THE CYPRUS INSTITUTE
RESEARCH•TECHNOLOGY•INNOVATION

THE CYPRUS ANTI-CANCER SOCIETY

EURO²

How to hack for a cause

Giorgos Kosta
Research Engineer
CaSToRC
eurocc.cyi.ac.cy

Co-funded by the European Union
Republic of Cyprus
RESEARCH & INNOVATION FOUNDATION
EuroHPC Joint Undertaking

A Bit About Myself

- Research Engineer at CaSToRC
- Awareness and Communications task leader for EuroCC2
- Contact & Info:
  - Email: g.kosta@cyi.ac.cy

Agenda

- What our datasets look like.
- What we expect from you
- The proposed methodology and tools.

How a patient file looks likes

Age	78		
Medicine	Dosage	Frequency	Route
Phenergan-Promethazine	25mg	prn , nocte	orally
Colchicine	0.5mg	b.d.	orally
DNS (dextrose & sodium chloride)	500ml	o.d.	iv
Human albumin 20g/100ml	2 vials	b.d.	iv
Meropenem	1gr	t.d.s.	iv
Bioflor	1 tablet	t.d.s.	orally
Loperamide	2mg	prn max dose 16mg/24h	orally
Magnesium eff.	300mg	b.d.	orally
Dexamethasone	4mg	o.d.	iv
Betaloc (metoprolol)	50mg	o.d.	orally
allopurinol	100mg	o.d.	orally
Crestor (rosuvastatin)	5mg	nocte	orally
Paracetamol	1gr	prn / max dosage 4gr/24h	iv
Metoclopramide	10mg	prn	iv
Ondasetron	4mg	b.d.	iv or orally
Pantoprazole	40mg	o.d.	iv or orally
Kcl - potassium chloride	30 cc	o.d.	iv
Atrovent 500mg (ipratropium)	1 ampoule	t.d.s.	neb.
Tramadex / Tramadol	50mg	o.d.	orally
Pregabalin 50mg	50mg	b.d.	orally

Allergies
N/A

Co-existing conditions
Hypokalemia
Pulmonary toxicity from chemotherapy
acute kidney injury
thrombocytopenia
Arterial hypertension
Diabetes
dyslipidemia

interactions
Tramadol + Pregabalin
KCl + Metoprolol
Tramadol + Ondansetron

Starting point for going through medicine files…

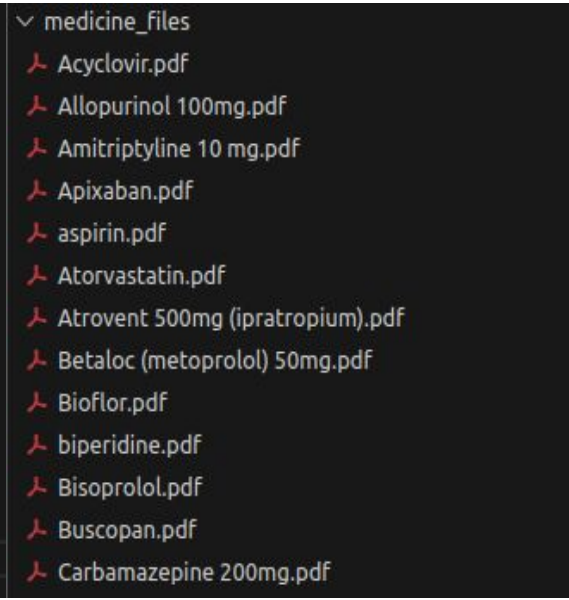
```python
from langchain_community.document_loaders import PyPDFLoader
from tqdm import tqdm
directory='medicine_files'
files = [f for f in os.listdir(directory) if f.endswith('.pdf')]

all_splits = []

for file_name in tqdm(files):
    file_path = os.path.join(directory, file_name)
    # print(f"Processing {file_path}...")  # Optional: for tracking progress

    # Load the PDF
    loader = PyPDFLoader(file_path)

    ...
```



You are free to download any more medicine information from the web!

Dataset caveats

- Patient files not entirely consistent - same medicine names are written in a different way (e.g haloperidole vs haloperidol).
  - You will be given a script to read the patient files though!

- The medicine leaflets that will be provided contain a lot of useless information.

What we expect from you

- A list of all explained interactions per patient found in the patient files. Deadline Sunday 13/10/2024 at 23:59
- The code that generates your the above list. Deadline Sunday 13/10/2024 at 23:59
- A presentation on Monday showcasing and explaining your solution.
- Winner will be announced in the future when we evaluate your results.

Main deliverable

- We expect the major interactions of the whole patient file. For patient 2, it should look something like this:


```
{ "PATIENT 2":
        [
                {       "interaction":["tramadol","pregabalin"],
                        "severity":"Major",
                        "Explanation":      "Using narcotic pain medications together with other medications that cause central
                                            nervous system depression can lead to respiratory distress, coma, and even death."
                },
                {       "interaction":["X","Y"],
                        "severity":"Minor",
                        "Explanation":      "There were no major interactions found in the dataset,
                                            however for medicine X it is suggested to consult with your doctor.”
                }
        ]
 }
```

Warning about the main deliverable.

You must use most of the patient information because:

- If you iterate through all combinations of possible interactions, you will run out of time (nCr):
  - Medicine to medicine
  - Medicine to allergy
  - Medicine to co-existing conditions
  - Etc.
- You may generate a lot of false positives because in the medicine leaflets they mention the phrase *“Consult with your doctor if …”* a lot!

