SOUTHERN BANK (FICTIONAL)

AI governance evidence report

Against the RBI FREE-AI Framework

Institution	Southern Bank (fictional), scheduled commercial bank, India
Reporting period	1 Jul – 11 Sep 2026 (Jul–Sep 2026 to date)
Figures as of	11 Sep 2026, 09:41 IST
Framework	RBI FREE-AI committee report, 13 Aug 2025 · 7 Sutras · 6 pillars · 26 recommendations
Prepared with	Winsen Rho

READ THIS FIRST

- **This is a sample.** Southern Bank, its people, vendors, agents and every figure in this report are fictional. No real institution is described.
- **FREE-AI is not a binding direction.** It is a report of recommendations by a committee set up by the Reserve Bank of India. It creates no obligation on its own.
- **No certification.** This report maps the bank's evidence to the recommendations. It does not certify, validate or attest compliance, and neither does Winsen Rho.
- **The mapping is the bank's own.** Which controls answer which recommendation is Southern Bank's interpretation, maintained by its compliance team. The bank decides and remains accountable.

CONTENTS

About this report

A record of what Southern Bank can show, from its own signed evidence, against each FREE-AI recommendation that is addressed to a regulated entity. Where the evidence is incomplete, the report says so and names who owns the fix.

How to read it

Each recommendation gets one status. It is the weakest status among the controls and requirements the bank maps to it. One open item is enough to mark a broad recommendation as partial or gap, so read the status together with the evidence listed beneath it.

Every figure carries its period and as-of date, and resolves to a record ID in Annex A. Ages are days since the item was opened, counted to 11 Sep 2026.

STATUS	MEANING
Evidence in place	Every mapped control has current evidence.
Partial	Evidence exists but is incomplete for part of the estate or period.
Gap	A mapped control has no evidence for part of the estate. It has an owner and an age.
Not applicable	Addressed to RBI, government or industry, not to a bank.

> Language used throughout: "association" or "coincides with", never a claim that one thing caused another. Rho supplies evidence. The bank reads it, decides, and remains accountable (FREE-AI Sutra 5; para 4.4.78).

SECTION 1

Summary for the Board

Southern Bank runs 187 AI agents, 41 of them high-risk. Of the 13 FREE-AI recommendations addressed to a bank, 3 have evidence in place, 5 are partial and 5 have at least one gap. Every gap has a named owner.

ESTATE AT A GLANCE · AS OF 11 SEP 2026

187	41	9	14
AI agents in the register	rated high-risk	vendor arrangements	open findings
A 62 · B 41 · C1 71 · C2 13 +11 since 30 Jun (176)	183 of 187 observed by Rho 4 registered, never connected	2 material · 5 non-material 1 licensed · 1 under review	4 high · 6 medium · 4 low all owned · oldest 102 days

Classes: A built by Southern · B vendor technology run by Southern · C1 vendor-operated AI service · C2 AI embedded in a SaaS product. Source: register export inventory_2026-09-11.csv; pk_board_q3.

Position against FREE-AI

13 recommendations addressed to a regulated entity

3 Evidence in place · **5** Partial · **5** Gap

The other 13 recommendations are addressed to RBI, government or industry and are not applicable to a bank (section 2).

Control obligations evidenced FREE-AI, outsourcing Directions, DPDP	**62 of 72 (86%)**
Control gaps · partials	**6 · 4**
Supervisor questions answerable internal question set v3, 118 questions	**97 of 118 (82%)**
Open findings with a named owner	**14 of 14**

CAPACITY	Rec 10 Partial			
GOVERNANCE	Rec 14 Gap	Rec 15 Partial	Rec 16 Gap	Rec 17 In place
PROTECTION	Rec 18 Gap	Rec 19 Partial	Rec 20 In place	Rec 21 Gap
	Rec 22 In place			
ASSURANCE	Rec 23 Gap	Rec 24 Partial	Rec 25 Partial	

The Infrastructure and Policy pillars hold no recommendation addressed to banks.

Three things the board should know

a **Vendor X claims 94% task success for the Collections Voice Agent. Measured on Southern's own customers, it is 71%.**

61,131 of 86,100 calls, 21 Jan – 31 Mar 2026, outcomes matured. Matched against Southern's own collectors on 8,412 accounts, the agent is cheaper per good outcome (₹37.70 against ₹60.29) but keeps fewer promises (61% against 68%) and draws more complaints (14.2 against 9.1 per 10,000 contacts).

f_002 · Anjali Rao · open 41 days · to be raised at the Vendor X renewal, 31 Jan 2027

b **Vendor X changed the agent's model on 14 Aug without the notice its contract requires.**

Version 4.1.8 became 4.2.1 at 01:30. The nightly probe caught it at 03:00, before the first call of the day. Tier-2 promise-to-pay conversion has since fallen from 0.51 to 0.44, below the pre-registered floor of 0.45. The fall is associated with the version change; customer mix accounts for 1 of the 7 points (inv_0418). A move to recommend-only was proposed on 5 Sep and has waited 6 days for the owner.

