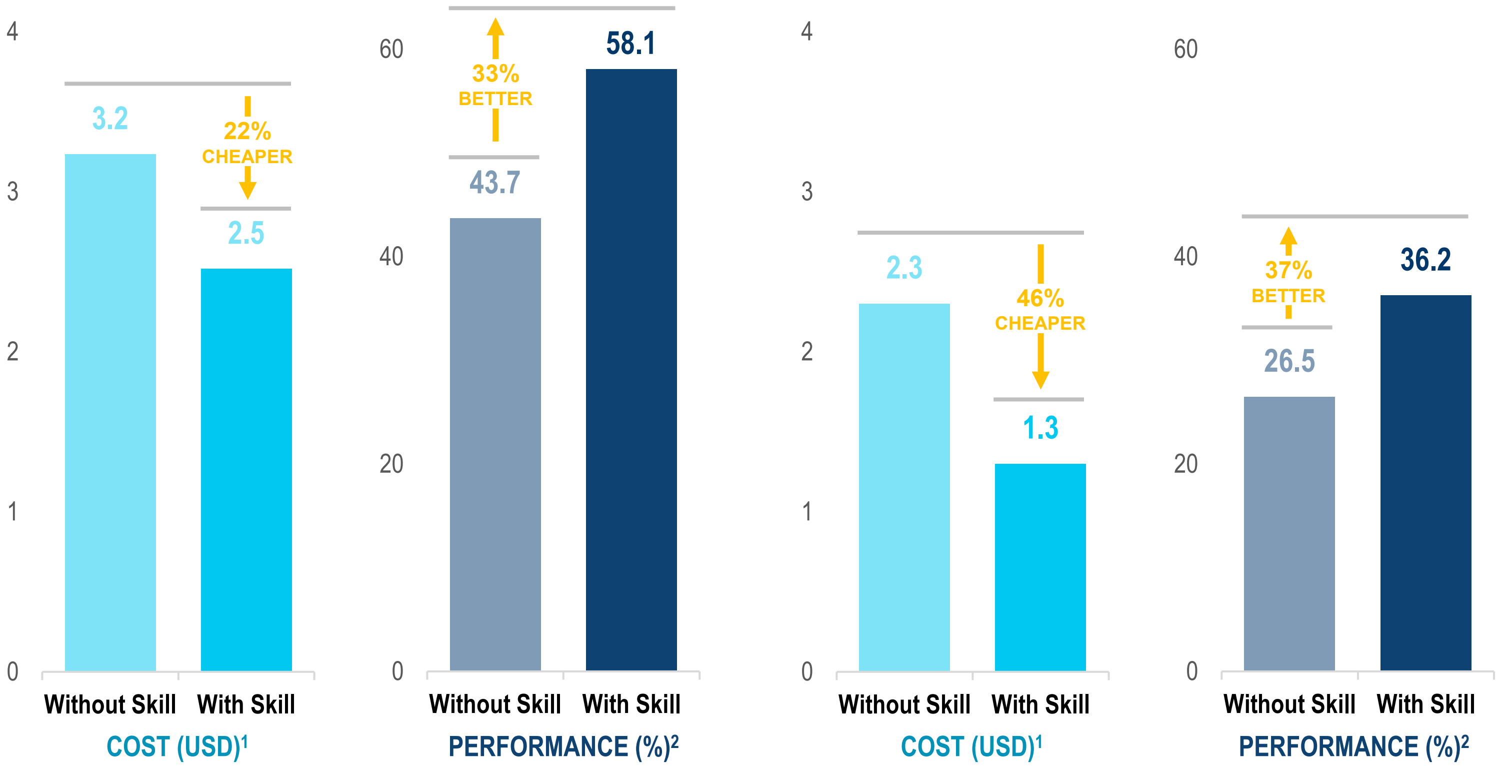
Source: Quinlan & Associates analysis

SKILLS (2/2) – COST IMPLICATIONS

On average across the model lineup, skills increase per-trial token consumption, while reducing per-trial cost in practice as a greater share of tokens shifts towards cheaper input-side overhead and improvements in first-pass accuracy with fewer retries

Cost & Performance Improvement

Skills vs. Without Skills, Performance-focused vs. Efficiency-focused Model



PERFORMANCE-FOCUSED MODELS
(GPT-5.5, Opus 4.8, and Gemini 3.1 Pro)

EFFICIENCY-FOCUSED MODELS
(GPT-5.4 mini, Sonnet 4.6, and Gemini 3.1 Flash Lite)

Key Reasons

Token Economics

OBJECTIVE FUNCTION

The curated skill provided the model with the code template or algebraic structure across steps / calls that it can then attempt to optimise

→ ***Focused intelligence = less waste / attempt***

STRUCTURED PROCEDURE

Skills break a complex task into structured steps, increasing tokens from context-reading and acting but fewer total trials are need to achieve outcome

→ ***Higher success rate with fewer trial counts***

REASONING / OUTPUT CAPABILITY

While skills introduce additional low-cost input tokens (e.g., workflows, templates), it reduces expensive output, retries, and reasoning loops

→ ***Less costly reasoning & correction cycles***

The skills context allows models to reach a solution with significantly less explanatory work

[1]Price per 1 million tokens; [2]Task-macro success rate by agent configuration

Source: SkillsBench, Quinlan & Associates analysis

QUINLAN & ASSOCIATES

KNOWLEDGE BASE AND MEMORY

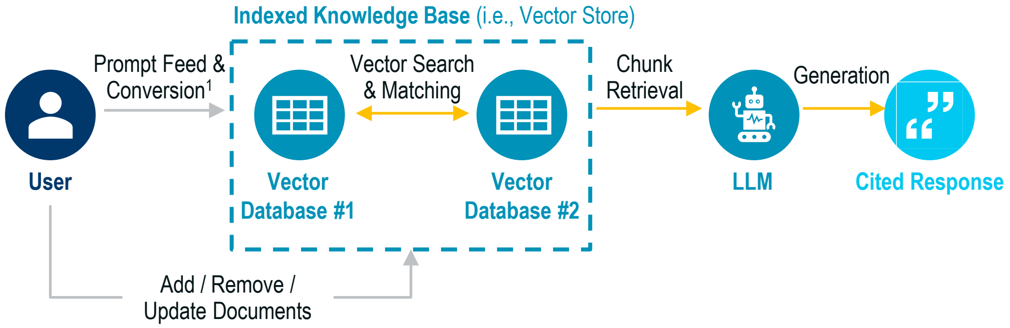
The right context engine should be driven by the intended use case, with RAG excelling at retrieving and grounding responses in the document repositories, while LLM Wiki can embed institutional know-how at scale, making it a powerful source of differentiation

