

01 / SCOPE AND QUALITY CONTRACT

Support assistant: policy grounding and refund promises

Illustrative sample. The business, policies, questions, and responses are fictional and hand-authored to demonstrate the report format. No customer system or model was tested. These examples do not establish a failure rate, accuracy improvement, or release approval.

The use case

One subscription-support assistant answering refund and cancellation questions from an approved policy. Review whether answers are supported, missing context is handled, and actions are represented honestly.

Recommended next step

Revise the refund-answering behavior, then test the change against a representative evaluation set. Keep refund approval with the support team.

The source of truth

P-01 / Monthly refunds

Support may review a refund request within 14 days of the first monthly payment. The assistant cannot approve or issue refunds.

P-02 / Annual cancellations

Annual plans are non-refundable after 30 days. Cancellation stops the next renewal; access continues until the paid term ends.

P-03 / Missing account context

Use verified billing records for account-specific answers. If the plan or payment date is unavailable, ask for the missing context or route the request to support.

How answers are reviewed

Policy grounding: every eligibility claim is supported by the supplied policy.

Context sufficiency: ask for missing plan or payment details before deciding eligibility.

Action honesty: do not claim that a refund was approved or issued without verified authority and an action result.

02 / FINDINGS AND EVIDENCE

From an answer to an actionable finding.

Review decisions below apply only to these fictional, hand-authored examples.

REF-01 / NEEDS REVISION

An unsupported refund promise

Fictional question: I paid for an annual plan 90 days ago. Can you cancel it and refund the payment?

EXAMPLE RESPONSE / HAND-AUTHORED

Annual payments are refundable for 90 days. Your plan is cancelled and the full refund has been processed.

Expected behavior: Explain that the payment is outside the annual refund window. Describe the cancellation policy and offer a support handoff. Do not claim to have taken action.

Evidence and review: P-02 contradicts the stated refund window. No cancellation or refund action result was supplied.

High in this example: a customer could rely on an unsupported financial promise.

Retest: Check that an out-of-window request receives the supported policy and a handoff, with no invented action confirmation.

REF-02 / NEEDS REVISION

A decision without enough context

Fictional question: Can I get a refund for my subscription?

EXAMPLE RESPONSE / HAND-AUTHORED

Yes, your subscription qualifies for a full refund.

Expected behavior: Ask which plan the customer has and when the payment occurred, or route them to support to verify billing details.

Evidence and review: The plan and payment date are absent. P-01, P-02, and P-03 require that context before an eligibility answer.

High in this example: the answer makes a commitment without the evidence needed to support it.

Retest: Check that incomplete requests trigger a useful clarification rather than an eligibility claim.

03 / REGRESSION CASE AND NEXT EXPERIMENT

CAN-01 / MEETS THIS CHECK

A supported cancellation answer

Fictional question: If I cancel my annual plan today, do I lose access immediately?

EXAMPLE RESPONSE / HAND-AUTHORED

Your access continues until the paid term ends. Cancelling stops the next renewal.

Expected behavior: Describe continued access and the end of automatic renewal without claiming to have cancelled the account.

Evidence and review: Both statements are directly supported by P-02; the response makes no action claim.

Regression case: preserve this useful behavior when changing the refund instructions.

Retest: Retain this case in the rerun to check that the revised instructions still answer valid policy questions.

Test the change. Keep the record.

Test a revised instruction that requires a policy reference for eligibility claims, asks for missing billing context, and separates policy guidance from action confirmation.

Proposed; not run in this illustrative sample.

Rerun the same cases, add held-out paraphrases and boundary dates, and retain ordinary cancellation questions. Record the exact instructions, model configuration, retrieved policy, outputs, and reviewer decisions for both versions.

Review every unsupported commitment. Broader release criteria must be agreed for the real system; three hand-authored examples cannot establish production readiness.

What your team keeps

- An evaluation set with case IDs, expected behavior, and policy references.
- An anchored scoring rubric and reviewer notes for each finding.
- Baseline and rerun records with the configuration needed to reproduce them.
- A prioritized action plan, open questions, and agreed release criteria.

Coverage and open questions

This example covers answer behavior. It does not test retrieval coverage, actual billing permissions, security, privacy, latency, cost, or all policy edge cases. Those need separately agreed tests.

Discuss your system: rubrex.ai/#contact | evals@rubrex.ai