

```
! pip install openai
! pip install pinecone-client
! pip install pinecone
```

 [Show hidden output](#)

```
import os
import pinecone
from openai import OpenAI
from pinecone import Pinecone, ServerlessSpec
from google.colab import userdata
```

```
PINECONE_API_KEY = userdata.get('PINECONE_API_KEY')
pc = Pinecone(api_key=PINECONE_API_KEY)
```

```
OPENROUTER_API = userdata.get('OPENROUTER_API_KEY')
client = OpenAI(
    base_url="https://openrouter.ai/api/v1",
    api_key=OPENROUTER_API,
)
```

```
index_name = "business-qa"
```

```
if index_name not in pc.list_indexes():
    pc.create_index(
        index_name,
        dimension=768,
        metric="cosine",
        spec=ServerlessSpec(
            cloud="aws",
            region="us-east-1"
        ),
        deletion_protection="disabled"
    )
index = pc.Index(index_name)
```

```
pc.list_indexes()
```


 [

```
    {
        "name": "business-qa",
        "dimension": 768,
        "metric": "cosine",
        "host": "business-qa-z9js94f.svc.aped-4627-b74a.pinecone.io",
        "spec": {
            "serverless": {
                "cloud": "aws",
                "region": "us-east-1"
            }
        },
        "status": {
            "ready": true,
            "state": "Ready"
        },
        "deletion_protection": "disabled"
    }
]
```

```
from transformers import BertTokenizer, BertModel
import torch
```

using distilbert model for embeddings

```
tokenizer = BertTokenizer.from_pretrained('distilbert-base-uncased')
model = BertModel.from_pretrained('distilbert-base-uncased')
```

 The tokenizer class you load from this checkpoint is not the same type as the class this function is called from. It may result in errors. The tokenizer class you load from this checkpoint is 'DistilBertTokenizer'. The class this function is called from is 'BertTokenizer'. You are using a model of type distilbert to instantiate a model of type bert. This is not supported for all configurations of models. Some weights of BertModel were not initialized from the model checkpoint at distilbert-base-uncased and are newly initialized: ['embedder.weight', 'embedder.bias', 'embedder.running_mean', 'embedder.running_var']. You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.

```
def embed_text(text: str) -> list:
    """
    Embed a given text using the pre-trained BERT model, to upload to pinecone.
    """
    inputs = tokenizer(text, return_tensors="pt", truncation=True, padding=True, max_length=512)
```

```

with torch.no_grad():
    outputs = model(**inputs)
    embeddings = outputs.last_hidden_state.mean(dim=1).squeeze().tolist()

return embeddings

```

```

def read_txt_file(file_path: str) -> list:
    """
    Read a .txt file and split its content into smaller chunks. This file will contain info about AMD
    """
    with open(file_path, 'r', encoding='utf-8') as file:
        text = file.read()

    words = text.split()
    chunk_size = 500
    chunks = [" ".join(words[i:i + chunk_size]) for i in range(0, len(words), chunk_size)]
    return chunks

```

```

def upload_txt_to_pinecone(file_path: str, index):
    """
    Process and upload the .txt file to Pinecone as embeddings.
    """
    chunks = read_txt_file(file_path)
    print(f"Processing {len(chunks)} chunks from {file_path}")
    for i, chunk in enumerate(chunks):
        embedding = embed_text(chunk)
        index.upsert([
            {
                "id": f"{os.path.basename(file_path)}_chunk_{i}",
                "values": embedding,
                "metadata": {"content": chunk}
            }
        ])
    print(f"Uploaded {len(chunks)} chunks from {file_path} to Pinecone.")

```

```

def query_relevant_documents(query: str, top_k: int = 5) -> list:
    """
    Retrieve relevant documents (AMD info file) from Pinecone for the given query.
    """
    query_vector = embed_text(query)
    results = index.query(
        vector=query_vector,
        top_k=top_k,
        include_metadata=True
    )
    return [match["metadata"]["content"] for match in results["matches"]]

```

```

def generate_response(query: str, documents: list) -> str:
    """
    Generate a response based on the retrieved document from pinecone.
    """
    context = "\n".join(documents)
    messages = [
        {
            "role": "system",
            "content": "You are a helpful assistant for answering questions about business in AMD."
        },
        {
            "role": "user",
            "content": f"Context: {context}\n\nQuery: {query}"
        }
    ]
    completion = client.chat.completions.create(
        model= 'google/gemini-2.0-flash-exp:free',
        messages=messages
    )
    return completion.choices[0].message.content


```

uploading the file with information about AMD

```

file_path = "/content/amd.txt"
upload_txt_to_pinecone(file_path, index)

```

 Processing 21 chunks from /content/amd.txt
 Uploaded 21 chunks from /content/amd.txt to Pinecone.