f_001 · Neha Kapoor · 28 days · f_005 · Anjali Rao · 6 days

c **Four registered agents have never connected, so nothing they do can be measured.**

The oldest, the Mortgage Pre-qualification Assistant, has been silent for 102 days. Its kill switch was configured on 1 Jun and has never been drilled.

f_003 · Vikram Nair · 102 days · C-15 · Dev Subramanian · 102 days

Headline actions

#	ACTION	OWNER	FREE-AI
1	Decide the recommend-only proposal for the Collections Voice Agent (proposed 5 Sep).	Anjali Rao	Rec 16, 21
2	Name owners for the 12 unowned AI features inside SaaS products, and complete the 14 missing purpose records.	Vikram Nair · Priya Menon	Rec 14, 23
3	Connect the 4 silent agents, and drill the kill switch on the Mortgage Pre-qualification Assistant.	Vikram Nair · Dev Subramanian	Rec 21, 23
4	Complete the 2 overdue independent validations of high-risk agents, due 30 Sep.	Priya Menon	Rec 16, 24
5	Record customer AI disclosure for the CRM Next-Best-Action feature and Transaction Fraud Screening.	Anjali Rao	Rec 18
6	Obtain Vendor X's §8.4 reply and sub-contractor attestation; carry f_001, f_002 and f_009 into the 31 Jan 2027 renewal.	Neha Kapoor	Annex IV
7	Assign an owner to the 3 public AI claims that diverge from measured results.	Board to assign	Rec 25

SECTION 2

Scope and method

FREE-AI names who should act on each recommendation. This report covers the 13 recommendations where a regulated entity is named, in full or jointly with a regulator.

2.1 Which recommendations apply

The 26 recommendations sit under six pillars: three enable innovation (Infrastructure, Policy, Capacity) and three mitigate risk (Governance, Protection, Assurance). Each names an addressee and a time horizon (paras 4.4.3 to 4.4.5).

REC	RECOMMENDATION	ADDRESSED TO	IN THIS REPORT
APPLICABLE · A REGULATED ENTITY IS NAMED			
10	Capacity building within REs	REs	Partial
14	Board-approved AI policy	REs and industry	Gap
15	Data lifecycle governance	REs	Partial
16	AI system governance framework	REs	Gap
17	Product approval process	REs	In place
18	Consumer protection	REs	Gap
19	Cybersecurity measures	REs	Partial
20	Red teaming	REs	In place
21	Business continuity plan for AI systems	REs	Gap
22	AI incident reporting and sectoral risk intelligence	REs, regulators (bank part: detect and report)	In place
23	AI inventory and sector-wide repository	Regulators and REs (bank part: internal inventory)	Gap
24	AI audit framework	Supervisors and REs (bank part)	Partial
25	Disclosures by REs	REs, regulators (bank part)	Partial
NOT APPLICABLE · ADDRESSED TO RBI, GOVERNMENT OR INDUSTRY			
1-5	Data infrastructure, innovation sandbox, incentives and funding, indigenous financial-sector models, AI with DPI	Regulators, government, SROs, industry	n/a
6-9	Adaptive policies; AI-based affirmative action; AI liability framework; AI institutional framework. Under Rec 8 the bank still remains liable for customer loss (para 4.4.29); see Sutra 5.	RBI, regulators	n/a
11-13	Capacity for regulators; sharing best practices; recognising responsible innovation	RBI, industry bodies, SROs	n/a
26	AI toolkit. Southern maps one control here (C-10, judge calibration). Toolkits complement, not replace, the bank's own validation (para 4.4.78).	Regulators and industry	n/a

Also in scope: FREE-AI's note on third-party AI (para 3.2.3.1) and its suggested vendor clauses (Annexure IV). They are not numbered recommendations, so they are reported separately in section 4 and not counted in the 13. Southern's Dubai branch runs no AI agents; the UAE regime is watched, not mapped. Southern has no EU entity. RBI's draft Model Risk Management guidance (24 Jun 2026) is a draft and is not covered here.

2.2 How evidence is drawn

1. **Recorded where the agent acts.** The Rho Bridge runs inside Southern's network and records each decision: inputs, tools used, the authority in force and the action taken. Identifiers are tokenised in the bank before anything leaves.
2. **Outcomes from the bank's own systems.** Results come from Southern's loan management system, collections dispositions, CRM, complaints register, bureau snapshots and card disputes, not from vendor reports.
3. **Every figure is a counted rate.** Each carries numerator, denominator, cohort, period and as-of date.
4. **Signed and reproducible.** Evidence packs are signed by the named second-line officer, object-locked for 7 years from period end, and can be regenerated from stored inputs to show the same figures.
5. **Nothing is overwritten.** When a figure matures or is restated, both views stay signed and retrievable.

WHAT RHO DOES AND DOES NOT DO

Rho does	Inventory and discovery; decision records; outcome tracking; vendor scorecards; human-oversight records; alerts; recording control instructions for the bank's gateway; signed evidence packs.
Rho does not	Validate models, audit, certify or file with a regulator. Validation stays with the bank and parties independent of it. The bank's gateway enforces any pause or restriction; Rho records the instruction.
The bank does	Maps controls to FREE-AI, decides on every finding, signs every pack and remains accountable for every agent, including vendor agents.

