

Assistant metrics and SQL

Define metrics precisely and query them consistently.

Data and BI

Product

QA

DEFINITIONS

Metric	Definition
Containment	Share not escalated to a person
Resolution	Goal met, by confirmation or no repeat contact
Repeat contact	Same user and topic within 7 days
Fallback rate	Share of turns hitting not-understood or not-found
Groundedness	Share of sampled answers supported by sources
Latency	p50 and p95, not averages
Cost per resolved	All costs, including humans, divided by resolved conversations

LOG AT MINIMUM

- conversation_id, turn_id, hashed user id, timestamp
- channel, language, market
- route, confidence, topic
- model, prompt and knowledge-base versions
- retrieval ids and scores; tool calls and status
- latency parts, tokens, costs
- outcome, escalation reason, feedback, guardrail triggers

COST PER RESOLVED (SQL)

```
SELECT route,
       SUM(model_cost + tool_cost +
           retrieval_cost + human_minutes * 0.75)
       / NULLIF(SUM(CASE WHEN resolved THEN 1
                       ELSE 0 END), 0) AS cost_per_resolved
FROM conversations
GROUP BY route;
```

DASHBOARD RULES

- Segment by language, channel, topic and complexity
- Show distributions and sample sizes
- Mark release dates
- Every chart links to an owner and an action