Context Framework

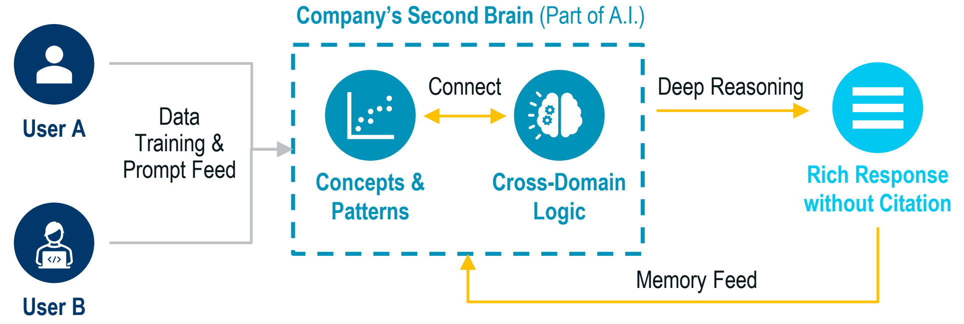
RAG vs. LLM Wiki

User | Knowledge Base / Source / A.I. Model | Output | Manual Flow | Automated Flow | ✓ High | - Moderate | × Low

Retrieval-Augmented Generation ("RAG")



Large Language Model ("LLM") Wiki



	RAG	RAG	LLM Wiki	LLM Wiki
Knowledge Depth & Reasoning Quality	×	Great at retrieval but not synthesis, resembling more of a well-organised filing cabinet limited to the documents that have been loaded into the knowledge base	✓	Broad and deep knowledge internalised (i.e., a second brain) to enable cross-domain reasoning and responses to nuanced, open-ended questions
Knowledge Updates	×	Requires manual addition / revision of the source document, which is subsequently re-chunked, re-run, and re-indexed in the vector store with no automatic feedback loop	✓	A memory feature can automatically produce summary of past interactions and responses, feeding them back into the second brain for any future prompts
Ease of Setup and Maintenance	×	Complex as it requires a chunking logic, an embedding model, and a vector database, and data updates can be automated via batch embedding	✓	Simple with no back-end infrastructure required, where users can drop files into context and update the specific markdown file
Protection from Hallucination Risk	✓	Less likely to fabricate as the source is clear and grounded in the retrieved text	-	Can confidently generate plausible-sounding but wrong specifics (e.g., rates, names, and dates) if no clear guardrails are provided

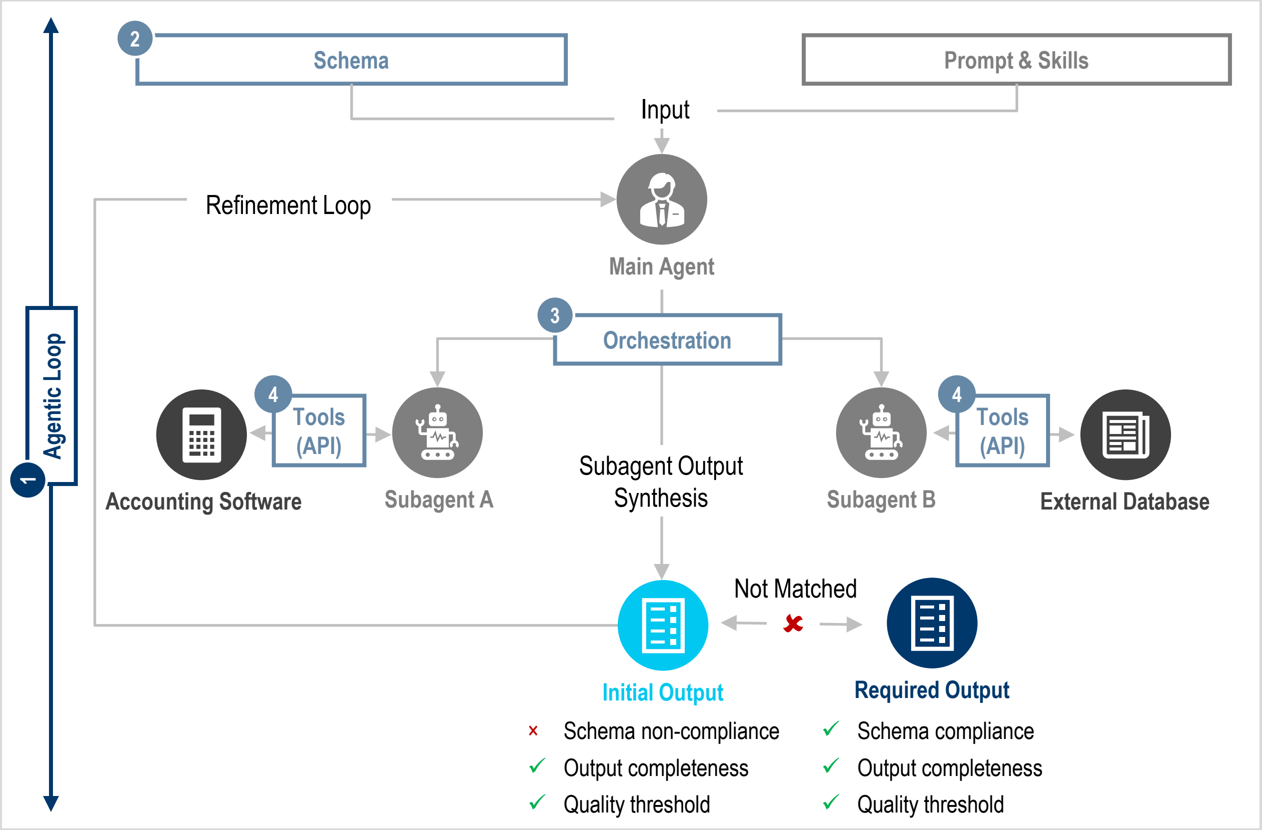
[1]Conversion to numerical embeddings in order to run similarity search, also known as a vector

