SAMPLE OUTPUT. A synthetic engagement against fictional evidence, generated to show what the workbench produces. “Meridian Financial” is not a real company and this is not a record of any real audit.

TRUSTRA CYBERSECURITY AUDIT WORKBENCH

Cybersecurity Vulnerability Audit Report

Meridian Financial | AI estate

Period 2026-01-01 to 2026-06-30

MATERIAL EXPOSURE IDENTIFIED

Report TRUSTRA-95FD-DEA7-F205-66FD-C060 Warehouse TCA-2026.07 Generated 27 Jul 2026

22	8	12	6	2
CONTROLS TESTED	PASSED	CONFIRMED FINDINGS	CRITICAL	ATTACK PATHS

Opinion

Material exposure identified. One or more material exposures identified.

A Critical-severity failure was confirmed (AIS-02, AIS-04, LLM-01, AGT-01, AGT-02, RAG-01). The opinion is forced out of the top band and requires explicit auditor commentary.

Auditor commentary. Critical failures concentrate in agent tool scoping and retrieval tenancy. Both are configuration changes rather than architectural ones, and are addressable within the remediation window.

Weighted risk 66.8% across 22 applicable controls, by severity (Low 1, Medium 3, High 7, Critical 20). The engine proposed each result; T. Vengal (lead) decided it. No automated result is final without sign-off.

The evidence chain

This history cannot be quietly rewritten. Every artifact, result, auditor decision, and sign-off is hash-linked at the moment it is recorded, across 56 entries, so a third party can re-verify the whole chain at any time.

`entry_hash = sha256(prev_hash + canonical_json(entry)). Edit anything and every later link breaks.`

`Chain head: a987821c51ae4de3e9b61ba8ee27cf2342b35416ff15527b11acbce75495f02b`

Confirmed findings

Ordered by severity. Each traces to a weakness in the knowledge warehouse, which supplies the remediation and the techniques that exploit it.

SEVERITY	CONTROL	FINDING	WEAKNESS	TIER
CRITICAL	AGT-01	**Every agent tool holds the minimum scope its function requires** 1 of 5 in-scope item(s) fail the control: crm_write	Unbounded tool credential scope	2
CRITICAL	AGT-02	**Irreversible actions require a human approval gate** 1 of 2 in-scope item(s) fail the control: crm_write	Missing approval gate on irreversible action	2
CRITICAL	AIS-02	**Model artifacts are not loaded from code-executing formats** 1 of 2 in-scope item(s) fail the control: code-assist-weights	Unsafe artifact deserialization	2
CRITICAL	AIS-04	**Serving dependencies carry no known exploited vulnerability** 2 component(s) carry vulnerabilities with confirmed exploitation in the wild: pkg:pypi/example-serve@2.1.0 (CVE-EXAMPLE-0001), pkg:npm/example-gateway@3.0.1 (CVE-EXAMPLE-0003)	AI supply chain compromise	1
CRITICAL	LLM-01	**Untrusted input is segregated from instruction context** 1 of 3 in-scope item(s) fail the control: code-assistant	Prompt injection	2
CRITICAL	RAG-01	**Retrieval enforces tenant scoping at query time** 1 of 1 in-scope item(s) fail the control: policy-index	Cross-tenant retrieval leakage	2
HIGH	AIG-01	**The AI asset inventory reconciles against observed traffic** 1 of 9 in-scope item(s) fail the control: code-assistant	Incomplete AI asset inventory	2
HIGH	AIS-01	**Every model artifact has a verified integrity digest** 1 of 4 in-scope item(s) fail the control: code-assist-weights	Unverified model artifact integrity	2
HIGH	LLM-03	**System prompts carry no secrets** 1 of 3 in-scope item(s) fail the control: copilot-system-prompt	System prompt leakage	2
HIGH	RAG-02	**Ingestion sources are authorised and integrity-checked** 1 of 3 in-scope item(s) fail the control: public-web-crawl	Unauthorised ingestion source	2
MEDIUM	AIG-02	**Every AI system has a named accountable owner** 1 of 9 in-scope item(s) fail the control: code-assistant	Incomplete AI asset inventory	2
MEDIUM	MOP-02	**Rollback has been exercised, not merely documented** 1 of 3 in-scope item(s) fail the control: code-assistant	No exercised rollback path	2

12 confirmed findings. Remediation and framework mapping for each are in the findings register and working papers, which accompany this report.

Ranked component exposure

Ranked by likelihood of exploitation against this organisation, not by intrinsic severity. Every score decomposes into its contributions, because a ranking that cannot be explained is not actionable.