2.3 How a status is set

● Evidence in place	Every control and requirement that Southern maps to the recommendation has current evidence in Rho for the estate it covers. Minor exceptions the bank has accepted are stated in the text.
● Partial	Evidence exists, but at least one mapped control is incomplete for part of the estate, a cohort or the period. The shortfall has an owner.
● Gap	At least one mapped control has no evidence for part of the estate. The gap has a named owner and an age.
● Not applicable	The recommendation is addressed to RBI, government or industry. It is listed, never removed.

Source of mapping: Southern's control framework (24 controls, C-xx) and requirement register (R-xx), installed with internal question set v3 on 1 Jul 2026. The recommendation-level status in this report is the weakest status among the items mapped to it. Where the bank maps no control to a recommendation, the status is set from the Rho records listed in the block.

SECTION 3

The Seven Sutras

The Sutras are the committee's guiding principles (para 4.3). They are not tested one by one; they are evidenced through the records below.

SUTRA	IN SHORT	HOW SOUTHERN EVIDENCES IT	SEE
1 · Trust is the foundation	Trust should remain uncompromised.	Expectations are written and signed before an agent goes live (178 of 187 agents). Evidence packs are signed, locked and reproducible. The January promise-kept figure was restated from 63% to 61% as outcomes matured, and both signed views are kept.	Rec 16, 24
2 · People first	AI augments human judgement; people can override; customers know when it is AI.	Humans review a stratified sample: 470 of 39,140 Collections Voice Agent calls in Aug (1.2%). When reviewers overrode the agent (Jan–Mar), they were right on 2,013 of 2,494 (80.71%). AI was disclosed within 10 s on 468 of 470 sampled calls. A move to a more restricted mode is proposed by rules but always decided by a person.	Rec 16, 18, 21
3 · Innovation over restraint	Responsible innovation with purpose.	Each agent carries a business case and is measured against it. The Collections Voice Agent is cheaper per good outcome than the in-house team (₹37.70 against ₹60.29), with weaker promise-keeping. Pilots run with a stated success criterion and an expiry date.	Rec 17
4 · Fairness and equity	Outcomes should not discriminate.	Outcomes are monitored by cohort against a ±8 point tolerance (C-12). Since 16 Aug, tier-2 Hindi promise-to-pay conversion sits 15 points below the tier-1 reference, outside tolerance. This is tracked under f_005 and investigated in inv_0418.	Rec 16, 24
5 · Accountability	Accountability stays with the deploying entity.	175 of 187 agents have a named owner; all 14 open findings have one. Vendor agents are owned inside the bank. Gap: 12 AI features inside SaaS products have no owner.	Rec 14, 23
6 · Understandable by design	Explainability and disclosure.	Any decision can be reconstructed with the authority and policy in force at the time. Every record states how it was observed: direct feed, adapter, inferred, or not at all. 94% of estate episodes were scored (1 Jul – 11 Sep); the rest are marked unscoreable, with a reason, not estimated.	Rec 15, 16, 24
7 · Safety, resilience, sustainability	Secure, resilient, energy efficient.	Nightly probes catch silent change; kill-switch drills on the Collections Voice Agent stopped it in 3.1 s (14 Jul); the BCP drill fell back in 6 m 40 s. Energy use is not measured in Rho and is not evidenced in this report.	Rec 19, 20, 21

SECTION 4

Recommendation by recommendation

Each block states what FREE-AI recommends, the status, the evidence Southern holds in Rho and any open item with its owner and age. Paragraph references are to the committee report.

Capacity

INNOVATION ENABLEMENT · 1 OF 4 RECOMMENDATIONS APPLIES

REC 10 Capacity building within REs 4.4.35–4.4.37

Partial

ASKS

Build AI governance skills for the Board and senior management, and give structured, continuing training to staff who use or oversee AI.

EVIDENCE

- Competency records for every reviewer of a high-risk agent. On the Collections Voice Agent, 3 of 4 reviewers hold current competency.
- Override accuracy by reviewer (Jan–Mar, matured). The reviewer whose competency has lapsed was right on 411 of 612 overrides (67.2%), the lowest of the four, against 80.71% overall.
- The Board Risk Committee receives AI estate packs each quarter (pk_board_q2, signed 14 Jul; pk_board_q3, signed 11 Sep for the 18 Sep meeting).

OPEN

Partial — 4 reviewers of high-risk agents without current competency: R. Das (expired 40 d), P. Nair (30 d), M. Iyer (17 d), T. Bose (never trained). — Priya Menon · 40 d

Board and senior-management training records are held outside Rho and are not assessed here.

Governance

RISK MITIGATION · 4 RECOMMENDATIONS APPLY

REC 14 Board-approved AI policy 4.4.43–4.4.46 · Annexure V

Gap

ASKS

A board-approved AI policy covering governance, accountability, risk appetite, safeguards, auditability, consumer protection, disclosure and the model lifecycle, with a low, medium and high risk classification that is reviewed periodically.