Source: Memgraph, Quinlan & Associates analysis

AGENTIC LOOP

When the components of harness engineering work together, they enable the agentic loop: the core mechanism by which Agentic A.I. iteratively executes, evaluates, and refines its own outputs autonomously to achieve defined objectives

Agentic Loop
Diagram



1

Agentic Loop

The autonomous cycle that the model executes to complete a specific task (including observing and matching output with the required conditions to instigate a refinement loop, if needed), all without human intervention

Powered by

2 **Schema**
The rulebook that defines the plan of attack and target required output that guides how agents should behave

3 **Orchestration**
The coordination of autonomous agents that reason and act with specific control logic and workflow sequencing

4 **Tools**
Solutions (e.g., API and MCP) that connect the model to external databases and applications

Source: Quinlan & Associates analysis

QUINLAN & ASSOCIATES

DYNAMIC WORKFLOW

At the centre of this loop is orchestration, where financial institutions can effectively organise agents to analyse the objective, coordinate skills and resources, delegate to subagents, test changes, and iterate on feedback to deliver the best outputs

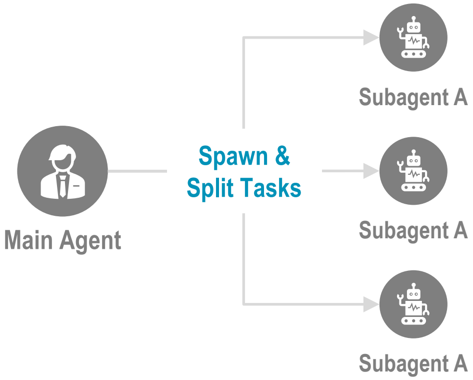
Dynamic Workflow

Description

FEATURE | **EXAMPLE**

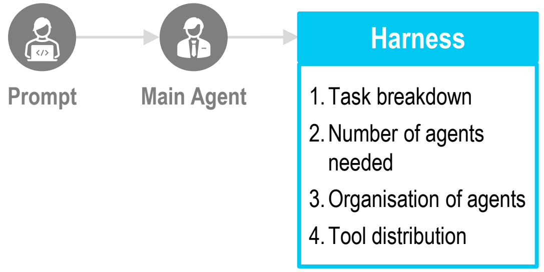
STATIC WORKFLOW

Create agents and orchestrate them to work in parallel so they can complete a designated goal together without human intervention



DYNAMIC WORKFLOW

Spontaneously generates a workflow and the surrounding harness that are tailored to the prescribed task and created on demand



Architectural Design Patterns

Six Major Workflow Patterns

Classify and Act

Classify a task for routing it to the most suitable agent or behaviour, or **classify at the end** to determine an output

Fanout and Synthesise

Split up a task into many smaller steps, each run on an agent with a **clean context window** to deliver results that are **synthesised into one** output

Adversarial Verification

For each spawned agent, run a separate agent to **adversarially verify** its output against a rubric or criteria

Generate and Filter

Generate a number of ideas on a topic and then **filter them by a rubric or by verification** to return only the highest quality, tested ideas

Tournament

Spawn multiple agents that **each attempt the same task** with different approaches to **judge the result** in a pairwise fashion until there's a winner

Loop Until Done

For tasks with an **unknown amount of work**, **loop spawning agents until a stop condition** is met (e.g., no new findings, no more errors, etc.)

Source: Anthropic, Quinlan & Associates analysis

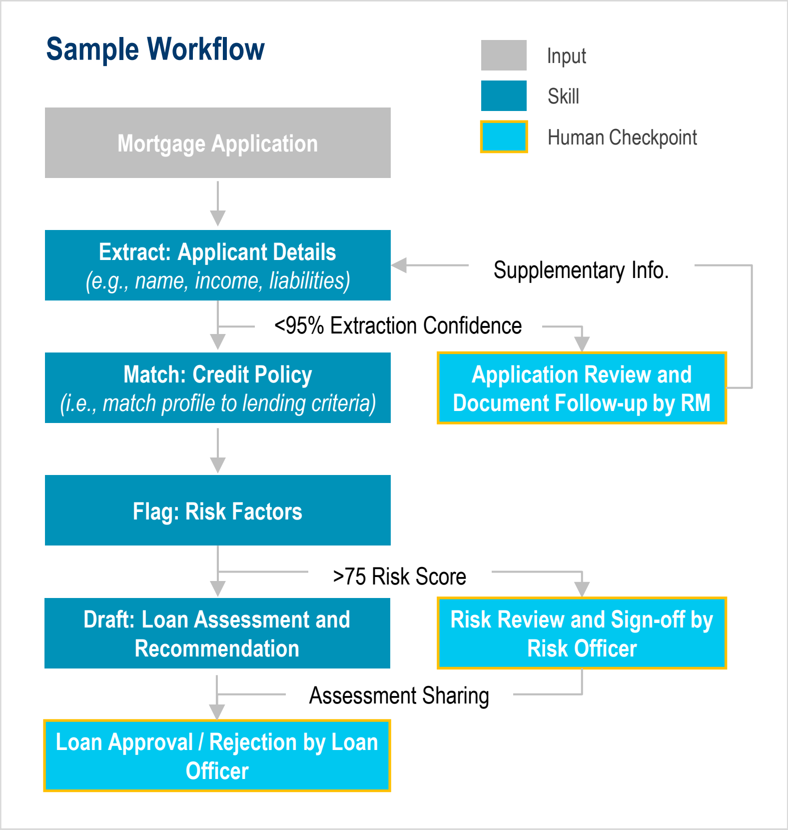
CONTROL MEASURES

With the right prompt inputs and context engine in place, the workflow governs how the A.I. operates to deliver outputs, anchored on (1) well-maintained reusable harness elements and (2) embedded human checkpoints for control

