

### ⚙ Configuration

▼ Search Service

Cortex Search Service: [?](#)

RAG\_DB.RAG\_SCHEMA

▼ Location

Database: [?](#)

RAG\_DB

Schema: [?](#)

RAG\_SCHEMA

Results to retrieve: [3](#)

▼ Stage Status

⟳ Recreating stage with server-side encryption...

✓ TruLens stage ready

⟳ Recreate Stage

# 🕒 LLM Evaluation & AI Observability

Evaluate your RAG application quality using TruLens and Snowflake AI Observability.

## ▼ Why Evaluate LLMs?

After building a RAG application (Days 21-22), you need to measure its quality:

### The RAG Triad Metrics:

1. **Context Relevance** - Did we retrieve the right documents?
2. **Groundedness** - Is the answer based on the context (no hallucinations)?
3. **Answer Relevance** - Does the answer address the question?

**TruLens:** An open-source library that integrates with Snowflake AI Observability to automatically evaluate LLM applications, track experiments, and store results for comparison.

Learn more: [Snowflake AI Observability](#)

⌚ TruLens packages are installed and ready!

## ⚙ Evaluation Configuration

App Name: [?](#)

customer\_review\_rag

App Version: [?](#)

v1

RAG Model:

claude-3-5-sonnet



## ☒ Test Questions

Questions (one per line):



What do customers say about thermal gloves?  
Are there any durability complaints?  
Which products get the best reviews?

## ☒ Run TruLens Evaluation

✓ Evaluation complete

- ✓ Importing required libraries...
- ✓ Preparing test dataset...
- ✓ Created dataset table:  
`RAG_DB.RAG_SCHEMA.CUSTOMER REVIEW_TEST_QUESTIONS`

- ✓ Setting up RAG application...
- ✓ Registering app with TruLens...
- ✓ Running evaluation on 3 questions...
- ✓ Question 1/3: What do customers say about thermal gloves?
- ✓ Question 2/3: Are there any durability complaints?
- ✓ Question 3/3: Which products get the best reviews?
- ✓ Waiting for all invocations to complete...
- ✓ Computing RAG Triad metrics...
- ✓ Metrics computed successfully!

[✓ Evaluation complete!]

## Evaluation Results

✓ Evaluation Run Complete!

### Run Details:

- App Name: **customer\_review\_rag**
- App Version: **v1\_1**
- Run Name: **v1\_1\_1769275674**
- Questions Evaluated: 3
- Model: **claude-3-5-sonnet**

**View Results in Snowsight:** Navigate to: AI & ML → Evaluations → **customer\_review\_rag**

### ▼ Generated Answers

**Question 1:** What do customers say about thermal gloves?

Based on these customer reviews, there are several consistent issues with the thermal gloves:

Major concerns include:

1. Durability problems - including unraveling stitching around the thumb after brief use
2. Sizing inconsistencies - gloves run small and sizing varies even within the same brand
3. Poor performance in cold weather - inadequate warmth, especially in temperatures below teens Fahrenheit
4. Functionality issues - inconsistent touchscreen capability and compromised dexterity
5. Comfort problems - poor breathability leading to clammy hands and clumsy fit

While one reviewer noted some positive aspects (durable leather palms and decent warmth once the correct size was obtained), the overall

sentiment is negative. Multiple reviewers felt the gloves offered poor value for their price and suggested there are better alternatives available in the market.

The reviews span from October to November 2023, indicating these are recent experiences from multiple customers.

### Question 2: Are there any durability complaints?

Yes, there are significant durability complaints across all three products:

1. Alpine Skis: Edges are delaminating after just 10 days of use and showing excessive topsheet wear.
2. Thermal Gloves: Stitching around the thumb unraveled after only two weeks of regular use.
3. Ski Goggles: Lenses developed multiple scratches within six days of skiing, despite careful use and proper storage, with the scratches affecting visibility.

All three products seem to have durability issues that customers feel are unacceptable for their price points, with problems appearing very early in their use.

### Question 3: Which products get the best reviews?

Based on the sentiment scores and review content, here's how the products rank from best to worst reviewed:

1. Pro Ski Boots (0.467 sentiment score) -  
Received the most positive review, praised for versatility, excellent downhill performance, and good uphill capability, despite some limitations for long tours.
2. Ski Goggles (0.417 sentiment score) -  
Generally positive review, with high-quality lenses and secure fit, though the lens-changing process was noted as somewhat difficult.
3. Thermal Gloves (-0.009 sentiment score) -  
Received the least positive review, mainly due to sizing inconsistencies and related shipping/return issues, though the gloves themselves were described as "decent" once the correct size is obtained.

The Pro Ski Boots appear to get the best overall review, with the most comprehensive positive feedback despite some minor drawbacks.