EVIDENCE

- A risk tier is recorded in the register; 41 of 187 agents are rated high-risk.
- 187 of 187 agents have an authority record stating what the agent may do alone and what it may not do (R-32).
- The second line (Priya Menon) approves pre-registrations, policy, promotions and risk class.
- The Board Risk Committee sees the estate quarterly: agents 176 to 187, open findings 9 to 14, readiness 79% to 82% since the last pack.

OPEN

Gap — Accountability: 12 of the 13 AI features embedded in SaaS products have no named owner, including the CRM Next-Best-Action feature (R-11, C-04, f_004). — Vikram Nair · 40 d

The policy document itself is a bank document, not a Rho record. This report evidences how the policy operates, not its text.

REC 15 Data lifecycle governance 4.4.47

Partial

ASKS

Controls for collecting, accessing, using, retaining and deleting data used by AI, in line with laws such as the DPDP Act, and keep sensitive data inside environments the bank controls.

EVIDENCE

- Retention by record type: decision records 7 years; credit decision records 8 years; evidence packs 7 years from period end, object-locked; sampled transcripts 3 years. Call audio stays inside the bank.
- All Rho Cloud storage is in ap-south-1 (Mumbai); nothing is held cross-region.
- In the last 24 hours the Bridge redacted 18,410 fields and forwarded 0 tier-0 fields.

- Data requests, 1 Jul – 11 Sep: 23 (19 erasure, 4 access), median 2.1 days. For dsr_0230, 159 fields were erased across 41 episodes, and the structural decision records kept with a tombstone; certificate signed 10 Sep.
- Lineage: every record states how it was observed (direct feed 106 agents, adapter 64, inferred 13, not connected 4), in line with para 3.3.12.

OPEN

Partial	The bank cannot yet say, for every system, what it does with personal data and on what basis: 14 purpose records are incomplete (f_013).	Priya Menon · 51 d

REC 16 AI system governance framework `4.4.48-4.4.49` — Gap

ASKS

Govern models across the lifecycle with documentation, validation and monitoring for drift. Before using autonomous AI, set safeguards, define what it may do alone and keep humans overseeing medium- and high-risk use. The bank stays liable for what the agent does.

EVIDENCE

- Expectations pre-registered before production for 178 of 187 agents (C-06). Collections Voice Agent: signed 14 Jan 09:00 by Anjali Rao, Rahul Iyer (Vendor X) and Priya Menon; `sha256 e1c9…7b2f`.
- Continuous evaluation with coverage stated: 1,351,626 of 1,437,900 estate episodes scored, 1 Jul – 11 Sep (94%) (C-09).
- Authority and breach detection (C-14): 4 settlements offered outside authority on 13 Aug were detected in 0.4 h and contained 14 minutes later (inc_0042).
- Drift: the nightly stability probe fell from 39/40 to 34/40 on 14 Aug, 1.5 h after an unnotified model change.
- Independent validation: 39 of 41 high-risk agents validated and current (pk_val_credit_q2 is one example).

OPEN

Gap	Purpose record incomplete on 14 agents (C-03, R-03, f_013).	Priya Menon · 51 d
Partial	2 independent validations overdue, due 30 Sep (C-22, R-22, f_008).	Priya Menon · 22 d
Partial	KYC Document Verification 21% unscoreable against a 15% ceiling since 8 Aug; the vendor sends no basis pointers (R-40, f_007).	Farid Sheikh · 34 d
Open	Tier-2 conversion red at 0.44; recommend-only proposed 5 Sep, awaiting the owner (f_005).	Anjali Rao · 6 d

REC 17 Product approval process `4.4.50-4.4.51` — Evidence in place

ASKS

Bring every AI-enabled product into the product approval framework, with AI-specific risk evaluation, and keep the evaluators independent of the developers.

EVIDENCE

- Promotion of the Collections Voice Agent on 21 Jan: dry run against the Dec 2025 pilot (0 would have failed), approved by the second line (Priya Menon), promoted by the platform team (Dev Subramanian), fallback configured.
- Change gate for vendor and internal changes: Vendor Z 2.3 to 2.4 passed 6 probes (8 Sep); credit policy v11 to v12 approved 9 Sep; Vendor Y 3.1 to 3.2 held in the gate.
- Draft policy v13 tested before approval: it would have failed 437 of 33,760 episodes (15 Aug – 10 Sep). Approver: Model Risk Committee.
- Pilots carry a success criterion and an expiry (Hardship Triage pilot expires 30 Sep; f_011, Anjali Rao, 3 d: promote, extend with reason, or retire).

NOTE

9 of 187 agents are not yet pre-registered; the bank has accepted this as a minor exception. AI found inside SaaS products after the fact is reported under Rec 14 and Rec 23.

