qktech

Reliable AI Reliable Decisions

Enterprise-grade AI Reliability for Banking & Finance

WHERE OUR WORK COMPOUNDS

1

High Cost of Failure

Regulated, customer-facing systems

2

AI Center of Reliability

Governance layer for the AI estate

3

Hybrid Agentic Workflows

Agent + deterministic systems of record

4

Voice Agents

Contact-center modernization, eval at scale

THE SHIFT

Capability has caught up. Confidence has not.

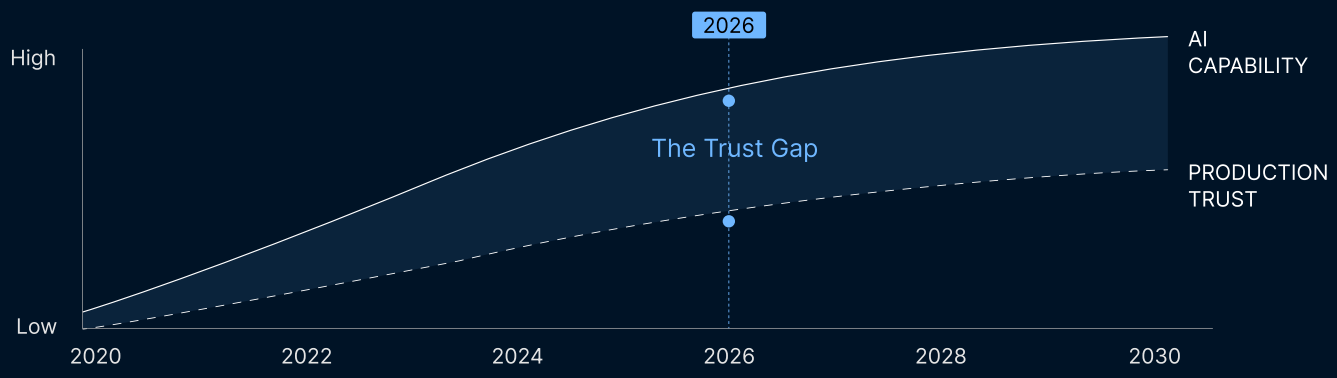
In banking, insurance, capital markets, healthcare and telecom, agents now answer the easy questions. They are being asked something harder: prove every decision can be audited, explained, and trusted at production load.

$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.

Domain-specific evals for regulated, customer-facing journeys.

Generic eval rubrics catch hallucinations. They miss the failure modes that drive regulatory exposure — policy drift, incomplete evidence, biased risk tiering, retries that mask root cause.

	Payments TRANSACTIONS	Underwriting RISK	Servicing CONV . AI
1	**Observe** Txn events, rails, decline codes	**Intake** Application, docs, signals	**Listen** Voice · chat · WhatsApp
2	**Ground** Merchant, BIN, issuer history	**Ground** Policy, KYC, prior decisions	**Ground** Account, exposure, regulation
3	**Decide** Retry · reroute · escalate · settle	**Assess** Risk score, gap analysis	**Resolve** Multi-turn, multi-lingual
4	**Recover** Auditable action, feedback loop	**Decide** Approve · refer · decline	**Escalate** Risk-flagged transfer + trace
	EVAL INJECTION Routing · decline reasoning · retry · latency	EVAL INJECTION Policy · evidence · risk-tier · bias	EVAL INJECTION Intent · risk-signal · transfer · edges
	↓ **Failed-txn rate** Smarter routing	↓ **Time-to-decision** Hours compressed	↓ **Human transfer** 27% → 8% (universal bank)
	↓ **Auth approval** Resolved in-flight	↑ **Decision consistency** Policy applied uniformly	↑ **AI accuracy** 62% → 95%
	↓ **Reconciliation gaps** Audit trail per decision	↑ **Audit-pass rate** Rationale per call	↑ **First-call resolution** Multi-turn handling

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

1 Business Alignment

Validates every agent achieves intended outcome.

2 Hallucination Resistance

Continuous tests for factual groundedness

3 Compliance & Governance

Policy, audit trails, regulatory evidence

4 Drift Detection

Tracks decay as data & environments evolve.

5 Resilience & Security

Hardens against adversarial inputs & failures.

