



Reliable AI Reliable Decisions

Enterprise-grade AI Reliability for Banking & Finance

WHERE OUR WORK COMPOUNDS

1	2	3	4
High Cost of Failure	AI Center of Reliability	Hybrid Agentic Workflows	Voice Agents
Regulated, customer-facing systems	Governance layer for the AI estate	Agent + deterministic systems of record	Contact-center modernization, eval at scale



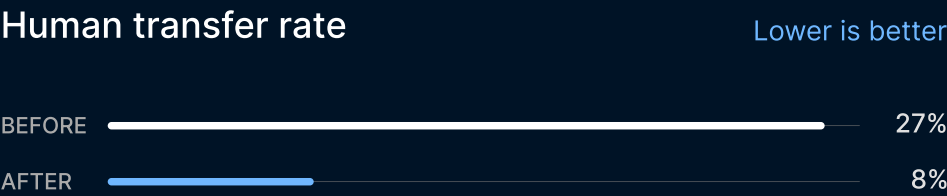
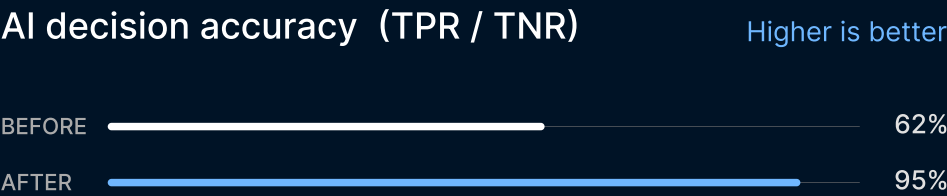
Reliability of Decisions.

Enterprise-grade AI Reliability for Banking & Finance



qktech.com

QK APPROACH — A MEASUREMENT SYSTEM FOR PRODUCTION AI



WHAT THE LOOP RETURNED - CONVERSATIONAL AI AT A UNIVERSAL BANK

120M+

Customers Served

27% → 8%

Human Transfer Rate

2% → 95%

AI Decision Accuracy

175K+

Live Transactions Handled

25 YEARS OF ENGINEERING DIGITAL TRUST — NOW APPLIED TO AI

25+

Years

300+

Enterprise Clients

12K+

Apps Assured

1.2M+

Automation Scripts

25M+

Users

THE SHIFT

Capability has caught up. Confidence has not.

In banking, insurance, capital markets, healthcare and telecom, agents now answer the easy questions. They are being asked something harder: prove every decision can be audited, explained, and trusted at production load.

\$30–40B

Enterprise spend

On generative AI per year

40%+

Project cancellation

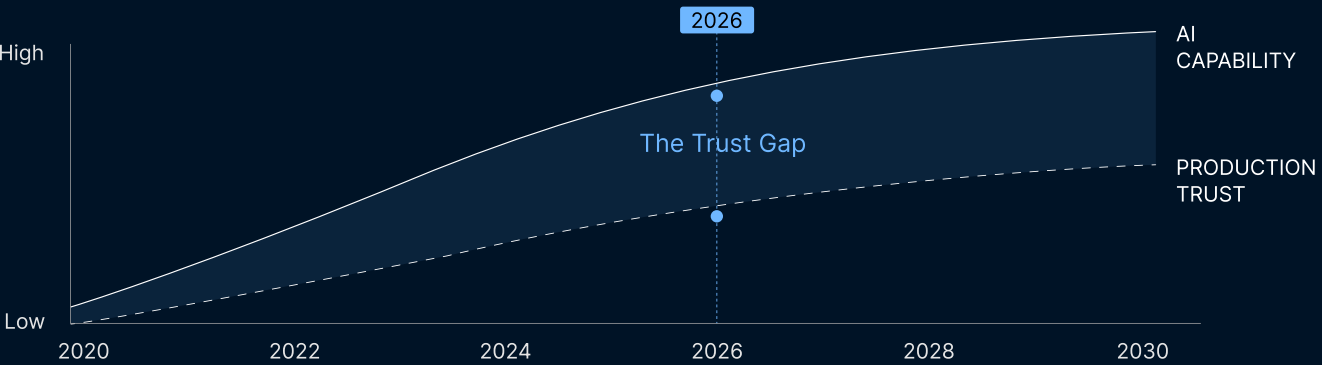
Agentic AI projects expected to be canceled by end-2027

~5%

Measurable yield

Organizations with P&L return from generative AI

WE CLOSE THE TRUST GAP



Sources: McKinsey · MIT Project NANDA, State of AI in Business 2025 · Gartner, Jun 2025.

