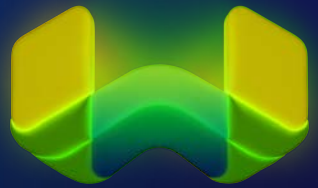


Weaviate

Introduction to Vector Databases

Conf42 Machine Learning 2023

Zain Hasan

Zain Hasan

Senior Developer Advocate

Weaviate

From keyword search to semantic search

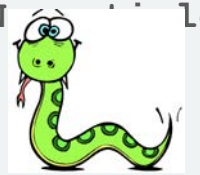
```
{"title": "How to build a REST API?"}

{"title": "Programming languages for data scientists"}
```

semantically similar, using ML

"Python"

N les found ...



Is python a poisonous snake?

Traditional, keyword-based search

"Python"

Programming languages for data scientists

ML-based (semantic) search

It is difficult to process, understand and search through unstructured data

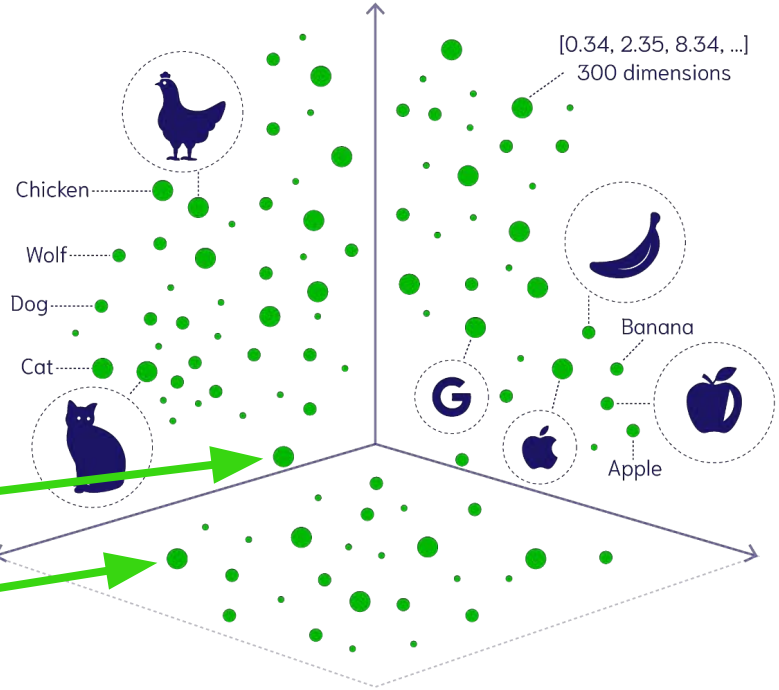
It is difficult to process, understand and search through unstructured data in a scalable and secure way

It is difficult to process, understand and search through unstructured data

1. We use Machine Learning to understand the context of unstructured data.

Machine Learning models

...

