

# New York Unstructured Data Meetup

Presented by:



# Unstructured Data Meetup | Host



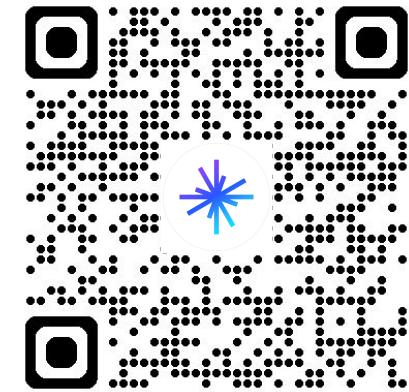
## Tim Spann

Principal Developer  
Advocate, Zilliz

[tim.spann@zilliz.com](mailto:tim.spann@zilliz.com)

<https://www.linkedin.com/in/timothyspann/>

<https://x.com/PaaSDev>



# Code of Conduct



## Be respectful and kind

When communicating with all event participants, speakers, and hosts. Be considerate

## All ideas are welcome

Be present and participate actively in discussions. Ask questions and reach out for help when needed.

## Report inappropriate behavior

Any inappropriate behavior is not tolerated at this event. Inform a Zilliz team member immediately if you see any behavior deemed inappropriate

# Getting Started with Vector Databases

**Milvus**

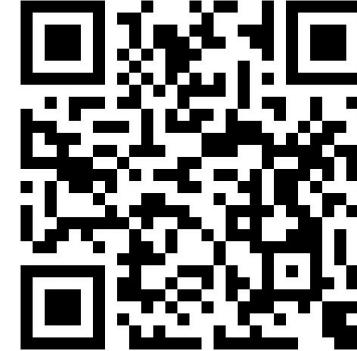
Open Source Self-Managed



[github.com/milvus-io/milvus](https://github.com/milvus-io/milvus)

**Zilliz Cloud**

SaaS Fully-Managed



[zilliz.com/cloud](https://zilliz.com/cloud)

# Zilliz is Hiring!

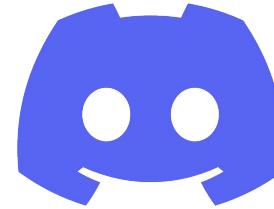
Join our  
Team

[Zilliz.com/careers](https://Zilliz.com/careers)

- Developer Advocate
- Senior Software Engineer
- Staff Software Engineer
- Solutions Architect



# Join the Milvus Discord!



# Become a Speaker!



Interesting in speaking at and/or sponsoring a Zilliz Unstructured Data Meetup? Fill out this form!



**Have you built  
something cool  
using Milvus or  
Zilliz? We want to  
hear all about it.**

Share Your Story



Star Milvus  
for a chance  
to win a prize  
tonight!





@Zilliz, @Milvus



@zilliz\_universe, @milvusio

# Share your photos!

#ZillizUnstructuredData



# Welcome Speakers

## TECH TALK 1

How Inkeep and Zilliz built  
an AI Assistant



**Robert Tran**

Founder, CTO @ Inkeep

## TECH TALK 2

Introduction to the Data  
Prep Kit



**Santosh Borse**

Senior Engineer, Watsonx Data  
Engineering at IBM Research

## TECH TALK 3

RGB-X Model Development:  
Exploring Four Channel ML  
Workflows



**Daniel Gural**

Machine Learning and  
DevRel, Voxel 51



Join us at our next meetup!  
[lu.ma/unstructured-data-meetup](https://lu.ma/unstructured-data-meetup)



