truera

LLM Evaluation & Observability

Avinash Joshi
Product Lead and Director, TruEra (India)

truera

Who are the founders of TruEra?

TruEra was founded in 2014 by Anupam Datta, Somesh Jha, and Tomer Geva. Some key details about the founders:

- Anupam Datta - Professor of Computer Science and Electrical and Computer Engineering at Carnegie Mellon University. He served as TruEra's President and Chief Scientist.
- Somesh Jha - Professor of Computer Science at the University of Wisconsin-Madison. He was the Vice President of Research at TruEra.
- Tomer Geva - Previously worked at Microsoft Research
  TruEra's CEO.

The three founder

puter

ns that could help

with regulations. The company was

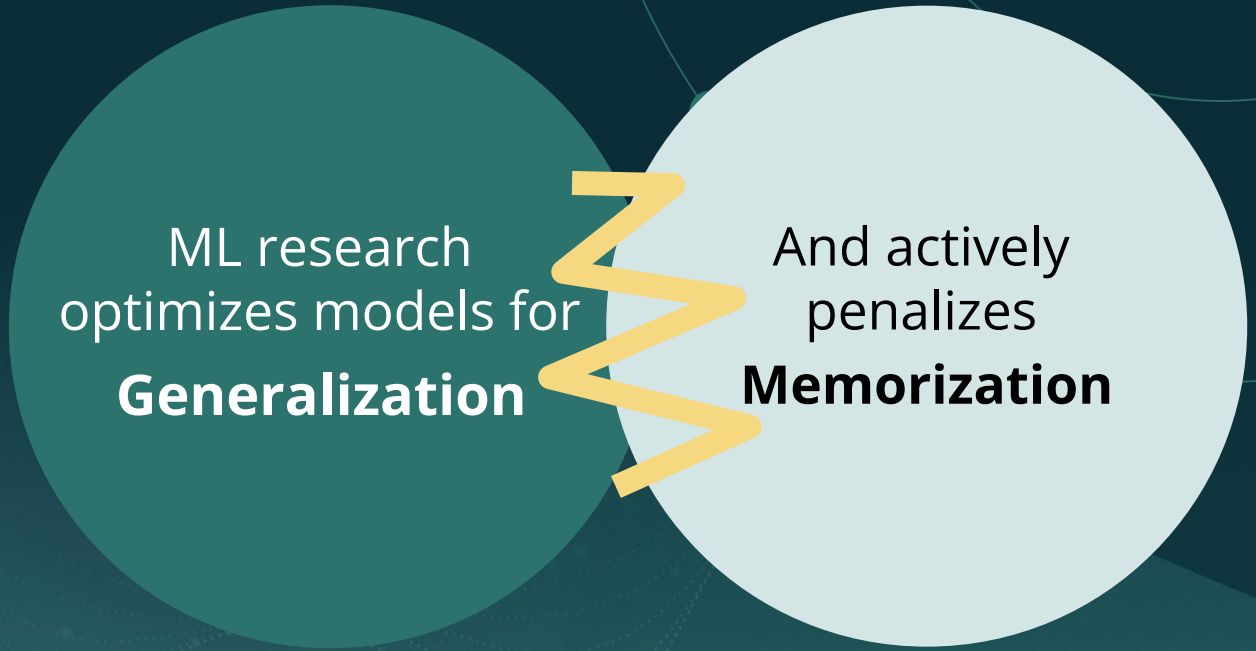
y, California.

, TruEra was acquired by SAS, the analytics software company, for an undisclosed sum. This allowed SAS to incorporate TruEra's AI auditing and risk intelligence capabilities into its software offerings for financial services, healthcare, government, and other regulated industries. While the TruEra brand was absorbed into SAS, the founders

Consider LLMs hallucinatory…
unless proven otherwise!

