Gen AI Engineer Assistant – Progress

LAHRACHE, ABDELGHAFOUR 3/25/2024

Plan

1. Introduction

a. *Problem statement*

b. *Opportunity and solution*

2. Internship and project

a. *Internship*

b. *Roadmap*

3. Progress

4. Work to be done

5. References

1. Introduction

1- Introduction

a- Problem Statement

Engineers often face challenges in efficiently accessing, analyzing, and synthesizing information from a vast array of presentations(pptx), technical documents, manuals, and research papers.

Traditional methods of document crawling/Scraping, and analysis are time-consuming and prone to errors, hindering productivity and decision-making processes.

Use case with existing LLMs (Foundation Models):

An example of such AI models is the large language model GPT-4 (Generative Pre-trained Transformer) by OpenAI, BERT, or SAM developed by META.

Strength	Weaknesses
• Generalization • Scalability • **Versatility** (Foundation, can be fine tuned for a specific field) • Cost-Efficiency (already trained on wide range of data)	• **Lack of Understanding** • **Inconsistency** • **Difficulty with long context** • **Difficulty with Specific Tasks**

B- Opportunity and Solution

Opportunity:

Leverage advanced AI technologies to develop a solution that empowers engineers with intelligent assistance in their day-to-day tasks through a chatbot interface.

Solution:

Propose a Gen-AI powered Engineer assistant that aims to address these challenges by offering a platform that automates document processing, extracts key insights, and provides intelligent responses to user queries.

For this I will be basing the project on LLMs, the model will be fine-tuned with features of RAG. Possibility of creating a **Multimodal LLM (OCR-Free document analyzer).**

2. Internship and Project

2-Internship and project

a- Internship

The internship is part of the end of studies pre-employment internships provided By Capgemini Engineering.

The duration of the internship **is 6 (six) Months**, in Casablanca. My internship is fully on-site.

I am part of the Tech Lab team, aiming to develop new solutions and projects by implementing AI and VR technologies.

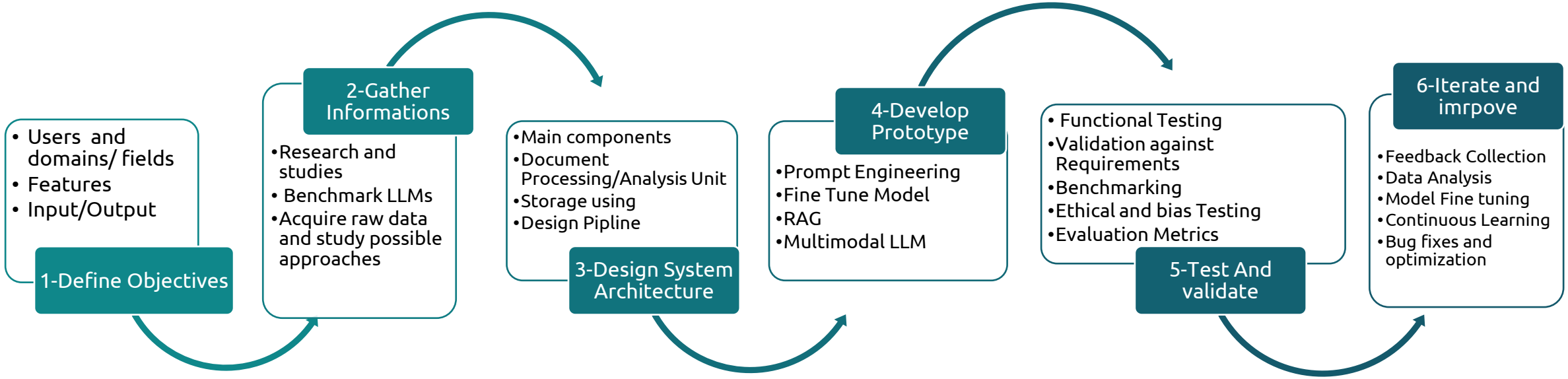
The training environnement is set to be either on the cloud or on local Machine with GPU. (Specs to be determined later)

After completion, the projects are to be deployed and used Onsite/locally.