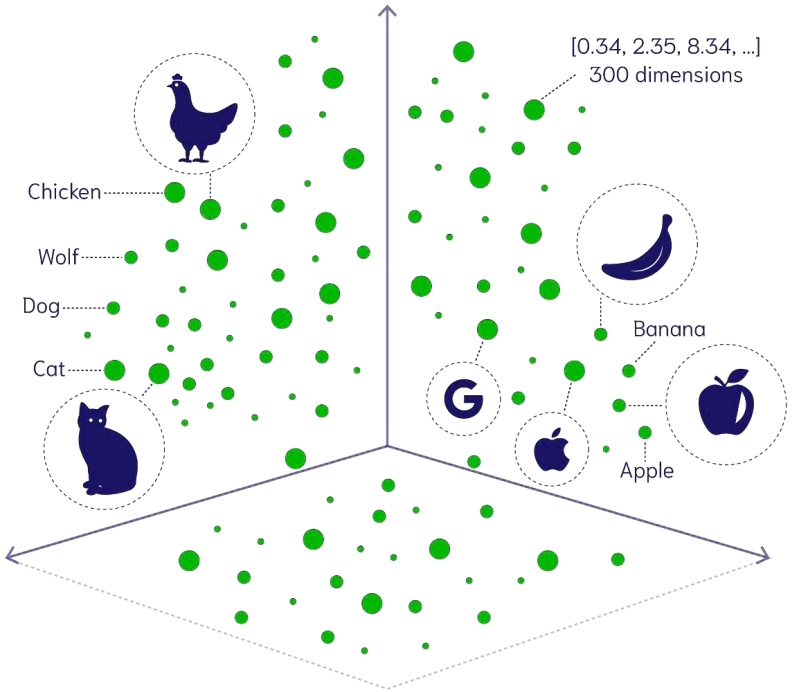


[0.23, 0.45, ..., 0.84, 0.23],

[0.63, 0.32, ..., 0.34, 0.92]

... create vector representations

Vector representations



Millions of data entries & vectors & many **queries** per second

How to search through it on large scale?

Store & Retrieve large amounts of vectors & non-vector data

Scale ML models to work reliably in production

Real-time full CRUD support

It is difficult to use machine learning models in a scalable and secure way

2. Vector Databases are the solution to these challenges

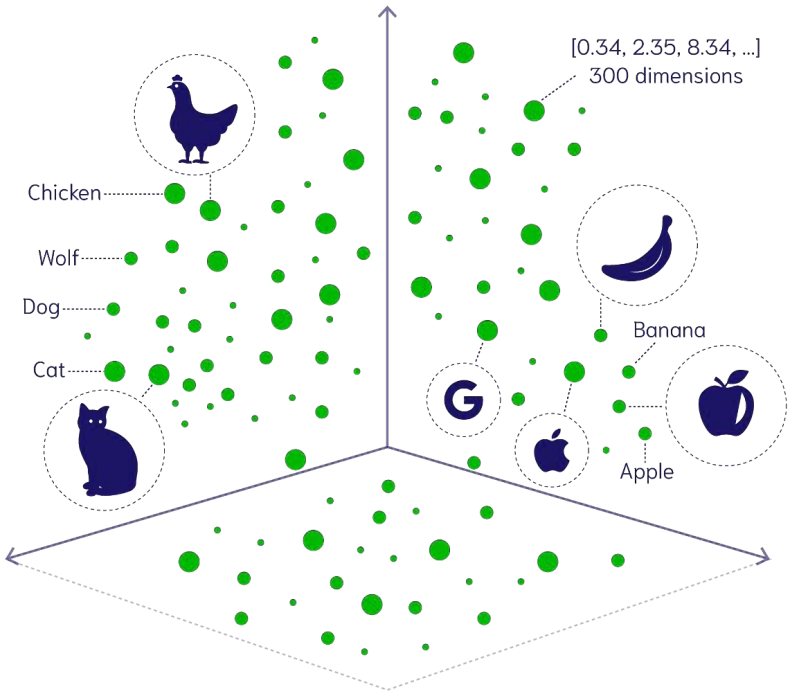
A vector database indexes and stores data objects *and* its vector embeddings enabling efficient similarity search combined with structured filtering in a scalable way.

What is a vector database?

Stores **data objects** & **vector embeddings**

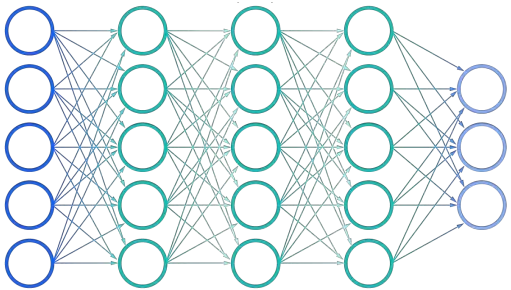
Similarity Search combined with **Structured Filtertering**