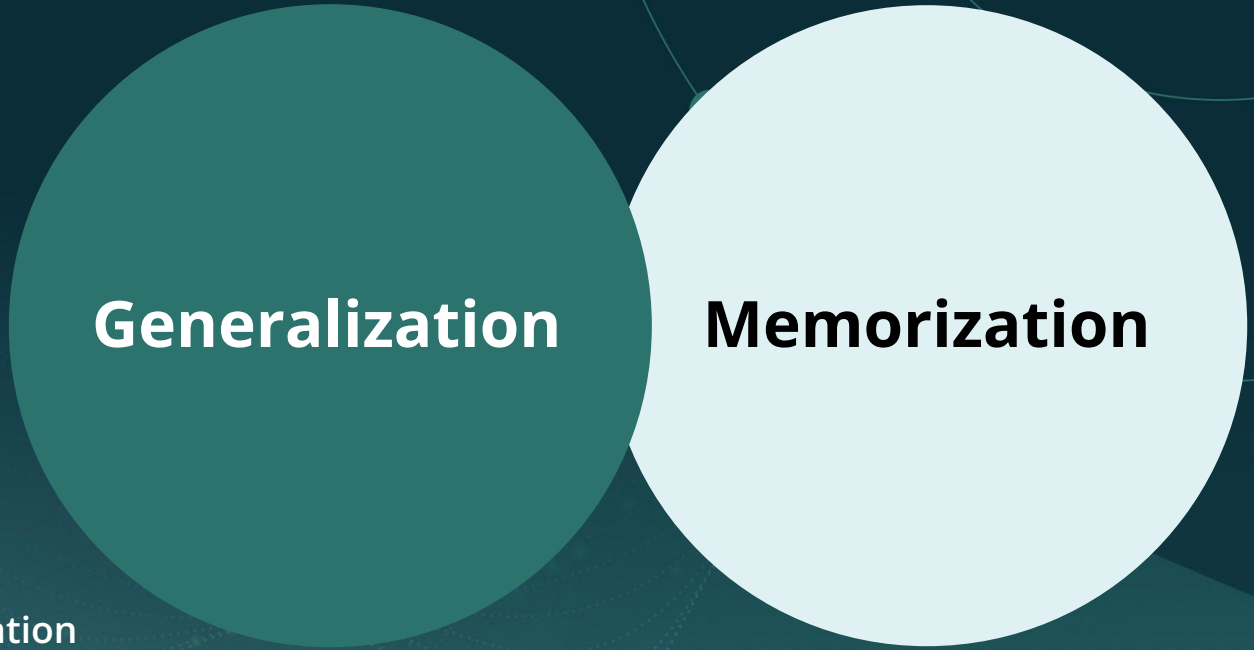
truera

This overlap is really murky



truera

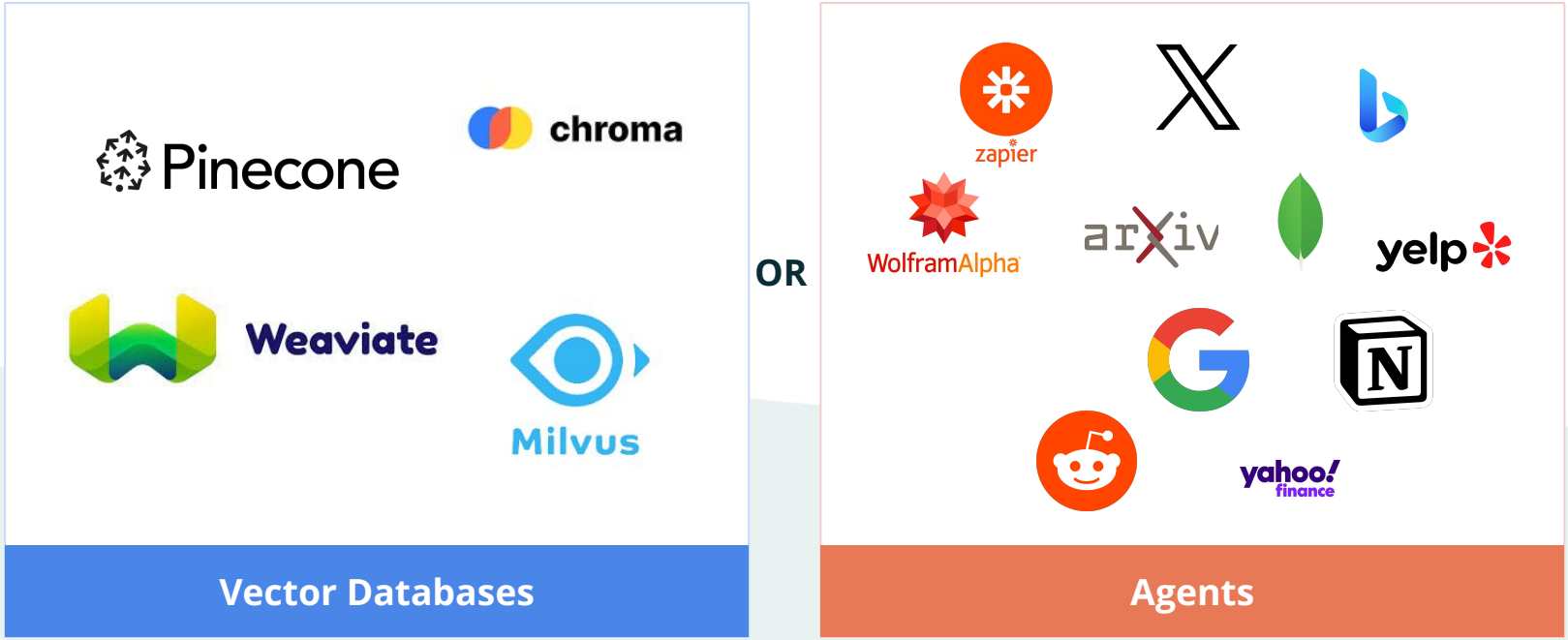
Focus LLMs on ‘General’ Tasks



- ✓ Summarization
- ✓ Text Embedding
- ✓ Inference
- ✓ Planning

Leave memorization to something else

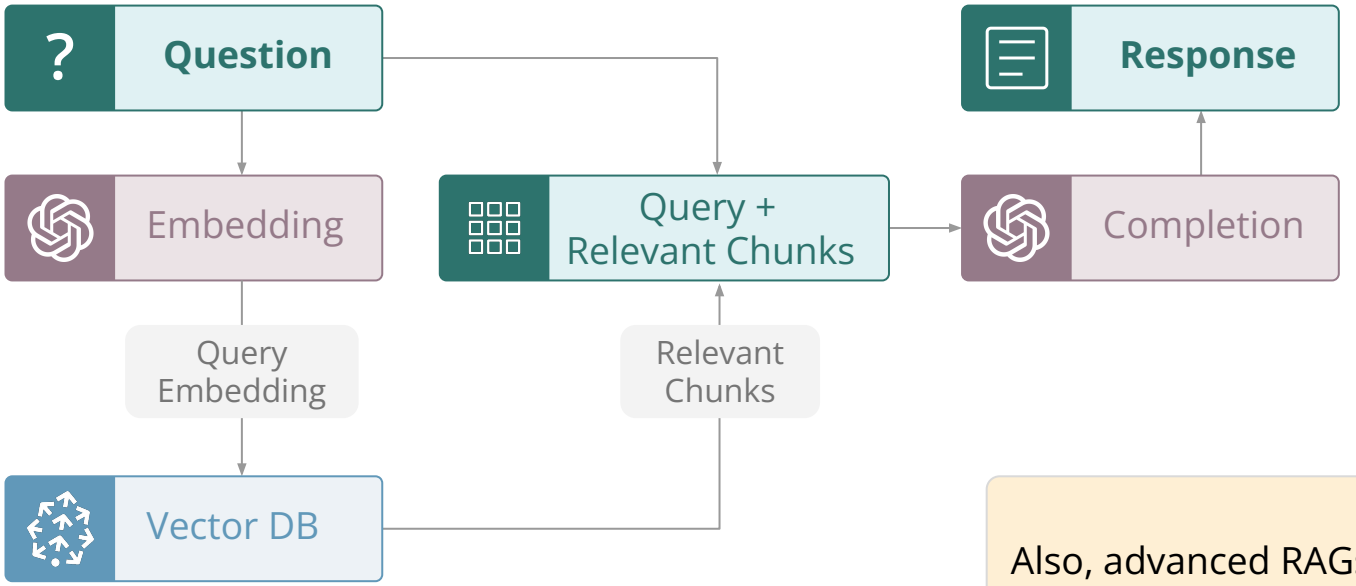
LLMs Need a Knowledge Source



truera

Enter Retrieval Augmented Generation (RAGs)