WHAT'S EASY TO BUILD

- 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

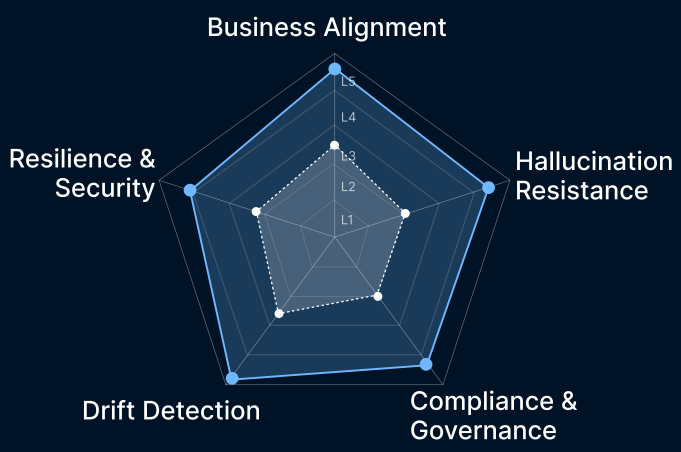
WHERE REAL ENGINEERING BEGINS

- → **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



Baseline today

Where most enterprises score on Day 0 — uneven, audit-exposed.

CoR target state

Audit-ready, drift-monitored, adversarially hardened — engineered to score.

L1 - Reactive L2 - Measured L3 - Governed
L4 - Engineered L5 - Self improving

HOW IT IS STRUCTURED — CENTRAL CoE + EMBEDDED DOMAIN SMES

Governance & Policy

Standards owner

- AI Governance Lead
- Policy Analyst
- Compliance Eng

Architecture & Standards

Design authority

- AI Principal Architect
- AI Engineers

Eval & Certification

Quality gate

- AI Eval Lead
- Domain Eval Specialist
- Eval Eng

AI Observability

Production signal

- Observability Lead
- AI Eng
- Domain Eng

THE SHIFT CoR DELIVERS

Without AI CoR

- × Hallucinated answers reach customers
- × Agents can't explain decisions — audit failure
- × 50 agents running unmonitored
- × Each team rebuilds reliability alone

With AI CoR

- → Agents ship with quality + compliance evidence
- → Audit-ready by design, traceable per decision
- → Every new agent inherits reliability standards
- → 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	Policy Drift	API / Contract	KPI MISS
CHANNEL	AGENT	ORCHESTRATION	SYSTEMS / KPI
Customer	**AI Agent**	**Routing + policy**	**Core · CRM · ERP**
Voice · chat · app · WhatsApp	LLM · tools · memory · guardrails	Probabilistic + deterministic	Conversion · CSAT · NPV
PROBABILISTIC	PROBABILISTIC	HYBRID	DETERMINISTIC

WHERE PROBABILISTIC MEETS DETERMINISTIC



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.

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

Eval suites · regression banks · guardrails · hallucination control — wired into CI/CD.

Backend & Integration

Two decades across core banking, payments, claims, CRM, ERP. Contract & chaos testing.

Performance Engineering

Latency budgets, load, resilience for journeys serving millions — proven on 25M+ user systems.

Domain-Native Test Design

BFSI, insurance, capital markets, telco. We speak the regulator's language. Audit-grade evidence.

THE OPERATING FUNCTION

The troubled AI that almost went live.

A South Asian universal bank deflecting calls across inbound voice, outbound voice, in-app chat, WhatsApp, and credit-risk task agents. Lab dashboards were green. CUG testing — 100 real customer journeys — was not.

44 / 100
Wrong Intent Recognition

31 / 100
Missed Risk Signals

27 / 100
Required Human Transfer

Nothing crashed. No alerts fired. Logs were clean — and decisions were confidently wrong. The exact failure mode that erodes customer trust fastest in regulated industries.

QK APPROACH — A MEASUREMENT SYSTEM FOR PRODUCTION AI

14K+
Domain evals
Real banking utterances. Automated pass/fail. Multiple formulations per intent.

21 + 14
AI + business metrics
TPR · TNR · transfer rate · regulatory edge cases — engineers + domain SMEs.