Real-time full **CRUD** support & **ANN** search algorithms



Vector databases

1. Vectorize and index data, using ML



ResNet-50
PRETRAINED MODEL

co:here

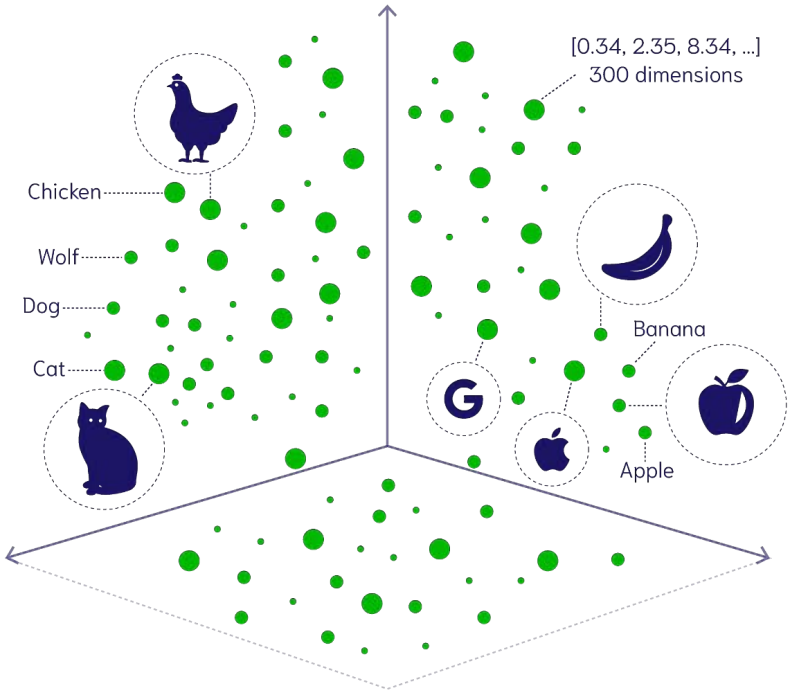
HUGGING FACE

OpenAI

Vector databases

1. Vectorize and index data, using ML

Example search

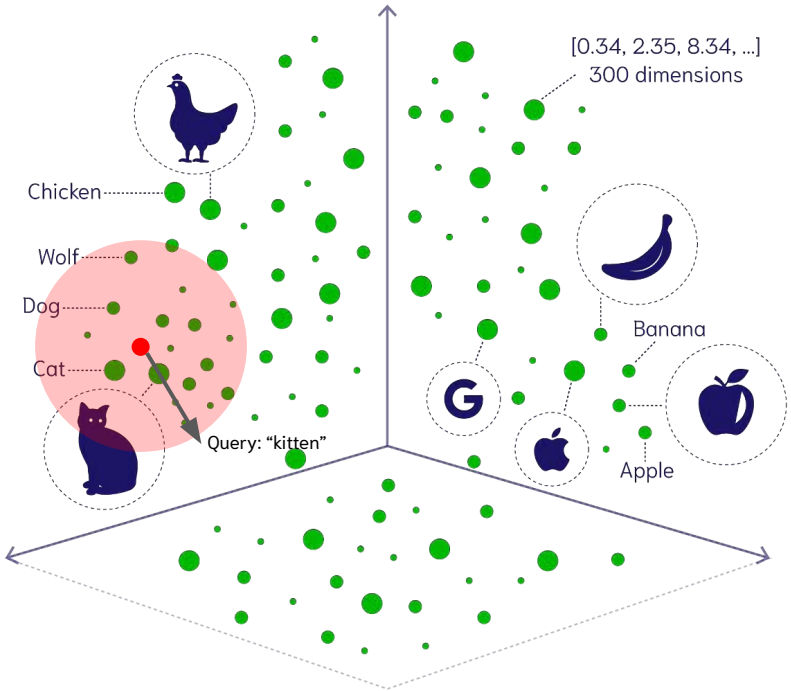


