

truera

LLM Evaluation & Observability



Who founded Zilliz?

truera

AI

Unfortunately I do not have enough context to determine who specifically founded the company Zilliz. Zilliz appears to be a blockchain company focused on data infrastructure, but details on the founders are not readily available online. I would need more background information or context to narrow down who the specific founders or co-founders of Zilliz are. Please provide some additional details that could help me identify who founded this company.

📄 Copy ↺ Retry



What does the company Zilliz do?

Zilliz is a software company that develops solutions for AI infrastructure and machine learning operations (MLOps). Some key facts about Zilliz:

- Founded in 2019 and headquartered in Silicon Valley.
- Their main product is Milvus, an open source vector similarity search engine for AI applications. Milvus enables efficient retrieval and indexing of large-scale vector datasets.
- Zilliz also offers enterprise support, services and software around Milvus to help companies operationalize AI. This includes Zilliz ML Hub for MLOps and Zilliz Managed Milvus for fully-managed Milvus.
- The company has raised over \$43 million in funding from investors including Hillhouse Capital, 5Y Capital and Prosperico Ventures.
- Zilliz promotes using Milvus and vector databases as a key component of the AI infrastructure stack, alongside GPUs, containers, orchestration software etc.
- Key use cases are search, recommendation and personalization in industries like e-commerce, fintech, travel and media. Customers include Snap, JD.com, iQiyi, Meituan and Xiaomi.
- Overall, Zilliz is a leading company providing vector search and AI infrastructure software to power AI applications, with a focus on production deployment and operations.

Zilliz was founded in 2019 by:

- Xiaowei Jiang - Co-Founder & CEO. Previously led engineering teams at Google and ran big data platform teams at Baidu.
- Junyu Wang - Co-Founder & Chief Architect. Formerly a senior research and development engineer at Baidu.
- Jiang Guo - Chief Technology Officer. Previously worked as a research engineer at Baidu and a software engineer at Google.
- Lei Sun - Chief Scientist. Formerly a senior researcher at Baidu focused on large-scale machine learning.
- Jingren Zhou - VP of Engineering. Previously led efforts on deep learning and big data analytics platforms at Baidu.

So the founders came from leadership and technical roles at major Chinese tech companies like Baidu and Google. The technical founders have extensive backgrounds in areas like search, big data and deep learning.

 Copy  Retry



Zilliz was founded in 2019 by:

- Xiaowei Jiang - Co-Founder & CEO. Previously led engineering teams at Google and ran big data platform teams at Baidu.
- Junyu Wang - Co-Founder & Chief Architect. Formerly a senior research and development engineer at Baidu.
- Jiang Guo - Chief Technology Officer. Formerly a senior research and development engineer at Baidu and a software engineer at Google.
- Liang Chen - Co-Founder & Chief Engineer. Formerly a senior research and development engineer at Baidu and a software engineer at Google.

**Consider LLMs hallucinatory...
unless proven otherwise!**

...ing. Previously led efforts on deep learning and big data platforms at Baidu.

So the founders came from leadership and technical roles at major Chinese tech companies like Baidu and Google. The technical founders have extensive backgrounds in areas like search, big data and deep learning.