Example: Question Answering ChatBot



Also, advanced RAGs, Agents

But RAGs can Hallucinate too

truera

Input

Who is Shayak?

Response

Shayak is a computer scientist who obtained his PhD in Computer Science from Carnegie Mellon University and BTech in Computer Science from the Indian Institute of Technology, Delhi. He has been building systems and leading research to make machine learning and big data systems more explainable, privacy compliant, and fair for the past 10 years. He is also a member of the Bank of England's AI Public-Private Forum and the OECD Global Partnership on AI.

For second meetings start here

truera

Gen AI is Taking over the World

truera

The World is Divided in Two Camps

1

Teams that are building cool prototypes

- Very easy to build apps that do the right thing 80% of the time.

2

Teams that are deploying apps in production

- Much harder to go beyond that

truera

Three Examples from Camp #2

Education Tech

Student support for assignments

Customer Support

Summarization of call transcripts with action items

Financial Services

Customer facing chatbot

What Sets these Camps Apart?

truera

Systematic iteration to improve LLM apps in development and in production

GenAI Observability Platform Requirements

✓ Feedback loop for debugging

Feedback Loop

✓ Broad App Support

Evaluate

✓ Reliable, comprehensive, extensible evals

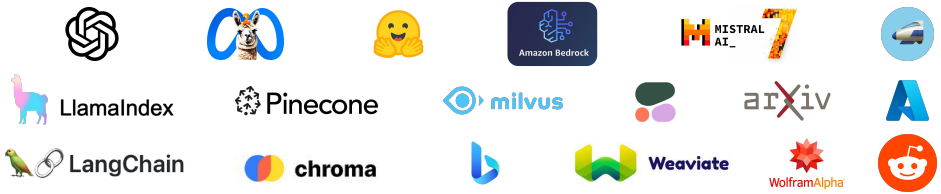
Track

✓ Experiment tracking to select best app config

Monitor

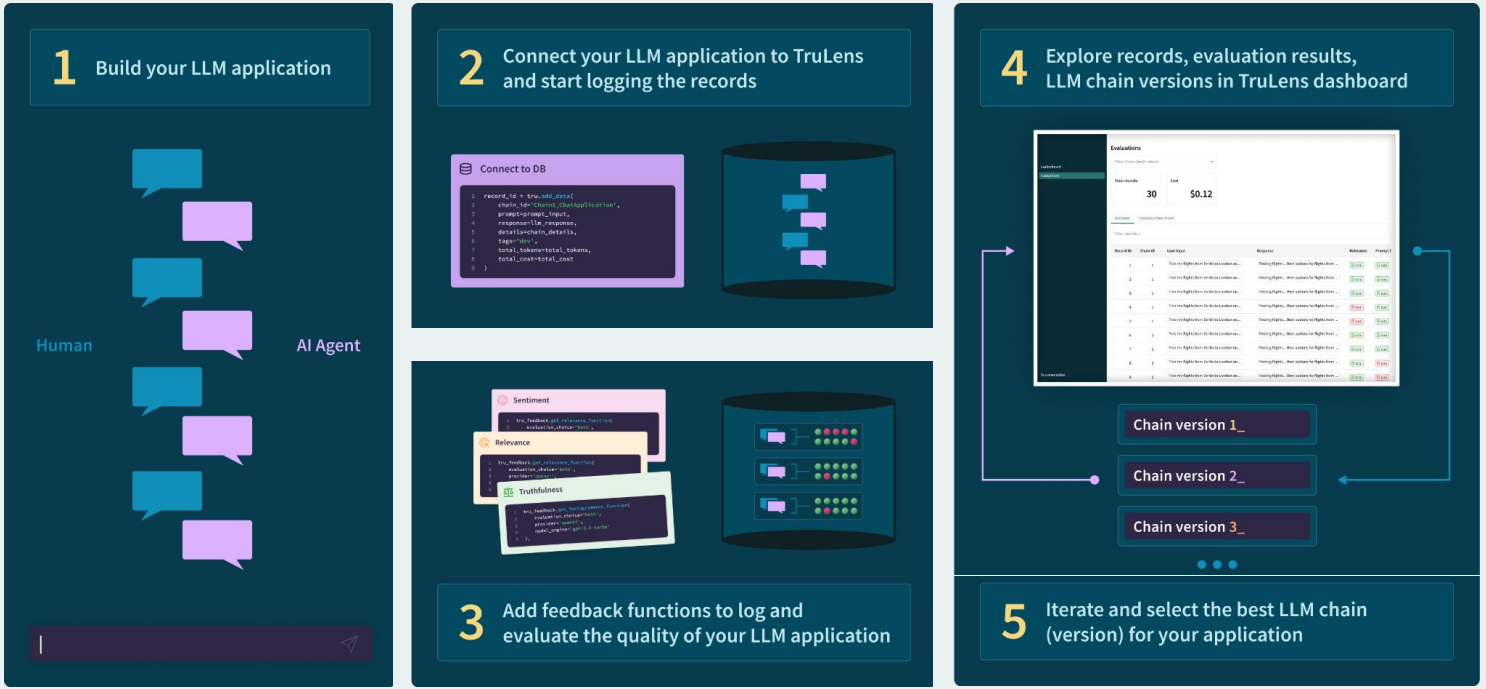
✓ Highly scalable
✓ Cost effective
✓ Low latency

✓ Easy integrations with emerging tech stack



TruLens : Track and Evaluate LLM Experiments

github.com/truera/trulens



TruLens has Broad Support for LLM Apps

Broad App Support

- RAGs
- Advanced RAGs (query planning, sentence window, auto merging)
- Agents
- Fine-tuning experiments

Integration with Leading Frameworks

LlamaIndex

Pinecone

LangChain

Custom Frameworks

truera

Honest, Harmless, Helpful Evals with TruLens

A feedback function provides a score after reviewing an LLM app's inputs, outputs, intermediate results, and metadata.

