

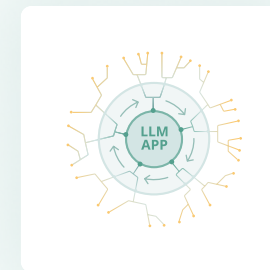
# TruEra for LLM Observability

Comprehensively test, debug, and monitor LLM apps, from development through production.

## Building LLM apps is easy. Making them reliable can be a lot harder.

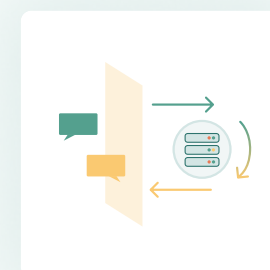
New applications and use cases based on Large Language Models (LLMs) are rapidly proliferating. This enthusiasm, however, is tempered by caution about the risks that LLM apps could potentially bring, including hallucinations, toxicity, and bias.

TruEra provides LLM app testing, debugging, and monitoring capabilities that can help to minimize risks, allowing you to create and maintain reliable and powerful LLM apps, faster.



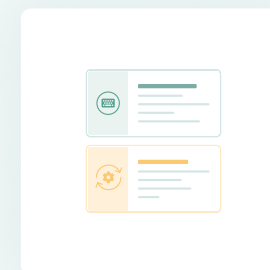
### Test

Easily integrate testing to evaluate LLM apps early in the development process



### Debug

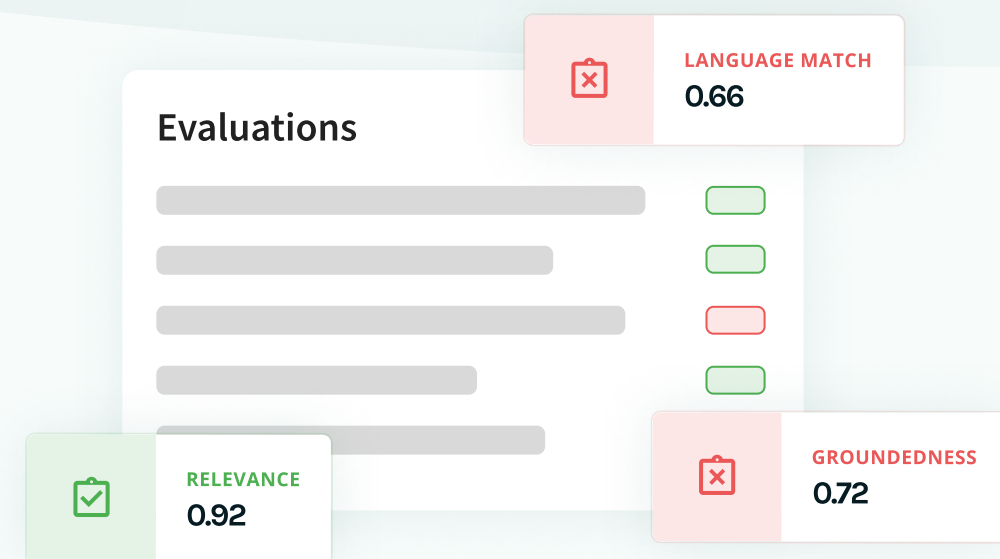
Efficiently debug retrieval augmented generation (RAG) and agent apps



### Monitor

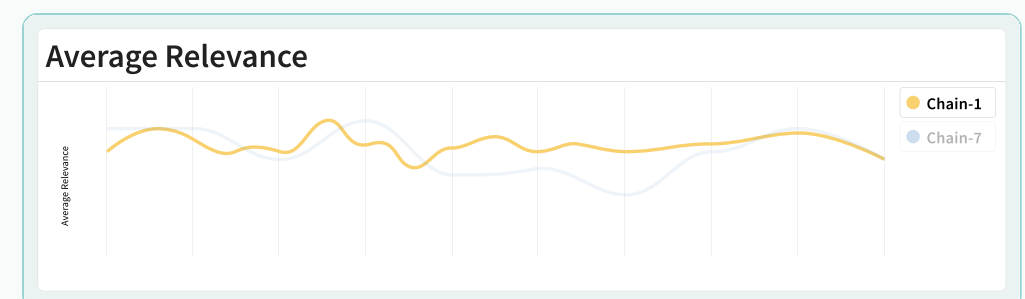
Track human and automated feedback at scale, in one place