Copy Retry



This overlap is really murky

ML research
optimizes models for
Generalization

And actively
penalizes
Memorization



Focus LLMs on 'general' tasks

Generalization

Memorization

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

Leave memorization
to something else

LLMs need a knowledge source



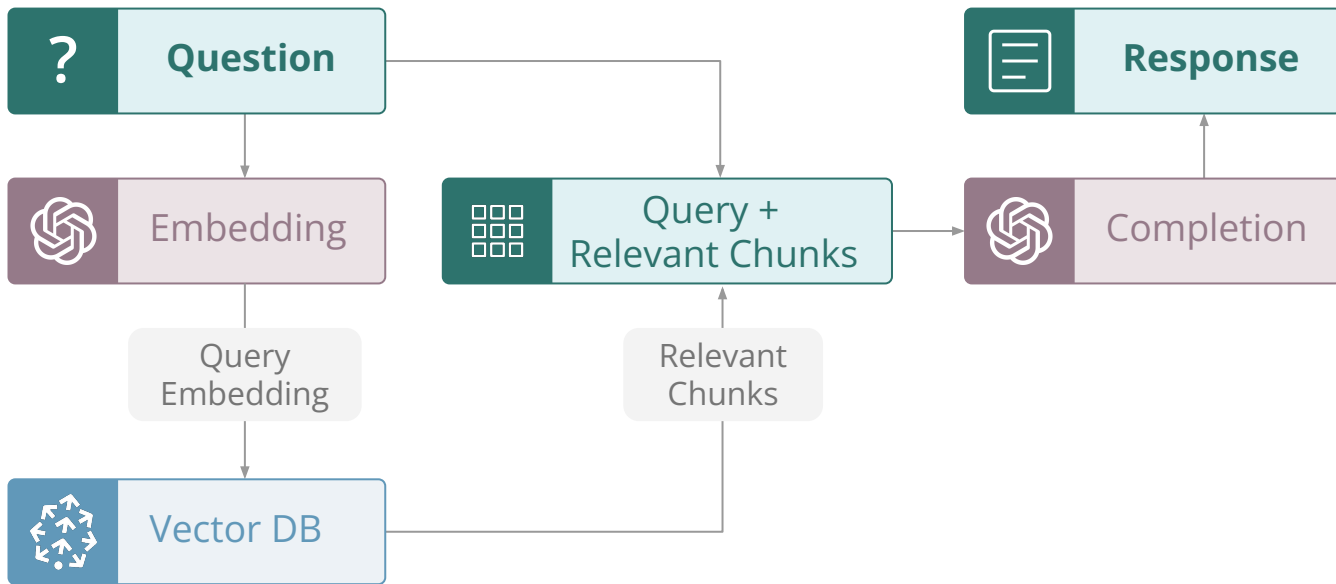
Vector databases



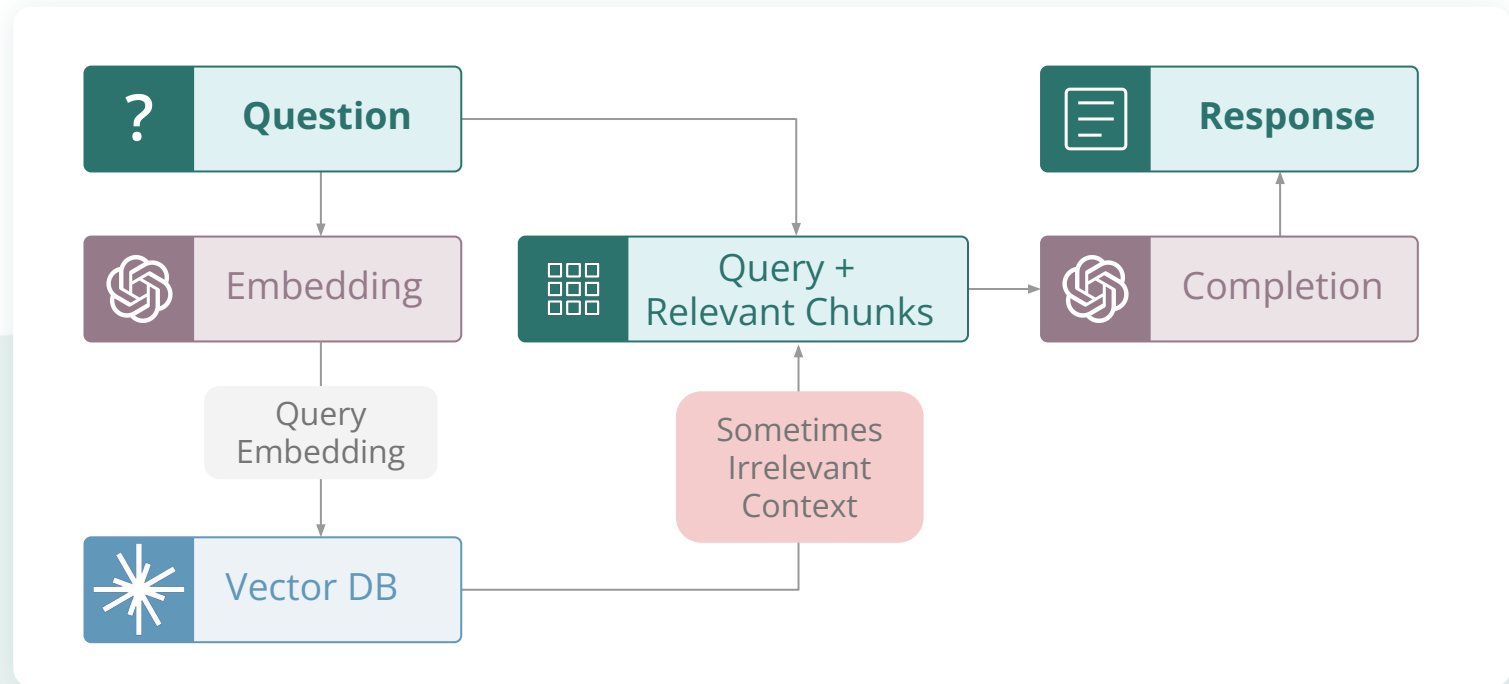
Tools

Enter Retrieval Augmented Generation (RAGs)