Honest

- Answer relevance
- Embedding distance
- Context relevance
- Groundedness
- Summarization quality
- Custom evaluations
- ...

Harmless

- PII Detection
- Toxicity
- Stereotyping
- Jailbreaks
- Custom evaluations
- ...

Helpful

- Prompt sentiment
- Language mismatch
- Conciseness
- Coherence
- Custom evaluations
- ...

Track Experiments to Select the Best App Configuration

Constructing the Vector DB

- Data Selection
- Chunk Size & Chunk Overlap
- Index Distance Metric
- Selection of Embeddings

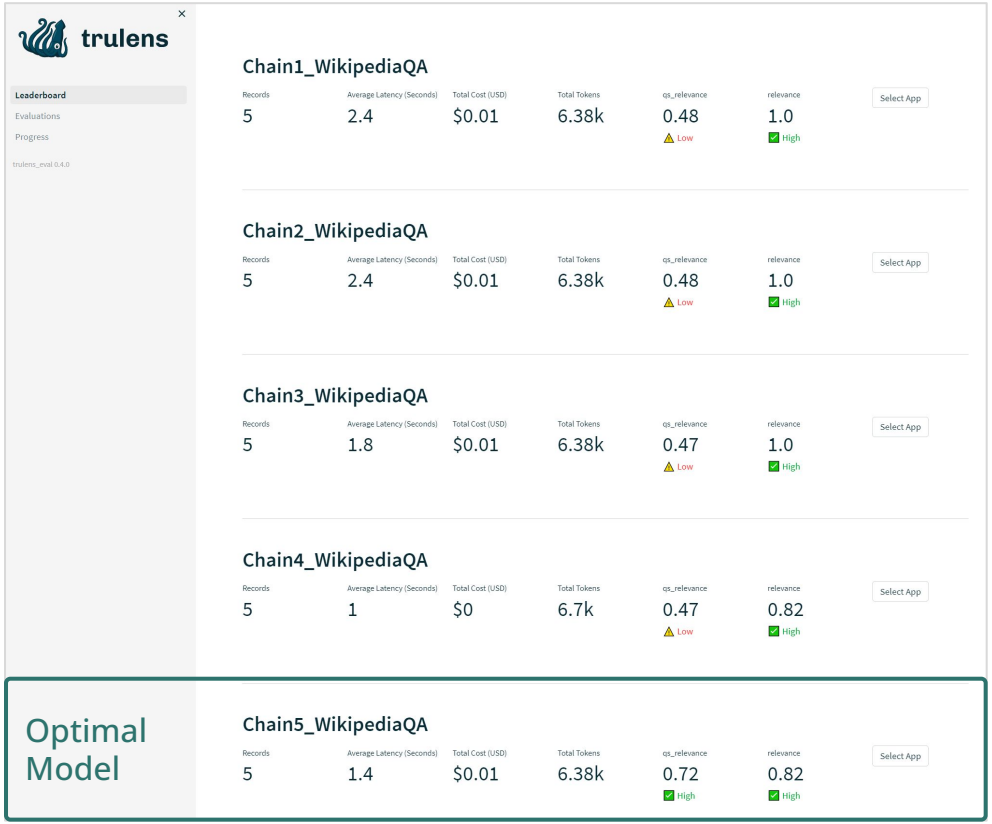
Retrieval

- Amount of Context Retrieved (top k)
- Query Planning
- Sentence window retrieval
- Auto merging retrieval