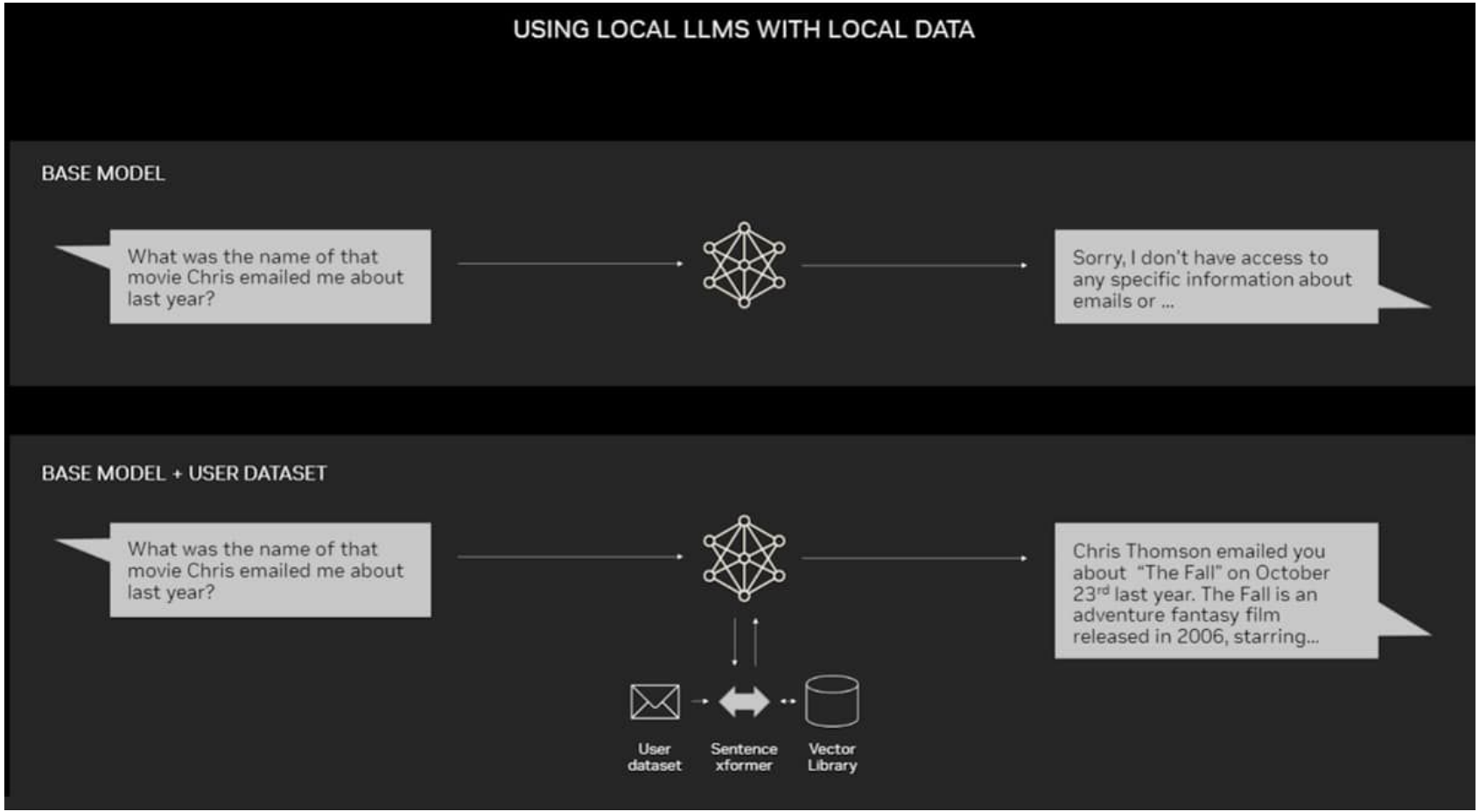
Solving the problem

1. Read through patient files and extract important information

2. Match this information with the appropriate medicine leaflets

3. Use an LLM to give an answer based on the information you gathered.

How to use RAG to win this Hackathon



Useful tools for RAG

- Langchain - Create chain of prompts
- Langgraph - Create sophisticated LLM agents
- Langsmith - Track your chains for debugging
- Ollama - Your local LLM inference server

Constraints

- 1 GPU per team - 32GB V100 from NVIDIA

- Limited available models:
  - Gemma2:27b (Google)
  - Llama3.1:8b (Meta)
  - Phi3.5:3.8b-mini-instruct-fp16 (Microsoft)
  - Nemotron-mini:4b-instruct-fp16 (Nvidia)

- Limited context size depending on model

Conclusion and Questions

- Discord server with private channels for each team.
  - If you need any help mention me @costacis21
- Github repository with our slides and more.

Thank you for the attention!

More information:

https://castorc.cyi.ac.cy/
https://eurocc.cyi.ac.cy/

Contact us at:
eurocc-contact@cyi.ac.cy

Co-funded by the European Union
Republic of Cyprus
RESEARCH & INNOVATION FOUNDATION
EuroHPC Joint Undertaking

Funded by the European Union. This work has received funding from the European High Performance Computing Joint Undertaking (JU) and Germany, Bulgaria, Austria, Croatia, Cyprus (co-funded by the EU within the framework of the Cohesion Policy Programme "THALIA 2021-2027"), Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Türkiye, Republic of North Macedonia, Iceland, Montenegro, Serbia under grant agreement No 101101903.