

What Observable Governance delivers in financial services

Benefits, growth and savings. Continuum as the path. Written in the voice of a working AI programme note.

GOVERNANCE FLOWS DOWN · OBSERVABLE FLOWS UP

Start here

Banks already have chat tools and first agents. People already use unofficial ones. Supervisors now ask two things: who may touch the customer file, and how do you know?

Use is common. Profit is not. McKinsey finds 88 in 100 firms use AI. Only 37 in 100 can show any effect on operating profit. About 6 in 100 see a material effect.

In banking they once sized generative AI at \$200–340 billion a year if the uses were actually live. Most of that is still in trials.

Why? The official tool cannot be halted. It cannot connect to the core system. Nobody can look up last month's yes or no. So risk will not bless it. Staff keep the unofficial tool.

What you get if you fix that: the assistant is allowed to finish the job. Growth is the process you can put it on. Saving is the waiting and the duplicate tool you stop running.

This note does not invent “saves \$X per bank.” McKinsey does not publish that. Size the first process. Then copy it.

Words used here

- Observable Governance — the rule that applies to a live step can be shown, and last month's allow or deny can be looked up, with the rule version.
- Continuum — the shared path: owner, rules, door into the official system, record.
- Core / system of record — payments, policy admin, claims, core banking, custody.
- Critical or important function (DORA) — work whose failure hurts customers or the firm's ability to run.

1. What it delivers — three tests, in a bank

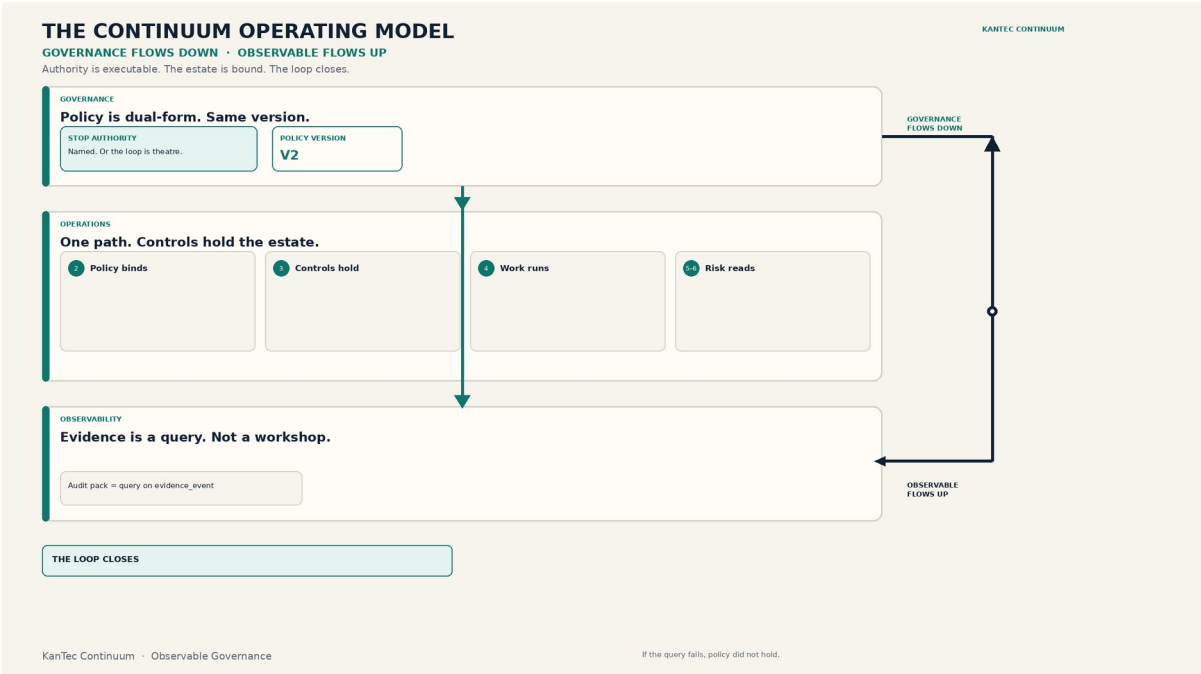


Figure. Policy down onto the payment, claim or advice step. Proof back up to risk, audit and the board.