2-Internship and project

b- Roadmap

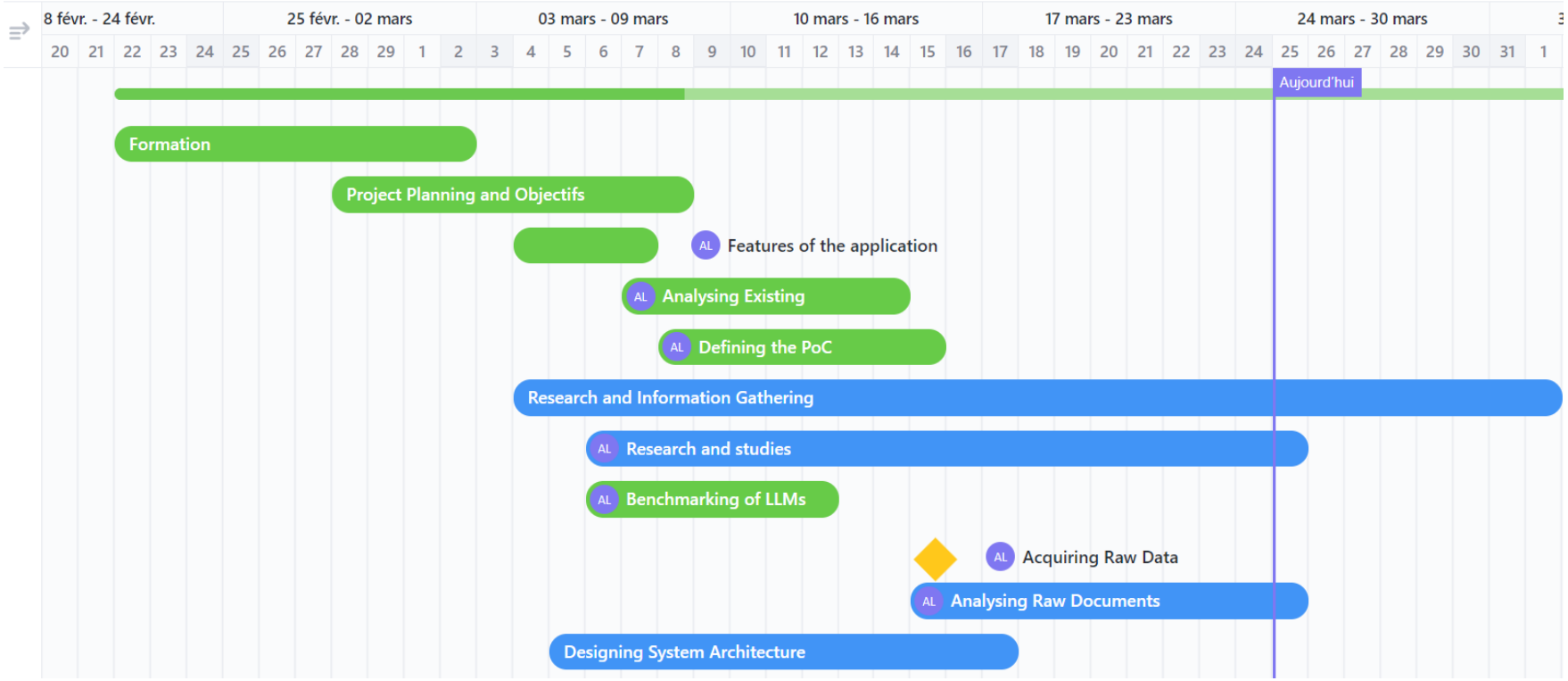
This is a roadmap and planning of the project (yet to be updated):



3. Progress

3- Progress

Gantt:
Here is a gantt chart showcasing the progress of the project.