# Quick Intro to Unstructured Data, Edge AI and Milvus



**Tim Spann**  
Principal Developer Advocate, Zilliz

# Welcome to New York!

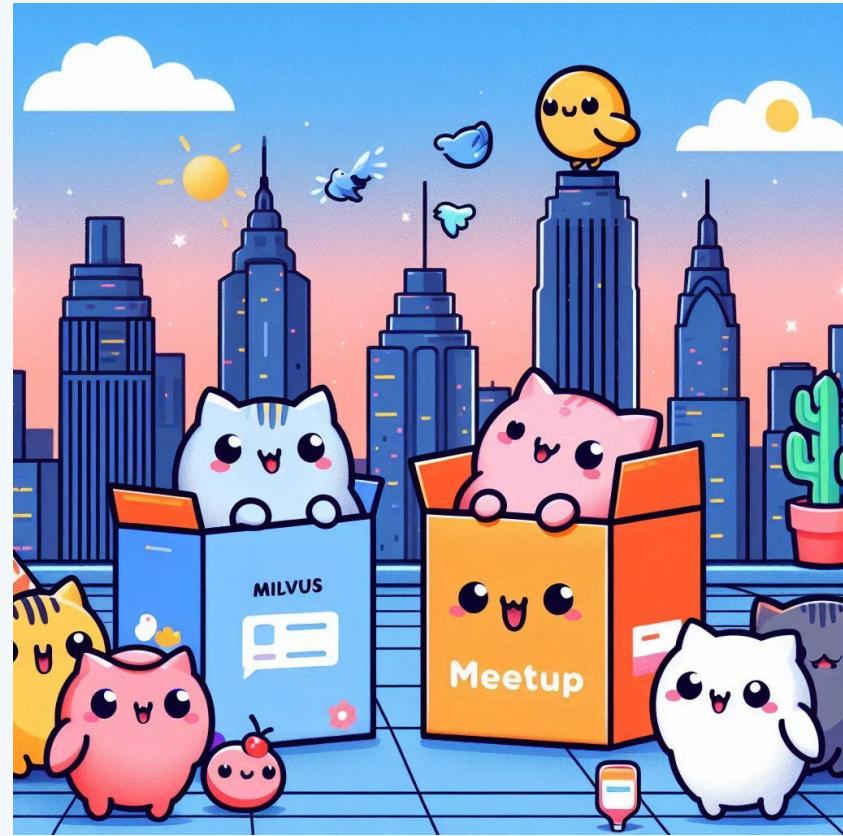
Tim Spann @ Zilliz



# These Slides



fdfdf



# 01

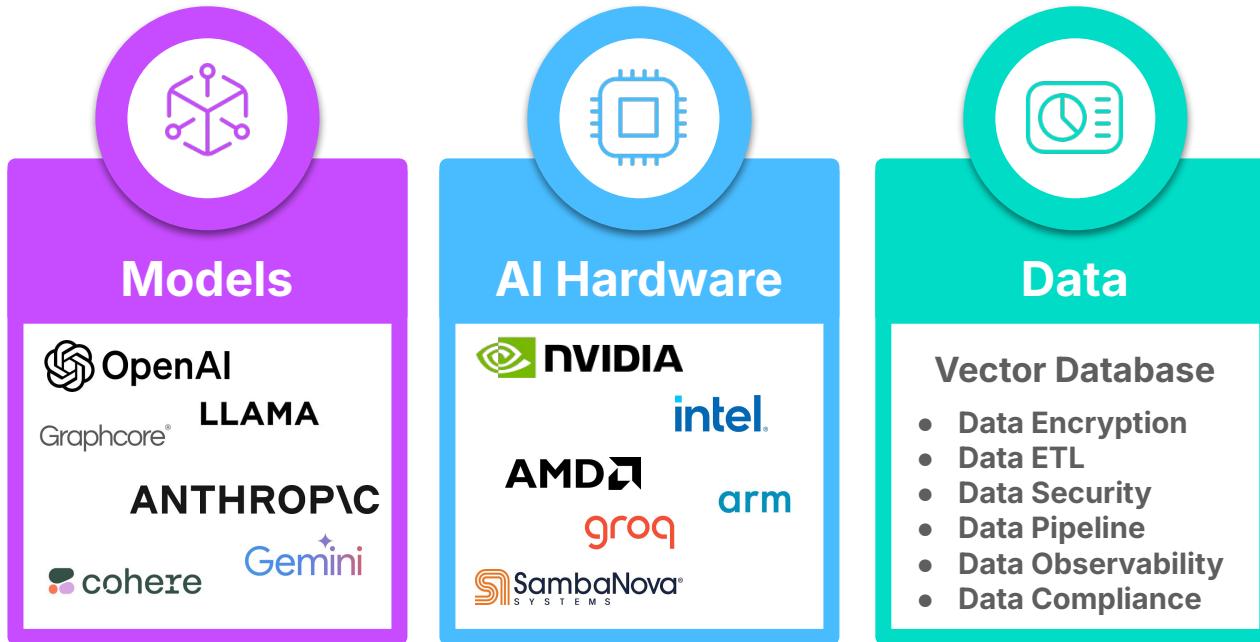
## Introduction

# MILVUS HAS VECTORS

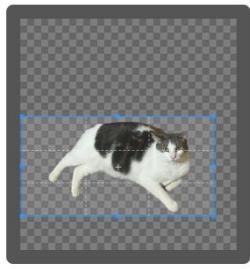


## I SEE ONLY CAT PICTURES

# Three Pillars of GenAI & the opportunities they bring



[Back to Demo](#)



[Upload Image](#)

Search Result:  
Duration: 35.93 ms



Sorted by Similarity metric



Similarity Metric: 0.470426



Similarity Metric: 0.485934



Similarity Metric: 0.520233



Similarity Metric: 0.534563



Similarity Metric: 0.536320



Similarity Metric: 0.540569



Similarity Metric: 0.539228



Similarity Metric: 0.544466



Similarity Metric: 0.550427



Similarity Metric: 0.537707



<https://milvus.io/milvus-demos/reverse-image-search>

# Show Me



Lots  
of Slides

Cool Demo

# Show Me Another Demo

zilliz

What is Milvus used for?

Ask Random Question Ask

The High-Performance Vector Database Built for Scale

ID	SCORE	QUESTION	ANSWER
450933285167028738	0.8684593439102173	What is Milvus?	The High-Performance Vector Database Built for Scale
450933285167028426	0.33861032128334045	The benefits from anti aging creams are the following except that	they make wrinkle faces rough
450933285167028302	0.3009645640850067	To prevent from catching a cold or flu , it 's good for you	to wash all parts of your hands
450933285167027798	0.2908312678337097	What is the function of our immune system ?	To destroy the troublemakers in order that they might not hurt the body .
450933285167028154	0.2765984535217285	Why is art therapy useful to patients with physical or emotional illnesses ?	It helps improve their social skills .
450933285167028454	0.26567766070365906	When Sarah went to school , she was wearing	blue jeans and a Jonas Brothers T - shirt
450933285167028658	0.2627016305923462	What is the special infill used to do ?	To collect water from rainfall .
450933285167028235	0.26243868470191956	The previous study found that	children 's math achievement is related to parents ' attitude about math
450933285167028102	0.25130075216293335	What do we know about TeliaSonera ?	It is in the charge of Pasi Kostinten
450933285167028620	0.2511159682273865	The MMORPGs are created by Jane McGonigal to	develop gamers ' problem - solving skills



<https://zilliz-semantic-search-example.vercel.app/>

# About Milvus

Milvus is an open-source vector database for GenAI projects. pip install on your laptop, plug into popular AI dev tools, and push to production with a single line of code.