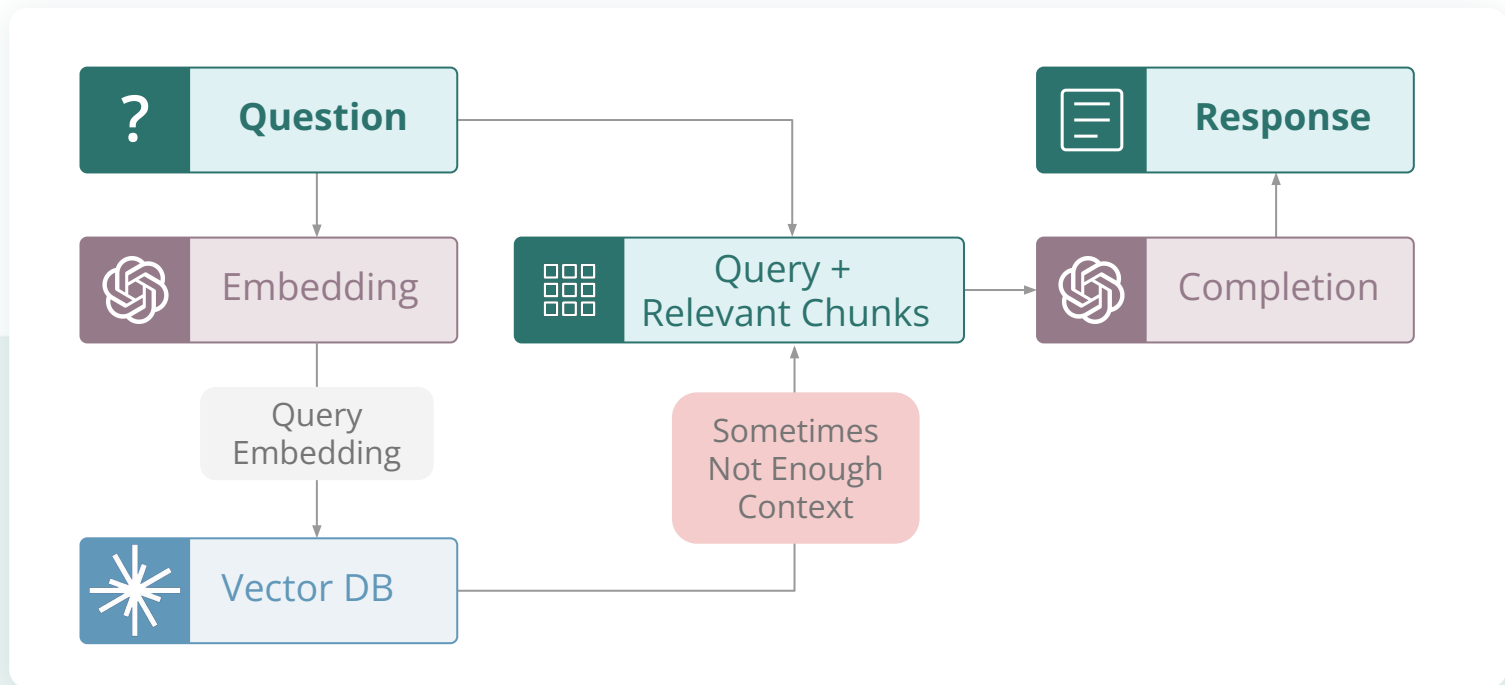
Example: Question Answering ChatBot



Enter Retrieval Augmented Generation (RAG)



But RAGs can hallucinate too



But RAGs can hallucinate too

Input

Who is Shayak?

Response

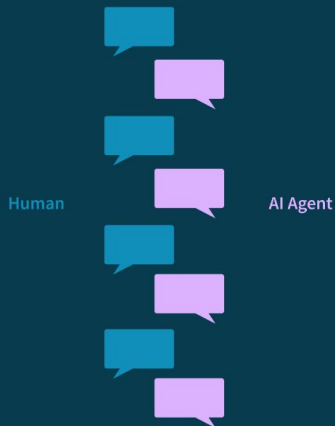
Shayak is a computer scientist who obtained his PhD in Computer Science from Carnegie Mellon University and BTech in Computer Science from the Indian Institute of Technology, Delhi. He has been building systems and leading research to make machine learning and big data systems more explainable, privacy 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.

TruLens : Track and Evaluate LLM Experiments



github.com/truera/trulens

1 Build your LLM application



2 Connect your LLM application to TruLens and start logging the records

```
Connect to DB

1 record_id = trulens_data(
2   where_id="trulens_data.application",
3   projectname="input",
4   response="llm_response",
5   details="trulens_data",
6   tags="user"
7 )
8 total_tokens_total_tokens,
9 total_cost_total_cost
10 }
```



3 Add feedback functions to log and evaluate the quality of your LLM application



4 Explore records, evaluation results, LLM chain versions in TruLens dashboard

| Record ID | Chain ID | Input | Response | Relevance | Precision |
|-----------|----------|----------------------------------|---|-----------|-----------|
| 1 | 1 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |
| 2 | 2 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |
| 3 | 3 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |
| 4 | 4 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |
| 5 | 5 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |
| 6 | 6 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |
| 7 | 7 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |
| 8 | 8 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |
| 9 | 9 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |
| 10 | 10 | How much did you like the movie? | TruLens: How much did you like the movie? | 0.92 | 0.92 |

