LLM – Resources

Prompt Engineering

- https://aman.ai/primers/ai/prompt-engineering/

Prompt Fine Tuning

- https://huggingface.co/docs/peft/task_guides/clm-prompt-tuning
- https://developer.nvidia.com/blog/an-introduction-to-large-language-models-prompt-engineering-and-p-tuning/
- https://www.union.ai/blog-post/fine-tuning-vs-prompt-tuning-large-language-models
- https://cobusgreyling.medium.com/prompt-tuning-hard-prompts-soft-prompts-49740de6c64c#
- https://medium.com/international-school-of-ai-data-science/optimizing-performance-with-peft-a-deep-dive-into-prompt-tuning-2b9a17bc9851

INSTRUCTION FINE TUNING :

- https://medium.com/@ud.chandra/instruction-fine-tuning-llama-2-with-pefts-qlora-method-d6a801ebb19
- https://heidloff.net/article/instruct-tuning-large-language-models/
- https://www.linkedin.com/pulse/beginners-guide-fine-tuning-large-language-models-vaidheeswaran/
- https://dr-bruce-cottman.medium.com/part-1-eight-major-methods-for-finetuning-an-llm-6f746c7259ee
- https://www.linkedin.com/pulse/get-insight-from-your-business-data-build-llm-application-jain-1f/

PEFT Methods

- https://towardsdatascience.com/parameter-efficient-fine-tuning-peft-for-llms-a-comprehensive-introduction-e52d03117f95
- https://abvijaykumar.medium.com/fine-tuning-llm-parameter-efficient-fine-tuning-peft-lora-qlora-part-1-571a472612c4
- https://huggingface.co/blog/trl-peft

RLHF

- https://medium.com/@kanikaadik07/reinforcement-learning-from-human-feedback-rlhf-d528a6a53156

LORA and QLORA

- https://dataman-ai.medium.com/fine-tune-a-gpt-lora-e9b72ad4ad3
- https://abvijaykumar.medium.com/fine-tuning-llm-parameter-efficient-fine-tuning-peft-lora-qlora-part-1-571a472612c4
- https://www.linkedin.com/pulse/step-by-step-guide-llm-fine-tuning-using-peft-kosaraju-ph-d-/
- https://github.com/ashishpatel26/LLM-Finetuning

VECTOR DB and RAG

- https://aman.ai/primers/ai/RAG/
- https://medium.com/predict/crafting-an-ai-powered-chatbot-for-finance-using-rag-langchain-and-streamlit-4384a8076960
- https://www.kaggle.com/code/peremartramanonellas/use-a-vectorial-db-to-optimize-prompts-for-llms?scriptVersionId=142791188
- https://m.youtube.com/playlist?list=PLfaIDFEXuae2LXbO1_PKyVJiQ23ZztA0x
- https://www.anyscale.com/blog/a-comprehensive-guide-for-building-rag-based-llm-applications-part-1
- https://www.deeplearning.ai/short-courses/building-applications-vector-databases/
- https://datasciencenerd.us/build-a-chatbot-using-local-llm-6b8dbb0ca514
- https://medium.com/predict/crafting-an-ai-powered-chatbot-for-finance-using-rag-langchain-and-streamlit-4384a8076960
- https://deepchecks.com/practical-guide-to-crafting-your-first-llm-powered-app-using-rag-framework/
- https://medium.com/mim-solutions-blog/fine-tuning-bert-model-for-arbitrarily-long-texts-part-1-299f1533b976
- **https://medium.com/predict/crafting-an-ai-powered-chatbot-for-finance-using-rag-langchain-and-streamlit-4384a8076960**
- https://towardsdatascience.com/retrieval-augmented-generation-intuitively-and-exhaustively-explain-6a39d6fe6fc9
- https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/
- https://towardsdatascience.com/retrieval-augmented-generation-intuitively-and-exhaustively-explain-6a39d6fe6fc9
- https://towardsdatascience.com/build-industry-specific-llms-using-retrieval-augmented-generation-af9e98bb6f68
- https://www.linkedin.com/posts/andreasmwelsch_artificialintelligence-machinelearning-generativeai-activity-7134571640100691968-iWmD?utm_source=share&utm_medium=member_ios
- https://vinija.ai/nlp/RAG/