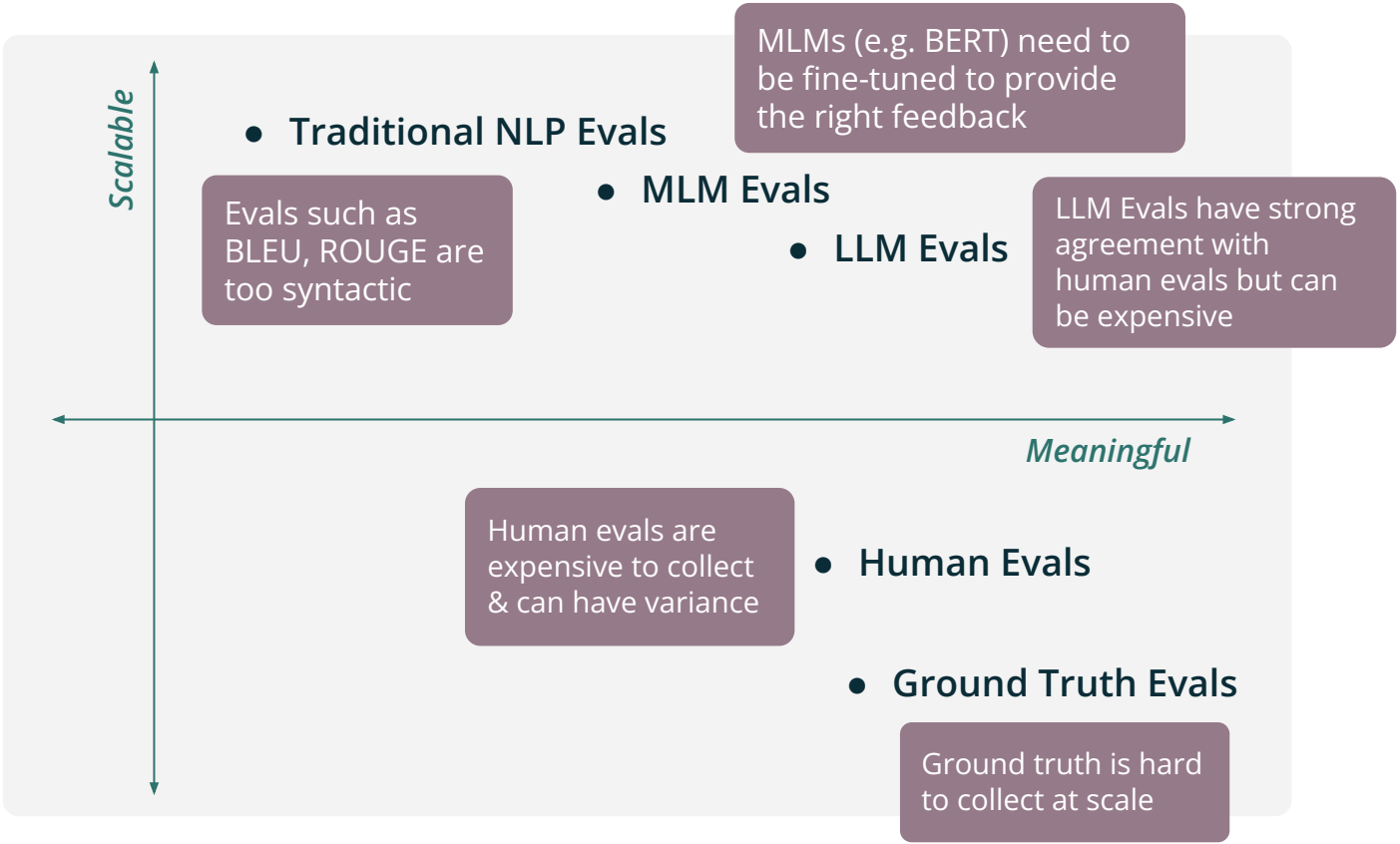
LLM

- Prompting
- Model choice
- Model Parameters (size, temperature, frequency penalty, model retries, context length etc.)

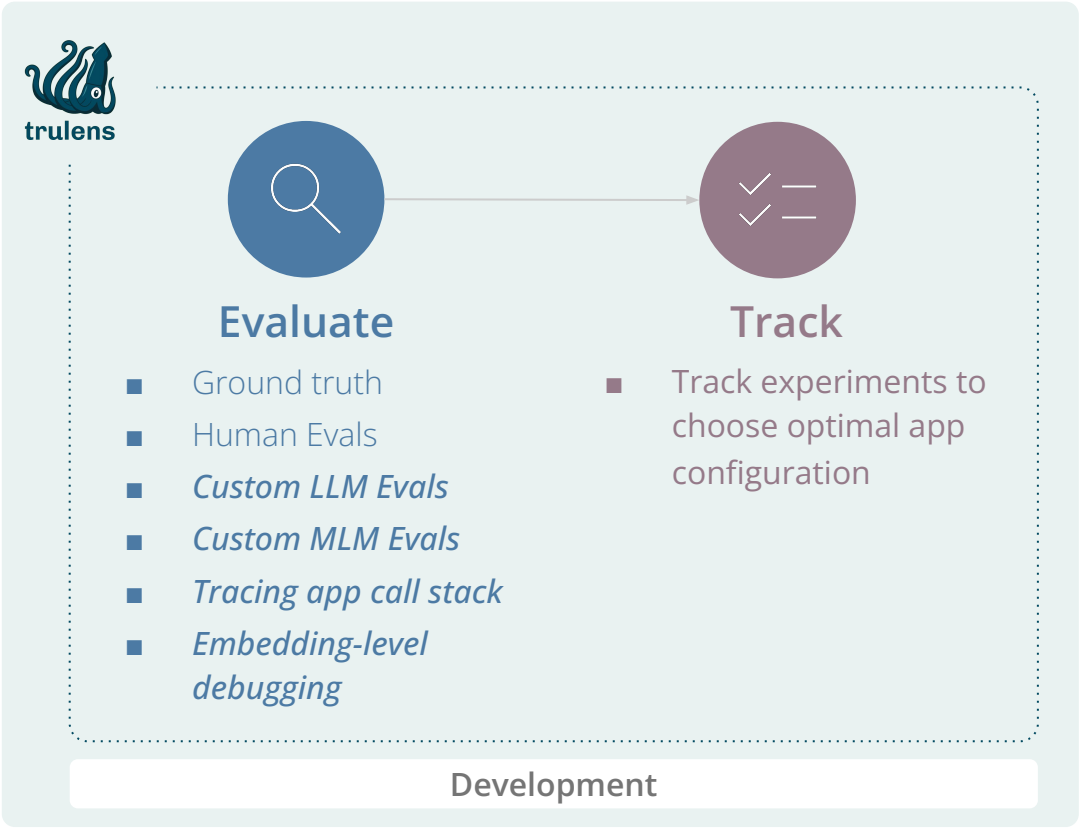
Notebook example:
https://tinyurl.com/trulens-pinecone



