



# Vector Embeddings

Applicable AI

# AI MODELS – GENERAL PURPOSE WORKERS

---

Large language models, or the AI models that power AI applications like ChatGPT, are built to be general purpose. They have a wide array of general knowledge but will never know the details of your organization or enterprise data. To deploy use cases of AI that are applicable to your business, workflows must be set up to house enterprise data in a manner that it can be retrieved for a given use case and served to a large language model when end business users / processes interact with them.



# HOW DO WE ACHIEVE THIS

---

We can set up applied language models by utilizing what is called a **vector embedding**.

## In a Nutshell

A vector embedding is simply an index to relevant information. Under the hood they are a long list of decimal points that represent the meaning behind a given chunk of information. The long list of decimal points serve as 'coordinates' to the relatedness or semantic meaning behind its actual context.

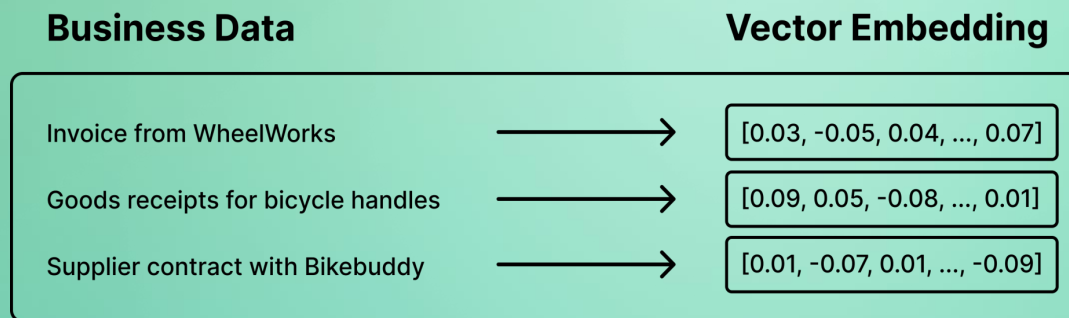
## Vector Embeddings Applied

In a business setting, we can take a large corpus of data required for a given AI use case, create vector embeddings out of the data, and now as business users or processes interact with an AI model the most relevant business data for the request can be served to an AI model before it ever generates an output. This is how AI models can be transformed from **general purpose workers** -> **applicable business assets that drive real value**.

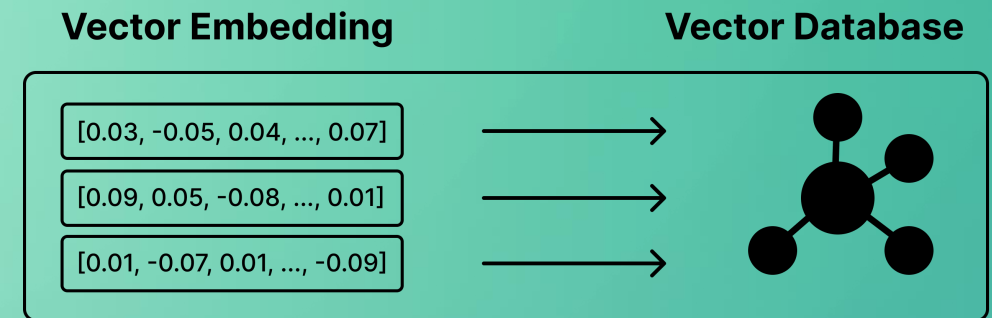
# GETTING STARTED

There are two key components of vector embeddings:

