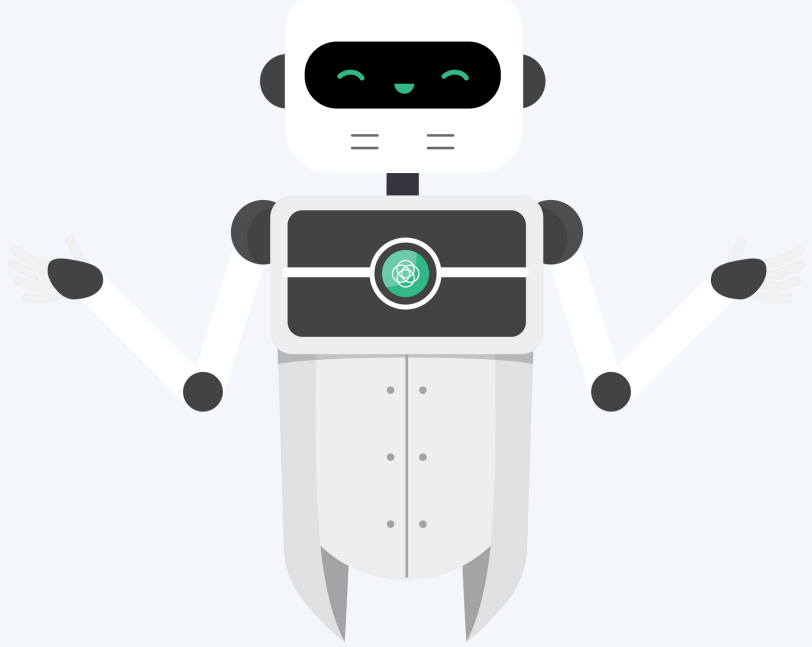
Building LLM Chatbots:
A Step-by-Step Guide



Swipe >>>

datasciencedojo
data science for everyone

1

Organizing the Knowledge Base

- Break down your knowledge into bite-sized chunks. Each chunk should be a clear, concise answer to a specific question.

- Gather data from a variety of sources, such as company documents, articles, and FAQs.

- Define clear boundaries for each chunk to ensure relevant information is easily accessible.

Swipe >>>

datasciencedojo
data science for everyone

2

Text into Vectors

- Use an embedding model to convert each chunk of text into a numerical representation. This allows the computer to understand the meaning of the text.

- Choose an embedding model that is trained on a large corpus of text. This ensures the vectors capture the nuances of language.

- Keep the vectors a fixed length for easy storage and querying.

Swipe >>>

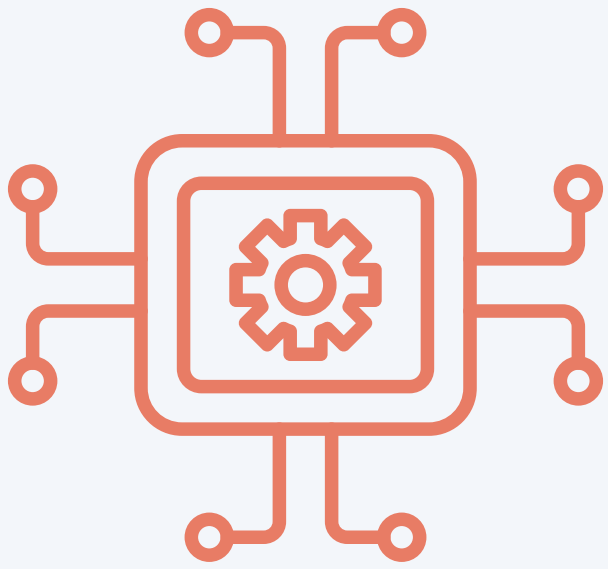
datasciencedojo
data science for everyone

3

Storing Vector Embeddings

- Save the vector embeddings in a Vector Database for efficient storage and retrieval.

- Index the vectors so they can be searched by keyword.



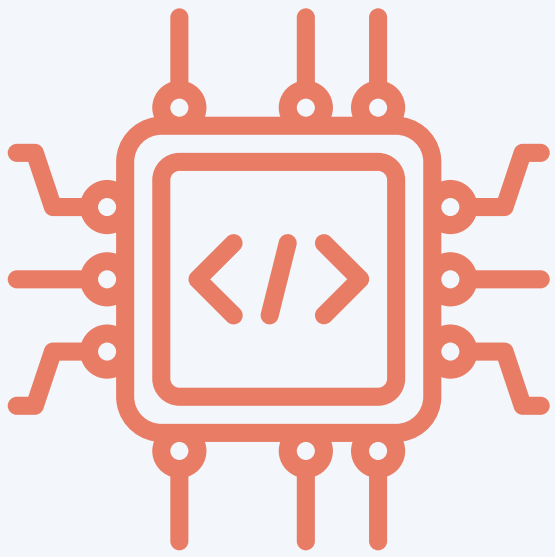
Swipe >>>

datasciencedojo
data science for everyone

4

Embedding the Question

- Use the same embedding model to convert the user's question into a vector representation.
- The question vector should be similar to the vectors of the chunks of text that contain the answer.



Swipe >>>

5

Querying the Database

- Query the Vector Database with the question vector.

- Retrieve a set of relevant context vectors that may contain the answer.

Swipe >>>

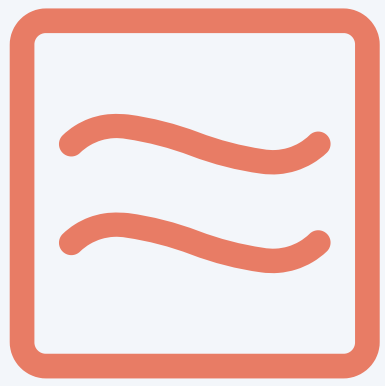
datasciencedojo
data science for everyone

6

Retrieving Similar Vectors

- Perform an Approximate Nearest Neighbor (ANN) search to find the vectors that are most similar to the query embedding.

- Extract the most relevant information from these vectors.



Swipe >>>

datasciencedojo
data science for everyone

7

Mapping Vectors to Text Chunks

- Match the retrieved vectors to their corresponding text chunks. This allows the LLM to access the original text.

Swipe >>>

8

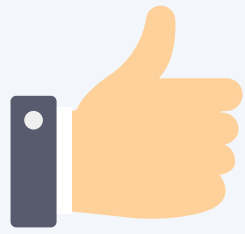
Generating the Answer

- Provide the question and retrieved context text chunks to the LLM.
- Prompt the LLM to use the context to generate an answer that aligns with the expected boundaries.
- The LLM will generate a natural language answer that is relevant to the question.

Swipe >>>

datasciencedojo
data science for everyone

#LargeLanguageModels

Like this Post? 

Check out our blogs for more interesting information