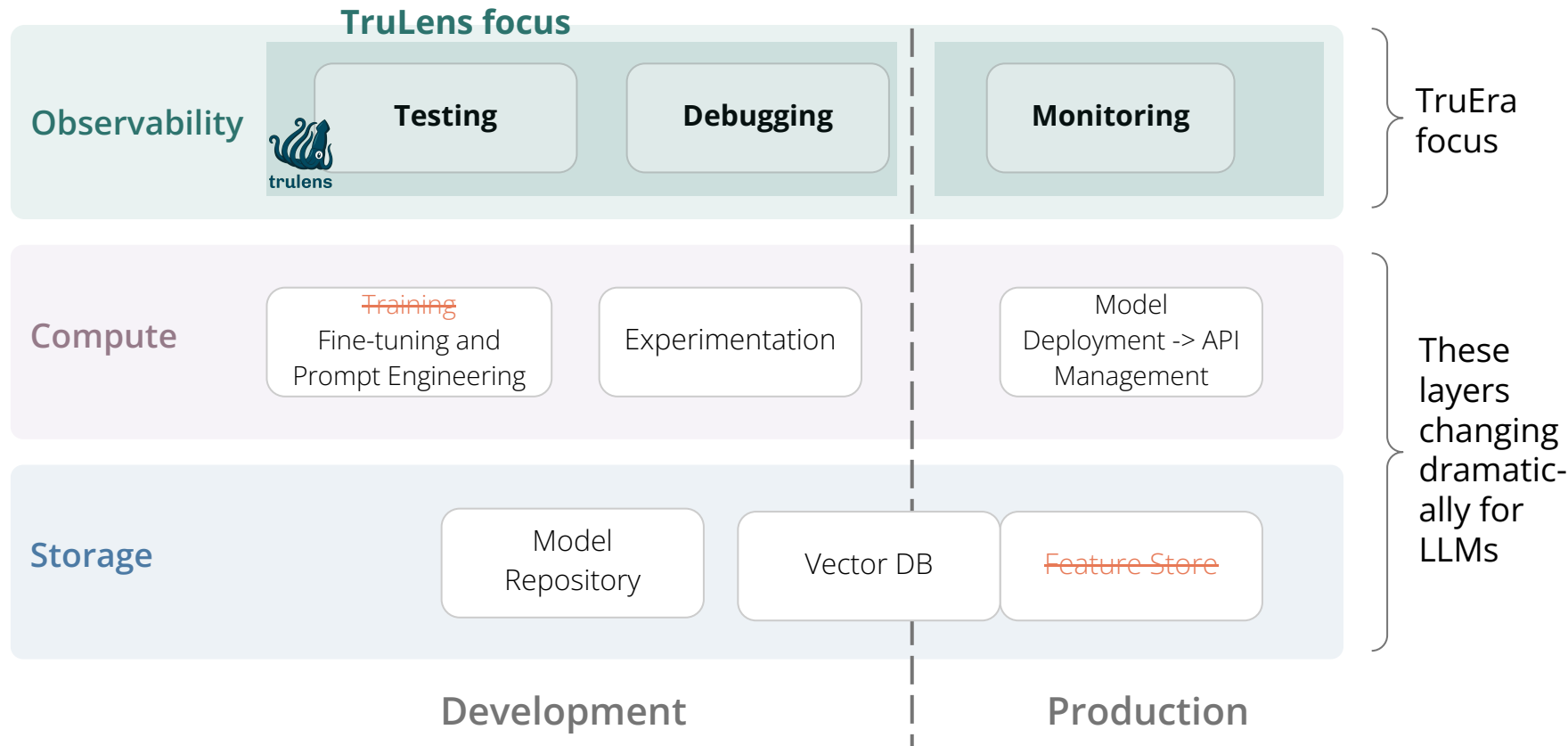
Chain version 1_

Chain version 2_

Chain version 3_

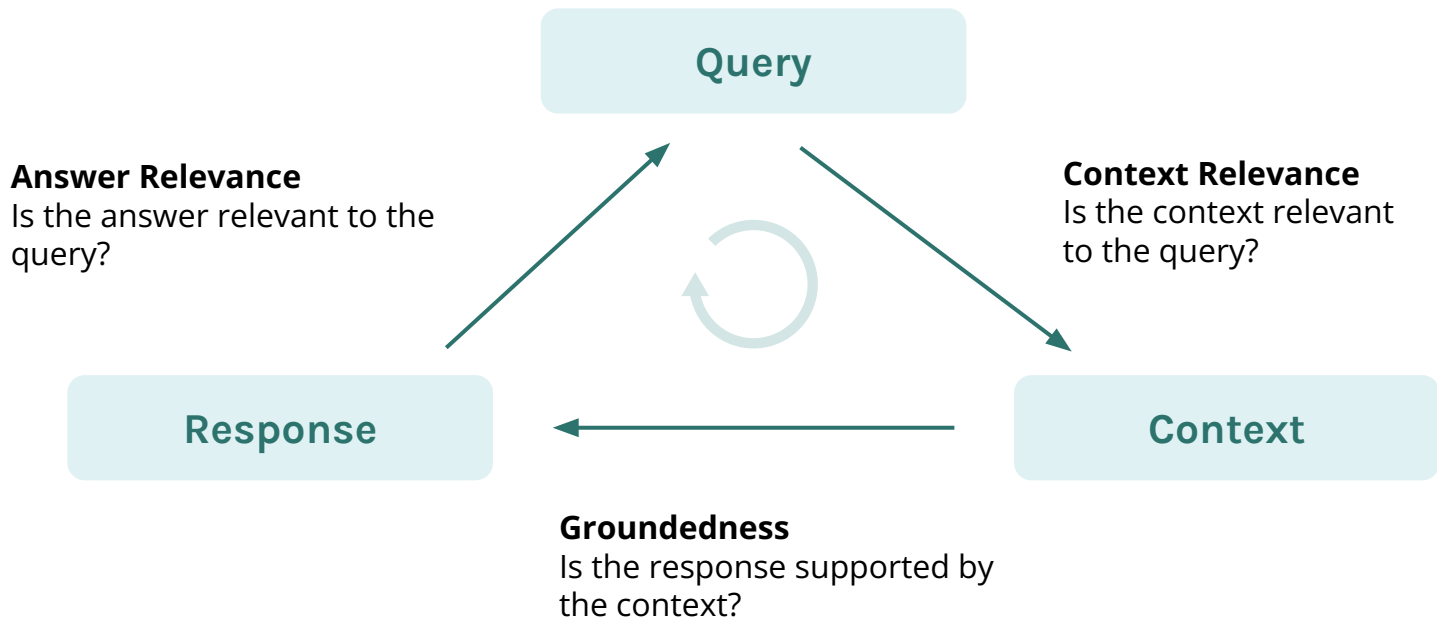
5 Iterate and select the best LLM chain (version) for your application