WHAT WE ENGINEER

Audit-grade evals

Decision rationales captured, sampled, replayable — regulator ready.

Domain reliability libraries

Pre-built eval banks for payments, underwriting, KYC, claims, voice servicing.

Production observability

Live traces · drift · hallucination · cost · CSAT — fed back into governance.

Region-tuned policy agents

Interpret local regulation: RBI, MAS, OCC, EU AI Act, and beyond.

Trustworthy enterprise AI = reliable agent + reliable backend + engineered performance + domain rigour.

Most teams cover Layer 1 with eval tooling. The remaining three layers — backend, performance, domain — are where journeys actually break under enterprise load.

KPI	Business KPI & customer journey	The only metric the business cares about. Every layer rolls up here.
LAYER 1	AI Agent Reliability	Eval · observability · guardrails · tool-use accuracy · hallucination control.
LAYER 2	Backend & Integration Assurance	Core banking · CRM · claims · ERP · API contracts · data integrity.
LAYER 3	Performance Engineering	Load · latency budgets · resilience · scale to millions of journeys.
LAYER 4	Domain & Regulatory Rigour	BFSI · insurance · capital markets · telco · audit-grade test design.

WHAT QK BRINGS TO EACH LAYER

AI Reliability Engineering	Backend & Integration	Performance Engineering	Domain-Native Test Design
Eval suites · regression banks · guardrails · hallucination control — wired into CI/CD.	Two decades across core banking, payments, claims, CRM, ERP. Contract & chaos testing.	Latency budgets, load, resilience for journeys serving millions — proven on 25M+ user systems.	BFSI, insurance, capital markets, telco. We speak the regulator's language. Audit-grade evidence.

THE OPERATING FUNCTION

Every division is building agents. No one is engineering trust.

Copilot Studio, Bedrock, OpenAI — a business unit can deploy an agent in days. Most large enterprises now run dozens, with no shared visibility into what they decide, where they drift, or whether they pass audit. Reliability is the missing institutional function.

AI CENTER OF RELIABILITY - FIVE DISCIPLINES

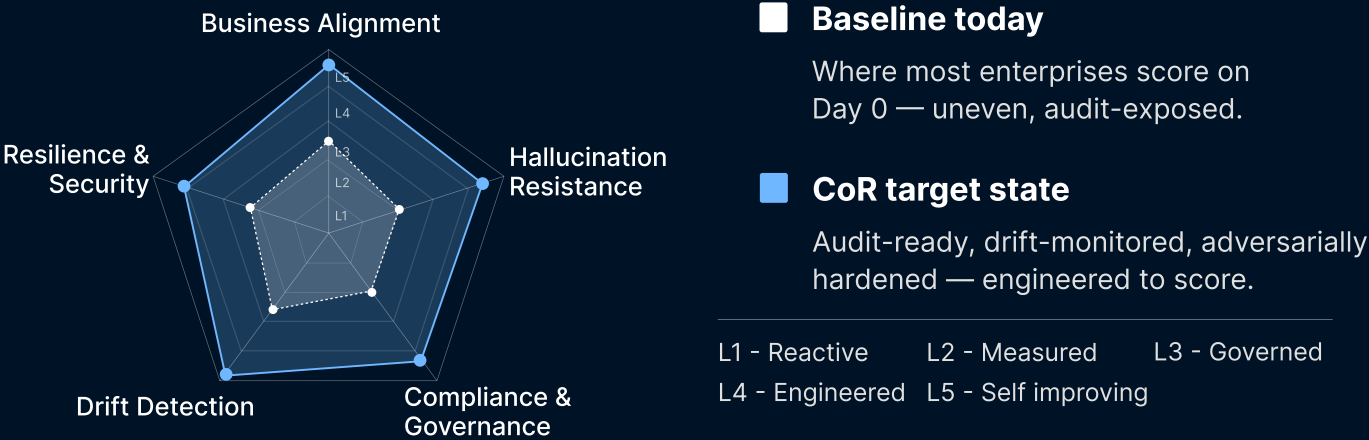
<div>1 Business Alignment</div> <div>Validates every agent achieves intended outcome.</div>	<div>2 Hallucination Resistance</div> <div>Continuous tests for factual groundedness</div>	<div>3 Compliance & Governance</div> <div>Policy, audit trails, regulatory evidence</div>
<div>4 Drift Detection</div> <div>Tracks decay as data & environments evolve.</div>	<div>5 Resilience & Security</div> <div>Hardens against adversarial inputs & failures.</div>	