**29K+**

GitHub Stars



**2,600+**

Forks



**25M+**

Downloads



**250+**

Contributors



## Easy Setup

Pip-install to start coding in a notebook within seconds

## Reusable Code

Write once, and deploy with one line of code into the production environment

## Integration

Plug into OpenAI, Langchain, LlmalIndex, and many more

## Feature-rich

Dense & sparse embeddings, filtering, reranking and beyond

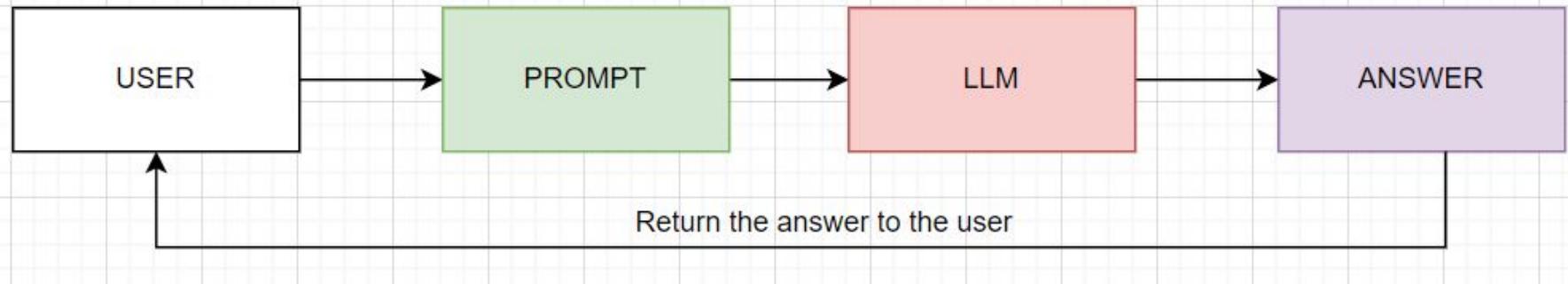
## Higher scalability

**10B vectors**

of 1536 dimensions  
in a single Milvus/Zilliz Cloud  
instance

**100B vectors**

in one of the largest deployment

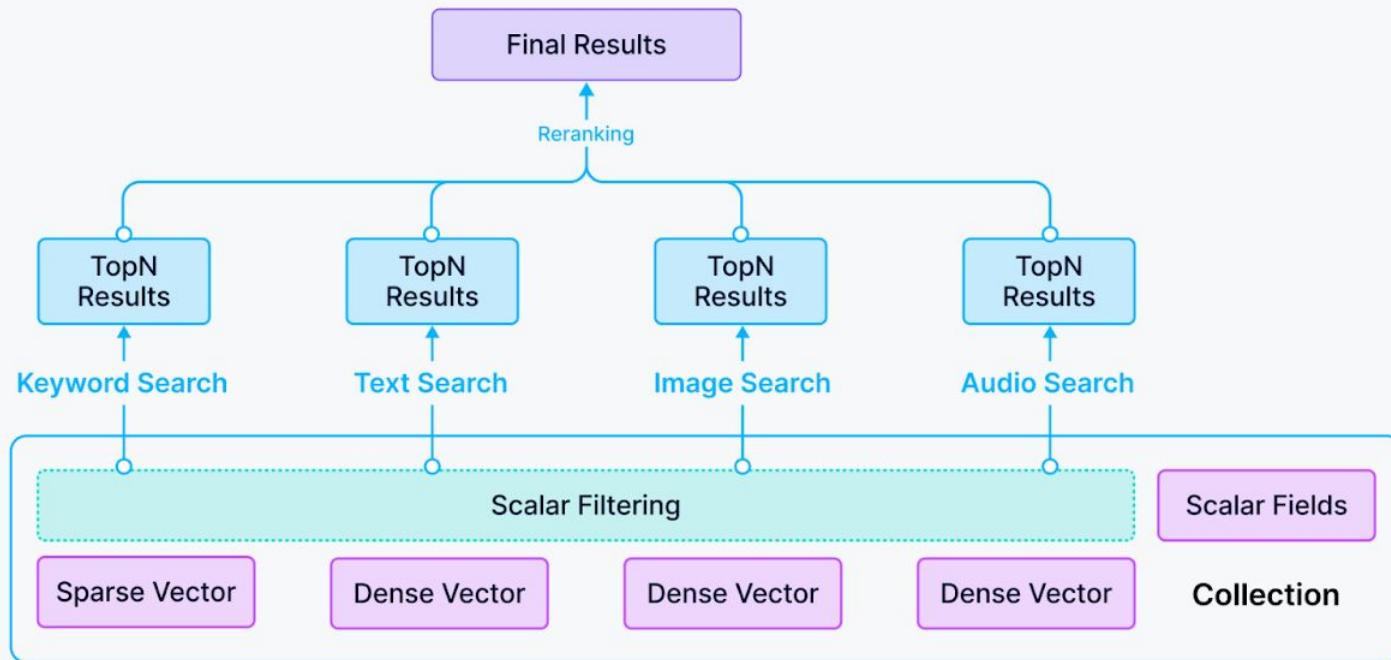


<https://zilliz.com/learn/large-language-models-and-search>

# 02

## Hybrid Search

# How Hybrid Search Works In Milvus



# Hybrid Search

Support the fusion of **vector search** and **full-text search**

Support the fusion of **multimodal vectors** from various unstructured data types such as **images**, **videos**, **audio**, and **text** files

Utilize various types of vector embeddings. This includes **dense embeddings** from models like **BERT** and **Transformers** and **sparse embeddings** from algorithms like **BM25**, **BGE-M3**, and **SPLADE**.