Vector databases

1. Vectorize and index data, using ML

2. Vectorize search query, using ML

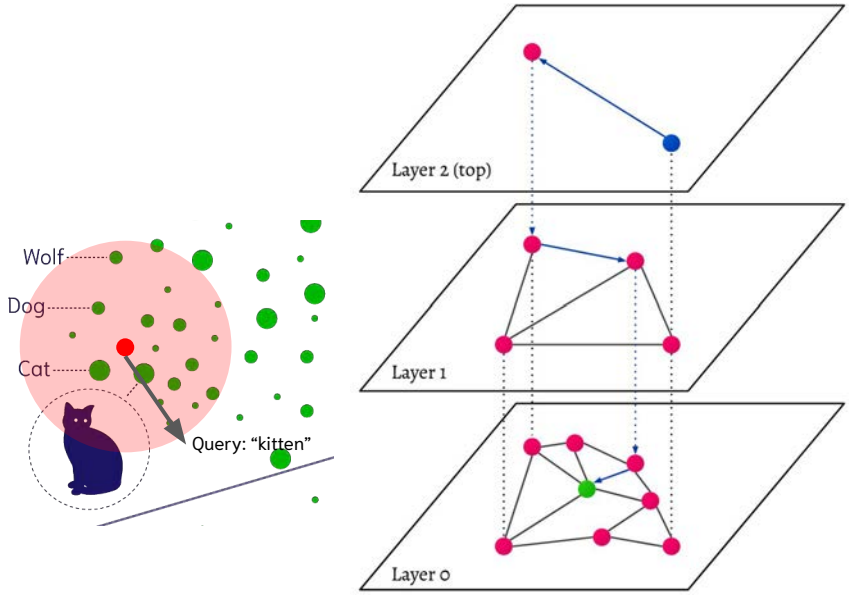
Example search



Vector databases

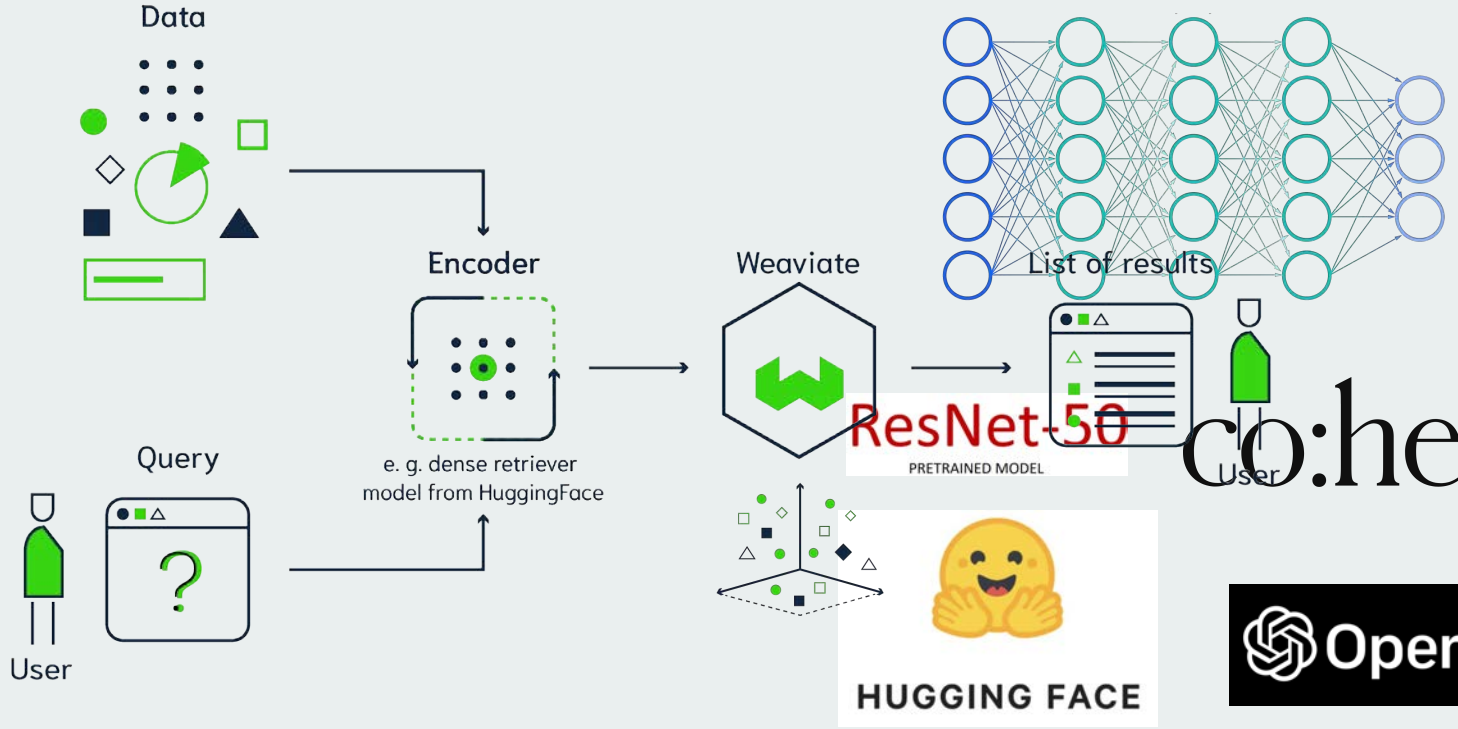
1. Vectorize and index data, using ML
2. Vectorize search query, using ML
3. Retrieve ANN results (using HNSW)