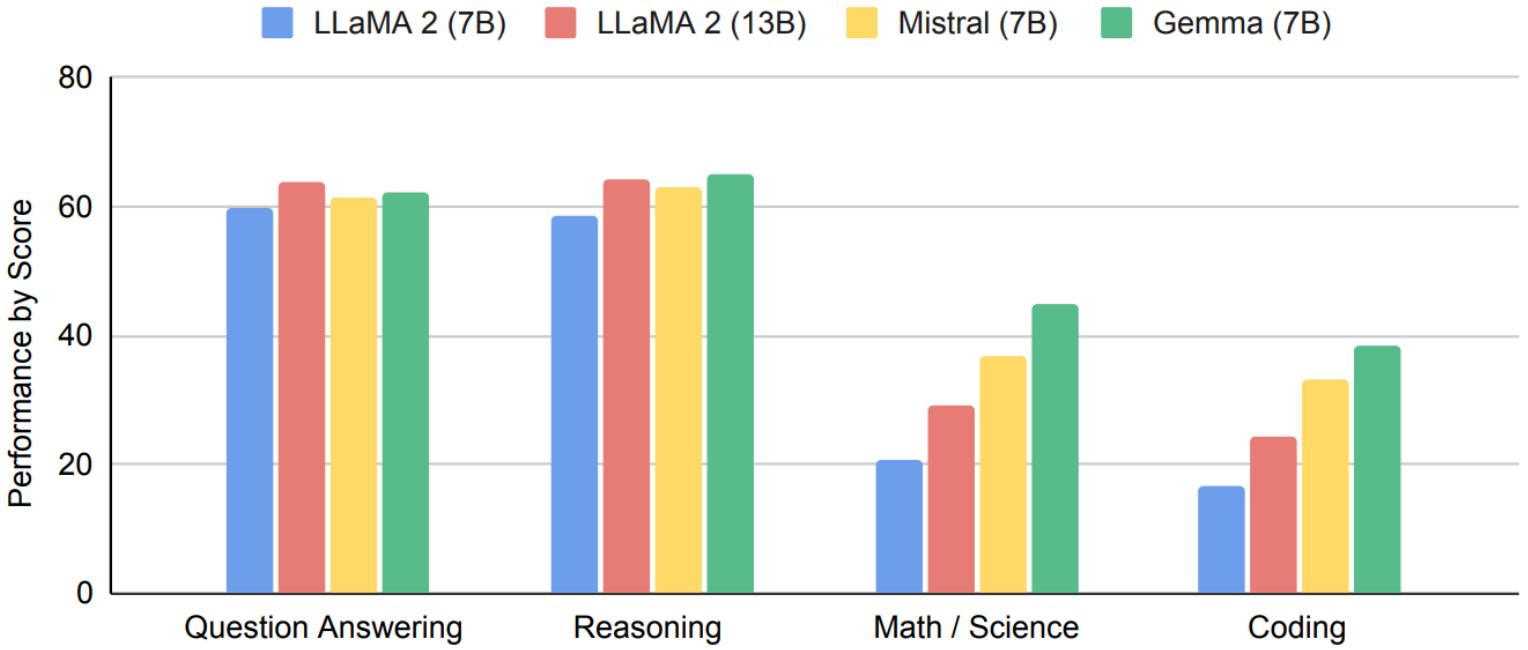


Formation: Capgemini Corporate related.

Github Project: https://github.com/users/Abdelghafour2001/projects/1

b- Model Benchmarking

After benchmarking many Open Source LLMs, Results showed that only 3 current models can excel at delivering the best performances for our domain specific assistant. But we will only focus on LLaMA and Mistral for the upcoming tasks.



C- Mistral vs LLaMA

Model	Developer	Params	Context Window	Use-case	Licence
LlaMa 2	Meta	7B 13B 70B	4k	–Further fine-tuning; –Assistant-like chats; –Various natural language generation tasks in English.	Meta licence
Mistral	Mistral.ai	7B	8k	–Affordable for further fine-tuning; –Chat applications; –Instruction following; –Code understanding.	Apache 2.0

d- Documents

After initial analysis of the documents recieved, many problems and difficulties could be visualy detected before processing them.

The documents contain :

- Text
- Tables
- Data Points
- Images / Screenshots
- Videos containing naration/presentations of the capitalisations.