# Hybrid Search

- Milvus supports the creation of up to **10 vector fields** for the same dataset within a single collection. Based on this support, **hybrid search** allows users to search across multiple vector columns simultaneously. This capability allows for combining **multimodal search**, **hybrid sparse** and **dense search**, and **hybrid dense** and **full-text search**, offering versatile and flexible search functionality.
- These vectors in different columns represent diverse facets of data, originating from different embedding models or undergoing distinct processing methods. The results of hybrid searches are integrated using **various re-ranking strategies**.

# Hybrid Search

This feature enables different columns to:

- **Represent multiple perspectives of information.** For instance, in e-commerce, product images include front, side, and top views. Different views can be represented with different types or dimensions of vectors.
- **Utilize various types of vector embeddings.** This includes dense embeddings from models like [BERT](#) and [Transformers](#) and sparse embeddings from algorithms like [BM25](#), [BGE-M3](#), and [SPLADE](#).
- **Support the fusion of multimodal vectors from various unstructured data types such as images, videos, audio, and text files.** For example, in criminal investigations, suspects can be represented through biometric modalities such as fingerprints, voiceprints, and facial recognition, aiding in identifying individuals across different modalities.
- **Support the fusion of vector search and full-text search.**

<https://milvus.io/docs/multi-vector-search.md>

# When is Hybrid Search Recommended?

Hybrid search is ideal for complex situations demanding high accuracy, especially when an entity can be represented by multiple, diverse vectors. This applies to cases where the same data, such as a sentence, is processed through different embedding models or when multimodal information (like images, fingerprints, and voiceprints of an individual) is converted into various vector formats. By assigning weights to these vectors, their combined influence can significantly enrich recall and improve the effectiveness of search results.

# Hybrid Search - FAQ 2

- **How does a weighted ranker normalize distances between different vector fields?**

A weighted ranker normalizes the distances between vector fields using assigned weights to each field. It calculates the importance of each vector field according to its weight, prioritizing those with higher weights. It's advised to use the same metric type across ANN search requests to ensure consistency. This method ensures that vectors deemed more significant have a greater influence on the overall ranking.

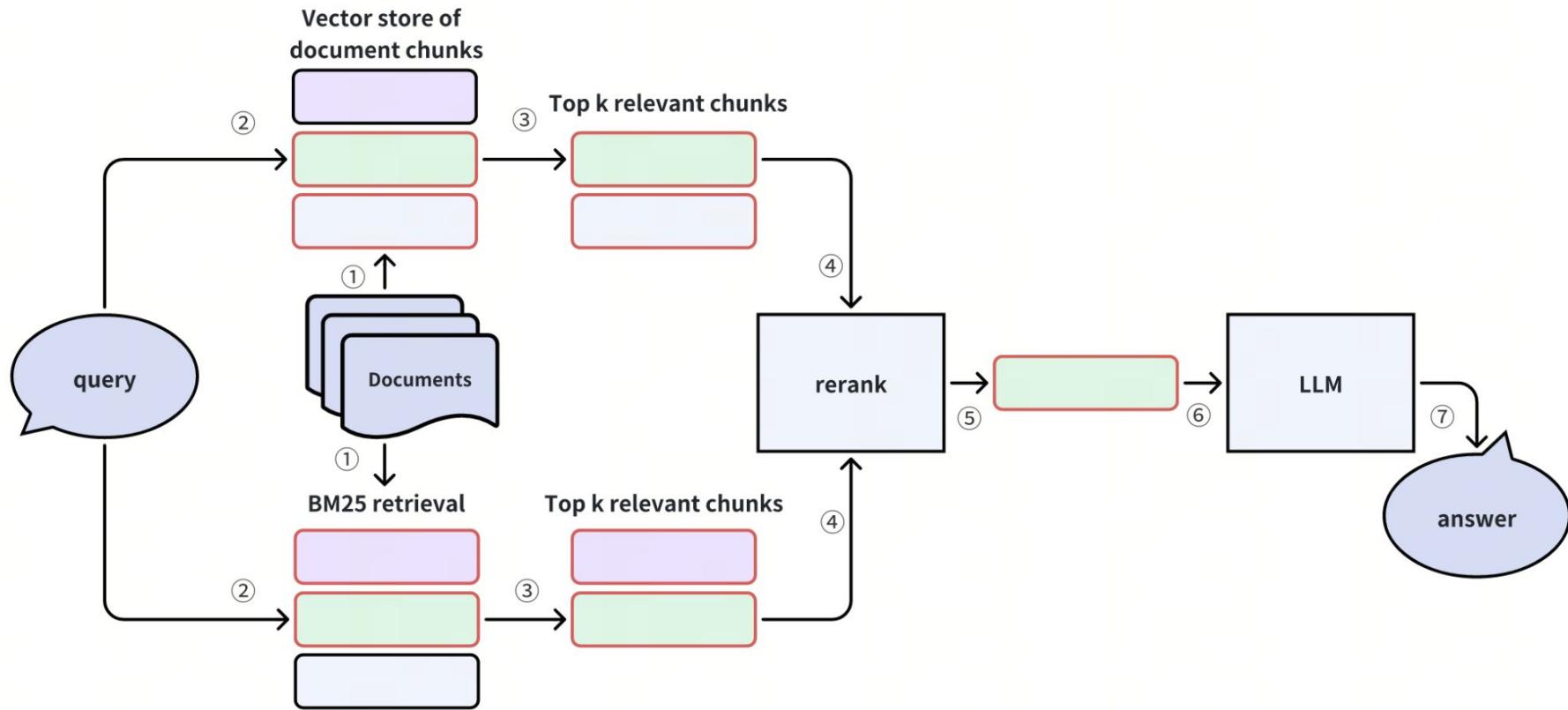
- **Is it possible to conduct multiple hybrid search operations at the same time?**

Yes, simultaneous execution of multiple hybrid search operations is supported.