Example search

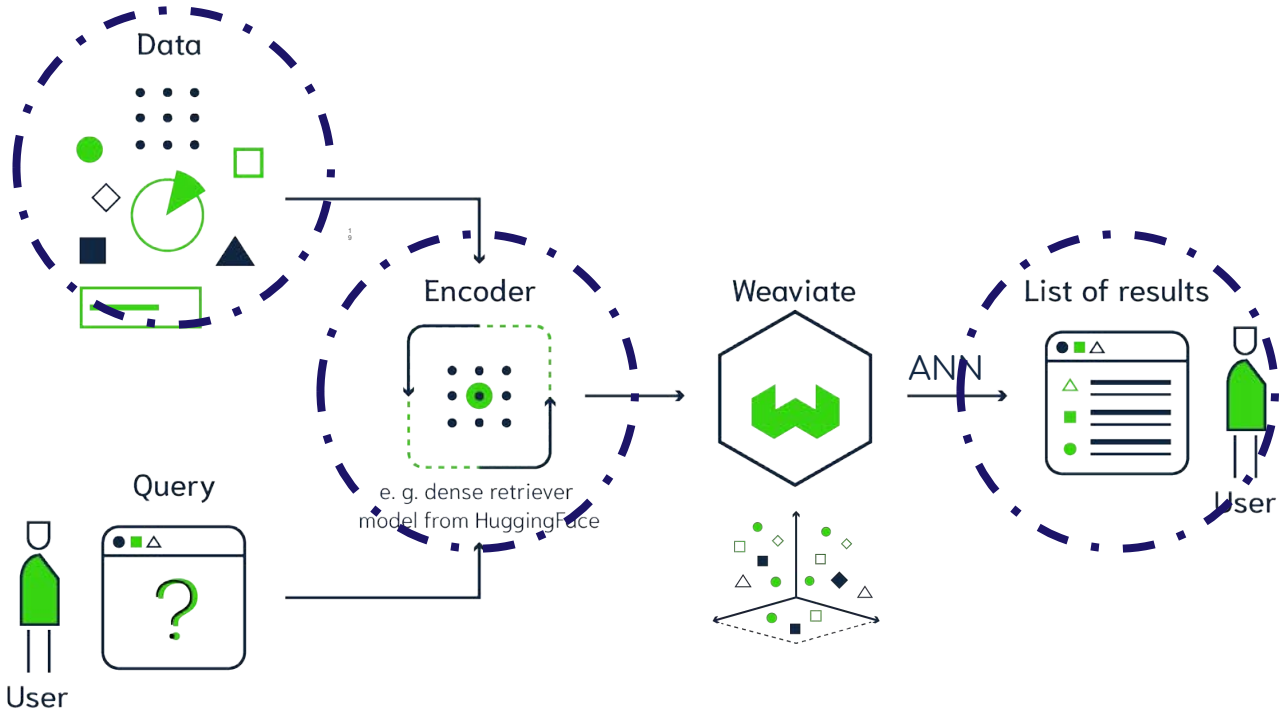


Weaviate: An open-source **vector database** that **understands** your data