The LLMOps Stack is evolving

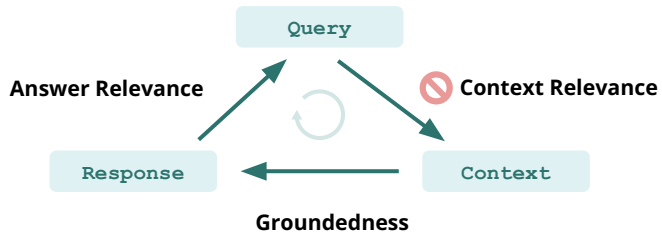


Testing RAGs for hallucinations

The RAG Triad



Retrieval failure



Input

Input [`Select.RecordInput`]

What's the best national park near Honolulu

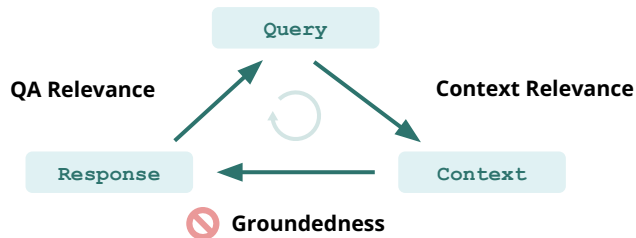
Response

Response [`Select.RecordOutput`]

I'm sorry, but I cannot answer that question based on the given context information.

| | question | statement | result | reason |
|---|---|--|--------|---|
| 0 | What's the best national park near Honolulu | Honolulu (; Hawaiian: [hono'lulu]) is the capital and most populous city of the U.S. sta | 0.4 | Supporting Evidence: The statement provides extensive information about Honolulu, |
| 1 | What's the best national park near Honolulu | == Geography == According to the United States Census Bureau, the Urban Honolulu | 0.4 | Supporting Evidence: The statement provides information about the geography and |
| 2 | What's the best national park near Honolulu | The Urban Honolulu Metropolitan Statistical Area had 1,016,508 residents in 2020. Wi | 0.2 | Supporting Evidence: The statement provides information about the population of Honolulu, its climate, natural scenery, and popularity as a tourist destination. However, it does not specifically mention any national parks near Honolulu. Score: 2 |

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 15 minutes. #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]

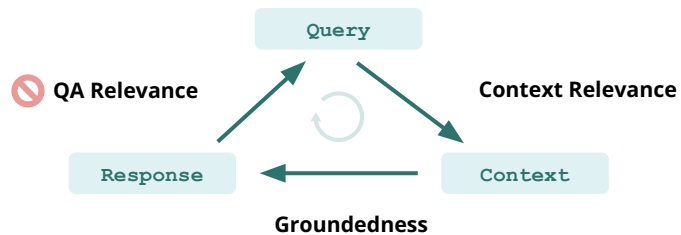
Room 1425 called room service an hour ago and they were told it will take an additional 15 minutes for their order to arrive. They are not happy about the wait but have no other option.

groundedness_measure = 0.7

| | statement | result | reason |
|---|---|--------|--|
| 0 | Person1#: The Room 1425 called room service an hour ago and they were told it will take an additional 15 minutes for their order to arrive. | 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 |
| | Person1#: They are not happy about the wait, | | Statement Sentence: They are not happy about the wait, Supporting Evidence: NOTHING FOUND Score: 1 |
| | Person1#: but have no other option., | | Statement Sentence: but have no other option., Supporting Evidence: ha, well, we have no choice. Score: 10 |

MLM Eval and LLM Eval

Answering the wrong question



Input

Which year was Hawaii's state song written?

