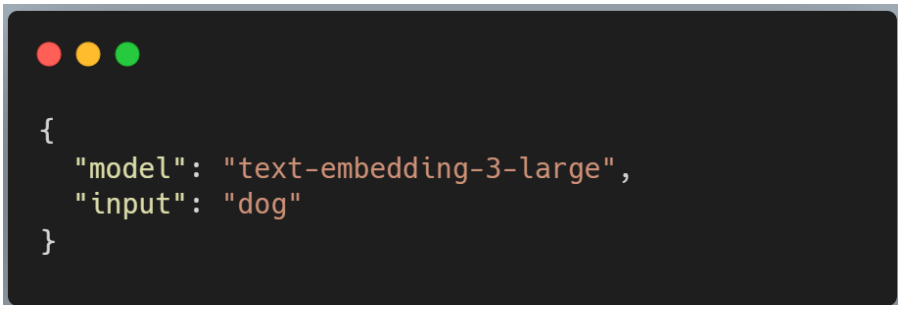
Lecture 17: Embedding using API Client

Introduction

In this lecture, we explore how to create embeddings using an API client instead of directly writing code. Embeddings are numerical vector representations of text (or other inputs) that help AI systems understand relationships and similarities.

Steps to Create Embeddings with API Client

- **API Endpoint:**
  - Send a request to: https://api.openai.com/v1/embeddings
- **Authorization:**
  - **Add a header field:**Authorization: Bearer <API_KEY>
  - Replace **<API_KEY>** with your valid OpenAI key.
- **Headers**
  - Content type should be JSON:Content-Type: application/json
- **Body (JSON)**
  - Specify the model:



```
{
  "model": "text-embedding-3-large",
  "input": "dog"
}
```

  - Common Models:
    - text-embedding-3-small
    - text-embedding-3-large
- **Response**
  - The API will return a large vector (embedding).
  - Example: “dog” gives an array of thousands of dimensions.

Working with Dimensions:

- Default embeddings are large (e.g., 3000+ dimensions).
- You can specify fewer dimensions (like 2) for simple plotting:

- **Example comparisons:**
  - "Laptop" and "Java" are plotted far apart.
  - "India" and "Russia" appear close together, showing semantic similarity.

Key Observations:

- Closer embeddings = more semantic similarity.
- Example: "India" is closer to "Russia" than to "Laptop."
- Two-dimensional embeddings can show relationships visually, but for accurate AI tasks, higher dimensions are used.

Summary

- Embeddings can be created easily with an API client.
- Steps:
  - Use a POST request to /v1/embeddings.
  - Add **Authorization** with the API key.
  - Specify the model and input in **Body (JSON)**.
- The response gives vector representations that can be used for semantic similarity, clustering, or visualization.