## Embedding Models

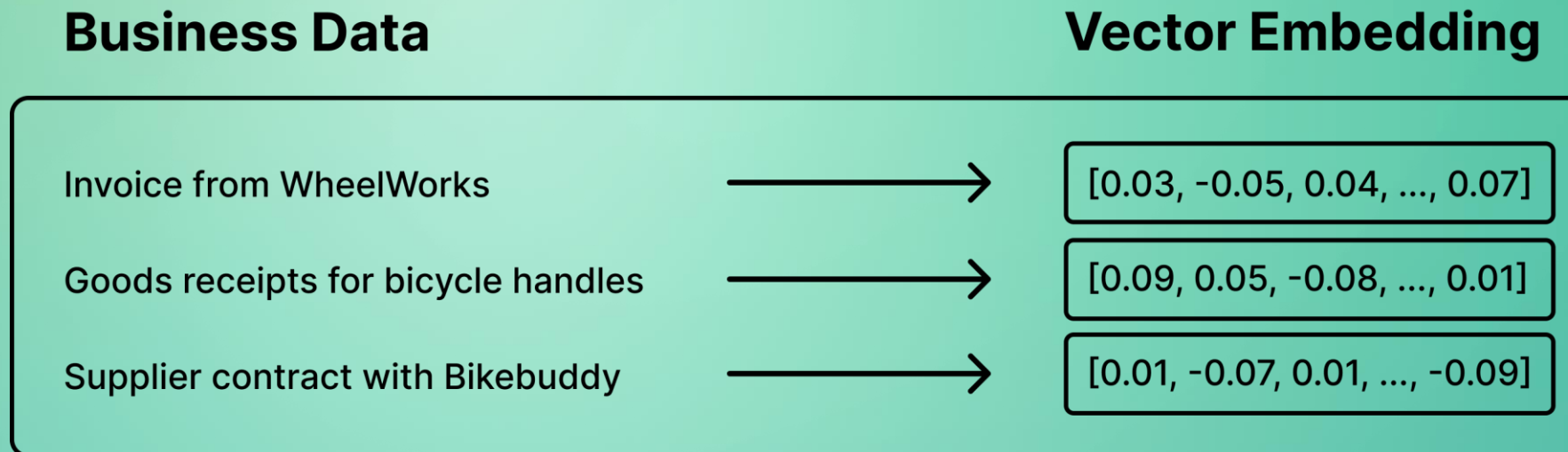


## Vector Databases



# Embedding Models

An **embedding model** is the AI model that transforms a body of data sitting in its original, natural language form to a vector embedding.



# Vector Databases

Once business data has been transformed into vector embeddings, these embeddings can be stored in a **vector database**. Vector databases serve as the infrastructure for storing and managing a large number of embeddings. They operate in a high-dimensional space, allowing embeddings to be stored as relevancy indexes to a large corpus of business data. After storing the embeddings for a given use case, the database can efficiently retrieve the most relevant data in response to a query, thereby supporting AI models in generating accurate and contextually appropriate responses.

## Vector Embedding

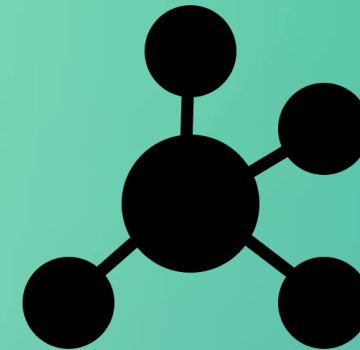
[0.03, -0.05, 0.04, ..., 0.07]

[0.09, 0.05, -0.08, ..., 0.01]

[0.01, -0.07, 0.01, ..., -0.09]



## Vector Database



# MOVING TO PRODUCTION

Vector databases alone are just touching the foundations of a production ready AI business application and hurdles come with every unique use case. Just to name a few:

- Lengthy documents
- Structured vs unstructured data
- Losing context from chunk to chunk
- Similar or related content in the same corpus
- Multimodal content
- Complex files



# WHY IT MATTERS



## ***General purpose AI → Tangible and Applicable Business Value***

General purpose assistant like ChatGPT and Copilot can realize some productivity gains... help wordsmith an email .. provide some generic code... provide some summary bullets...

But until AI assistants are integrated with enterprise data and workflows that drive your organization forward, AI remains a general-purpose assistant rather than a strategic business asset.



# REACH OUT

---

## If you want to:

- Move AI use cases beyond general-purpose assistants
- Improve your applications already utilizing vector embeddings
- Accelerate through the hurdles that come with production-ready application
- Minimize overhead and maintenance of unique applications

If you want continuous insights into vector embeddings, follow our page or reach out:

**LinkedIn** | **Contact us**