## Part-1

### Introduction

* Accessing LLM on cloud (Groq) using API
* Prompt techniques: zero shot, few shot, chain of thoughts
* Sentence embedding
* Similarity search, re-rankers
* Retrieval Augmented Generation (RAG) basic implementation

## Part-2

### Offline Models

* Data types in pytorch
* Loading models from Hugging Face by data type
* Quantization, model size and resource requirement estimation
* Fine-Tuning: PEFT + QLoRA
* Ollama

## Part-3

### Llama-index, LLM Observability and Evaluation

* RAG using llama-index
* Evaluating a RAG framework
* Tools and Agents: ReAct Agent using Llama-index

## Notebooks

### Part-1

[Workshop: 1.1 LLM Prompting.ipynb](https://colab.research.google.com/drive/1nBbD7r4eXtkrodM1pEFxaWJhI-nECdjf?usp=sharing)

[Workshop: 1.2 RAG.ipynb](https://colab.research.google.com/drive/1sRZgZKShqQx_gPVqeId5cA9uOLvjW3jv?usp=drive_link)

### Part-2

[Workshop 2.1 - Data Types And Quantization.ipynb](https://colab.research.google.com/drive/10jwjbxrCLdUe2OCuRAuYAydwmBHe_TwQ?usp=drive_link)

[Workshop 2.2 - llama3-8B.ipynb](https://colab.research.google.com/drive/1YKaJzKuWQHcLft22XItu_8NjyolMk-XB?usp=drive_link)

[Wrokshop 2.3 - Fine-Tuning LLM: PEFT + QLoRA.ipynb](https://colab.research.google.com/drive/1N8d0kEjKpLBBVUlrnc_gmidNrReCqFuc?usp=drive_link)

### Part-3

[Wrokshop 3.1 - RAG using LLama-index.ipynb](https://colab.research.google.com/drive/1G3QF-uGUTLebrgev6MsEbWre9K7Wv9e_?usp=sharing)

[Workshop 3.2 - LLM Observabilty, Evaluation.ipynb](https://colab.research.google.com/drive/1aiMLWn9sbj02iZAFc0x8cbvwAcf20eRY?usp=sharing)

[Workshop 3.3 - Tools and Agents.ipynb](https://colab.research.google.com/drive/1Poqql3I2JJyFOj4v9jaZFJJhx-lqcJEG?usp=sharing)

## Project Ideas

* Tech support chatbot
  + Customer Support Emails - Ticket System - Helpdesk - <https://www.kaggle.com/datasets/tobiasbueck/multilingual-customer-support-tickets>
  + bitext/Bitext-customer-support-llm-chatbot-training-dataset - <https://huggingface.co/datasets/bitext/Bitext-customer-support-llm-chatbot-training-dataset>
* A tool for HR - The manager's guide to HR (book): index books that have information useful to assist HR.
* Legal advisor - legal data
* News summarizer and QA tool (live data)
* Sales Automation (math tool)
* Text Classification on medical data (<https://www.kaggle.com/datasets/matthewjansen/pubmed-200k-rtc>)
* Text Classification and information extraction on Youtube reviews (h[ttps://www.kaggle.com/datasets/ahsenwaheed/youtube-comments-spam-dataset](http://www.kaggle.com/datasets/ahsenwaheed/youtube-comments-spam-dataset))

some useful datasets: (Medical Transcriptions) <https://www.kaggle.com/datasets/tboyle10/medicaltranscriptions/data>