RELATED Third-party AI and vendor clauses `3.2.3.1 · Annexure IV` — Gap

ASKS

When a vendor uses AI to deliver an outsourced service, that is outsourcing. The committee suggests adding AI-specific clauses to outsourcing agreements, including disclosure of AI use by vendors and their sub-contractors. Not a numbered recommendation.

EVIDENCE

- Vendor register: 9 arrangements, each with its regime recorded. AI clauses present: 27 of 44 across the four vendors assessed.

- Change notification (C-17): Vendor X notified 2 changes (May, June) but not the 14 Aug model change. Vendors Y and W notified all 7 of their changes this quarter.
- Vendor claims measured on Southern's own outcomes: Vendor X 94% claimed, 71% measured; Vendor Y 89% and 81%; Vendor W 91% and 88%.
- Where Vendor X withheld training data provenance, 12,000 Southern calls were scored instead; residual risk accepted by Priya Menon on 2 Jul (C-19).

OPEN

- Gap — Vendor X speech sub-contractor seen in Bridge traffic but not disclosed; attestation due 6 Sep, not received (C-16, f_009). — Neha Kapoor · 19 d
- Gap — Unnotified change under MSA §8.4; letter sent 15 Aug, no reply by the 29 Aug due date (C-17, f_001). Gap under the outsourcing Directions. — Neha Kapoor · 28 d

Protection

RISK MITIGATION · 5 RECOMMENDATIONS APPLY

REC 18 Consumer protection 4.4.53-4.4.55

Gap

ASKS

A board-approved consumer protection framework: tell customers when they deal with AI, let them reach a person, give accessible grievance and recourse, and run awareness campaigns.

EVIDENCE

- Disclosure measured on calls: the Collections Voice Agent disclosed AI within 10 s on 468 of 470 sampled calls (99.6%, Aug). Near miss nm_0071: disclosure late on 38 calls, corrected in the call.
- Complaints linked to the decision they concern: 129 of 142 AI-related complaints (1 Jul – 11 Sep). Of the 142, 33 led to a change, 70 were assessed with no change and 39 are open.
- By nature: contact after a complaint 51, tone or pressure 35, wrong amount 24, not told it was AI 18, other 14.
- Contest: a customer declined for credit in March asked for a re-examination; the decision is being re-run on the version then in force (cmp_2203).

OPEN

- Gap — No customer disclosure record for CRM Next-Best-Action and Transaction Fraud Screening (C-23, R-62). — Anjali Rao · 12 d
- Partial — 13 of 142 complaints cannot be linked to a decision (C-24, R-64). — Anjali Rao · 24 d

Awareness campaigns are run outside Rho and are not assessed here.

REC 19 Cybersecurity measures 4.4.56-4.4.58

Partial

ASKS

Identify AI-specific security risks such as adversarial attacks, data poisoning and model manipulation; keep assessing after deployment; be able to stop an AI system quickly if harm is likely.

EVIDENCE

- The Rho Bridge runs inside the bank and tokenises identifiers before anything leaves; 3 records quarantined in the last 24 hours.
- Each production agent has its own credential with a rotation date (Collections Voice Agent: rotates 20 Jan 2027).
- Access breadth is watched against a baseline: the Collections Voice Agent reached 1,140 distinct customers in one hour against a 90-day baseline of 335 (3.4 times), although each call was individually permitted.
- Dependency scanning found a transitive vulnerability in Transaction Fraud Screening (as_112); the fix is scheduled.
- Discovery watches for unapproved AI use, such as an LLM API key read by 3 identities (see Rec 23).

OPEN

- Open — Scope creep: access breadth 3.4 times baseline (f_006). — Dev Subramanian · 11 d
- Open — Dependency vulnerability, medium, fix scheduled (as_112). — Dev Subramanian · 20 d

REC 20 **Red teaming** 4.4.59–4.4.60 — Evidence in place

ASKS
Structured red teaming across the lifecycle, in proportion to risk: at least half-yearly for medium- and high-risk AI, and again after major changes. Findings documented for audit.

EVIDENCE
- Collections Voice Agent: half-yearly run on 30 Aug; follow-up on 2 Sep; a triggered run on 9 Sep after the model change.
- Red-team register: 9 open findings (4 high, 5 medium), each with an owner, kept apart from the 14 open findings. Scope: the 116 high- and medium-risk agents.
- rt_041 (high): 0 occurrences in 95,240 production episodes, 1 Jul – 11 Sep, checked by shadow replay.

NOTE
This report does not list red-team dates for each of the 116 agents in scope; they are in the red-team register.

REC 21 **Business continuity plan for AI systems** 4.4.61–4.4.62 — Gap

ASKS
Extend BCP to AI-specific degradation: fallback paths, an AI that can declare itself unavailable, drills that simulate drift, and human checks on a sample of decisions (1% is given as an example) to catch silent degradation.