Query

```
user_query = "amd did partnership with which companies?"
relevant_docs = query_relevant_documents(user_query)
```

retrieving response from model (testing)

```
response = generate_response(user_query, relevant_docs)
```

```
print("User Query:", user_query)
print("Response:", response)
```

```
➡ User Query: amd did partnership with which companies?
Response: Based on the provided text, here are the companies that AMD has partnered with:

* **Fujitsu:** For flash memory manufacturing, leading to the creation of Spansion.
* **Advanced Technology Investment Co. (ATIC):** A joint venture to form GlobalFoundries for manufacturing operations.
* **Lenovo:** Rory Read, a former Lenovo executive, became CEO of AMD.
* **Hindustan Semiconductor Manufacturing Corporation (HSMC):** For the production of AMD products in India.
* **Sun Microsystems:** To enhance OpenSolaris and Sun xVM on the AMD platform.
* **Meta:** to make the chips used in the Metaverse
* **Samsung:** to develop a mobile processor.
* **Microsoft** and **Sony** to power their next-generation game consoles.
* **U.S. Department of Energy, Oak Ridge National Laboratory, and Cray Inc.:** to develop the Frontier exascale supercomputer
* **U.S. Department of Energy, Lawrence Livermore National Laboratory, and HPE:** to develop the El Capitan exascale supercomputer
```

creating function to retrieve response

```
def question (question: str):
    """
    Input question from user
    """
    user_query = question
    relevant_docs = query_relevant_documents(user_query)
    response = generate_response(user_query, relevant_docs)

    print("User Query:", user_query)
    print("Response:", response)
```

```
question('What is AMD?')
```

```
➡ User Query: What is AMD?
Response: Based on the provided text, here's a breakdown of what AMD is:

**AMD (Advanced Micro Devices) is a technology company that designs and produces a variety of hardware and software products,** print

**1. CPUs and APUs (Central Processing Units and Accelerated Processing Units):**
* AMD is a major player in the microprocessor market, producing CPUs for desktops, laptops, and servers.
* They also develop APUs, which combine a CPU and a GPU (graphics processing unit) on the same chip, designed to improve performance.
* They have families like the Zen series, which is currently in its 5th generation, and previously the Bulldozer series.

**2. GPUs (Graphics Processing Units):**
* AMD produces graphics cards (primarily under the Radeon brand) for gaming, professional graphics, and other applications.
* They also provide software tools like GPUOpen and ROCm for graphics and compute tasks.

**3. Software:**
* AMD has made a focus of opening their software for both graphics and computing.
* They developed suites like, AMD Radeon Software, AOCC, and AMDuProf.

**Key aspects of AMD's business:**

* **Innovation:** AMD has a history of innovation, constantly developing new architectures and technologies for both CPUs and GPUs.
* **Competition:** They are a major competitor to Intel in the CPU market and Nvidia in the GPU market.
* **Acquisitions:** AMD has grown through strategic acquisitions of companies like ATI Technologies (for graphics) and Xilinx (for Open Source).
* **Open Source:** AMD has demonstrated commitment to the open-source community with projects like GPUOpen and ROCm.
* **Partnerships:** AMD has made significant partnerships with other tech companies such as Samsung and Meta.
* **Global Presence:** They have a presence in various markets, ranging from consumers to embedded systems to high-performance computing.

In short, AMD is a technology company that focuses on designing and producing high-performance CPUs, GPUs, and the required support tools.
```

```
question("Who is current CEO of AMD")
```

```
➡ User Query: Who is current CEO of AMD
Response: Based on the provided text, the current CEO of AMD is **Lisa Su**.
```

```
question("What is Zen 2 microarchitecture?")
```

```
➡ User Query: What is Zen 2 microarchitecture?
Response: Based on the provided text, here's what we know about the Zen 2 microarchitecture:

**Key Features and Release:**

* **7nm Process:** Zen 2 is built using a 7nm manufacturing process. This smaller process node generally leads to improved energy efficiency and performance.
```

- * **June 2019 Launch:** AMD launched CPUs based on the 7nm Zen 2 microarchitecture in June 2019.
- * **Epyc Update:** The Epyc line of server processors was also updated in August 2019 to use the Zen 2 microarchitecture.
- * **Mobile Processors:** In 2020, AMD announced the Ryzen Mobile 4000 series, which was their first 7nm x86 mobile processor and the first Zen 2 mobile processor.
- * **Gaming Consoles:** The Steam Deck, PlayStation 5, Xbox Series X, and Series S all utilize chips based on the Zen 2 microarchitecture.

Key Characteristics

- * **Improvements over Zen+:** The text implies Zen 2 offered substantial architectural improvements over the prior Zen+ microarchitecture.

In Summary

The Zen 2 microarchitecture was a very important release for AMD. It was launched during 2019 and 2020. It represented a significant

vague question

question("dd")



User Query: dd

Response: Okay, I will respond to your query with information from the provided text.

You asked me to clarify the meaning of "dd". However, the term "dd" by itself is *not mentioned* in the provided text.

Therefore, I cannot provide a definition for "dd" specifically in the context of AMD or its business, according to the text you have

Without additional context, "dd" is ambiguous and could refer to a few things. More information is needed to understand the context

Start coding or [generate](#) with AI.