# ASSIGNMENT 4- Building a Simple Rag Pipeline
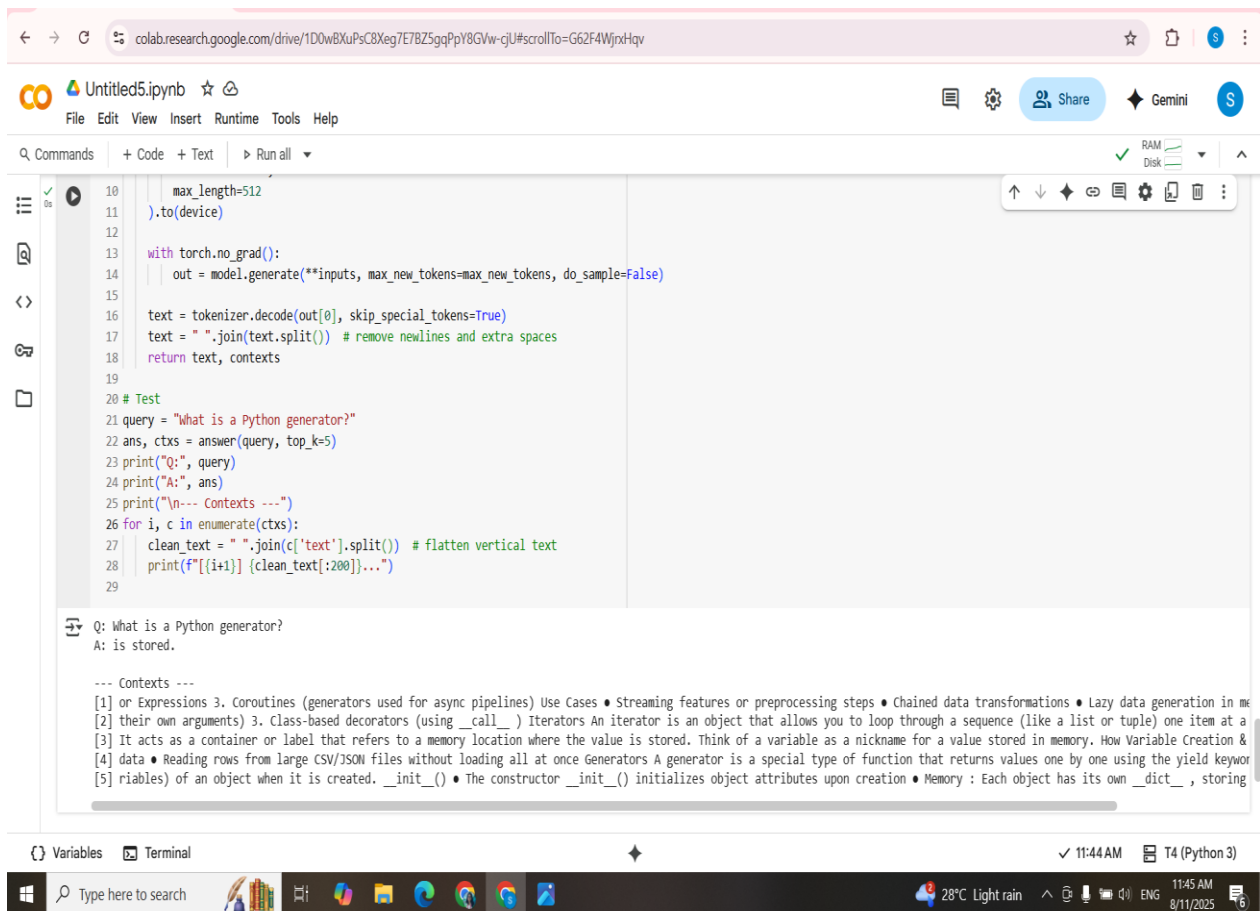
This is the Google colab link below:

https://colab.research.google.com/drive/1D0wBXuPsC8Xeg7E7BZ5gqPpY8GVw-cjU?usp=sharing

Experimenting on Rag text generation:

Untitled5.ipynb  ☆  Saving...