EVIDENCE
- BCP drill, 14 Jul (Vendor X region outage): passed; fallback to the in-house team in 6 m 40 s.
- Kill-switch drills on the Collections Voice Agent: 9.8 s (14 Jan, shadow), 4.2 s (14 Apr), 3.1 s (14 Jul). A combined kill-switch and fallback drill is scheduled for 14 Oct.
- Human sample (C-11): 470 of 39,140 Collections Voice Agent episodes in Aug (1.2%, 8 strata). Across the estate, 14,380 of 1,437,900 episodes were human-calibrated, 1 Jul – 11 Sep (1%).
- Silent degradation caught: the 14 Aug model change was flagged at 03:00, before the first call of the day.

OPEN

Gap	Kill switch on the Mortgage Pre-qualification Assistant configured 1 Jun, never drilled (C-15).	Dev Subramanian · 102 d
Partial	Oversight in practice: 361 of 2,494 overrides made the outcome worse (14.47%, Jan–Mar, matured) (R-38).	Anjali Rao

---

REC 22 **AI incident reporting** 4.4.63–4.4.64 · Annexure VI — Evidence in place

ASKS
Detect and report AI incidents promptly, including agents acting beyond their scope. Regulators are to set up the reporting framework; FREE-AI leaves timing to severity and sets no deadline.

EVIDENCE
- inc_0042: 4 settlements outside authority (₹2.3 L), detected in 0.4 h, contained 14:40, initial notice 20:10, internal report filed 20 Aug by Priya Menon (pk_inc_0042). 3 settlements honoured; 1 reversed, and that complaint resolved.
- inc_0043: unnotified model change, detected in 1.5 h; post-incident review signed 28 Aug (pk_rca_0043); internal target 13 Sep.
- Open: inc_0044 (KYC, 212 cases, detected in 0.2 h, containment, Farid Sheikh) and inc_0045 (retry storm, 0.1 h, assessment, Dev Subramanian).
- Median time to detect: 0.4 h in Jul–Sep 2026, against 31 h in Oct–Dec 2025, before Rho. 7 near misses recorded this quarter.

NOTE
Southern files against its own internal targets. No statutory hours are shown for AI quality incidents.

REC 23 AI inventory 4.4.67-4.4.71

Gap

ASKS

A complete internal inventory of models, use cases, target groups, dependencies, risk tier and grievances, including whether a system was changed after complaints. Update at least half-yearly and keep it ready for supervisors (para 4.4.68).

EVIDENCE

- Half-yearly inventory report pk_rec23_h1 (1 Jan – 30 Jun 2026, 176 agents, signed 1 Jul); current register export, 187 agents as of 11 Sep 09:41, signed by Priya Menon.
- Fields against para 4.4.68: base model recorded for 157 of 187 agents (30 unknown); use case and target population per agent; vendor, sub-contractor and data dependencies; risk tier; linked complaints and whether they led to a change.
- Declared against observed: 187 declared, 196 observed; 13 undeclared uses found, 4 declared agents never seen. 7 discovery candidates open.

OPEN

Gap	Undeclared AI not yet in the register: service account svc-acct-8841, 4,102 calls to an inference endpoint (C-05, f_014).	Dev Subramanian · 39 d
Gap	Purpose record incomplete on 14 agents (C-03, f_013).	Priya Menon · 51 d
Partial	12 embedded SaaS features without a named owner (C-04, f_004).	Vikram Nair · 40 d
Open	4 registered agents never connected; oldest silent 102 days (f_003).	Vikram Nair · 102 d

REC 24 AI audit framework 4.4.72-4.4.75

Partial

ASKS

Risk-based audit of input data, model and outputs. Internal audit in proportion to risk; independent third-party audit for high-risk or complex cases (24(b)). Confirm the ability to stop, pause or unwind, and human override. Review the framework at least every two years.

EVIDENCE

- Reproducible packs (C-18): pk_2026q1_coll_r2 regenerated from stored inputs on 11 Sep 09:24; signature verified and figures matched in 41 s.
- Restatement kept: the January promise-kept rate read 63% in the pack signed 7 Apr and 61% on 11 Sep, fully matured. Both packs stay locked.
- Three layers: inputs through lineage; the model through probes and version fingerprints; outputs through rubric scoring by judge j7.3 (κ 0.68 against human raters, threshold 0.65) and cohort fairness (C-12).
- Third line: Internal Audit (Imran Khalid) reads all evidence read-only; the external auditor read a pack on 9 Sep. Where Vendor X withheld evidence, compensatory testing stands in (C-19, 24(b)).
- Stop and pause: on 13 Aug the agent moved to recommend-only 14 minutes after detection; settlement was then hard-blocked at the bank's dialler gateway (18 Aug).

OPEN

Partial	2 independent validations of high-risk agents overdue, due 30 Sep (C-22, f_008).	Priya Menon · 22 d
Gap	One kill switch never drilled. Counted under Rec 21 (C-15).	Dev Subramanian · 102 d

REC 25 Disclosures by REs 4.4.76

Partial

ASKS

Include AI disclosures in the annual report and on the website: governance, where AI is used, consumer protection, and complaints handled.

EVIDENCE

- Claims register: 11 public claims about Southern's AI tested against measured results: 6 supported, 3 diverging, 1 outdated, 1 withdrawn. The withdrawn claim, "all of our AI models are built in-house" (62 of 187 agents are class A), was corrected by an annual-report erratum on 2 Jul.