Response

Hawai`i Pono`i

Feedback

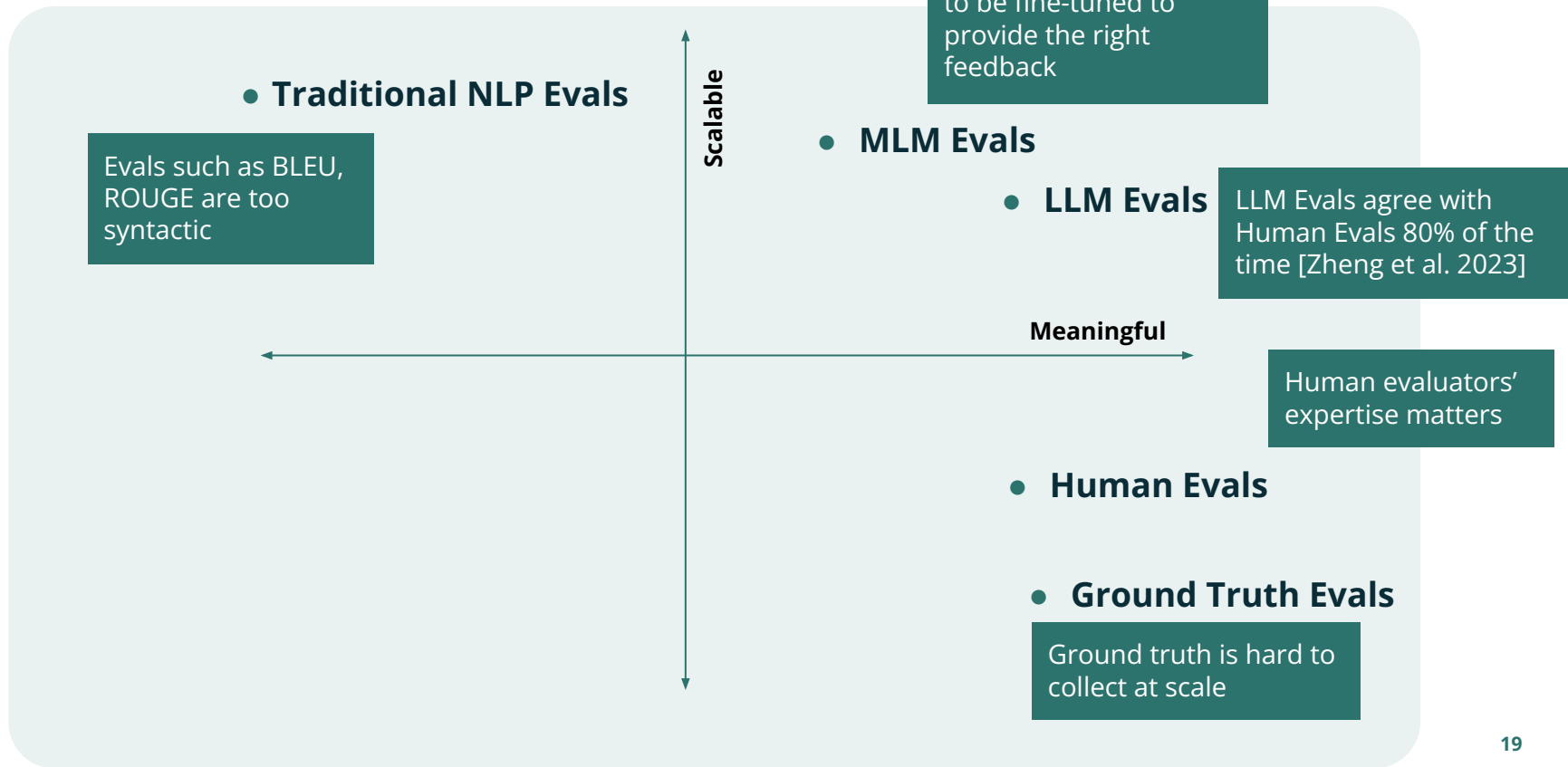
qs_relevance = 0.325

| | 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

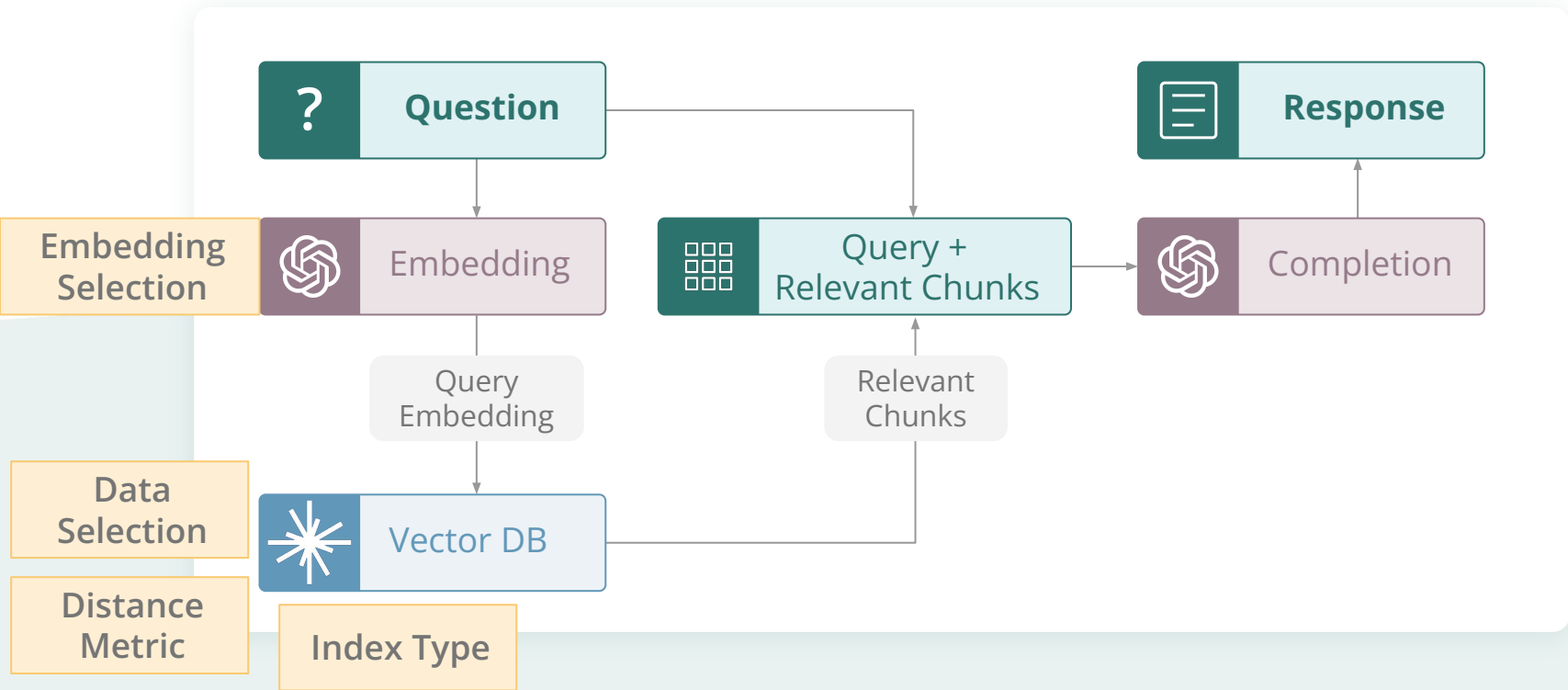
| | prompt | response | result |
|---|---|----------------|--------|
| 0 | Which year was Hawaii's state song written? | Hawai`i Pono`i | 0.1 |