File  Edit  View  Insert  Runtime  Tools  Help

Share    Gemini

Commands    + Code    + Text    Run all

RAM Disk

```
10        max_length=512
11    ).to(device)
12
13    with torch.no_grad():
14        out = model.generate(**inputs, max_new_tokens=max_new_tokens, do_sample=False)
15
16    text = tokenizer.decode(out[0], skip_special_tokens=True)
17    text = " ".join(text.split())  # remove newlines and extra spaces
18    return text, contexts
19
20 # Test
21 query = "What is a Python only in 2 lines?"
22 ans, ctxs = answer(query, top_k=5)
23 print("Q:", query)
24 print("A:", ans)
25 print("\n--- Contexts ---")
26 for i, c in enumerate(ctxs):
27    clean_text = " ".join(c['text'].split())  # flatten vertical text
28    print(f"[{i+1}] {clean_text[:200]}...")
29
```

```
Q: What is a Python only in 2 lines?
A:

--- Contexts ---
[1] erent memory) Indentation Indentation in Python refers to the spaces or tabs at the beginning of a line of code. It defines code blocks such as the body of functions, loops, condit
[2] Take user input len() Find length of string, list, etc. type() Get data type of a variable int(), float(), str() Convert between types max(), min() Find maximum/minimum c) Operator
[3] meaning in Python. • Cannot be used as variable names or identifiers. • Python has 35+ keywords (in Python 3.10+). Use help("keywords") to list all. Eg: if, else, while, for, def,
[4] in a sequence (like list, tuple, string, etc). Operation Symbol Example Result In in 'a' in 'apple' True Not In not in 'x' not in 'box' False 7. Identity Operators Used to compare
[5] It acts as a container or label that refers to a memory location where the value is stored. Think of a variable as a nickname for a value stored in memory. How Variable Creation &
```

{} Variables    Terminal    ✓ 11:45 AM    T4 (Python 3)

---

Untitled5.ipynb  ☆  Saving...

File  Edit  View  Insert  Runtime  Tools  Help

Share    Gemini

Commands    + Code    + Text    Run all

RAM Disk

```
10        max_length=512
11    ).to(device)
12
13    with torch.no_grad():
14        out = model.generate(**inputs, max_new_tokens=max_new_tokens, do_sample=False)
15
16    text = tokenizer.decode(out[0], skip_special_tokens=True)
17    text = " ".join(text.split())  # remove newlines and extra spaces
18    return text, contexts
19
20 # Test
21 query = "What is a EC2?"
22 ans, ctxs = answer(query, top_k=5)
23 print("Q:", query)
24 print("A:", ans)
25 print("\n--- Contexts ---")
26 for i, c in enumerate(ctxs):
27    clean_text = " ".join(c['text'].split())  # flatten vertical text
28    print(f"[{i+1}] {clean_text[:200]}...")
29
```

```
Q: What is a EC2?
A: virtual computing platform

--- Contexts ---
[1] u a virtual machine, where you can run websites, apps, or even big data tasks. • Choose the memory, storage, and CPU you need, and stop the instance when done. • EC2 offers secure,
[2] ail Address": After entering the code, click on "Verify Email Address" to move to the next step. Step 3: Set the Root User Password • Create a Password: Choose a strong password fo
[3] WS EC2 (Elastic Compute Cloud) The following are the features of AWS EC2: 1. AWS EC2 Functionality EC2 Offers a virtual computing platform where users can run operations and launch
[4] ystems Amazon EC2 includes a wide range of operating systems to choose from while selecting your AMI. Not only are these selected options, but users are also even given the privile
[5] d computing service on the AWS cloud platform called instances. It lets you rent virtual computers to run your applications. You pay only for what you use. Amazon EC2 (Elastic Comp
```

{} Variables    Terminal    ✓ 11:46 AM    T4 (Python 3)

CO  Untitled5.ipynb  Saving...

File  Edit  View  Insert  Runtime  Tools  Help

Commands  + Code  + Text  ▷ Run all

```python
        max_length=512
    ).to(device)

    with torch.no_grad():
        out = model.generate(**inputs, max_new_tokens=max_new_tokens, do_sample=False)

    text = tokenizer.decode(out[0], skip_special_tokens=True)
    text = " ".join(text.split())  # remove newlines and extra spaces
    return text, contexts

# Test
query = "Summarize S3?"
ans, ctxs = answer(query, top_k=5)
print("Q:", query)
print("A:", ans)
print("\n--- Contexts ---")
for i, c in enumerate(ctxs):
    clean_text = " ".join(c['text'].split())  # flatten vertical text
    print(f"[{i+1}] {clean_text[:200]}...")
```

```
Q: Summarize S3?
A: AWS CLI

--- Contexts ---
[1] n S3 works? Amazon S3 works on organizing the data into unique S3 Buckets, customizing the buckets with Acccess controls. It allows the users to store objects inside the S3 buckets
[2] ( AWS C onsole >> Amazon S3 >> S3 Buckets ) 2. AWS CLI Commands In this methods firstly you have to install the aws cli software in the terminal and try on configuring the aws acc
[3] : Amazon S3 is often considered as data lake because of its capacity to store large amounts of both structured and unstructured data offering seamless integration with other AWS A
[4] different storage classes based on the age of the object which is uploaded into the S3 bucket. Step 5: You need to turn to enable the services to monitor and analyze S3. You need
[5] nature of the machine, many organizations like Flipkart, and Amazon rely on these days whenever humongous traffic occurs on their portals. Amazon S3 What is Amazon S3? Amazon S3 is
```

{} Variables    Terminal                                           ✓ 11:47 AM    T4 (Python 3)