Control Measures

Key Elements, Illustrative Example

	Description
WORKFLOW CONTROLS	Govern how each A.I. model **operates at runtime**: • Permitted input types • Output formats and post-processing validation • Checkpoint trigger conditions (e.g., accuracy threshold breach) and action taken (e.g., human review, hold)
ESCALATION CONTROLS	Define what happens when **something goes wrong**: • Human intervention triggers and exception handling (e.g., responsible person, timeline, and authority) • Risk-based escalation paths, routing different severity levels to different reviewers
RESILIENCE CONTROLS	Ensure **contained damage and smooth recovery** when an A.I. model fails / produces harmful outputs: • Kill switches and its fallback procedures • Recovery procedures (e.g., resolution evidence, authority to reinstate, and monitoring)



HUMAN CHECKPOINT

Defined points in the workflow where human intervention is required, whether it be for approval, escalation, or fallback

Best Practices:

- ✓ Adopt a “design for failure” approach (e.g., log for compliance, escalation triggers, etc.)
- ✓ Define clear trigger conditions for intervention (e.g., risk score, confidence threshold, etc.)
- ✓ Ensure right-sizing of human involvement

*An enforceable control mechanism with **clear accountability***

Source: Quinlan & Associates analysis

SECTION 4

THE ENTERPRISE A.I. OPERATING MODEL

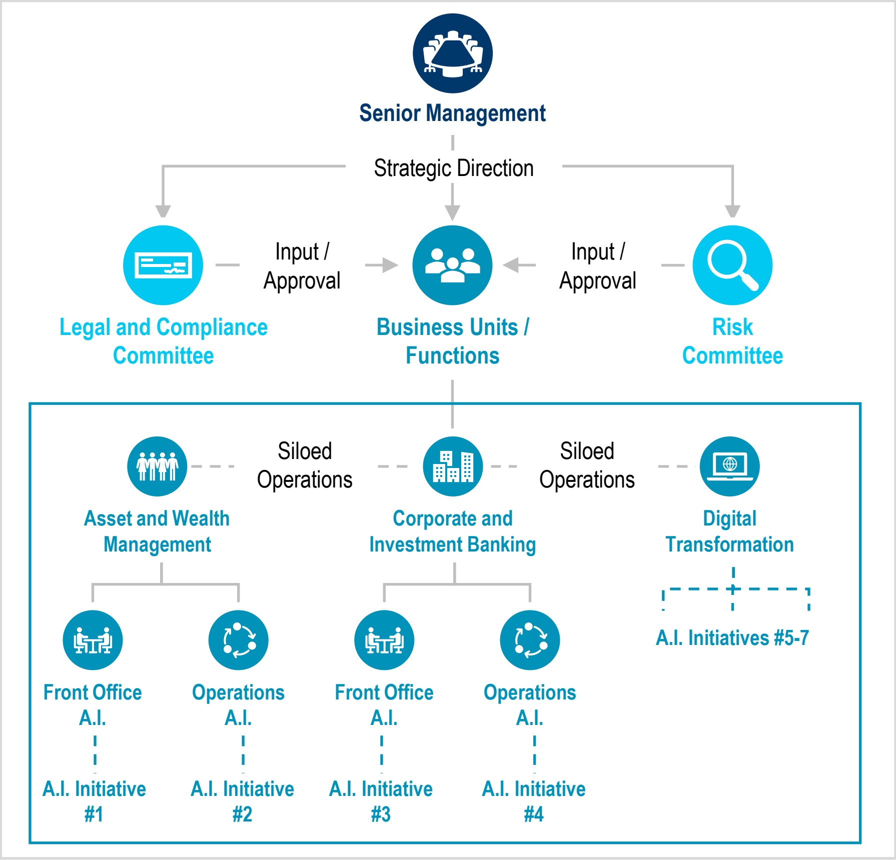
QUINLAN
&ASSOCIATES

THE WIDENING GAP BETWEEN INTENT AND DELIVERY

Despite significant ambitions and budget commitments to embed A.I. across banking workflows, many institutions are experiencing fragmented deployments that generate more internal friction than business value, reflecting several underlying structural problems

A.I. Governance Model

Illustrative Structure



OBSERVATION	IMPACT
Pressure for "More A.I." vs. Scalable A.I. Senior management is under pressure to deliver rapid A.I.-driven ROI, but often lacks visibility into scaling constraints	Pressure translates into demand for more projects and faster delivery, equating activity with impact, flowing downwards
Risk-first Decisions in A.I. Projects As pressure trickles down, governance functions default to a risk-control mindset with strong approval hierarchies and risk committees that simply classify and sign off on initiatives, without a full understanding of the nuances of A.I. use cases	Initiatives are rejected early due to perceived risk / complexity, or approved because they are basic, low-value use cases (e.g., taken on-premise from 3^{rd} parties) with constrained deployment
Siloed Execution and Adoption While experimentation is accelerating, A.I. use cases are developed independently across business units, with limited interaction during execution across different teams to provide feedback / inputs	Pilots remain as pilots without the ability to scale and those that do reach deployment does not scale, resulting in low overall adoption and impact

Creates a recurring cycle that reinforces top- layer expectations.