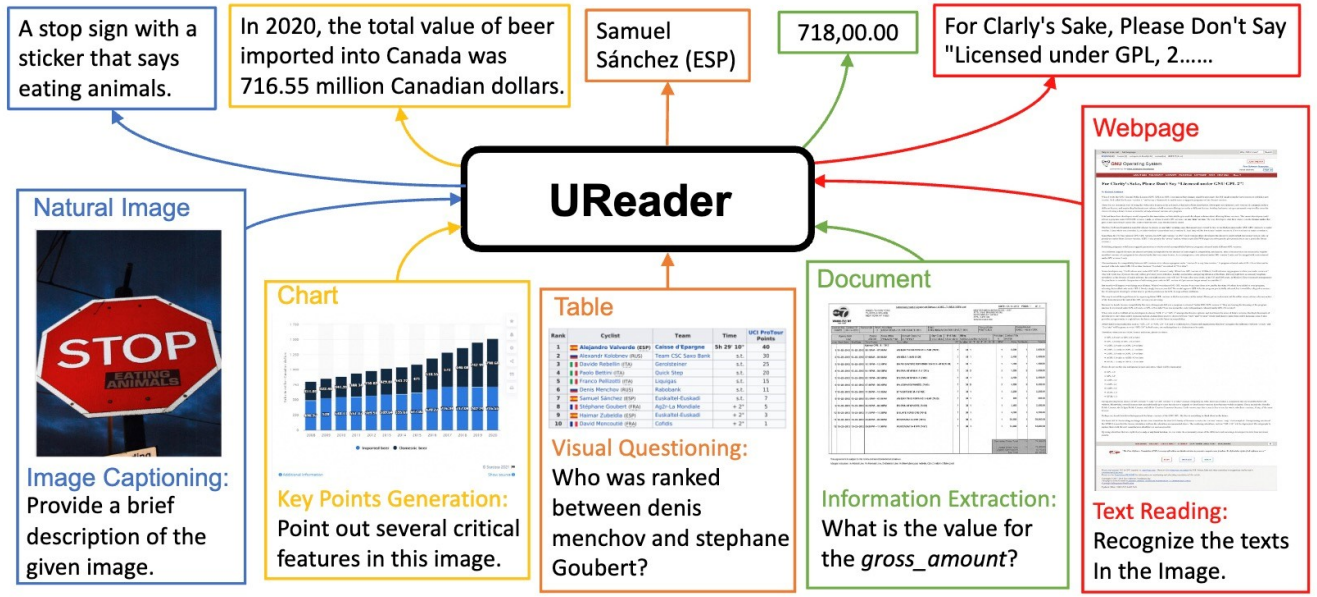
This will require creating a new approach/pipeline for the raw data to be mined and extracted as useful informations to be, then used to fine-tune the model or in the Retrieval-Augmented Generation approach.

4. Work to be Done

4- Work to be done

After acquiring the documents, comes the part of processing them, in order to extract valuable, structured informations so that we can start training the model.

The approach that i am currently studying is the implementation of Multimodal LLMs, through research papers examples like **Ureader** by Jiabo Ye and Colleagues, **mPLUG-DocOwl** by Anwen Hu and Colleagues and **TextMonkey** by Yuliang Liu and Colleagues.



Example of Ureader

5. References

6- References

- https://platform.openai.com/docs/assistants/overview
- https://www.promptingguide.ai/research/rag
- https://huggingface.co/
- https://github.com/AGI-Edgerunners/LLM-Agents-Papers
- Retrieval-Augmented Generation for Large Language Models: A Survey(opens in a new tab) (Gao et al., 2023).
- UReader: Universal OCR-free Visually-situated Language Understanding with Multimodal Large Language Model (Jiabo Ye et al., 2023).
- mPLUG-DocOwl 1.5: Unified Structure Learning for OCR-free Document Understanding (Anwen Hu et al., 2024)
- Monkey: Image Resolution and Text Label Are Important Things for Large Multi-modal Models (Zhang Li et al.,2023)
- https://github.com/users/Abdelghafour2001/projects/1