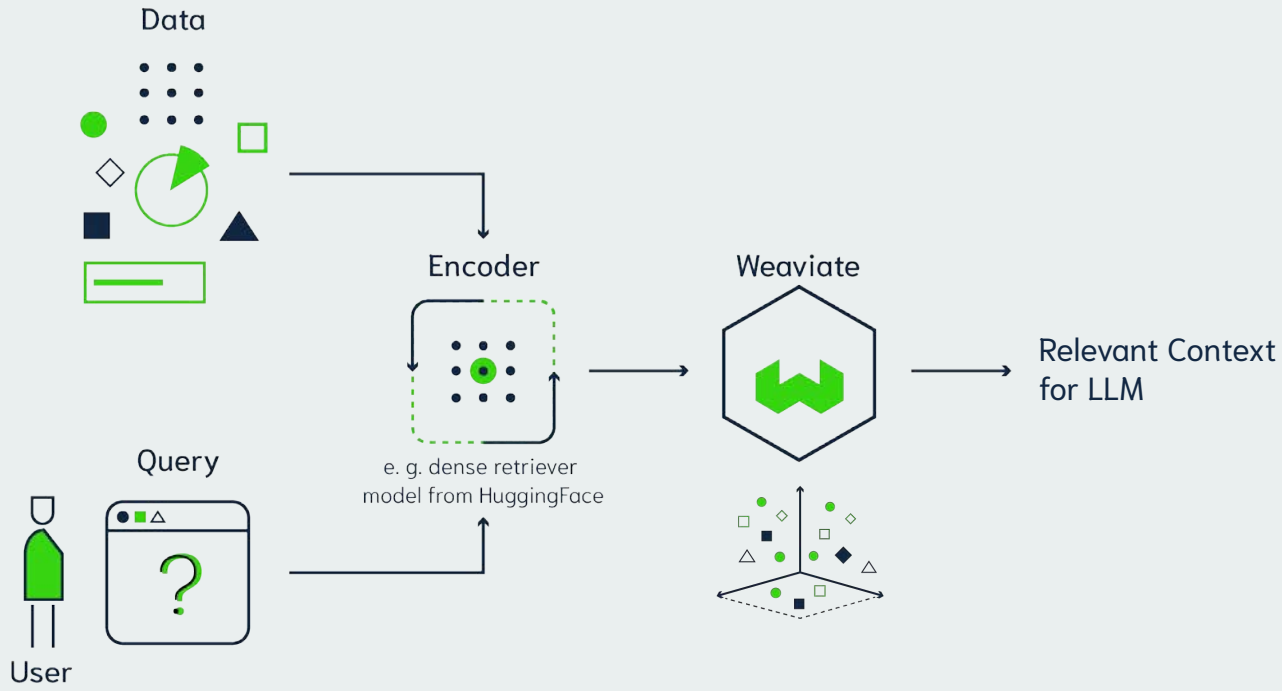
A vector search pipeline



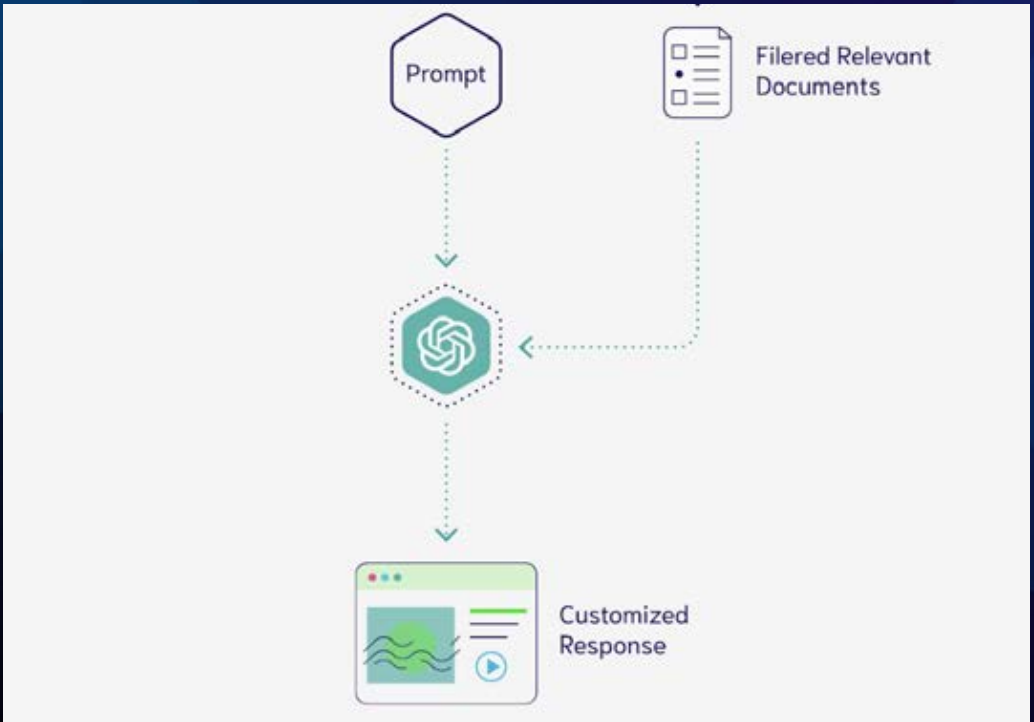
Weaviate is modular: flexible search pipelines