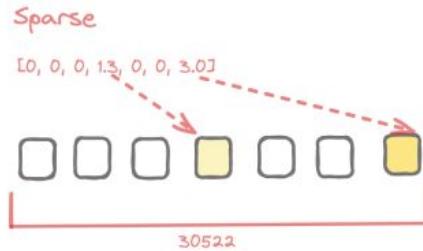
- **Can I use the same vector field in multiple AnnSearchRequest objects to perform hybrid searches?**

Technically, it is possible to use the same vector field in multiple AnnSearchRequest objects for hybrid searches. It is not necessary to have multiple vector fields for a hybrid search.

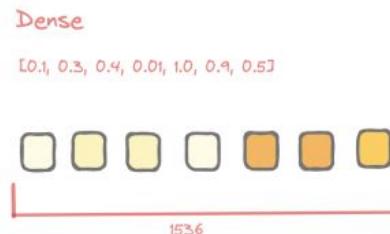
# Hybrid Retrieve & reranking



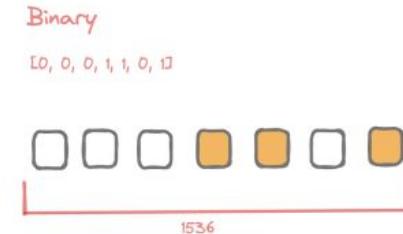
# Choosing Vector Embedding Types



Distance : IP  
Models: Splade, BGE-M3  
Index: Wand, Graph



Distance : IP, L2, Cosine  
Models: OpenAI, BGE, Cohere  
Index: Faiss, HNSW



Distance : Hamming,  
Superstructure, Jaccard, Tanimoto  
Models: Cohere, Meta ESM-2  
Index: Faiss