LLAMA Applications

- https://towardsdatascience.com/fine-tune-your-own-llama-2-model-in-a-colab-notebook-df9823a04a32
- https://www.kaggle.com/code/heyytanay/gpt-from-scratch-using-lightning-and-lance
- https://blog.briankitano.com/llama-from-scratch/ - code LLAMA on Tiny Shakespeare
- https://towardsdatascience.com/fine-tune-your-own-llama-2-model-in-a-colab-notebook-df9823a04a32
- https://stackabuse.com/guide-to-fine-tuning-open-source-llms-on-custom-data/
- https://medium.com/analytics-vidhya/question-answering-system-with-bert-ebe1130f8def

DOCLLM

- https://medium.com/@basics.machinelearning/discover-docllm-the-new-llm-from-jpmorgan-for-working-with-complex-documents-5f54ea287d52
- https://www.dataiku.com/solutions/catalog/llm-enhanced-demand-forecast/
- https://www.ey.com/en_us/coo/how-generative-ai-in-supply-chain-can-drive-value

STREAMLIT

- **https://www.geeksforgeeks.org/a-beginners-guide-to-streamlit/**
- https://pub.towardsai.net/build-and-deploy-a-bert-question-answering-app-using-streamlit-1aba1d76b84
- https://chatbotslife.com/conversational-chatbot-using-transformers-and-streamlit-73d621afde9
- https://medium.com/predict/crafting-an-ai-powered-chatbot-for-finance-using-rag-langchain-and-streamlit-4384a8076960

GEMINI

- https://medium.com/@sathishkumar.sk236/gemini-streamlit-end-to-end-llm-and-large-image-model-application-using-gemini-pro-cd30824c9058
- https://www.analyticsvidhya.com/blog/2023/12/google-gemini-api/
- https://www.kaggle.com/code/youssef19/text-summarization-information-retrieval/notebook

LANGCHAIN

- https://python.langchain.com/docs/get_started/installation
- https://www.sitepoint.com/langchain-python-complete-guide/
- https://www.geeksforgeeks.org/introduction-to-langchain/
- https://www.python-engineer.com/posts/langchain-crash-course/
- https://medium.com/databutton/getting-started-with-langchain-a-powerful-tool-for-working-with-large-language-models-286419ba0842

QA Applications

- https://skimai.com/fine-tuning-bert-for-sentiment-analysis/
- https://towardsdatascience.com/question-answering-with-a-fine-tuned-bert-bc4dafd45626
- https://krishnayogi.medium.com/building-a-question-answering-system-using-llm-50904793ae07

Sentiment Analysis

https://skimai.com/fine-tuning-bert-for-sentiment-analysis/

Valuable GITHUB /REPOS

- https://github.com/8806667653/finetune-business-data
- https://github.com/ianand/spreadsheets-are-all-you-need
- vision transformer
- https://github.com/ashishpatel26/LLM-Finetuning
- https://github.com/peremartra/Large-Language-Model-Notebooks-Course/tree/main
- https://github.com/rasbt/LLMs-from-scratch/blob/main/ch03/02_bonus_efficient-multihead-attention/mha-implementations.ipynb
- https://github.com/chetnakhanna16/CoQA_QuesAns_BERT/blob/main/CoQA_BERT_QuestionAnswering.ipynb
- https://mltechniques.com/2023/12/01/10-genai-notebooks-openai-llm-rag-gpt-and-more/

TEXT Summarization

- https://seekinginference.com/applied_nlp/T5.html
- https://www.analyticsvidhya.com/blog/2023/07/build-a-text-summariser-using-llms-with-hugging-face/?utm_source=related_WP&utm_medium=https://www.analyticsvidhya.com/blog/2023/07/build-your-own-translator-with-llms-hugging-face/
- https://towardsdatascience.com/text-summarization-using-deep-neural-networks-e7ee7521d804
- https://huggingface.co/learn/nlp-course/chapter7/5?fw=tf
- https://medium.com/@anyuanay/fine-tuning-the-pre-trained-t5-small-model-in-hugging-face-for-text-summarization-3d48eb3c4360
- https://github.com/Ryota-Kawamura/Generative-AI-with-LLMs/blob/main/Week-2/Lab_2_fine_tune_generative_ai_model.ipynb

Language Translation

- https://www.analyticsvidhya.com/blog/2023/07/build-your-own-translator-with-llms-hugging-face/
- https://www.analyticsvidhya.com/blog/2023/07/build-your-own-translator-with-llms-hugging-face/