OPEN

Diverging	3 claims diverge from measured results: "every AI decision affecting a customer is reviewed by a human" (measured 1.2%); "our AI resolves 90% of customer queries" (74%); "AI-assisted collections recover 30% more" (matched difference −10%).	No owner recorded · 29–41 d

SECTION 5

Open gaps and remediation plan

Every gap and partial in this report, with owner and age in days to 11 Sep 2026 (n/r: none recorded). All 14 findings are in Annex B.

REF	ITEM	FREE-AI	OWNER	AGE	NEXT STEP
CONTROL GAPS · 6					
C-15	Kill switch on the Mortgage Pre-qualification Assistant never drilled	Rec 21, 24	Dev Subramanian	102 d	Drill once the agent is connected (f_003, Vikram Nair)
C-03	Purpose record incomplete on 14 agents (f_013)	Rec 15, 16, 23	Priya Menon	51 d	Complete the purpose fields; owner signs
C-05	Undeclared AI: svc-acct-8841, 4,102 calls (f_014)	Rec 23	Dev Subramanian	39 d	Register or retire; inv_0397 merged into discovery
C-17	Vendor X model change not notified under MSA §8.4 (f_001)	Annex IV · outsourcing	Neha Kapoor	28 d	Obtain the overdue reply; carry to the 31 Jan 2027 renewal
C-16	Vendor X speech sub-contractor not disclosed (f_009)	Annex IV	Neha Kapoor	19 d	Obtain the attestation (due 6 Sep); add the clause at renewal
C-23	No disclosure record for CRM Next-Best-Action and Transaction Fraud Screening (R-62)	Rec 18	Anjali Rao	12 d	Record each customer touchpoint and its disclosure
CONTROL PARTIALS · 4					
C-04	12 embedded SaaS AI features without a named owner (f_004)	Rec 14, 23	Vikram Nair	40 d	Assign owners from the discovery evidence; owners attest quarterly
C-24	13 of 142 complaints cannot be linked to a decision	Rec 18	Anjali Rao	24 d	Link each complaint to its decision record
C-22	2 independent validations of high-risk agents overdue (f_008)	Rec 16, 24	Priya Menon	22 d	Complete by 30 Sep
C-10	Judge j7.3 below threshold on the hardship criterion since 4 Sep (κ 0.58)	Rec 26 (bank's mapping)	Priya Menon	7 d	Human scoring on all hardship cases until 200 revalidation cases are done (126 so far)
OTHER OPEN ITEMS AFFECTING THE FREE-AI POSITION					
f_003	4 registered agents never connected; oldest 102 days	Rec 23	Vikram Nair	102 d	Connect, or retire from the register
S33	4 reviewers of high-risk agents without current competency	Rec 10	Priya Menon	40 d	Retrain or reassign the four reviewers
f_007	KYC Document Verification 21% unscoreable against a 15% ceiling	Rec 16	Farid Sheikh	34 d	Obtain basis pointers from Vendor Y (inv_0402 accepted)
as_112	Transitive dependency vulnerability, Transaction Fraud Screening	Rec 19	Dev Subramanian	20 d	Fix scheduled
f_006	Collections Voice Agent access breadth 3.4 times baseline	Rec 19	Dev Subramanian	11 d	Review scope against the authority record
f_005	Tier-2 conversion red at 0.44; recommend-only proposed	Rec 16, 21	Anjali Rao	6 d	Owner decides the proposal
R-38	361 of 2,494 overrides made the outcome worse (Jan–Mar)	Rec 21	Anjali Rao	n/r	Act on value-losing overrides through the ladder
S49	3 public AI claims diverge from measured results	Rec 25	No owner recorded	29–41 d	Board to assign an owner; correct or substantiate each claim

SECTION 6

Annex

A Evidence index

8 evidence packs were signed or filed from 1 Jul to 11 Sep 2026 (7 signed, 1 filed). Each is signed by Priya Menon, Head of AI & Model Risk (second line), stored in ap-south-1 and object-locked for 7 years from the end of its period.

