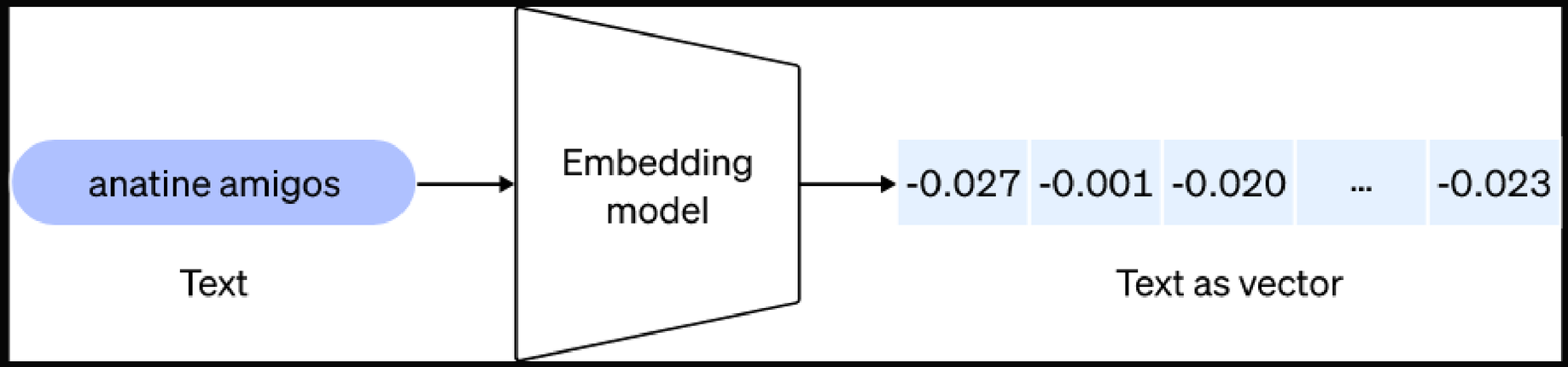
Vector Databases

Embeddings

Embeddings are the core of building LLMs applications.

Text embeddings are numeric representations of text and are used in NLP and ML tasks.



LangChain

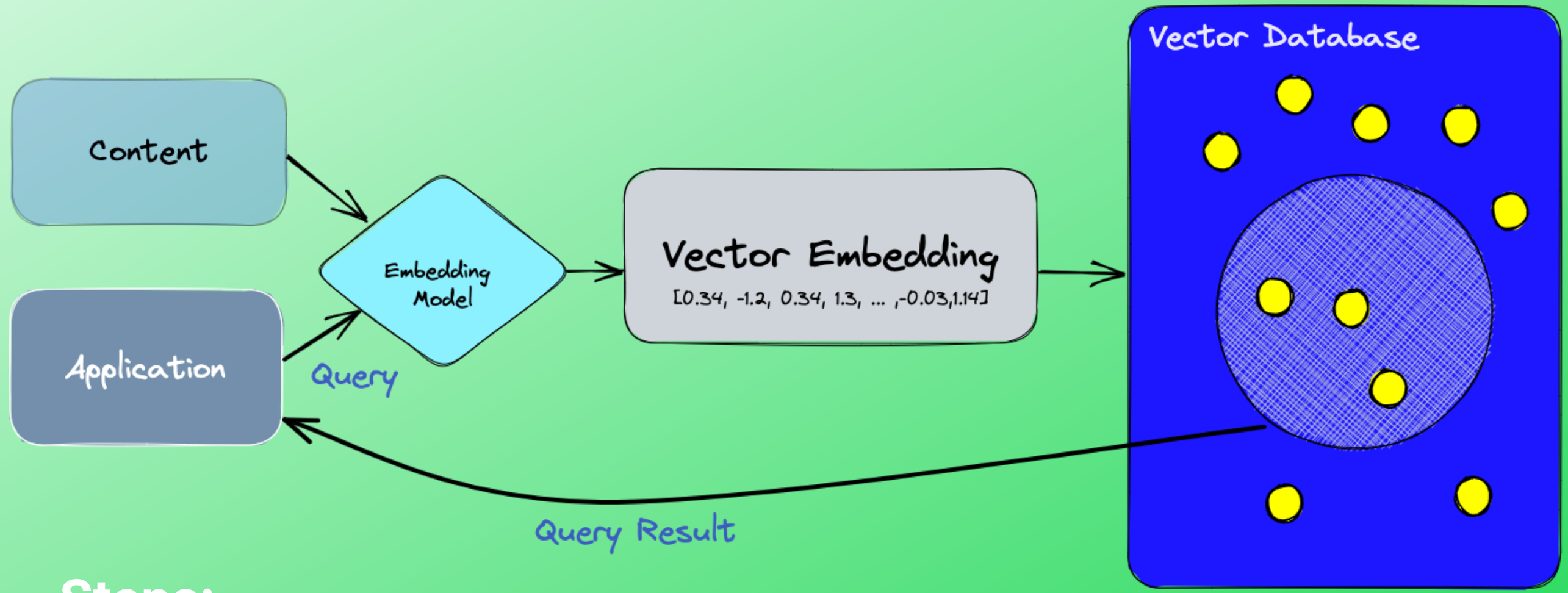


Image from pinecone.io

Steps:

1. **Embedding**
2. **Indexing**
3. **Querying**

Challenges

One of the biggest challenges of AI Applications is efficient data processing.