# TruEra provides critical capabilities for driving high-performance LLM apps.



Record ID	Chain ID	User input	Response	Relevance	Prompt Sentiment	Groundedness
001	001	Find me flights from Berlin to London on...	Finding flights... Best options for flights from ...	0.76	0.99	0.75
002	001	Flights next week from Paris to Copenha...	Finding flights... Best options for flights from ...	0.72	0.88	0.76
003	001	Flights from Paris to Grenoble next week...	Finding flights... Best options for flights from ...	0.41	0.99	0.71
004	001	Cheapest flights to New York a week afte...	Finding flights... Best options for flights from ...	0.88	0.72	0.99
005	001	Cheap flights from Barcelona to Paris lat...	Finding flights... Best options for flights from ...	0.88	0.92	0.87
006	001	2 tickets for flights from New York a week...	Finding flights... Best options for flights from ...	0.82	0.92	0.52
007	001	Los Angeles to New York next week chea...	Finding flights... Best options for flights from ...	0.81	0.96	0.88
008	001	Find me cheap destinations from Berlin...	Finding flights... Best options for flights from ...	0.55	0.93	0.56
009	001	Flights from Vilnius to Copenhagen next...	Finding flights... Best options for flights from ...	0.92	0.98	0.99
010	001	Best flights to Norway late summer 2024	Finding flights... Best options for flights from ...	0.89	0.97	0.92

Version comparison					
App version	Average latency	Total Cost (USD)	Avg. relevance	Groundedness	Language_match
● GPT4	12.67	\$16.15	0.92	0.91	0.83
● Local-1	6.02	\$9.12	0.63 ⚠️	0.89	0.55 ⚠️



## Test

- Build your initial LLM app and connect TruEra with a few lines of code to starting logging records.
- Leverage an extensible library of “feedback functions” to programmatically test and evaluate your LLM app.

## Debug

- Debug app weaknesses to inform iteration on prompts, hyperparameters, and more.
- Understand your top performing app versions and make informed decisions before moving into production.

## Monitor

- Monitor and report on app performance changes over time.
- Perform comparisons and A/B testing to analyze multiple approaches to understand tradeoffs between cost, latency and performance.

## What's a Feedback Function?

A feedback function provides a score after reviewing an LLM app's inputs, outputs, intermediate results, and metadata. TruEra can evaluate your LLM app feedback functions to increase performance and minimize risk including the following:

- Context relevance
- Answer relevance
- Groundedness
- Embedding Distance
- Language mismatch
- Prompt sentiment
- PII Detection
- Stereotyping Language
- Other custom feedback functions you provide



# Getting Started is Easy!

Use TruEra for a wide range of LLM-based apps.

## Start with a few lines of code

You can start using TruEra to evaluate and track your LLM apps quickly and easily, with just a few lines of code.

```
1 record_id = tru.add_data(  
2     chain_id='Chain1_ChatApplication',  
3     prompt=prompt_input,  
4     response=llm_response,  
5     details=chain_details,  
6     tags='dev',  
7     total_tokens=total_tokens,  
8     total_cost=total_cost  
9 )
```

## Easily integrate with a wide range of LLM apps

Evaluate a wide range of LLM apps including retrieval augmented generation (RAGs) and agents. These can be built on LangChain, LlamaIndex, or other custom frameworks.



### App use cases

- ✓ Retrieval-augmented generation
- ✓ Query planning
- ✓ Data agents
- ✓ Streaming
- ✓ Summarization
- ✓ Marketing/sales copy
- ✓ Tuning experiments

### Apps built with leading frameworks, such as

- ✓ LangChain
- ✓ LlamaIndex
- ✓ Custom frameworks

# Resources

Start testing & debugging your LLM apps today with the help of these resources & our community.

## Quickstarts

TruEra Full Lifecycle LLM Observability is powered by the TruLens open source package so it's free and easy to get started testing and debugging your LLM apps. These simple quickstart examples illustrate TruLens integration across a number of popular LLM frameworks.



**LlamaIndex**

[Code Example](#)



**Pinecone**

[Code Example](#)



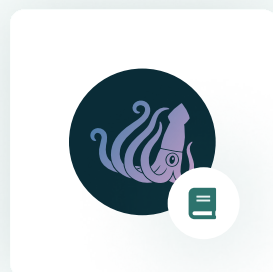
**LangChain**

[Code Example](#)

**No Framework**

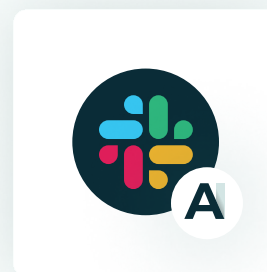
[Code Example](#)

## Additional Resources



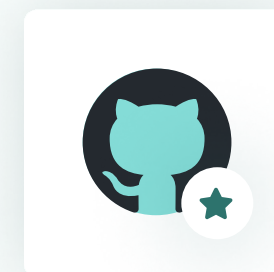
**Review Documentation**

Read about TruLens [here](#).



**Join the Community**

Join the [TruLens community](#) via the [AI Quality Forum](#).



**TruLens on GitHub**

Visit [TruLens on GitHub](#).