Name : Mayur Mankar
Roll No. : 20169
Department : Data Science and Engineering (4th year)

Assignment 3

Aim :
Use these GPT codes and apply them for the databases related to biopharma.

Material :
a) Code:
https://github.com/PromtEngineer/localGPT/tree/main
https://github.com/imartinez/privateGPT
b) Dataset:
Here we downloaded the first 10 full-text review papers published in last year on topic of "triple negative breast cancer" from PubMed

1: Schettini F, Venturini S, Giuliano M, Lambertini M, Pinato DJ, Onesti CE, De Placido P, Harbeck N, Lüftner D, Denys H, Van Dam P, Arpino G, Zaman K, Mustacchi G, Gligorov J, Awada A, Campone M, Wildiers H, Gennari A, Tjan-Heijnen V, Bartsch R, Cortes J, Paris I, Martín M, De Placido S, Del Mastro L, Jerusalem G, Curigliano G, Prat A, Generali D. Multiple Bayesian network meta-analyses to establish therapeutic algorithms for metastatic triple negative breast cancer. Cancer Treat Rev. 2022 Dec;111:102468. doi: 10.1016/j.ctrv.2022.102468. Epub 2022 Sep 28. PMID: 36202026.

2: Kudelova E, Smolar M, Holubekova V, Hornakova A, Dvorska D, Lucansky V, Koklesova L, Kudela E, Kubatka P. Genetic Heterogeneity, Tumor Microenvironment and Immunotherapy in Triple-Negative Breast Cancer. Int J Mol Sci. 2022 Nov 29;23(23):14937. doi: 10.3390/ijms232314937. PMID: 36499265; PMCID: PMC9735793.

3: Liu S, Li J, Gu L, Wu K, Xing H. Nanoparticles for Chemoimmunotherapy Against Triple-Negative Breast Cancer. Int J Nanomedicine. 2022 Nov 7;17:5209-5227. doi: 10.2147/IJN.S388075. PMID: 36388877; PMCID: PMC9651025.

4: Leon-Ferre RA, Goetz MP. Advances in systemic therapies for triple negative breast cancer. BMJ. 2023 May 30;381:e071674. doi: 10.1136/bmj-2022-071674. PMID: 37253507.

5: Nasiri F, Kazemi M, Mirarefin SMJ, Mahboubi Kancha M, Ahmadi Najafabadi M, Salem F, Dashti Shokoohi S, Evazi Bakhshi S, Safarzadeh Kozani P, Safarzadeh Kozani P. CAR-T cell therapy in triple-negative breast cancer: Hunting the invisible devil. Front Immunol. 2022 Nov 22;13:1018786. doi: 10.3389/fimmu.2022.1018786. PMID: 36483567; PMCID: PMC9722775.

6: van den Ende NS, Nguyen AH, Jager A, Kok M, Debets R, van Deurzen CHM. Triple-Negative Breast Cancer and Predictive Markers of Response to Neoadjuvant Chemotherapy: A Systematic Review. Int J Mol Sci. 2023 Feb 3;24(3):2969. doi: 10.3390/ijms24032969. PMID: 36769287; PMCID: PMC9918290.

7: Lee J. Current Treatment Landscape for Early Triple-Negative Breast Cancer

(TNBC). J Clin Med. 2023 Feb 15;12(4):1524. doi: 10.3390/jcm12041524. PMID: 36836059; PMCID: PMC9962369.

8: Lu B, Natarajan E, Balaji Raghavendran HR, Markandan UD. Molecular Classification, Treatment, and Genetic Biomarkers in Triple-Negative Breast Cancer: A Review. Technol Cancer Res Treat. 2023 Jan-Dec;22:15330338221145246. doi: 10.1177/15330338221145246. PMID: 36601658; PMCID: PMC9829998.

9: Manoochehri M, Borhani N, Gerhäuser C, Assenov Y, Schönung M, Hielscher T, Christensen BC, Lee MK, Gröne HJ, Lipka DB, Brüning T, Brauch H, Ko YD, Hamann U. DNA methylation biomarkers for noninvasive detection of triple-negative breast cancer using liquid biopsy. Int J Cancer. 2023 Mar 1;152(5):1025-1035. doi: 10.1002/ijc.34337. Epub 2022 Nov 8. PMID: 36305646.

10: Loizides S, Constantinidou A. Triple negative breast cancer: Immunogenicity, tumor microenvironment, and immunotherapy. Front Genet. 2023 Jan 12;13:1095839. doi: 10.3389/fgene.2022.1095839. PMID: 36712858; PMCID: PMC9879323.

Methodology :

a) LocalGPT

1: Clone the repo using git:

```
git clone https://github.com/PromtEngineer/localGPT.git
```

2: Install conda for virtual environment management. Create and activate a new virtual environment:

```
conda create -n localGPT python=3.10.0
conda activate localGPT
```

3: Install the dependencies using pip. To set up your environment to run the code, first install all requirements:

```
pip install -r requirements.txt
```

4: Installing LLAMA-CPP:

LocalGPT uses [LlamaCpp-Python](#) for GGML (you will need llama-cpp-python <=0.1.76) and GGUF (llama-cpp-python >=0.1.83) models.

For NVIDIA GPUs support, use cuBLAS

```
CMAKE_ARGS="-DLLAMA_CUBLAS=on" FORCE_CMAKE=1 pip install llama-cpp-python==0.1.83
--no-cache-dir
```

5: Ingesting your OWN Data:

Put your files in the SOURCE_DOCUMENTS folder. You can put multiple folders within the SOURCE_DOCUMENTS folder and the code will recursively read your files.

6: Run the following command to ingest all the data:

```
python ingest.py
```

7: In order to chat with your documents, run the following command (by default, it will run on cuda):

```
python run_localGPT.py
```

8: This will load the ingested vector store and embedding model. You will be presented with a prompt:

```
> Enter a query:
```

Once the answer is generated, you can then ask another question without re-running the script, just wait for the prompt again.

Type `exit` to finish the script.

b) PrivateGPT

We were facing many issues running the setting up the PrivateGPT.

Conclusion:

LocalGPT works really well on new dataset and is easy to use. In the case of privateGPT, there are some issues in code given in documentation.

****