



Truveil

AI ACCOUNTABILITY AUDIT REPORT

Customer Retention Assistant

Audit of Customer Retention Assistant

REPORT ID	SR-135
GENERATED	Sun, 06 Sep 2026 21:51:39 GMT
SUBJECT AGENT	Customer Retention Assistant
AGENT CATEGORY	1 (Declared at Registration)
RISK CLASSIFICATION	LOW
FRAMEWORKS APPLIED	NIST AI RMF, NIST AI 600-1

Report Contents

I	Auditor Opinion
II	Scope and Methodology
III	Agent System Description
IV	Findings
V	Recommended Actions
VI	Technical Detail

Auditor Opinion

A

Accountability Grade · **85.1/100**

LOW RISK

This Category 1 agent scored 85.1/100 (Grade A, LOW risk) across a churn triage and retention discount workflow. Transparency and Reversibility are perfect. Accountability is reduced by absent model owner identification and no documented risk assessment, both relevant under NIST AI RMF GOVERN-2.1 and NIST AI RMF MAP-1.5. Data Trust is reduced by missing freshness and groundedness checks. No customer-facing action was taken; all outputs remained text recommendations pending human approval.

TRANSPARENCY

100

ACCOUNTABILITY

78.9

DATA TRUST

70

REVERSIBILITY

100

ISSUED BY

This report has been generated by Truveil, an AI agent accountability platform. The accountability scoring is rule-based and deterministic, calibrated against primary regulatory text. The narrative interpretation in this report is produced by Anthropic Claude under structured constraints maintained by Truveil. No portion of the scoring is delegated to a generative model.

Scope and Methodology

What this report covers

This report covers a single execution of the agent identified on the cover page. The audit assesses the logged actions, decisions, validations, and human checkpoints recorded during that execution against four accountability dimensions: *Transparency*, *Accountability*, *Data Trust*, and *Reversibility*. Each dimension corresponds to a category of regulatory obligation recurring across the frameworks Truveil applies.

How Truveil scores

Truveil's scoring engine is rule-based and deterministic. The same evidence, scored against the same framework, produces the same result every time. Each accountability signal is anchored to specific provisions in primary regulatory text. Caps and floors apply where a binding obligation is unmet, these are not weighted heuristics but explicit thresholds defined in the underlying regulation.

Narrative interpretation in this report is produced by a language model under structured constraints. The model summarises findings already determined by the engine. It does not score, classify, or rank the agent. Citations in this report are restricted to provisions present in the engine's citation index for this run.

Regulatory anchors for this audit

NIST AI RMF, NIST AI 600-1

What this report does not certify

This report is an evidence-based assessment of one execution of one agent. It does not constitute a regulatory filing, a legal opinion, or a certification of conformity. It is not a substitute for accredited third-party assurance such as ISO/IEC 42001 certification. Truveil's role is to surface accountability gaps that would otherwise remain undocumented, against the regulatory provisions identified in the engine's citation index.

Agent System Description

Agent and run metadata

AGENT	Customer Retention Assistant
CATEGORY	1 (Declared at Registration)
CATEGORY BASIS	Category 1 declared at agent registration.
RUN ID	run_20260902_e966b2
EXECUTION	2026-09-02T14:41:17.378Z
TIMESTAMP	
TOTAL LOGGED	10
EVENTS	

Step-by-step execution

Reconstructed from 10 logged events via the SDK; SDK instrumentation observes events the agent code explicitly logs.

01 AI Disclosure & Registration

Agent disclosed AI-generated nature of outputs to user before any analysis began.

OK

02 Source Data Ingestion

Ten-account synthetic churn dataset received via user-pasted table; no external systems queried.

WATCH

03 Churn Risk Scoring & Discount Recommendation

Three accounts ranked highest-risk and proposed for 20% retention discounts; seven accounts declined with reasons.

OK

04 Approval Request Submission

Human sign-off requested for three pricing concessions totalling \$33,840 annualised; no action taken pending approval.

OK

05 Operator Override & Final Approval

Operator substituted Winnow Retail AI for Tandem Grid; final discount set approved and documented.

OK

06 Run Close-Out

Offer emails drafted as plain text only; no billing system, CRM, or email client accessed.

OK

Findings

Findings index

#	SEVERITY	FINDING	FRAMEWORK
1	HIGH	No Formal Risk Assessment Documented	NIST AI RMF MAP-1.5, voluntary standard · NIST AI RMF MEASURE-1.1, voluntary standard +1
2	MEDIUM	Model Owner Not Named in Any Log	NIST AI RMF GOVERN-2.1, voluntary standard · NIST AI 600-1 GV-2.1-001, voluntary standard
3	MEDIUM	No Data Freshness Check Logged	NIST AI RMF MEASURE-2.5, voluntary standard · NIST AI 600-1 MP-4.1-002, voluntary standard
4	MEDIUM	Groundedness Not Assessed	NIST AI RMF MEASURE-2.3, voluntary standard · NIST AI 600-1 MS-2.3-002, voluntary standard

Findings in full

Finding 1No Formal Risk Assessment Documented HIGH

No risk assessment was logged for this Category 1 agent, leaving a required governance control absent per NIST AI RMF MAP-1.5 and NIST AI RMF MEASURE-1.1.