A structural change is required to break this cycle

Source: Quinlan & Associates analysis

COMMON BLOCKERS ACROSS THE A.I. VALUE CHAIN

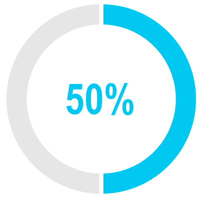
Across the A.I. lifecycle, a consistent pattern emerges: there are multiple blockers at different stages, all tracing back to a single structural root cause: governance and operating models that are no longer fit to support A.I. at scale

A.I. Value Chain
Common Blockers

	INFORMATION INGESTION	WORKFLOW DEFINITION	MODEL SELECTION	AGENT ASSEMBLY	OUTPUT EVALUATION	DEPLOYMENT & OPERATION
Description	**Fragmented and Unreliable Data** A.I. systems often ingest inconsistent, outdated, or over-permissioned data across repositories, resulting in hallucination risks	**Automation Without Clear Accountability** A.I. is embedded into workflows without redesigning ownership, leading to ambiguity over accountability and potentially more rework	**Trade-off in Capability vs. Explainability** Institutions may default to less capable, pre-approved models, prohibiting the use of outperforming, frontier models at the outset due to model risk requirements	**Decentralised Build Without Standards** When capability is handed to different functions, teams assemble their own prompts and agents, resulting in duplicative efforts and resources	**Box-Ticking Evaluation Exercise** Output evaluation relies on a mere box-ticking exercise that fail to reflect the standards / expectations of a regulator or a client	**Scale Limitations Due to Human Approvals** A.I. output is generated instantly but constrained by independent human review and sign-off spanning across multiple teams
Example	**Insight Personalisation** Each model draws from a single source (e.g., loans, investments) with no unified view of a customer	**Insurance Policy Memo** Share policy memo to multiple teams, but it is unclear who is accountable (agent, legal, underwriter)	**Internal GPT Copilot** Deploy an on-premise proprietary copilot with limited reasoning ability from outdated LLMs[1]	**Client Communications** Regional teams develop the same assistant, all with misaligned messaging and use of tones	**Portfolio Reporting** Generate monthly portfolio commentary for wealth clients that do not meet regulatory disclaimers	**Marketing Collateral** Create product brochures that may be approved by marketing team but rejected by legal team
Underlying Problem	**Poor Data Gov.[1]** × Multiple repositories × Inconsistent ownership	**Opaque Decisioning** × Unclear RACI[2] framework for A.I.-influenced decisions	**"Simplistic" Gov.[1]** × Binary approval / prohibitions (vs. risk-tiered)	**Lack of Coordination** × Decentralised governance × Limited interoperability	**No Shared Standards** × Inconsistent judgment of what constitutes as "good"	**Unclear Approval Flow** × Siloed sign-offs × Sequential vs. parallel governance

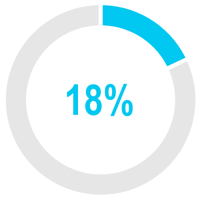
POINTS TO A SINGLE STRUCTURAL PROBLEM – THE GOVERNANCE AND OPERATING MODEL IS NO LONGER FIT TO SUPPORT A.I. ADOPTION AT SCALE

GOVERNANCE & COMPLIANCE BARRIERS:



…of banking institutions consider such barriers to limit A.I. performance

A.I. GOVERNANCE AND CONTROLS:



…of banking institutions are fully confident in their A.I. governance and controls

[1]Large Language Models; [2]Governance, [3]Framework to clarify roles and responsibilities for specific tasks, covering Responsible, Accountable, Consulted, and Informed,

Source: Grant Thornton, Quinlan & Associates analysis

RISE OF GREATER OPERATIONAL DEMANDS

This challenge is set to intensify as roadmaps across major organisations point toward Agentic A.I. with higher autonomy that current operational frameworks cannot accommodate, making A.I. governance revamp no longer an option

Evolution of A.I.

From Traditional to Agentic A.I. Automation

New Features
✓ Offered
- Dependent
✗ Not Offered

COPILOT TO AUTO-PILOT SET TO CO-EXIST

	GEN 1 TRADITIONAL AUTOMATION	GEN 2 GENERATIVE A.I.	GEN 3 A.I. AGENTS	GEN 4 AGENTIC OS / TEAM	Operational Demands
Workflow / Process Automation Execution of task-related steps	✓	✓	✓	✓	With traditional A.I. adoption, governance is largely focused on ensuring outputs are fair, compliant, and explainable, supported by training data quality controls and post-output review processes, making the overall operational footprint relatively contained
Context Awareness Contextual information digestion	✗	✓	✓	✓	
Generative Output Production of content	✗	✓	✓	✓	
Autonomy Takes action independently	✗	✗	✓	✓	Extends beyond human review to pre-authorised action limits and fallback mechanisms, since agents can now execute decisions
API Connectivity Connecting with and using other tools	✗	✗	✓	✓	Requires tight system access controls, with identity, permissions, and authority boundaries beyond traditional output-level checks
Multi-step Implementation Planning & execution of multiple steps	✗	✗	✓	✓	Requires workflow-level governance, requiring runtime constraints, step-level monitoring (i.e., kill switch), and escalation triggers
Orchestration Managing other A.I. agents	✗	✗	✗	✓	Requires end-to-end auditability and traceability across interacting agents and actions

As A.I. solutions across generations co-exist, **governance and the operating model become increasingly important and complex**, not only to adhere to the latest compliance requirements, but from a strategic standpoint that ultimately determines how institutions scale, prioritise, and extract value from A.I.