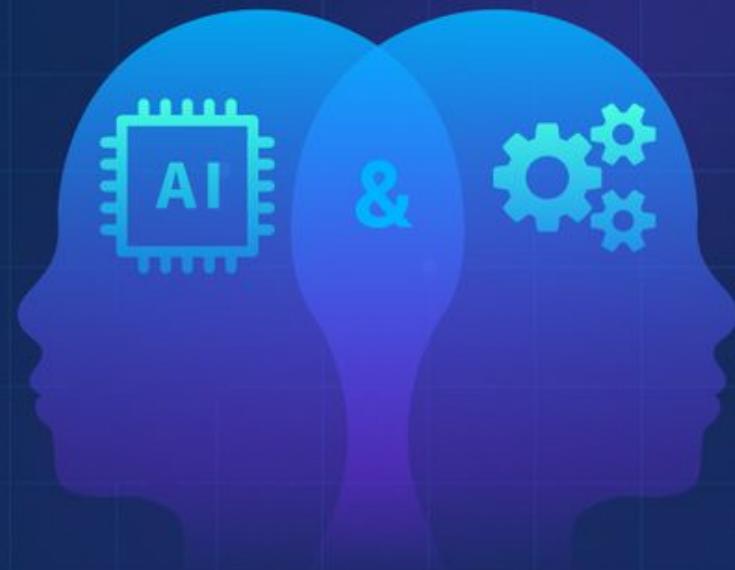
# RESOURCES



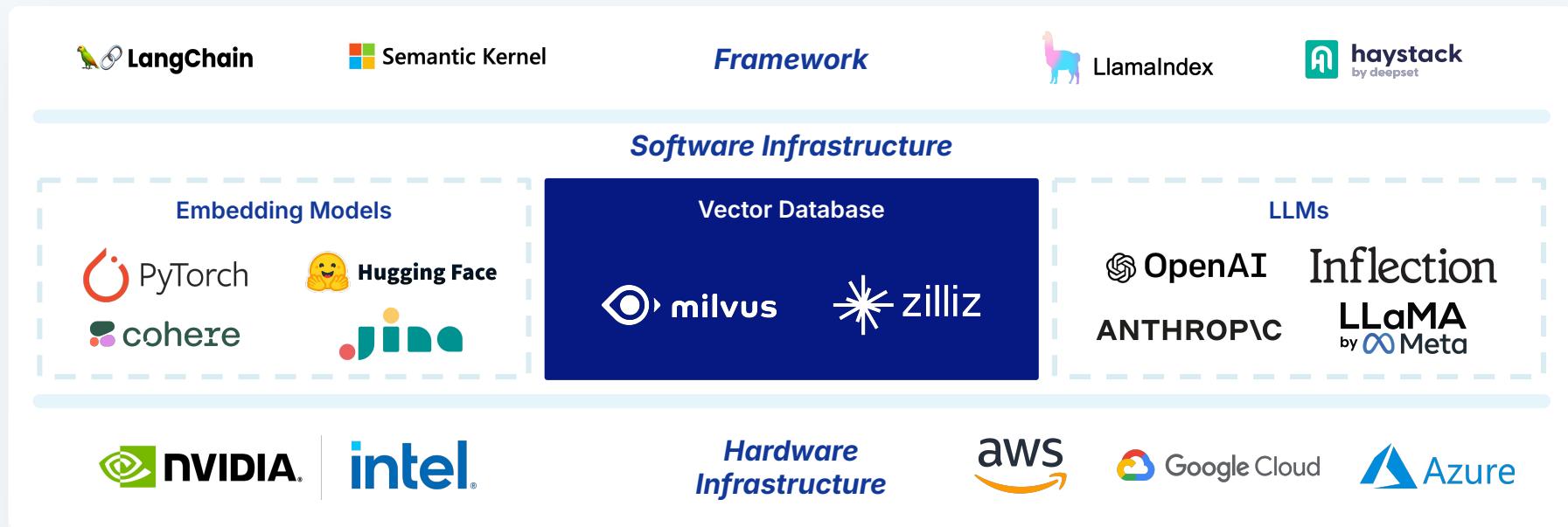


# The Critical Role of VectorDBs in Building Intelligent AI Agents

[Learn More](#)



# Well-connected in LLM infrastructure to enable RAG use cases



Raspberry Pi AI Kit - Hailo  
Edge AI



Milvus



<https://medium.com/@tspann/unstructured-data-processing-with-a-raspberry-pi-ai-kit-c959dd7fff47>



<https://github.com/tspannhw/ AIM-XavierEdgeAI>



<https://medium.com/@tspann/edgeai-edge-vector-database-6a9b5238bffb>

# Vector Database Resources

Give Milvus a Star!



<https://github.com/milvus-io/milvus>

Chat with me on Discord!



# Generative AI Resource Hub

Tutorials, Code Examples, and Best Practices for Developing and Deploying GenAI Applications.



Learn



Build

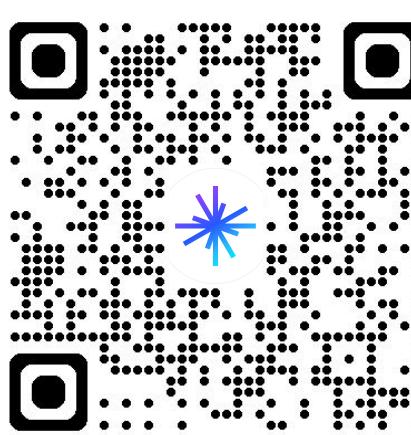


Explore

<https://zilliz.com/learn/generative-ai>



# Unstructured Data Meetup



<https://www.meetup.com/unstructured-data-meetup-new-york/>

This meetup is for people working in unstructured data. Speakers will come present about related topics such as vector databases, LLMs, and managing data at scale. The intended audience of this group includes roles like machine learning engineers, data scientists, data engineers, software engineers, and PMs.

This meetup was formerly Milvus Meetup, and is sponsored by [Zilliz](#) maintainers of [Milvus](#).

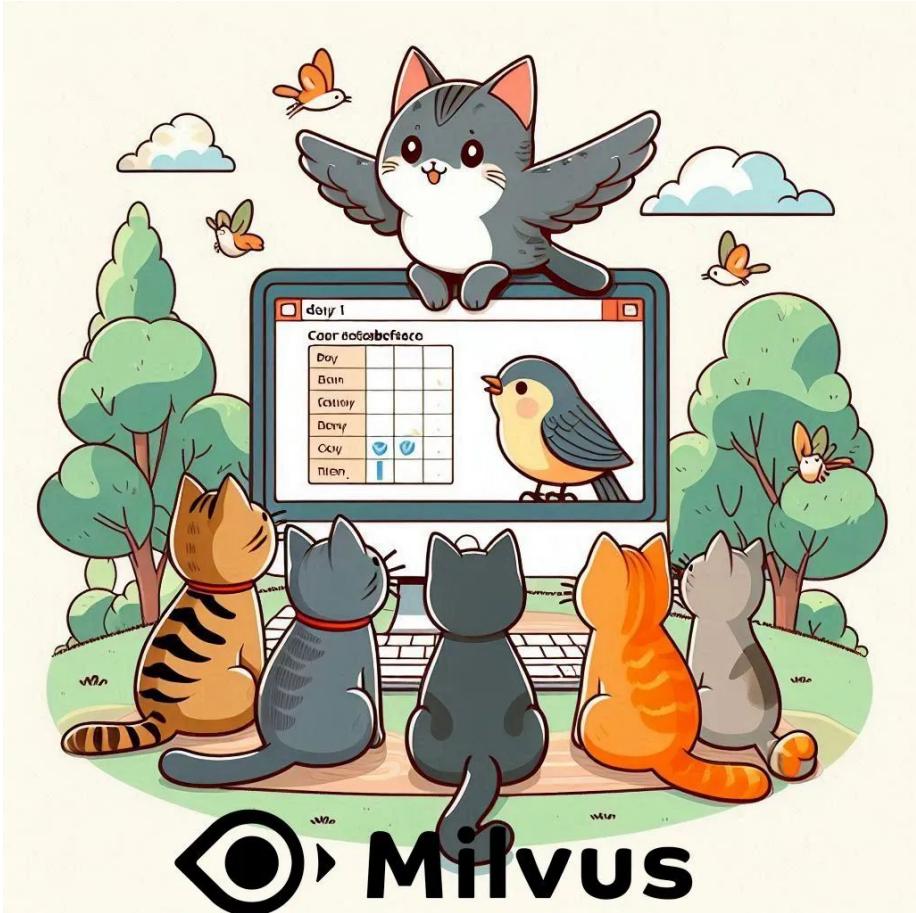




<https://medium.com/@tspann/unstructured-street-data-in-new-york-8d3cde0a1e5b>



<https://medium.com/@tspann/not-every-field-is-just-text-numbers-or-vectors-976231e90e4d>



 Milvus

<https://medium.com/@tspann/shining-some-light-on-the-new-milvus-lite-5a0565eb5dd9>



# Extracting Value from Unstructured Data

## Example

- A company has 100,000s+ pages of proprietary documentation to enable their staff to service customers.