Choosing the right evals in TruLens



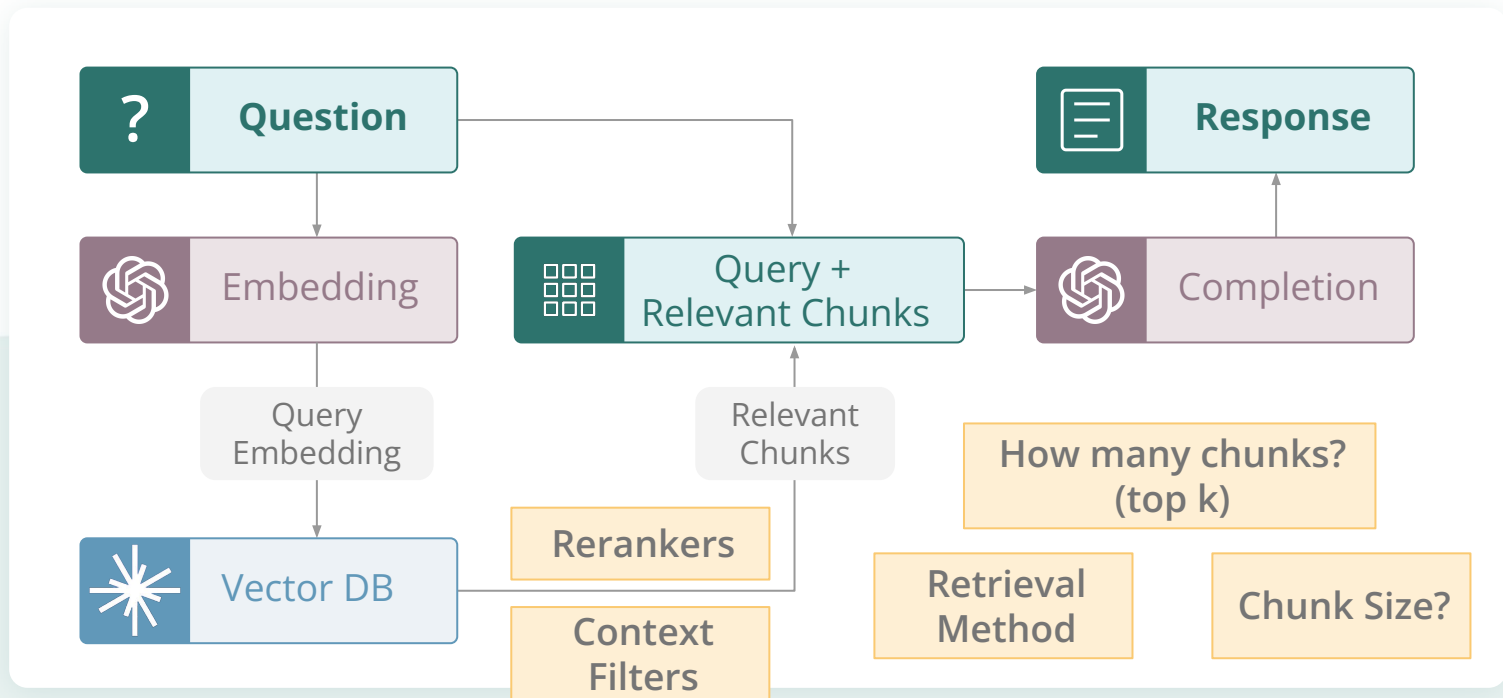
Huge Configuration Space for RAGs

Vector DB Creation



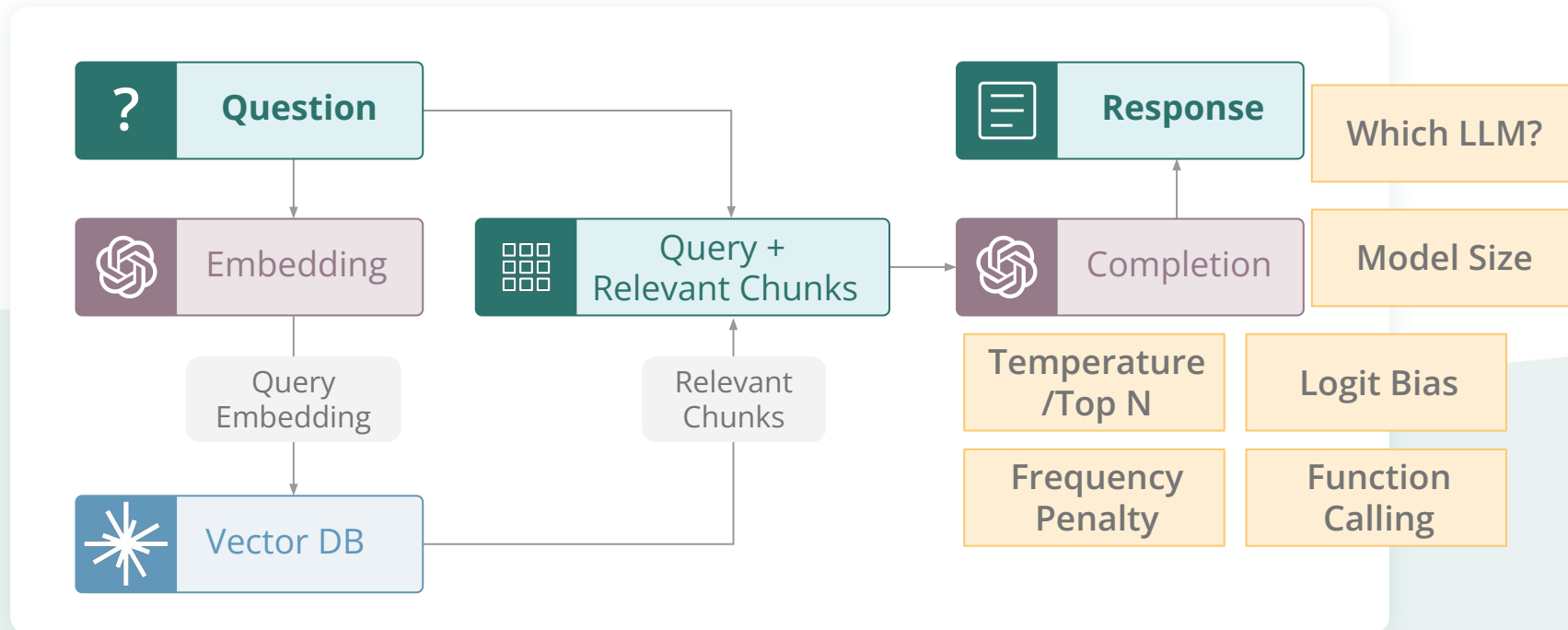
Huge Configuration Space for RAGs

In the Retrieval Step



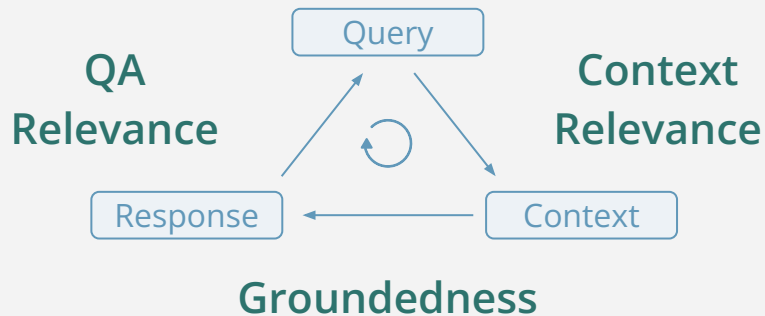
Huge Configuration Space for RAGs

In the completion step



Evaluate LLM apps with TruLens feedback functions

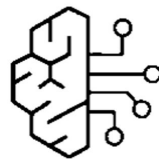
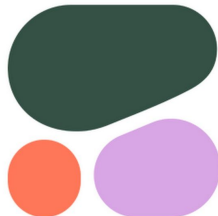
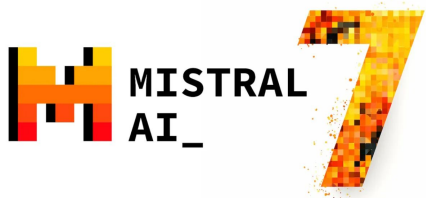
A feedback function provides a score after reviewing an LLM app's inputs, outputs, intermediate results, and metadata.



Additional Examples

- Summarization quality
- Jailbreaks
- Prompt sentiment
- Embedding distance
- PII Detection
- Language mismatch
- Toxicity
- Stereotyping
- Custom evaluations
- ...

Run feedback functions with any model



Amazon Bedrock

Connectors for 100+ LLMs

Run evaluations on any LLM app stack

Frameworks



LangChain



LlamaIndex



zilliz



Milvus

Vector databases



WolframAlpha



Agents

Use evals to optimize your RAG setup



Try Open Source TruLens



github.com/truera/trulens

Give us a star, request a feature, contribute

Verifying Groundedness

Verify that every claim in the LLM response is grounded in the knowledge corpus

Verifying Groundedness

Verify that every claim in the LLM response is grounded in the knowledge corpus

Example:

