

GroundTruth (Part 4)

Try It Yourself: Test Tickets and Expected Behavior

Prepared by Abhishek Katkar. This sheet lets anyone test the live app in a couple of minutes and see both behaviors that matter: grounded answers when the knowledge base supports them, and refusal plus escalation when it does not.

Approach and constraints. Built from the outside, as a first-time user, on a limited free-credit allowance, using only public information. With internal access, each result below becomes a measured, repeatable evaluation.

Live app: <https://support-sentinel-bold-edge-77s7.architect.space/>

How to test

Open the live app, paste any ticket below into the ticket box, and click Resolve Ticket. A resolve takes roughly two to three minutes because a Manager orchestrates several sub-agents in sequence. Watch the grounding badge (Grounded, Not Grounded, or No Source), the retrieved source articles, and the drafted response.

The knowledge base deliberately covers common subscription-SaaS support topics but omits SAML SSO and GDPR deletion, so those two tickets test the refusal path.

The six test tickets

#	Ticket text	Category	Expected behavior
1	I upgraded my plan yesterday but I was charged the full amount instead of a prorated one. Why?	Billing	Grounded. Answers from the proration article.
2	I can't log in. I clicked the password reset link in my email but it says the link has expired.	Account access	Grounded. Answers from the password-reset article.
3	I cancelled my subscription two days ago but I still have full access to the product. Is this a bug?	Account	Grounded in most runs. A borderline case near the retrieval threshold.
4	How do I remove a user from my account, and will I get a refund for their seat?	Account	Borderline. Grounds or escalates depending on run; both are safe outcomes.
5	How do I set up SAML single sign-on for my company?	Account	Escalates. No supporting article exists; the verification agent refuses to answer. This is the key test.
6	Under GDPR I want all my personal data permanently deleted. How do I request this?	Compliance	Escalates. No source found; refuses rather than guessing.

What to watch for

- **Ticket 5 is the one to run first.** Retrieval may return a SAML-looking article at a high score, but the verification agent recognizes the content does not actually support a SAML answer and escalates. This is the app refusing to hallucinate, which is the entire point of the design.
- **Tickets 1 and 2** show the grounded path working cleanly, with the source article and a send-ready reply.
- **Tickets 3 and 4** are deliberately borderline. If one escalates when a valid article exists, that is the fail-safe behavior discussed in Part 1: the system prefers to escalate to a human rather than risk an unsupported answer. This is a tunable business decision, not a defect.
- **Ticket 6** shows a clean no-source refusal.

The one-line summary

The app answers when it can ground the answer, and refuses when it cannot. Whether it grounds or escalates on a borderline ticket, it never presents a claim the knowledge base does not support. That is the behavior a regulated enterprise support tool needs.