FRAMEWORK	NIST AI RMF MAP-1.5, voluntary standard · NIST AI RMF MEASURE-1.1, voluntary standard · NIST AI 600-1 MP-1.1-001, voluntary standard
EVIDENCE	absent; required for Category 1 agents
AFFECTED STEP	REGISTRATION

Finding 2Model Owner Not Named in Any Log MEDIUM

No owner or deployer identity appears in any log detail, leaving accountability lineage incomplete under NIST AI RMF GOVERN-2.1.

FRAMEWORK	NIST AI RMF GOVERN-2.1, voluntary standard · NIST AI 600-1 GV-2.1-001, voluntary standard
EVIDENCE	no owner or deployer mention in any log detail
AFFECTED STEP	REGISTRATION

Finding 3No Data Freshness Check Logged **MEDIUM**

No freshness, data_age, or date metadata was recorded for the source dataset, leaving currency of inputs unverified per NIST AI RMF MEASURE-2.5.

FRAMEWORK	NIST AI RMF MEASURE-2.5, voluntary standard · NIST AI 600-1 MP-4.1-002, voluntary standard
EVIDENCE	no freshness, data_age, or ISO date in source_fetch detail
AFFECTED STEP	Step 02: Source Data Ingestion

Finding 4Groundedness Not Assessed **MEDIUM**

No groundedness or retrieval score was logged, leaving output fidelity to source data unverified per NIST AI 600-1 MS-2.3-002.

FRAMEWORK	NIST AI RMF MEASURE-2.3, voluntary standard · NIST AI 600-1 MS-2.3-002, voluntary standard
EVIDENCE	no groundedness or retrieval_score mention in log detail
AFFECTED STEP	Step 03: Churn Risk Scoring & Discount Recommendation

Recommended Actions

Actions are ordered by priority. *Now* actions remediate findings that breach binding obligations under the frameworks applied. *Soon* actions address material gaps that fall short of best practice without breaching a binding rule. *Later* actions improve accountability posture for forward-looking compliance.

01 Document and Log a Formal Risk Assessment

Produce and attach a structured risk assessment to agent registration and log its identifier at run start, satisfying NIST AI RMF MAP-1.5 and NIST AI RMF MEASURE-1.1.

NOW

02 Add Model Owner Identification to Registration

Populate a `model_owner` field at agent registration and include deployer identity in run-open log events per NIST AI RMF GOVERN-2.1.

NOW

03 Capture Freshness Metadata on Source Ingestion

Log `data_age` or `freshness_checked` timestamp whenever a dataset is ingested, per NIST AI RMF MEASURE-2.5 and NIST AI 600-1 MP-4.1-002.

SOON

04 Emit Groundedness Scores on Recommendation Logs

Add a `groundedness` or `retrieval_score` field to each recommendation event, supporting NIST AI RMF MEASURE-2.3 and NIST AI 600-1 MS-2.3-002.

SOON

Technical Detail

This section is intended for engineering and platform teams responsible for the agent. Each finding from Section IV is restated with its developer-view structured fields and suggested remediation steps. Affected steps refer to the trace numbering in Section III.

Finding 1 · No Formal Risk Assessment Documented

FAILURE TYPE	absent_risk_assessment_record
AFFECTED STEP	REGISTRATION
CONFIDENCE	HIGH

SUGGESTED FIXES

1. Add a structured risk_assessment field to agent registration for Category 1 agents.
2. Log a risk_assessment event at run start referencing the assessment document identifier.

Finding 2 · Model Owner Not Named in Any Log

FAILURE TYPE	missing_owner_identification
AFFECTED STEP	REGISTRATION
CONFIDENCE	HIGH

SUGGESTED FIXES

1. Populate a model_owner structured field at agent registration.
2. Include deployer name in run-open log event metadata.

Finding 3 · No Data Freshness Check Logged

FAILURE TYPE	missing_freshness_metadata
AFFECTED STEP	Step 02: Source Data Ingestion
CONFIDENCE	HIGH

SUGGESTED FIXES

1. Capture a data_age or freshness_checked timestamp on source ingestion logs.
2. Reject or flag datasets lacking an explicit as-of date.

Finding 4 · Groundedness Not Assessed

FAILURE TYPE	absent_groundness_score
AFFECTED STEP	Step 03: Churn Risk Scoring & Discount Recommendation
CONFIDENCE	HIGH

SUGGESTED FIXES

- 1. Emit a groundedness or retrieval_score field on each recommendation log.
- 2. Flag recommendations below a defined groundedness threshold for additional review.

Address risk assessment and model owner gaps to secure Category 1 compliance readiness under NIST AI RMF.