WHAT'S EASY TO BUILD	WHERE REAL ENGINEERING BEGINS
<ul style="list-style-type: none">• Write a prompt describing the agent's job• Connect to a data source or internal tool• Deploy to a business team in days• Scale to more agents as demand grows	<ul style="list-style-type: none">→ Business alignment — Does it deliver the promised outcome?→ Grounded — not hallucinating — Every output factually correct?→ Audit & policy survivable — Can it pass regulatory review?→ Resilient to drift — Still right three months later?→ Secure under adversarial input — Hardened against prompt manipulation?

Baseline today. Engineer the target state.

Every CoR engagement opens with a maturity baseline across the five disciplines. We score the enterprise's current state, then jointly agree the target state the AI Center of Reliability is built to reach.

MATURITY HEATMAP — FIVE DISCIPLINES



HOW IT IS STRUCTURED — CENTRAL CoE + EMBEDDED DOMAIN SMES

Governance & Policy Standards owner <ul style="list-style-type: none">AI Governance LeadPolicy AnalystCompliance Eng	Architecture & Standards Design authority <ul style="list-style-type: none">AI Principal ArchitectAI Engineers	Eval & Certification Quality gate <ul style="list-style-type: none">AI Eval LeadDomain Eval SpecialistEval Eng	AI Observability Production signal <ul style="list-style-type: none">Observability LeadAI EngDomain Eng
---	--	---	--

THE SHIFT CoR DELIVERS

Without AI CoR	With AI CoR
× Hallucinated answers reach customers	→ Agents ship with quality + compliance evidence
× Agents can't explain decisions — audit failure	→ Audit-ready by design, traceable per decision
× 50 agents running unmonitored	→ Every new agent inherits reliability standards
× Each team rebuilds reliability alone	→ Dashboards show business outcome, not uptime

FROM AGENT DEMO TO AGENT SYSTEM

Reliability of the agent is not enough. Reliability of the journey is the bar.

Enterprise AI converses with customers, executes work, and underwrites new offerings — but the customer's journey still runs through core banking, claims, CRM, and ERP. The model is one link. The chain moves the KPI.

9 / 10 Eval Score Of agents that still miss the business KPI in production	~5% Enterprise GenAI Pilots with measurable P&L impact · MIT, 2025	3 – 7 System of record Per autonomous agent journey
---	---	--

WHERE PROBABILISTIC MEETS DETERMINISTIC

Hallucination CHANNEL	Policy Drift AGENT	API / Contract ORCHESTRATION	KPI MISS SYSTEMS / KPI
Customer Voice · chat · app · WhatsApp	AI Agent LLM · tools · memory · guardrails	Routing + policy Probabilistic + deterministic	Core · CRM · ERP Conversion · CSAT · NPV
PROBABILISTIC	PROBABILISTIC	HYBRID	DETERMINISTIC

WHERE PROBABILISTIC MEETS DETERMINISTIC

1	Backend & integration	31%
2	Performance & load	27%
3	Agent quality	22%
4	Data quality	14%
5	UX / other	6%

72% of broken journeys are not the model.
Eval and observability are necessary. They are not sufficient. QK analysis · 40+ enterprise agent rollouts, 2024–26.