For **unstructured** data, **Vector embeddings** is a form of automatic feature engineering.

Let's break it down!

$(3,2)$ | vector | is a point

whaa....?
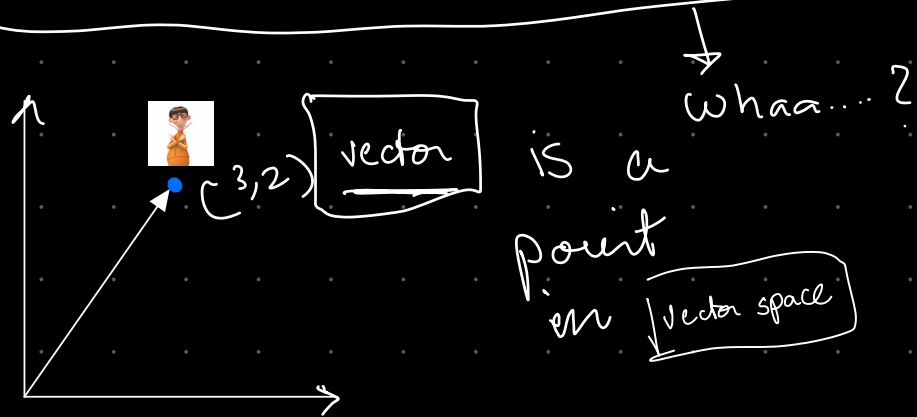
in | Vector space |

Think of it as an array $[0, -2, \ldots 4]$

why? Vectors in a vector space help understand relationships in data, eg. closer vectors are related, farther are not,

(We will understand more....)

How? To convert data to vectors we use **Embeddings model**

↓

pre-trained models which have a defined vector-space. Give them my data and it will just add to its space.

Data → (Embedding Model) → Vector Embeddings

Again, why?

Once relationships are established, your unstructured data has

MEANING

Now you can search not only for words but because it's in vetor space it looks for distance in the vector space.

So, if you embed your data with Open AI's embeddings, chat GPT can now understand what your document means)   * It's like talking in GPT's language!
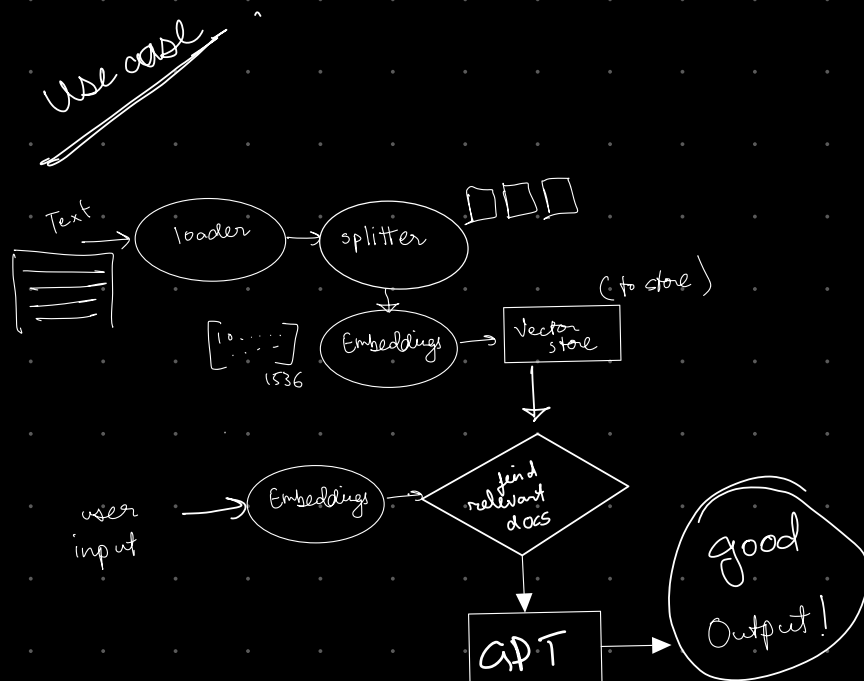
But still

Why?
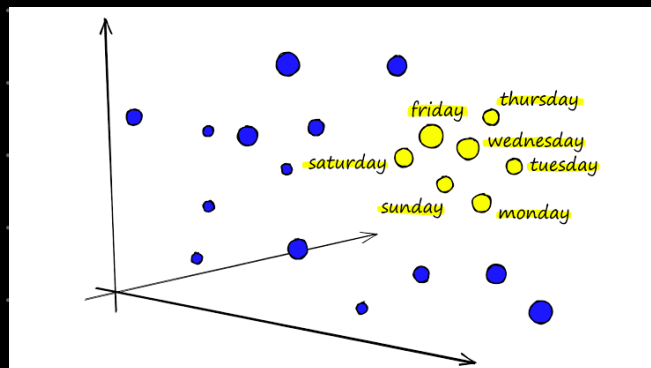
GPT is sometimes stupid!
① That's why.

You can train it on your
own data and tell it do
stuff from your own data!

② Lost effectiveness:

"Open" AI is expensive, you can
create your own vector store and
do a semantic search find relevant
docs to a question and send only
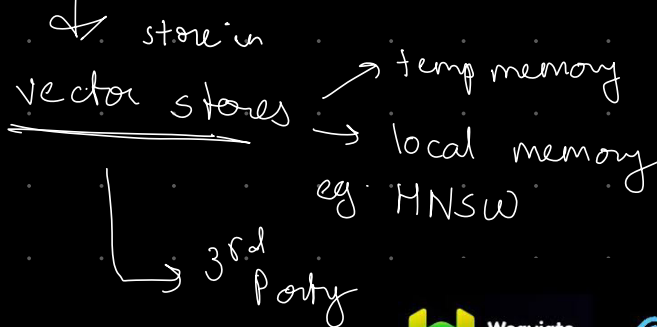those to the model to answer
as well!

Use case

Text → loader → splitter → ☐☐☐

[0...] → Embeddings → Vector store → (to store)
1536

user input → Embeddings → ⟨find relevant docs⟩ → GPT → (good Output!)

Good Theory ? How to make IRL!

for GPT,

OpenAI Embedding model

```
import openai
response = openai.Embedding.create(
    input="porcine pals say",
    model="text-embedding-ada-002"
)

print(response)
{
    "data": [
        {
            "embedding": [
                -0.0108,
                -0.0107,
                0.0323,
                ...
                -0.0114
            ],
            "index": 0,
            "object": "embedding"
        }
    ],
    "model": "text-embedding-ada-002",
    "object": "list"
}
```

OpenAI's embeddings model

store in vector stores → temp memory
→ local memory eg. HNSW

↳ 3rd Party

Weaviate   milvus

Pinecone

openSearch,
                  , prisma ,
supabase ....?

References:

https://platform.openai.com/docs/guides/embeddings/what-are-embeddings

https://www.pinecone.io/learn/vector-embeddings/

https://js.langchain.com/docs/modules/indexes/vector_stores/

JC