## Problem

- Searching can be slow, inefficient, or lack context.

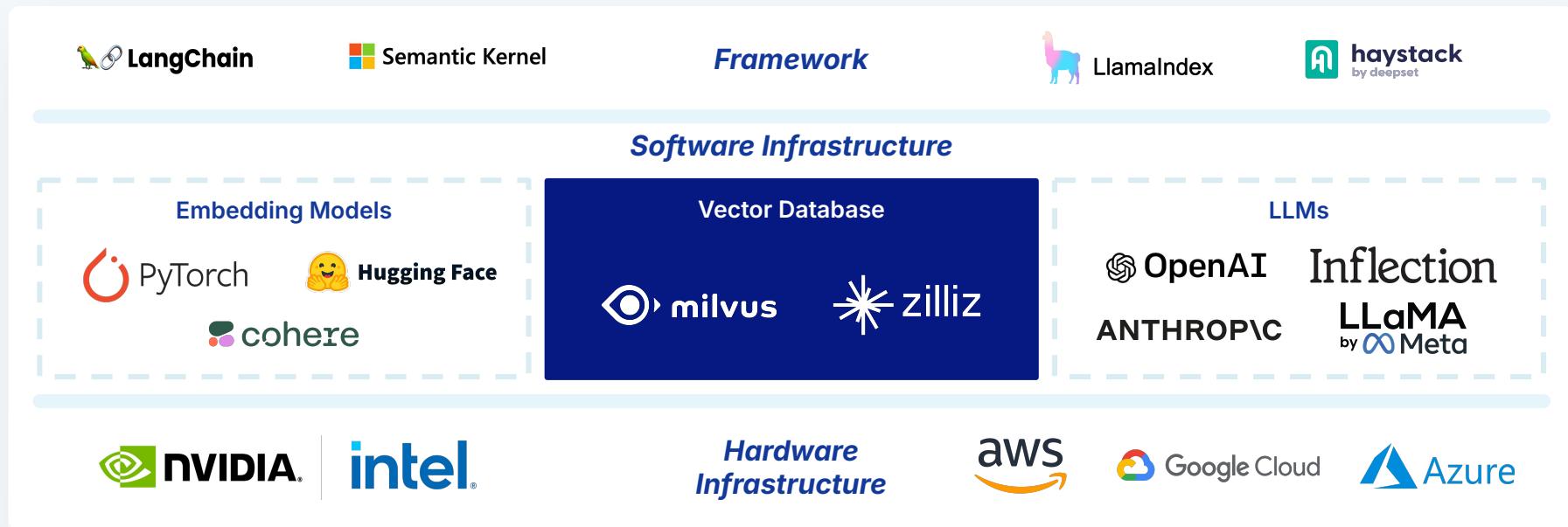
## Solution

- Create internal chatbot with ChatGPT and a vector database enriched with company documentation to provide direction and support to employees and customers.



<https://osschat.io/chat>

# Well-connected in LLM infrastructure to enable RAG use cases



# AIM Weekly by Tim Spann



<https://bit.ly/32dAJft>

<https://github.com/milvus-io/milvus>

This week in Milvus, Towhee, Attu, GPT Cache, Gen AI, LLM, Apache NiFi, Apache Flink, Apache Kafka, ML, AI, Apache Spark, Apache Iceberg, Python, Java, Vector DB and Open Source friends.

# Thank you!

---



[milvus.io](https://milvus.io)



[github.com/milvus-io/](https://github.com/milvus-io/)



[@milvusio](https://twitter.com/milvusio)

# Connect with me!

---



[@paasDev](https://twitter.com/paasDev)



[/in/timothyspann](https://www.linkedin.com/in/timothyspann/)



# These capabilities make us the perfect partner to uplifting your initiatives on vector search and AI/ML

## Milvus



### Operational burden & resources

**Tedious** configuration. **Manual & resource-intensive day-to-day operations** to deploy, manage, and scale clusters.



### Security & Availability

**Custom-built** security tools & integrations that **creates tech & operational debt**. **Resource-intensive** failover design & operations.



### Data processing & connectivity

**Siloed** solutions and **custom integrations, escalating complexity & costs** to manage and maintain platform as it scales

## Zilliz



**Instant** cluster provisioning and scaling. **Automated** capacity mgmt and upgrades. **Improved performance** for compute intensive use cases.



**Battle-tested and enterprise-grade** security tools and compliance ready out-of-the-box. **Highly available** and consistent access to data across all of your environments.



**Well-integrated** into AI and data ecosystems. **Out-of-box pipeline builders** transform unstructured data into searchable vectors efficiently.