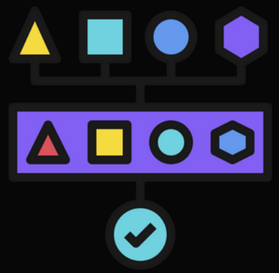
Many of the latest AI applications rely on vector embeddings. Chatbots, question-answering systems, and machine translation rely on vector embeddings.

Vector Databases

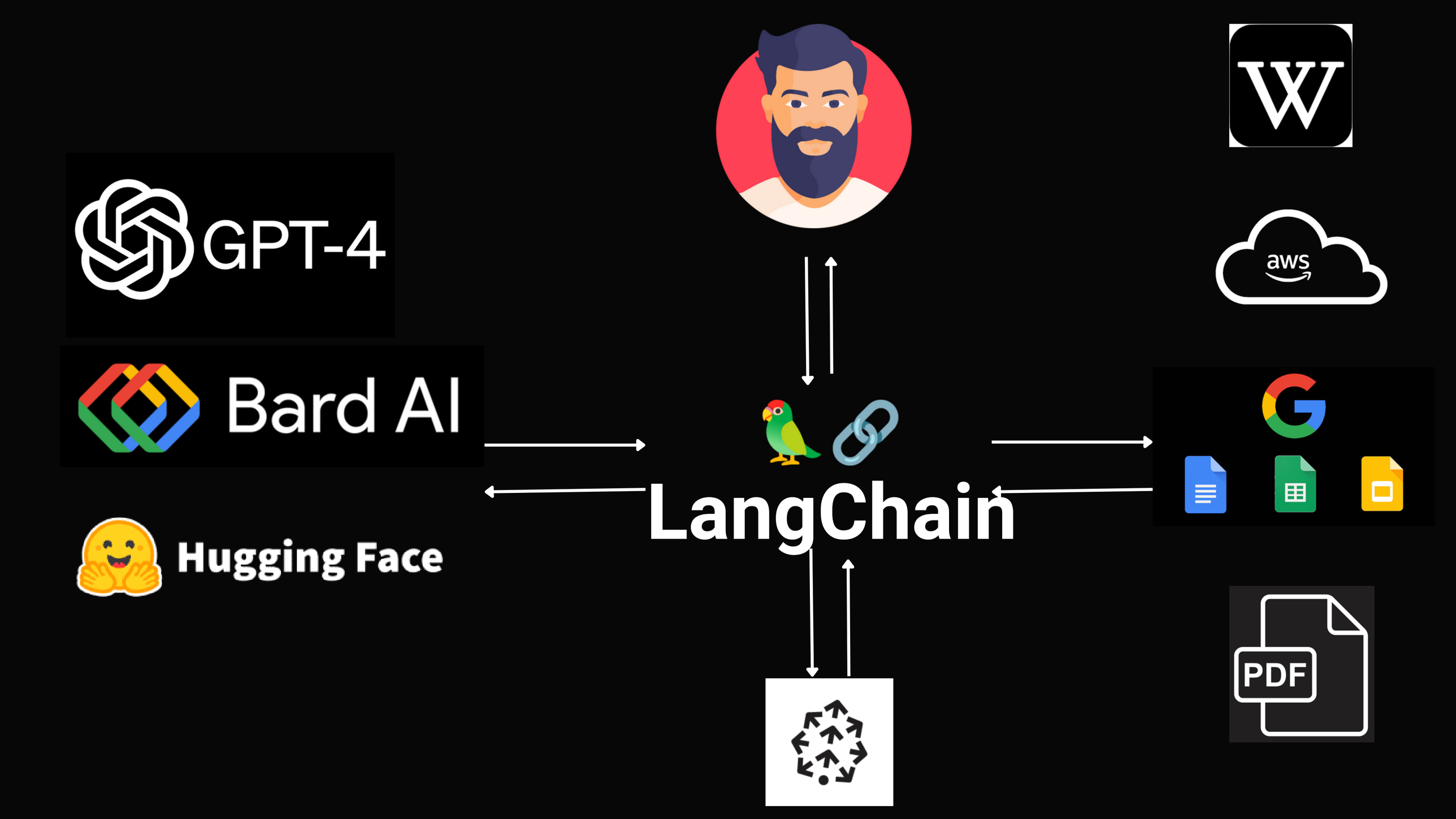
Vector databases are a new type of database, designed to store and query unstructured data.

Unstructured data is data that does not have a fixed schema, such as text, images, and audio.

Pipeline for Vector Databases



Vector databases use a combination of different optimized algorithms that all participate in Approximate Nearest Neighbor (ANN) search.


GPT-4
Bard AI
Hugging Face
LangChain
aws
PDF

Namespaces

- Pinecone allows you to partition the vectors in an index into namespaces.
- Queries and other operations are then limited to one namespace.

Namespaces

- Every **index** is made up of one or more namespaces.
- Every **vector** exists in exactly one namespace.
- **Namespaces** are uniquely identified by a namespace name.