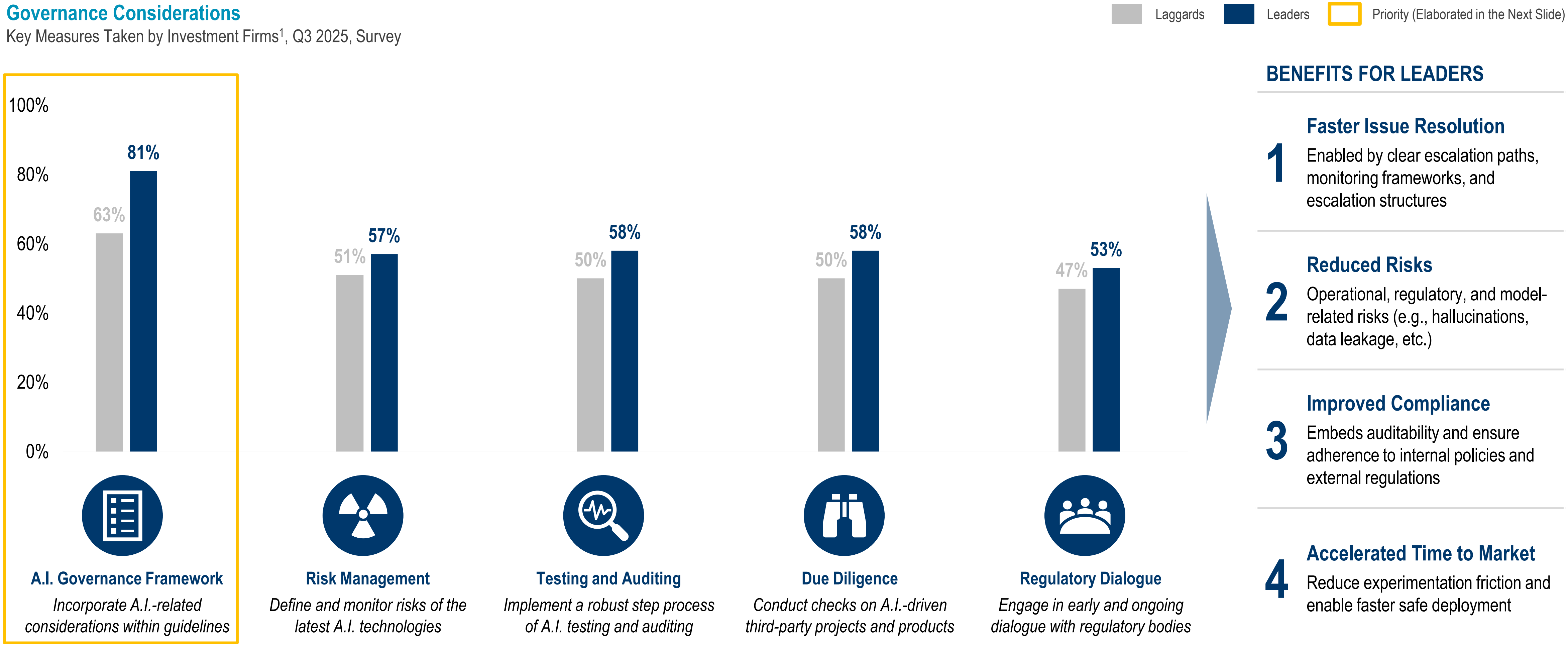
Source: Quinlan & Associates analysis

THE WINDOW FOR GOVERNANCE (1/2) – KEY MEASURES

Meanwhile, institutions that prioritise governance, particularly through the establishment of a clear A.I. governance framework, are able to realise more benefits compared to those with less mature governance structures

Governance Considerations

Key Measures Taken by Investment Firms[1], Q3 2025, Survey



BENEFITS FOR LEADERS

1. **Faster Issue Resolution**
   Enabled by clear escalation paths, monitoring frameworks, and escalation structures
2. **Reduced Risks**
   Operational, regulatory, and model-related risks (e.g., hallucinations, data leakage, etc.)
3. **Improved Compliance**
   Embeds auditability and ensure adherence to internal policies and external regulations
4. **Accelerated Time to Market**
   Reduce experimentation friction and enable faster safe deployment

[1]500 senior executives in asset / wealth management firms, private banks, hedge funds, family office, and other investment firms across top investment markets

Source: The AI-Powered Investment Firm, ThoughtLab, Grant Thornton, Quinlan & Associates analysis

THE WINDOW FOR GOVERNANCE (2/2) – A.I. GOVERNANCE FRAMEWORK

As the A.I. landscape is still evolving and best practice is still being set industry-wide, early governance is a compliance and competitive advantage, where institutions could direct their efforts on establishing the structure, assessment, and accountability

Governance Considerations

Key Factors

1 STRUCTURE

DECENTRALISED

A.I. governance is currently dispersed across functions with limited decision power, resulting in fragmented enforcement and oversight

→ **CENTRALISED**

A centralised decisioning layer with real authority is needed to ensure firm-wide visibility, coordination, and enforcement of A.I. priorities

2 ASSESSMENT

INCONSISTENT JUDGMENT

A.I. initiatives are still assessed case-by-case without a consistent evaluation logic, leading to subjective sign-offs and unclear outcomes

→ **SYSTEMATIC PLAYBOOK**

A standard playbook should guide a central governing body in assessing use cases systematically using objective criteria