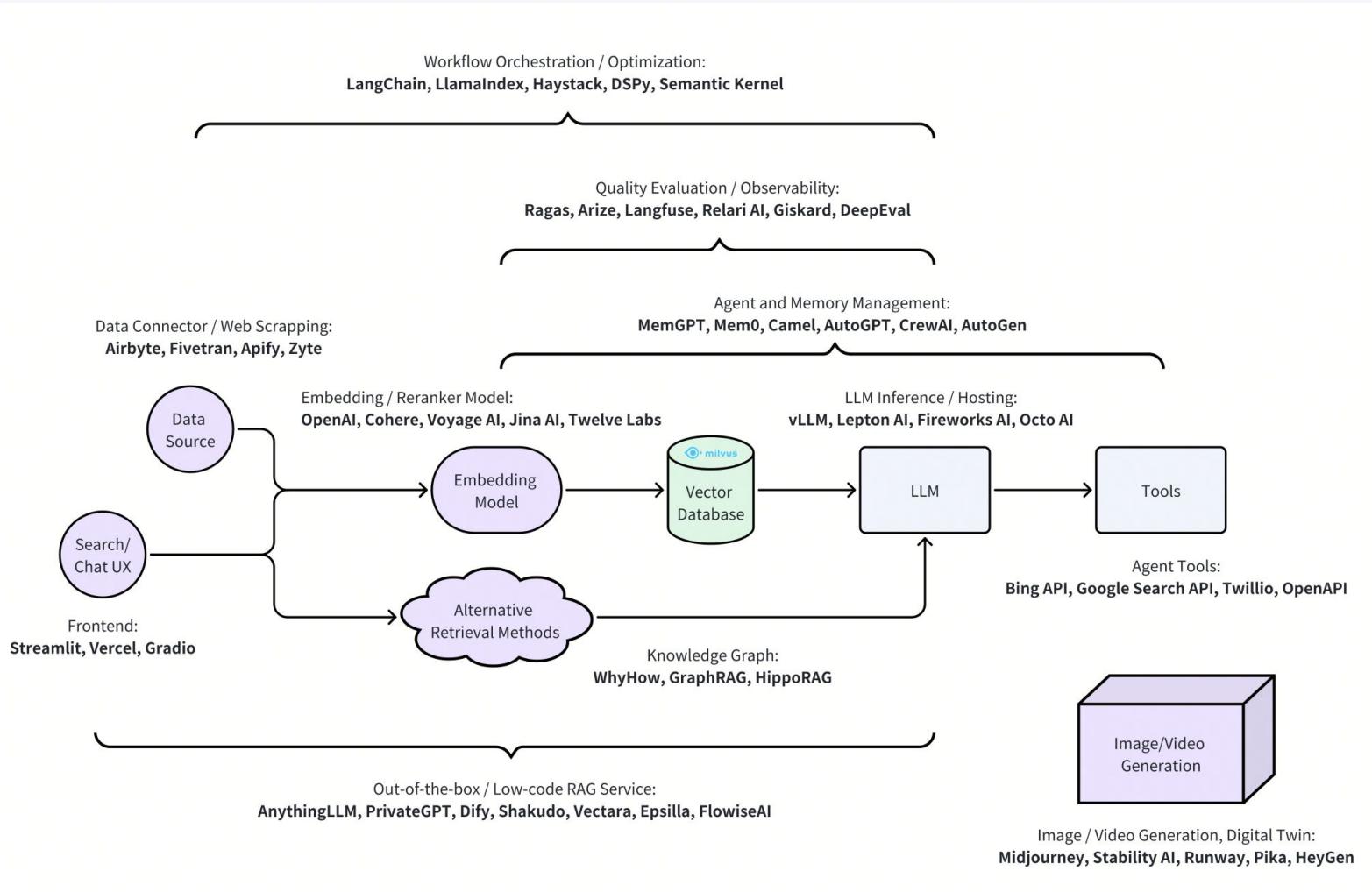
# Cool AI News

OpenAI and Thrive Capital recently backed **Chai Discovery**, a six-month-old AI biology startup founded by ex-OpenAI and Meta researchers that raised \$30 million to develop AI models for drug discovery.

## The details:

- Chai's AI model, Chai-1, predicts biochemical molecule structures, potentially speeding up drug development.
- The company claims Chai-1 outperforms Google DeepMind's AlphaFold on certain benchmarks.
- Chai-1 can work with proteins, small molecules, DNA, and RNA, making it versatile for various applications.
- Chai is making its first model free and open-source for non-commercial use.

<https://github.com/chaidiscovery/chai-lab>



# Number of AI-Created Images\*

EVERYPIXEL

DALL-E 2

**916 million**

Models based on Stable Diffusion

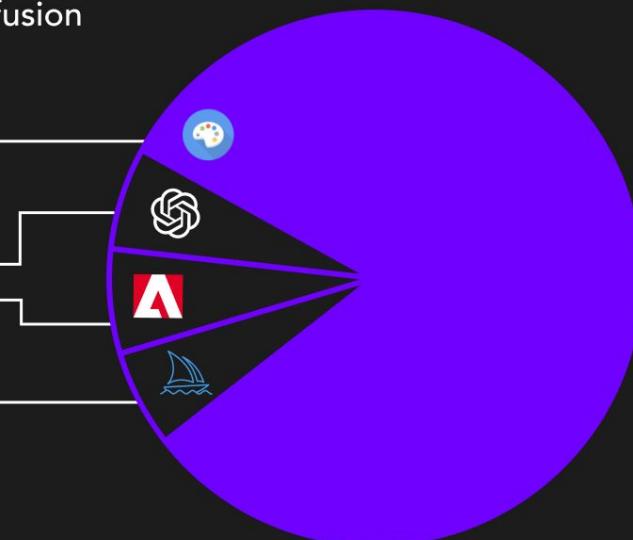
**12.590 billion**

Adobe Firefly

**1 billion**

Midjourney

**964 million**



Sources: Adobe;  
our estimates, based on Photutorial, OpenAI, Cividai

\*As of August 2023



Join us at our next meetup!  
[meetup.com/unstructured-data-meetup-new-york/](https://meetup.com/unstructured-data-meetup-new-york/)

THANK YOU