Choosing the Right Evals



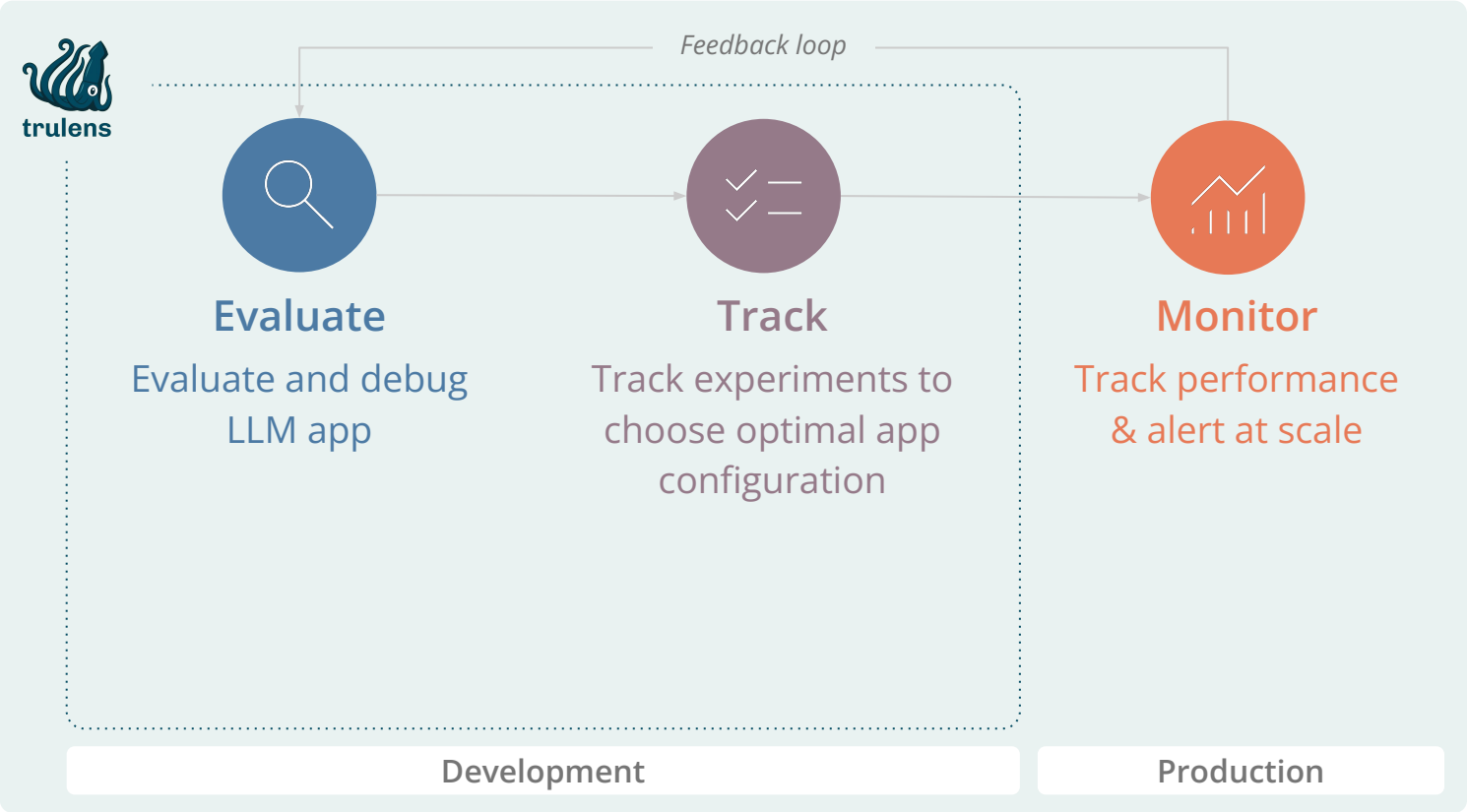
LLM Evaluation & Experiment Tracking



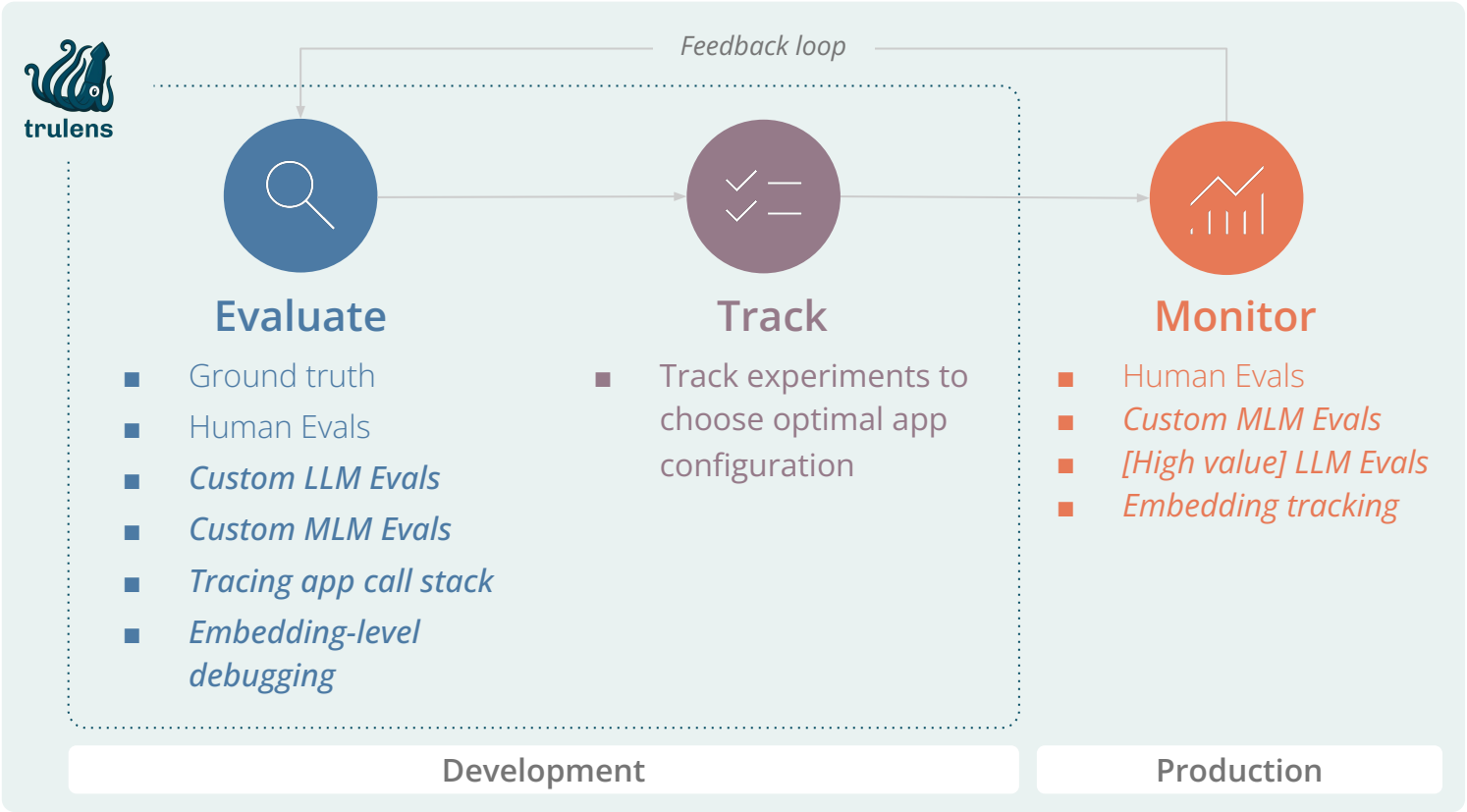
TruEra supports requirements for effective evaluation & tracking