3 ACCOUNTABILITY

UNCLEAR ROLES

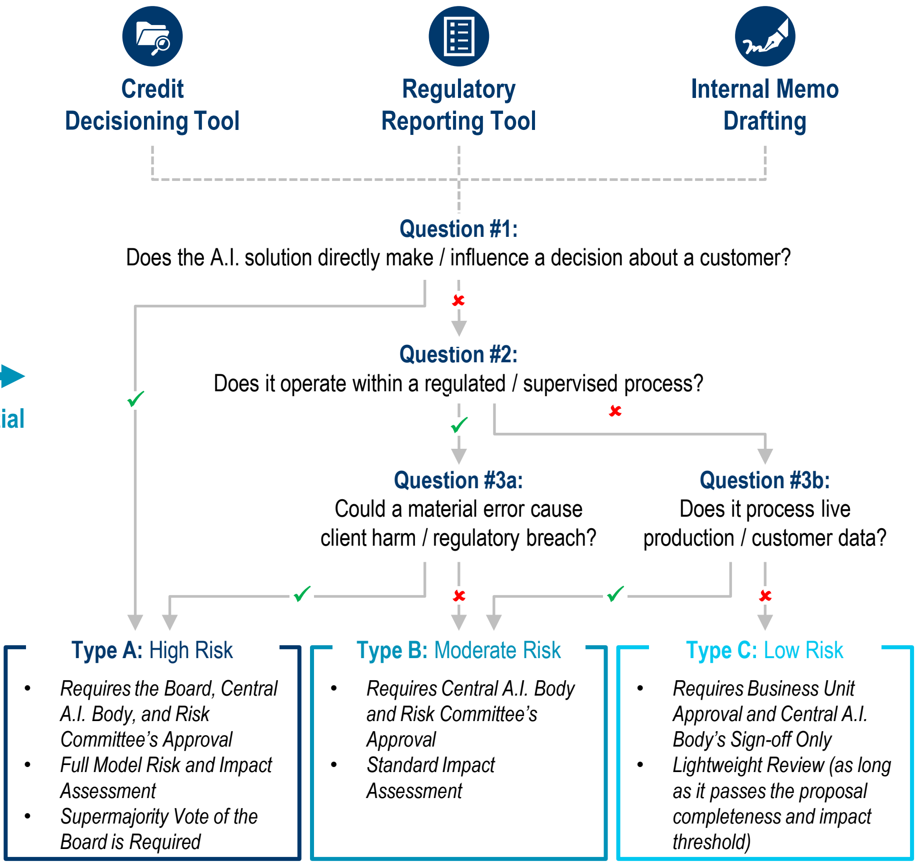
Ownership across A.I. governance is often unclear, with weakly defined RACI structures and inconsistent escalation pathways

→ **ENFORCEABLE ACCOUNTABILITY**

Clear decision rights must define who proposes, validates, and approves, ensuring governance is not just advisory but enforceable

Example of an Initial Risk Assessment →

SYSTEMATIC PLAYBOOK (ILLUSTRATIVE)



Source: Quinlan & Associates analysis

OPERATING MODEL

Every role is becoming managerial in nature due to A.I., with team members now directing and reviewing A.I. outputs, managers orchestrating mixed teams of humans and A.I., and leaders stewarding institutions through a shift they may not fully understand

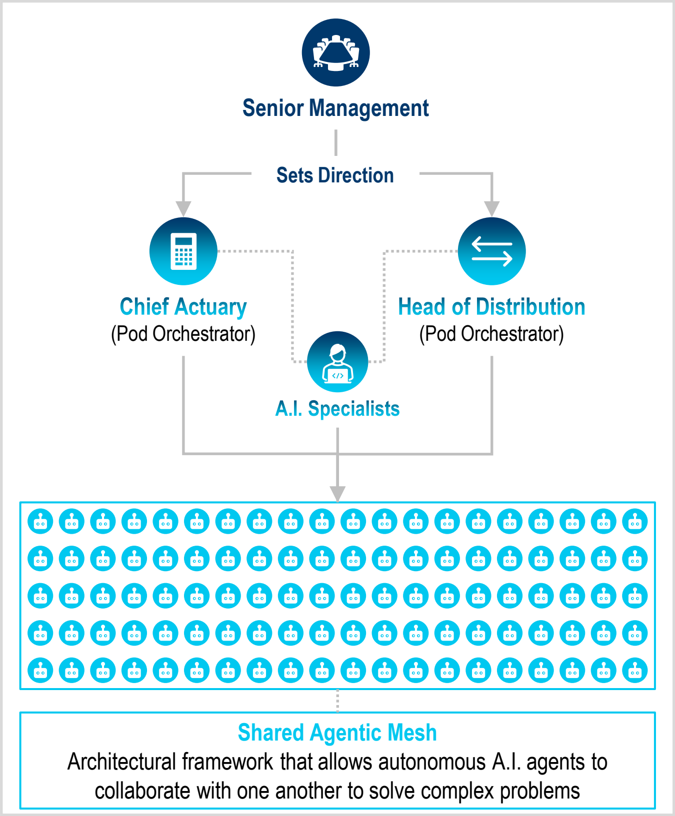
Operating Model

Traditional Team vs. A.I.-enabled Team

Managerial Role · Execution Role · Mixed Role

TRADITIONAL TEAM

A.I.-ENABLED TEAM



	Traditional	A.I.-Enabled
1	**Trust Specialist** Senior leadership set direction and relied on technical teams to recommend and manage technology decisions, given their domain expertise	**Trust Steward** Senior leadership is accountable for how A.I. is used across the organisation, the tone on responsible A.I. use, and leading teams who are often more A.I.-capable
2	**People Coordinator** Middle management allocates work to team members based on capacity and skills, reviews outputs, and manages performance	**Human and A.I. Orchestrator** Workflow design is not a core management skill, and middle management owns morale, ensuring team members feel secure, valued, and clear on changes in roles
3	**Task Executor** Team members act as skilled contributors, producing output directly and reporting to the department heads, with their performance measured	**Task Delegator and Reviewer** Team members decides what to prompt, review A.I. output critically, and make adaptations, along with adding the judgment layer that may be missing in A.I.

Source: Quinlan & Associates analysis

ENVISIONING UPLIFT IN OPERATING MODELS

To accommodate this shift, more broadly, a fundamental operating model shift is needed to unlock scalable near-term deployment, which also creates the groundwork for future agentic capability when the organisation is ready

