LLMs and Their Applications

# Topics

1. Fine-tuning Large Language models: PEFT techniques, evaluation and benchmarking.
2. Model Quantization.
3. Prompting techniques, RAG, evaluation frameworks for RAG.
4. Guardrails, PII masking techniques.
5. Data augmentation using LLMs.
6. Model benchmarking on domain specific data.
7. Tool/function calling and Agents.
8. Cost estimation and analysis for serving LLMs.

# Resources

1. **Gen AI Project** Using Llama3.1 [codebsics] [link](https://youtu.be/CO4E_9V6li0?si=BxUEcoCOiL7CZgC7); beginner friendly.
2. Generative AI with Large Language Models [Coursera] [link](https://www.coursera.org/learn/generative-ai-with-llms?utm_campaign=WebsiteCoursesGAIA&utm_medium=institutions&utm_source=deeplearning-ai): pretraining LLM, finetuning using PEFT and RLHF.
3. Quantization in Depth [Deeplearning.ai] [link](https://www.deeplearning.ai/short-courses/quantization-in-depth/).
4. Prompt Engineering Guide [link](https://www.promptingguide.ai/).
   1. Zero-shot, few-shot.
   2. Retrieval Augmented Generation (RAG).
   3. ReAct Prompting and other advanced techniques.

## Tools and Frameworks

1. **Lamma-index**: framework for building context-augmented generative AI applications with LLMs including agents and workflows [link](https://docs.llamaindex.ai/en/stable/getting_started/concepts/).
2. **Trulens:** LLM evaluation Framework [link](https://www.trulens.org/trulens/getting_started/). (related short course: [link](https://www.deeplearning.ai/short-courses/building-evaluating-advanced-rag/))
3. Neo4j Knowledge graph.

# Projects

|  |  |  |  |
| --- | --- | --- | --- |
| Title | Description | Dataset | Learning outcome |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |