TRIAGE: an AI agent for billing reconciliation

Transfer Reconciliation and Investigation Agent with Guarded Execution

Problem. After a billing system move, charges, claims and statements stop matching the clinical record.

Approach. Rules find every mismatch, an AI agent explains it, a person approves the fix, code applies it.

rudcorp.io

HOW IT WORKS

Built on Amazon Bedrock, SQS, Lambda, DynamoDB, Terraform and Python

1 Sources
Clinical record, old billing, new billing

→

2 Rules
Five SQL checks on every account. No AI.

→

3 AI agent
Amazon Bedrock. Finds the cause, cites policy.

→

4 Verifier
Code rejects invented IDs and citations.

→

5 Reviewer
A person sees a dry run, then approves.

→

6 Fix
Code applies it and re-runs the check.

Accuracy and cost on 40 held-out cases

Tiered routing matched the large Nova model at 36% of its cost.

Model setup	Cases right	Cost per 1,000 cases
Nova Lite	36 of 40	$0.51
Nova Pro	37 of 40	$7.21
Tiered (default)	**37 of 40**	**$2.62**
Claude Haiku 4.5	39 of 40	$23.85

Tiered sends simple checks to Nova Lite and policy questions to Nova Pro. Right means cause and action both match. One run each; at 40 cases the gaps are within sampling error. Claude lost one case to an AWS rate limit, none wrong.

Scale, measured

1,000,000

accounts checked by the rules in 3.75 seconds, on a laptop

Cases a minute on AWS

2 workers	20
4 workers	41
8 workers	**97**

174 new cases on AWS: 160 right, $0.60. Cost per case stays flat as workers are added.

Bounded by design

Per case: 10 turns, 16 tool calls, $0.05, 90 seconds.

Per run and per day: hard dollar caps, plus a kill switch.

No automatic retries: a failed case goes to a person.

Read-only agent: code makes every change. Patient identifiers never reach the model.

Audit: every step is logged in a tamper-evident chain.

Limits: all data is synthetic (fictional hospital, patients, payers and policies). Right answers are planted by code, not labelled by billing specialists. Not connected to live systems; not reviewed for use with patient data. Total model spend to date: under $5.