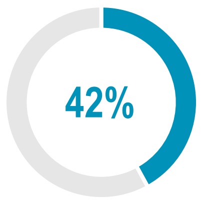
Key Layers

Operating Model

OPERATING MODEL CONSIDERATIONS	FOCUS AREAS	
Data Governance and Ownership The data supply chain problem undermines even well-designed models. For Gen. A.I. and Agentic A.I. that connect patterns across contexts and reasons, reliable core datasets a hard requirement	• Centralised data platform setup • Ownership for the dataset in A.I. workflows • Data quality thresholds for A.I. ingestion	• Rulebook for data access and feed to A.I. systems • Normalise, clean, and tag the data system
Decision-making Mechanism The binary risk tier is the source of time-to-production loss and suboptimal output, requiring making scrutiny proportionate to cases that warrant it, and adopting a faster pathway for cases that do not	• Pre-deployment framework (e.g., organised pilots, exit criteria, ethical considerations, etc.)	• Tiered approval pathways (e.g., fast-track, time-to-approval, and decision authority) • Clear decision rights allocation
Access Control The question of what the model is permitted to do and what it can access becomes a primary governance concern, which requires hard boundaries independent of the risk tier it sits in	• Identity and access management • Authority boundaries (what is autonomous vs. what requires authorisation)	• API controls
Terms of Reference and Standards Shared standards on what constitutes good A.I. output, business cases, or a successful deployment, coupled with clear ownership and accountability, could solve for the outcome / impact drift	• RACI framework for every use case (e.g., model, data, and operational ownership)	• Shared standards (e.g., output quality, value case, pilot success, etc.)
Coordination Structure The fragmentation of teams across functional areas may have merits, but often results in duplicative development and misallocated investment with no tangible impacts, raising coordination needs	• New committee setup (e.g., central coordination body)	• Structured communication and repository (e.g., shared use case registry, feedback loop)
Leadership Development The capability gap at the top of the organisation is the most overlooked aspect, undermining accountability at the point where it matters most	• Programme for senior leaders (focused on fluency to challenges and validations vs. technical depth)	• Decision-making for high-stakes A.I. scenarios without technical understanding • Accountability structures for leaders
Workforce Training and Capability Building The gap between an organisation that has A.I. tools and one that uses them well is a people problem, where without a structured capability, output remains unvalidated, and impact never materialise	• Role-based A.I. training curriculum (e.g., prompt engineering, A.I. output review, orchestrating human-A.I. workflows) with ongoing learning pathways	

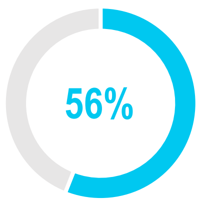
Governance   People & Culture

INSUFFICIENT DATA READINESS:



…of banks cite insufficient readiness as a cause of A.I. project failure

A.I.-RELATED SKILLS GAP:



…of finance leaders consider Gen A.I. as the biggest skills gap, with training as part of addressing this gap

Source: Backbase, Quinlan & Associates analysis

SECTION 6

HOW WE CAN HELP

QUINLAN &ASSOCIATES

QUINLAN & ASSOCIATES

HOW WE CAN HELP

Before pursuing further A.I. initiatives, particularly agentic A.I., we recommend a structured diagnostic to assess institutional readiness and strengthen the foundation for scalable A.I. adoption over the next 12–24 months, with support across the following key areas

How We Can Help
End-to-End Support

UNDERSTAND STATE → **BUILD FOUNDATION** → **ACTIVATE MODEL**

Best Practice Identification

What can we learn from others?

*Evaluate **leading industry practices** across the financial services sector while referencing key **regulatory standards and emerging supervisory expectations** for A.I.*

- Peer mapping
- Competitive benchmarking (e.g., A.I. use cases, governance structures, operating model, etc.)
- Regulatory review across key operating jurisdictions
- Trend analysis and case study development

Internal Diagnostic Assessment

Where are the critical gaps?

*Assess the state of A.I. maturity across strategic and operational dimensions within the institution to **identify gaps / whitespaces** and **inform initial recommendations***

- Use case mapping
- Data, model, and tooling review
- Governance and workflow assessment
- Gap, duplication, and bottleneck identification and analysis
- Severity assessment and prioritisation

Strategy Development / Enhancement

What should we prioritise?

*Define / refine the A.I. strategy, and prioritise **A.I. use cases** based on key criteria, and define the approach to develop those use cases via model selection*

- Business case development
- Use case prioritisation (e.g., by value, feasibility, and risk)
- Deployment approach (e.g., frontier, open-source, or hybrid)
- Service model selection (e.g., build, buy, or partner)
- (If partner) Potential partners evaluation and shortlisting

Target Operating Model ("TOM") Design

What needs to evolve?

*Design a **future-ready operating model** that enables scalable and safe A.I. adoption and prepare any **supporting documentation**, based on agreed recommendations*

- Governance framework design (e.g., data governance, decision-making, coordination, etc.)
- Improved operational workflow / rulebook documentation (e.g., role-based prompt templates, reusable A.I. skill library, human checkpoints, etc.)
- Technology design

Implementation Roadmap Creation

How do we get there?

*Develop an **implementation roadmap** based on TOM & prioritised use cases, sequenced for controlled delivery, capability readiness, and managed dependencies*

- Phased transformation roadmap
- Use case rollout roadmap
- Milestone and KPI definition
- Stakeholder mobilisation plan (including resources, ownership, and internal and external communications)

Execution Support

How do we make it happen?

*Provide **hands-on execution support** to translate the roadmap into tangible outcomes, ensuring initiatives are delivered on time, at quality, and achieve the intended impact*

- Leadership development
- Employee training programme facilitation
- PMO setup and execution tracking
- PoC / prototype development
- MVP build support
- Model / A.I. solution testing support

*Collaborate with **different functions / divisions** and facilitate **senior stakeholder alignment via workshops** to validate findings, surface priorities, and agree on recommendations to drive A.I. to scale across the organisation*

Source: Quinlan & Associates analysis

DISCLAIMER

The information and opinions in this document were prepared by Quinlan & Associates. This document is not financial or investment advice and should not be relied upon for such advice or as a substitute for professional accounting, tax, legal, or financial advice. Quinlan & Associates has made every effort to use reliable, up-to-date, and comprehensive information and analysis in this document, but all information is provided without warranty of any kind, express or implied.

Quinlan & Associates disclaims any responsibility to update the information or conclusions in this document. Quinlan & Associates accepts no liability for any loss arising from any action taken or refrained from as a result of information contained in this document or any documents or sources of information referred to herein, or for any consequential, special, or similar damages, even if advised of the possibility of such damages. This document is not an offer to buy or sell securities or a solicitation of an offer to buy or sell securities.

QUINLAN
&ASSOCIATES

STRATEGY WITH A DIFFERENCE