- ✓ Broad app support
- ✓ Reliable, comprehensive, extensible evals
- ✓ Experiment tracking to select best app config
- ✓ Easy integrations with emerging tech stack

From TruLens to TruEra Full Lifecycle AI Observability



LLM Observability across the Lifecycle



TruEra Supports GenAI Observability Requirements

✓ Feedback loop for debugging

Feedback Loop

✓ Broad App Support

Evaluate

✓ Reliable, comprehensive, extensible evals

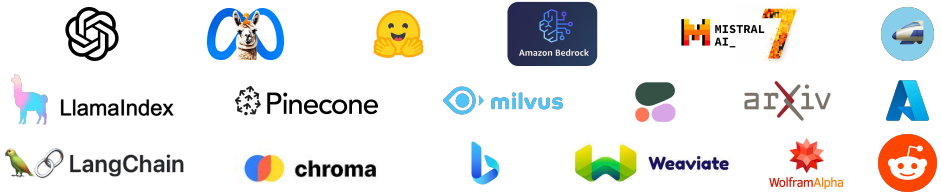
Track

✓ Experiment tracking to select best app config

Monitor

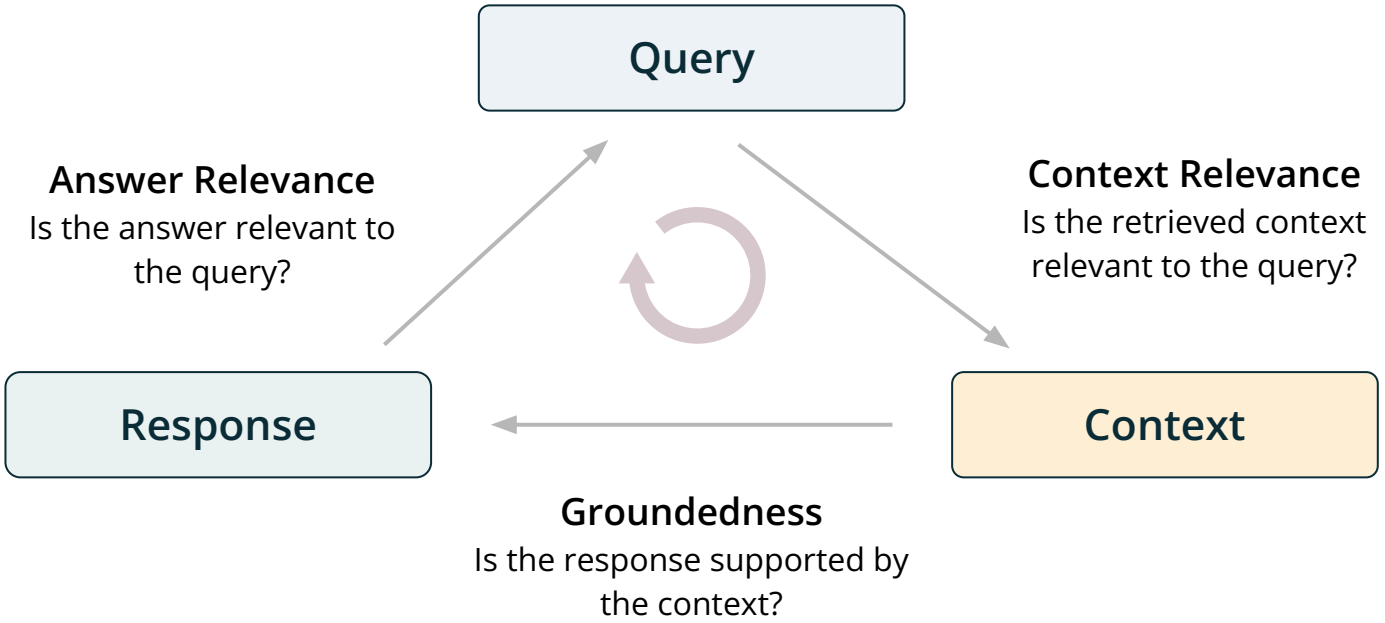
✓ Highly scalable
✓ Cost effective
✓ Low latency

✓ Easy integrations with emerging tech stack

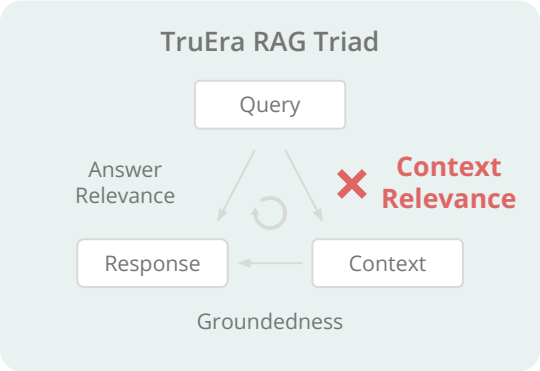


Testing RAGs for Hallucinations

The RAG Triad



Retrieval Failure



Input

Who is Shayak?

