# **Tokenization, Embeddings, and API Keys**

## **1. Tokenization**

### **What is Tokenization?**

Tokenization is the process of **splitting text into smaller units** called tokens.  
Tokens can be **words**, **characters**, or **subwords** depending on the method used.

### **Why Tokenize?**

* Machines do not understand raw text; they understand numbers.
* Tokenization converts text into a format that models can process and understand.

### **Example:**

Given the text:

"I love AI."

* **Word-level Tokenization**: ["I", "love", "AI", "."]
* **Character-level Tokenization**: ["I", " ", "l", "o", "v", "e", " ", "A", "I", "."]
* **Subword Tokenization** (used in models like GPT): ["I", " love", " AI", "."]

## **2. Embeddings**

### **What are Embeddings?**

Embeddings are **numerical vector representations** of tokens, words, sentences, or even documents.  
They capture the **meaning and relationships** between words based on context.

### **Why Use Embeddings?**

* Machines can process and compare numbers, not text.
* Embeddings allow models to understand semantic similarity, context, and meaning.

### **Example:**

| **Word** | **Example Embedding (4-D Vector)** |
| --- | --- |
| cat | `[0.4, 0.2, 0.1, 0.9]` |
| dog | `[0.5, 0.1, 0.2, 0.8]` |
| car | `[0.9, 0.5, 0.7, 0.2]` |

Notice:

* "cat" and "dog" embeddings are closer than "car", representing their semantic similarity.

### **Applications:**

* Semantic Search
* Chatbots
* Recommendations
* Text Classification

## **3. API Keys**

### **What is an API Key?**

An **API Key** is a **secret authentication token** that identifies who you are when you interact with an API.  
It acts like a **password for programs** to request access to services.

### **Why Are API Keys Needed?**

* To control and secure access.
* To monitor and limit usage.
* To prevent unauthorized users.

### **Example: OpenAI API Key Usage**

python

CopyEdit

import openai

openai.api\_key = "sk-XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"

response = openai.ChatCompletion.create(

model="gpt-4",

messages=[{"role": "user", "content": "Hello!"}]

)

* The "sk-..." string is your **OpenAI API Key**.
* Without the correct key, the service will reject requests.

### **Important Guidelines:**

* **Never share** your API keys publicly.
* **Store securely** using environment variables (.env files) or secret managers.
* **Rotate** keys periodically for enhanced security.

# 