Test	What a bank can show
Authority is executable	A named stop owner can halt an agent on a payments or claims path. The rule exists as intent and as code, same version.
The estate is bound	The assistant uses the same attested door as every other caller — API and, for agents, a governed MCP. Unscoped tools fail closed. DORA timeout is deny, not hope.
The loop closes	Allow or deny writes who (teller, bot or agent), what, which rule version, which door. The next rule cites those records. The board pack is that query.

2. Benefits by the people who feel them

Business line — growth they can ship

- The assistant can update the real file, not only draft in a side window.
- The unofficial tool becomes a named tool — or it stops.
- Risk can answer from a look-up, not from a workshop.

Risk, compliance, audit — a pack that exists

- One record serves DORA, the AI Act and the board.
- You can show last month's updates stayed on the path you promised.
- “Just this once” has an end date.

Operations and technology — one plane

- No second “AI SOC.” Identity, integration and the agent tool-firewall are PEP classes on the same path.
- Exit is designed (DORA Article 28). Fail-closed onto a named alternative, not a war-room.

Customer and colleague

- They are told when they are talking to an AI (Article 50, live from 2 August 2026).

- A wrong step can be found. The record is searchable.

3. Growth opportunities — where the path pays

McKinsey’s banking work put generative AI at \$200–340 billion a year if the uses were fully live. That is a pool, not a promise. It only shows up when the tool sits in the real process.

Domain	What Observable Governance unblocks
Customer operations	Servicing and complaints agents that may read and, where the rule allows, write the case — not a chatbot that stops at a summary.
Distribution	Advice and next-best-action tools that can be shown to conduct and product governance. Growth without a mystery model.
Claims and underwriting	A first-notice or triage agent on an attested door. The file stays in the residency tier. The deny is visible.
Financial crime / KYC	Analyst assistants on a halt-able path. High-risk Annex III clock is 2 December 2027 — build the path now.
Engineering	Coding agents that cannot land unsigned artefacts. Throughput without a shadow pipeline.

Growth rule: pick the process that is already leaking into unofficial tools. That is the demand signal. Do not start with a platform programme.

4. Savings — what you can defend, and what you cannot

What the research lets you say

- Waiting. On programmes McKinsey watched, 30 to 50 percent of build time was permission and rework. A shared path cut that wait. They have seen approval time drop by as much as 90 percent.
- Missing profit. 88 in 100 use AI. 37 in 100 show any profit effect. About 6 in 100 show a material one.
- When it is live. A selected McKinsey client group that changed how work runs saw about \$3 back per \$1 after the tool was live. Not a promise for a demo.
- The banking pool. \$200–340 billion a year if uses are implemented. A live path takes a slice. A stranded trial takes none.

Savings a bank can size on a beachhead

- Hours not spent rebuilding the same risk pack for each pilot.
- Unofficial licences, plugins and duplicate prompt-spend retired once the attested assistant does the job.
- Fewer dual-running teams (official bot plus the private one that actually works).
- Incident and audit cost: reconstruct-from-Slack versus query-the-store.
- Concentration and exit drills that are paper today and a switch tomorrow (DORA).

What this note will not invent

There is no honest industry-wide “Observable Governance saves 12 percent of opex” figure. Shadow-AI breach add-ons published by IBM are not McKinsey and are not a run-rate saving. Size the first process: hours in the permission queue, unofficial tools to retire, write-backs per week once live.

5. How the saving and the growth show up in 90 days

- Day 0 — stop owner named on one critical or important function (servicing, claims, onboarding).
- Days 1–30 — unofficial tools on that path listed. Dual-form rule. Residency and key tier set.
- Days 31–60 — attested door live (API + governed MCP). Unscoped tool-call fails closed.

- Days 61–90 — query works. First rule change cites live records. DORA register row and AI Act transparency file point at the same store. Unofficial twin retired or denied.

That is the first saving (permission time and a retired twin) and the first growth (write-back on a live file). Clone the path. Do not buy a second tower.

6. Decision

Put this under the next agent that wants to touch a customer file.

Fund one path. Not a new framework.

Growth is the process you are then allowed to scale. Saving is the queue and the unofficial twin you stop running.

If you cannot look up last month's yes or no on a payment or a claim, you are still in the group that uses AI and cannot show it in profit.

— KanTec Platform Engineering · Financial services brief —