PACK	CONTENTS	PERIOD · AS OF	SIGNING AND REPRODUCTION
pk_board_q3	AI estate, Board Risk Committee	1 Jul – 11 Sep 2026	Signed 11 Sep 09:39, for the 18 Sep meeting · locked to 30 Sep 2033
pk_2026q1_coll_r2	Collections agent estate, supervisory evidence, restated (19 agents; 86,100 episodes)	21 Jan – 31 Mar 2026 · as of 11 Sep 08:00	Signed 11 Sep 09:00 · sha256 4a9c…e17b · reproduced 09:24 in 41 s, figures matched · locked to 31 Mar 2033 · held for inspection
pk_rec23_h1	Internal inventory report, half-yearly (176 agents)	1 Jan – 30 Jun 2026 · as of 1 Jul	Signed
pk_board_q2	AI estate, Board Risk Committee	1 Apr – 30 Jun 2026 · as of 14 Jul	Signed
pk_val_credit_q2	Retail Underwriting Assistant, validation evidence	1 Apr – 30 Jun 2026 · as of 6 Aug	Signed
pk_inc_0042	Internal incident report: settlement offered outside authority	Occurred 13 Aug · as of 20 Aug	Filed 20 Aug by the named officer
pk_rca_0043	Post-incident review: unnotified model change	As of 28 Aug	Signed 28 Aug
pk_vy_renewal	Vendor Y renewal brief	As of 2 Sep	Signed
RETAINED AND RELATED RECORDS · NOT COUNTED AS PACKS			
pk_2026q1_coll	Collections supervisory evidence, as issued (January promise kept 63%)	21 Jan – 31 Mar 2026 · as of 5 Apr	Signed 7 Apr · superseded by r2, retained and valid as issued · locked to 31 Mar 2033
inventory_2026-09-11.csv	Register export, 187 agents	As of 11 Sep 09:41	Signed by Priya Menon · an export, not a pack
ltr_vx_0815	Letter to Vendor X citing MSA §8.4	15 Aug · reply due 29 Aug	From Neha Kapoor · no reply received
brief_vx_renewal	Vendor X renewal brief, carries f_001, f_002, f_009	Renewal 31 Jan 2027	Draft · Neha Kapoor
pre-registration	Collections Voice Agent expectations, 3 signatories	Signed 14 Jan 09:00	sha256 e1c9…7b2f · before the first production episode

Other record IDs cited: incidents inc_0042 to inc_0045 and near miss nm_0071; investigations inv_0418, inv_0402, inv_0397; red-team findings rt_041 and as_112; complaints cmp_2105 and cmp_2203; data request dsr_0230; authority record auth_collections_v7. Evidence reads, 1 Jul – 11 Sep: 386 (owners 355, Rho agents 28, non-owners 3), each logged.

B Open findings · 14, all owned

ID	SEVERITY	FINDING	OWNER	AGE
`f_003`	High	Registered agent has never emitted (Mortgage Pre-qualification Assistant)	Vikram Nair	102 d
`f_013`	Low	Purpose record incomplete on 14 agents	Priya Menon	51 d
`f_002`	High	Vendor claim diverges: 94% claimed, 71% measured	Anjali Rao	41 d
`f_004`	High	Unowned agent in production (CRM Next-Best-Action)	Vikram Nair	40 d
`f_014`	Low	Shadow inference endpoint: 4,102 calls	Dev Subramanian	39 d
`f_007`	Medium	Unscoreable rate 21% above 15% ceiling (KYC Document Verification)	Farid Sheikh	34 d
`f_001`	High	Unnotified model version change (Vendor X, 4.1.8 to 4.2.1)	Neha Kapoor	28 d
`f_008`	Medium	Independent validation overdue (2 high-risk agents)	Priya Menon	22 d
`f_009`	Medium	Undisclosed sub-contractor (speech provider)	Neha Kapoor	19 d
`f_010`	Medium	Rubber-stamp rate 31% against 15% expected (Retail Underwriting Assistant)	Vikram Nair	15 d
`f_006`	Medium	Scope creep: access breadth 3.4 times baseline	Dev Subramanian	11 d
`f_012`	Low	Policy currency: 12 episodes on v11 after v12	Dev Subramanian	8 d
`f_005`	Medium	Pre-registered threshold breached: red	Anjali Rao	6 d
`f_011`	Low	Pilot expiring in 19 days (Hardship Triage)	Anjali Rao	3 d

C Glossary

Agent	Any AI system that makes or shapes a decision, whether built by Southern or supplied by a vendor.
Class A, B, C1, C2	A: built by Southern. B: vendor technology run by Southern. C1: vendor-operated AI service. C2: AI embedded in a SaaS product.
Episode · decision record	One unit of agent work, such as a call, with the inputs, tools, authority and action recorded.
Pre-registration	Expectations, floors and early-warning levels written and signed before an agent's first production episode.
Authority record	What an agent may do alone, what needs approval and what is prohibited, in force at a given date.
Recommend-only	A mode in which the agent proposes and a person decides. A threshold breach can propose it; a person approves.
Evidence pack	A signed, object-locked set of records and figures for a period, which can be regenerated from stored inputs.
Rho Bridge · Rho Cloud	The Bridge runs inside the bank and records agent activity. Rho Cloud holds evidence, in ap-south-1 for Southern.
Judge j7.3 · κ	The model that scores episodes against a rubric. κ (kappa) measures its agreement with human raters.
Promise to pay · DPD	A customer's recorded commitment to pay; "kept" if paid within 45 days. DPD: days past due.
RE · Sutra	RE: regulated entity, such as a bank. Sutra: one of FREE-AI's seven guiding principles.

FREE-AI: Report of the Committee to develop a Framework for Responsible and Ethical Enablement of Artificial Intelligence in the Financial Sector, Reserve Bank of India, 13 August 2025. Recommendations are paraphrased; see the report for full text. This sample report is illustrative. Southern Bank and all figures are fictional.