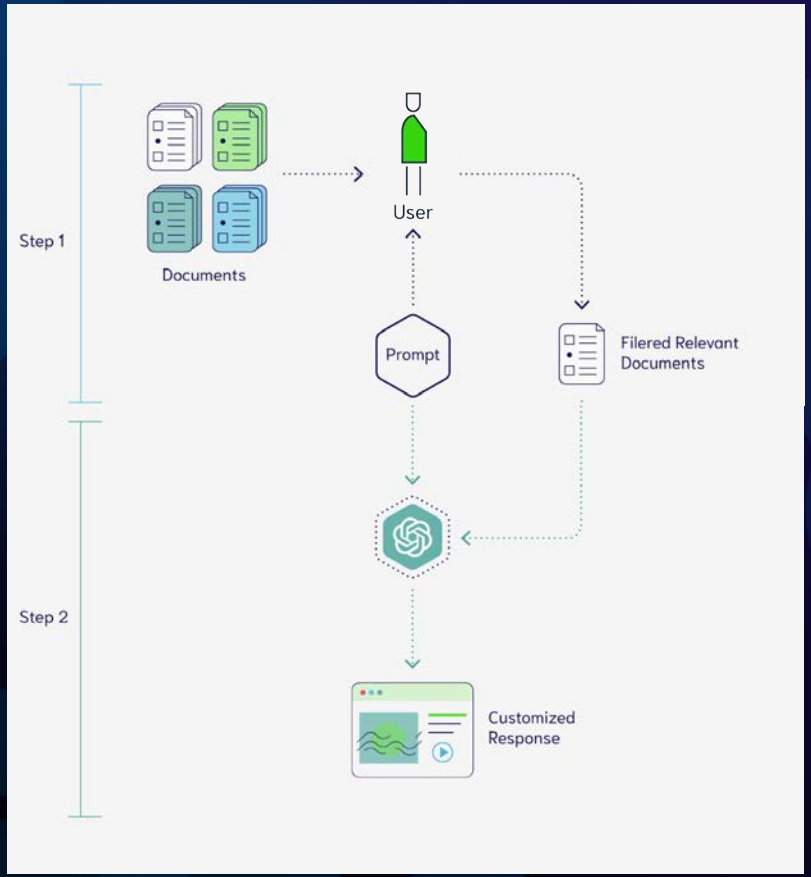
Using vector search to provide LLMs context