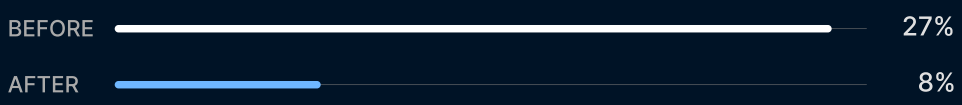
6
Indian languages
Real banking utterances. Automated pass/fail. Multiple formulations per intent.

Live
Continuous on traces
Real banking utterances. Automated pass/fail. Multiple formulations per intent.

QK APPROACH — A MEASUREMENT SYSTEM FOR PRODUCTION AI

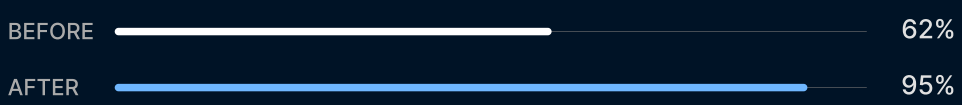
Human transfer rate

Lower is better



AI decision accuracy (TPR / TNR)

Higher is better



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

QK AI AGENT CATALOGUE

We don't just evaluate agents. We design and ship them.

20+ agents in production across BFSI, airlines, manufacturing, HR-tech, fintech, and our own delivery line.

	AGENT ARCHETYPE	AGENT NAME	AGENT SKILLS
1	Enterprise Search Agents	HR Support Agent	HR ticket triage and employee self-service. 16+ subtypes with AI severity classification, cross-timezone SLA management, Employee KB chatbot
2		Live Agent Copilot	Real-time agent-assist chatbot providing conversation summaries, response suggestions, and SOP-guided next steps.
3		3D Modeling KMS Agent	Index/retrieve 3D models and guidelines relevant to the design from 6 different enterprise systems
4	Conversational Agents	Repair Consultant Agent	Multilingual support support for complex long tail scenarios, Video citations for old models, Unstructured Data repository
5		Sales Agent for new model launch	Customer-first sales enablement system ingesting profiles and generating personalized pitch audio
6		Financial Advise Agent	Multi-agent orchestration platform for personalized financial advice via chat

	AGENT ARCHETYPE	AGENT NAME	AGENT SKILLS
7	Task Automation Agents	Trade Reconciliation Agent	Automated audio transcription, Reconciliation engine, Audit trail logging
8		Sales Insights Agent	Market data analyzer, competitive intelligence, promotions planner
9		Timesheet and Work Summary Agent	Automated timesheet lifecycle with custom objects, conditional logic, and workflows that auto-update time hours eliminating manual entry Aggregates weekly issues, summarizes project health, and auto-triggers Friday status emails with action items
10		Contract Lifecycle and Support Routing Agent	AMC contract management with automated expiry alerts, routing across 6 service categories, and self-service portal.
11		Recruitment Agent	JD Structuring, Candidate Matching, Interview Analysis and Feedback
12	Recommendation Agents	Digital twin - journey intelligence and routing	Journey categorization (same-day/next-day/future), contact metadata analysis, multi-channel routing, and AI-to-human handoff
13		Airline Refund Agent	Managing high volume complex refund processing for cancellations either by passenger or airline
14		Supplier Recommendation Agent	Analyze Inventory levels atw service workshops, Calculate predicted lead times from the suppliers for that zone, recommend suppliers, get human approval to place orders
15	AI in SDLC Agents	AiBuddy - QE Generator Agents	Requirements Creator, Test Case Generator, Workflow Modeler, Coverage & Change Impact Analyzer.
16		AiBuddy - QE Automation Agents	Functional Tester, API Tester, Service Virtualizer, Early CX Analyser, Auto Healer
17		AiBuddy - QE Reseliency Agents	Discovery Agent, Experiment Planner, Resiliency Analyser.
18		AiBuddy - QE Prod Insights Agents	Change Analyzer, Reliability Predictor, Action & Insights Router
19		Dhruva - Automation Migration Agent	Legacy automation framework analyzer, Manifest creation, Script translation to modern platforms
20		Anabot - SRE Agent	End-to-end journey assurance across deterministic and non-deterministic systems, Defect analysis for automated root cause identification and resolution recommendations

qktech

Reliability of Decisions.

Enterprise-grade AI Reliability for Banking & Finance

qktech.com