COMPONENT	ISSUE	SCORE	WHY IT RANKS HERE
pkg:npm/example-gateway@3.0.1	Authentication bypass in inference gateway	**0.833**	listed as exploited in the wild; EPSS 63%; sits on an exposed asset

COMPONENT	ISSUE	SCORE	WHY IT RANKS HERE
`pkg:pypi/example-serve@2.1.0`	Deserialization of untrusted model artifact in serving library	**0.825**	listed as exploited in the wild; EPSS 42%; sits on an exposed asset
`pkg:pypi/example-rag@1.7.2`	Server-side request forgery in retrieval connector	**0.273**	EPSS 11%; sits on an exposed asset
`pkg:pypi/example-vector@0.9.1`	Path traversal in vector store admin API	**0.055**	component not on the request path; sits on an exposed asset

Every entry above is a real vulnerability. The lowest-ranked one scores as it does because nothing on the request path reaches it - a conventional scanner would report it beside the others on CVSS alone.

Remediation plan

Ordered by **leverage**, not by severity: the finding with the worst label is not always the one to fix first. Every ordering decision shows its basis, so the sequence can be defended in a room.

4 of 12 findings trace to a single subject (code-assistant). Addressing it is one decision, not 4 pieces of work.

#	SUBJECT	URGENCY	EFFORT	CLOSES	WHY HERE
1	`crm_write`	**Immediate**	hours	2	breaks 2 path(s)
2	`pkg:npm/example-gateway@3.0.1`	**Immediate**	days	1	exploited in the wild
3	`pkg:pypi/example-serve@2.1.0`	**Immediate**	days	1	exploited in the wild
4	`code-assistant`	**This week**	days	4	Schedule into the current sprint.
5	`code-assist-weights`	**This week**	days	2	Schedule into the current sprint.
6	`copilot-system-prompt`	**This week**	hours	1	Schedule into the current sprint.
7	`policy-index`	**This week**	days	1	Confirmed Critical. Effort is non-trivial, but it cannot wait a cycle.
8	`public-web-crawl`	**Backlog**	days	1	Track and address on the normal cadence.

First actions

crm_write — Start today. High exposure closed for little effort.

- Enumerate the credential's effective permissions and compare them with the tool's declared function.
- Issue a narrowly scoped credential per tool; retire shared or admin ones.
- Rotate the over-scoped credential once it has been replaced.

pkg:npm/example-gateway@3.0.1 — Exploited in the wild. Patch or isolate before the next change window.

- Inventory every externally sourced model, dataset, and adapter with its supplier and licence.
- Pin each to an immutable revision and verify a published digest on load.
- Gate new suppliers through the same review as any other dependency.

pkg:pypi/example-serve@2.1.0 — Exploited in the wild. Patch or isolate before the next change window.

- Inventory every externally sourced model, dataset, and adapter with its supplier and licence.
- Pin each to an immutable revision and verify a published digest on load.
- Gate new suppliers through the same review as any other dependency.

Predicted exposure annex

Nothing in this annex is a confirmed finding. These are predictions: directed tests to run, not verdicts. A prediction becomes a finding only when an auditor collects independent evidence confirming it. Predictions carry no severity.

Attack paths

Composite exposure no single-asset scan can see. Every hop below individually passed its own controls.

PATH	HOPS	LIKELIHOOD	CHEAPEST BREAK
`support-chatbot → ops-agent → crm_write → crm-store`	3	0.4410	Reduce the credential scope held by ops-agent, or move the action behind a human approval gate.
`inference-gw → support-chatbot → ops-agent → crm_write → crm-store`	4	0.3087	Reduce the credential scope held by ops-agent, or move the action behind a human approval gate.

Latent weakness prediction

Switched off for this engagement. It learns from outcomes confirmed by auditors across prior audits, and stays disabled until a pattern is supported by enough distinct clients to be defensible. Below that threshold it returns nothing rather than guessing.

Controls forecast to lapse

CONTROL	EXPIRES	DAYS	BASIS
`MOP-02`	2026-06-01	-56	last evidenced 2026-02-01, cadence 120 days
`AGT-02`	2026-07-14	-13	last evidenced 2026-01-15, cadence 180 days
`AIS-01`	2026-07-30	+3	last evidenced 2026-05-01, cadence 90 days
`LLM-05`	2026-09-30	+65	explicit expiry

Scope and methodology

12 client exports were ingested read-only, each SHA-256 hashed, timestamped, attributed, and locked into the evidence ledger on arrival. The workbench holds no client credentials, opens no connection to client systems, and cannot modify anything it audits. Controls were tested against warehouse snapshot TCA-2026.07, pinned to this engagement so the result reproduces exactly.

Results are proposed across three match tiers: deterministic identity matching, declarative rules over parsed evidence, and assisted review of unstructured evidence. The code computing results, severity, and the opinion contains no model call and no network call. Every finding was confirmed or overridden by a named auditor, with rationale recorded.

This report is a product-generated audit instrument, not a security certification and not a legal opinion. Framework references are drafting aids verified against the source publications. Re-verification of the chain can be requested at any time against the report reference.

Trustra Cybersecurity Audit Workbench. Prove your AI to anyone.