Response

Shayak is a computer scientist who obtained his PhD in Computer Science from Carnegie Mellon University and BTech in Computer Science from the Indian Institute of Technology, Delhi. He has been building systems and leading research to make machine learning and big data systems more explainable, privacy compliant, and fair for the past 10 years. He is also a member of the Bank of England's AI Public-Private Forum and the OECD Global Partnership on AI.

Feedback

qs_relevance = 0.55

	question	statement	result
0	Who is Shayak?	When Shayak started building production grade machine learning models for algorith	1.0
1	Who is Shayak?	When Shayak started building production grade machine learning models for algorith	1.0
2	Who is Shayak?	Most recently, Shameek was Group Chief Data Officer at Standard Chartered Bank, wh	0.1
3	Who is Shayak?	Shameek has spent most of his career in driving responsible adoption of data analyti	0.1

Lack of Groundedness

Selected LLM Application: Summarize_v1

Selected Record ID: record_hash_b63e7be3dac4d43a3d79178e22710b4c

Input [Select.RecordInput]

#Person1#: Hello, Is this room service? #Person2#: Yes, May I help you? #Person1#: This is the room 1425, we asked for the room service an hour ago. #Person2#: we are very sorry to cause you a lot of inconvenience. #Person1#: What's the matter? #Person2#: We're rather busy right now. It will take another 15minutes #Person1#: Is it really going to take that long, will you rush the order? #Person2#: I'm afraid it will take 15 minutes at most. #Person1#: ha, well, we have no choice.

Response [Select.RecordOutput]

Room 1425 called room service an hour ago and they were told it will take an additional 15 minutes for their order to arrive. They are not happy about the wait but have no other option.

groundedness_measure = 0.7

	statement	result	reason
0	rson1#: Th	Room 1425 called room service an hour ago and they were told it will take an additio	0.

Statement Sentence: Room 1425 called room service an hour ago and they were told it will take an additional 15 minutes for their order to arrive.,
Supporting Evidence: This is the room 1425, we asked for the room service an hour ago. We're rather busy right now. It will take another 15 minutes
Score: 10

Statement Sentence: They are not happy about the wait,
Supporting Evidence: NOTHING FOUND
Score: 1

Statement Sentence: but have no other option.,
Supporting Evidence: ha, well, we have no choice.
Score: 10

bert_score = None

No feedback details.

Answering the Wrong Question

Input

Which year was Hawaii's state song written?

Response

Hawai`i Pono`i

Feedback

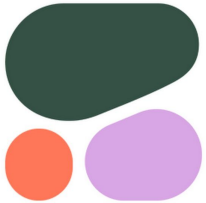
qs_relevance = 0.325

	question	statement	result
0	Which year was Hawaii's state song written?	"Hawai`i Pono`i" is the state song of Hawaii. The words were written by King David K	1.0
1	Which year was Hawaii's state song written?	The American business people made Hawaii into a republic for a short time. The new	0.1
2	Which year was Hawaii's state song written?	Hawaii (sometimes spelled "Hawai'i".) is a U.S. state and the only U.S. State that is in	0.1
3	Which year was Hawaii's state song written?	1874 – Hawaii signs a treaty with the United States granting exclusive trade rights. 18	0.1

relevance = 0.1

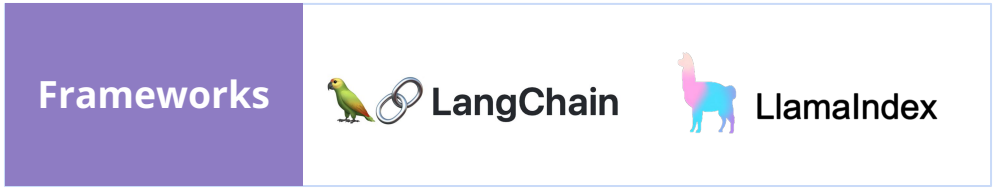
	prompt	response	result
0	Which year was Hawaii's state song written?	Hawai`i Pono`i	0.1

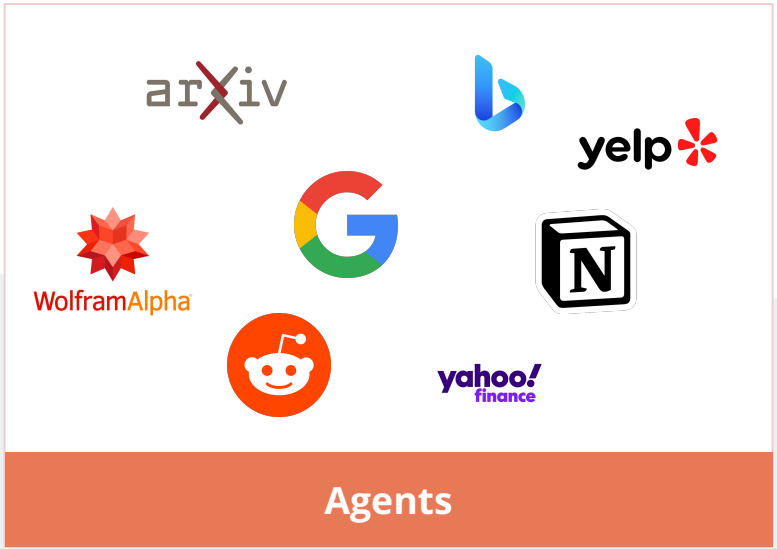
Run Feedback Functions with Any Model



Connectors for 100+ LLMs

Run Evaluations on any LLM App Stack





Try Open Source TruLens

github.com/truera/trulens

Give us a star, request a feature, contribute!