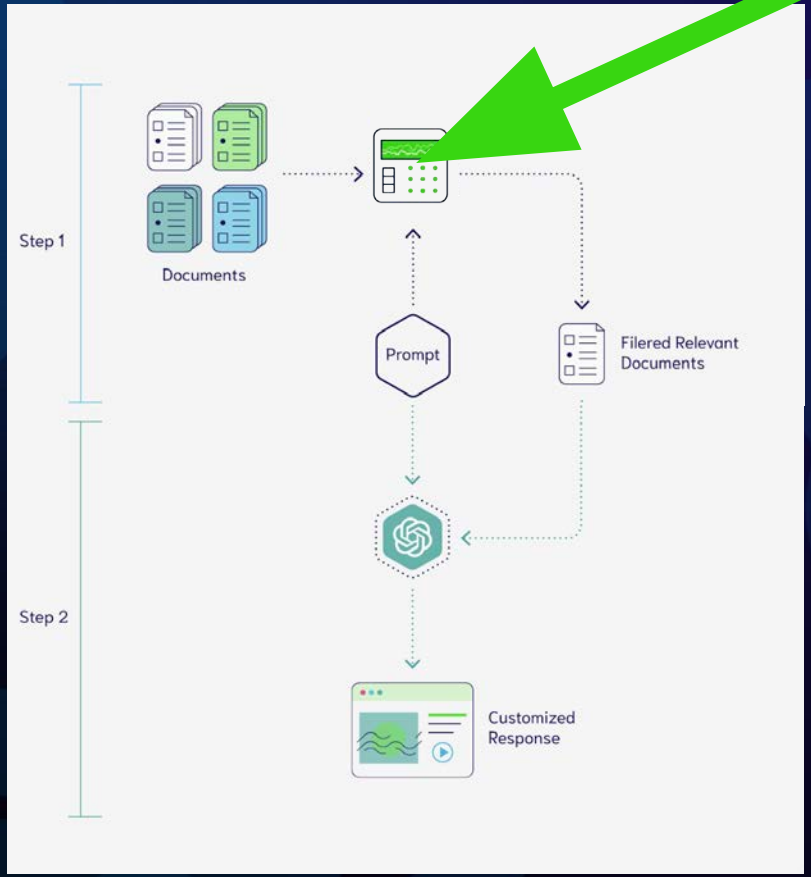
Say to ChatGPT: Answer my question, here's everything relevant you need to know.



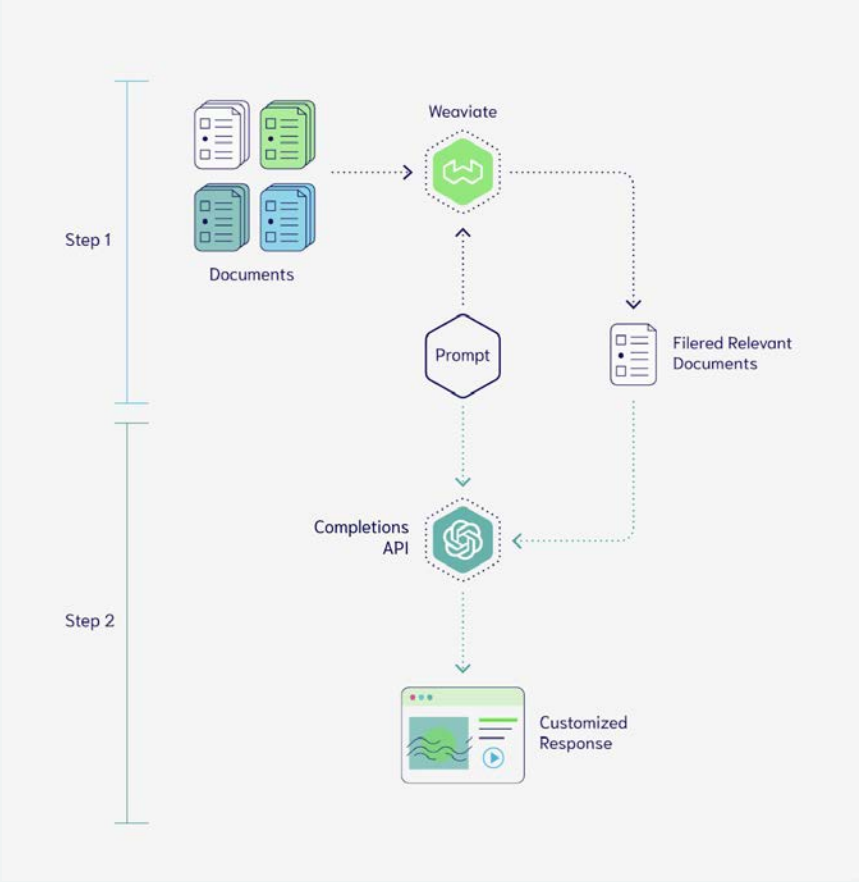
Manually filter documents to identify relevant context



To scale this approach we need a Vector Database!



Customizing Large Language Models using Vector Databases



Demo!

Connect with me!

Zain Hasan

@zainhasan6

linkedin.com/in/zainhas/

Thank you!

weaviate.io

weaviate/weaviate

@weaviate_io