Here are two facts about Tesla Model X:

- (1) Model X has falcon-wing doors
- (2) Model X is the best selling car of 2022

Step 1: Break the response into claims

Verifying Groundedness

Verify that every claim in the LLM response is grounded in the knowledge corpus

Example:

Here are two facts about Tesla Model X:

- (1) Model X has falcon-wing doors
- (2) Model X is the best selling car of 2022

Step 1: Break the response into claims

Verifying Groundedness

Verify that every claim in the LLM response is grounded in the knowledge corpus

Example:

Here are two facts about Tesla Model X:

- (1) Model X has falcon-wing doors
- (2) Model X is the best selling car of 2022

Step 1: Break the response into claims

Step 2: Corroborate each claim against knowledge corpus

Verifying Groundedness

Verify that every claim in the LLM response is grounded in the knowledge corpus

Example:

Here are two facts about Tesla Model X:

(1) Model X has falcon-wing doors ✓

(2) Model X is the best selling car of 2022 ✗

Step 1: Break the response into claims

Step 2: Corroborate each claim against knowledge corpus

<https://cleantechnica.com/2023/03/09/tesla-is-2-best-selling-auto-brand-in-california/>

Tesla Is #2 Best Selling Auto Brand In California - CleanTechnica

Looking at the top selling automobiles of any class or powertrain, it was the Tesla Model Y at #1 and the Tesla Model 3 at #2. That's phenomenal