

Before jumping into Week 4, please watch this video link:

https://www.youtube.com/watch?v=70arjDLn0_g

And share your comments on what you have learned from this in the WhatsApp group "Sign-ify 2.0".

THE ABOVE TASK IS COMPLISORY.

BELOW ARE OPTIONAL.

Week 4: Retrieval Augmented Generation App

In this week, you will gain an understanding of the tools and technologies involved in Retrieval Augmented Generation (RAG) applications, focusing on LlamaIndex, Langchain, and Vector Databases.

- Document your findings on advanced RAG in your slide deck. (VERY IMP)

1. Getting Started with RAG

Resources:

- Get a basic overview of RAG by following this article. Link:
<https://www.linkedin.com/pulse/what-rag-lets-dive-deeper-time-alex-wang-ln6dc/>

2. Llama Index

Resources:

- Follow the tutorial on "Building an LLM application" and go through individual steps of using LLMs, building RAG, agents and observability. Link:
<https://docs.llamaindex.ai/en/stable/understanding/>
- Compare the features offered with the LangChain ecosystem. It is important to know the strengths of each.
 - You can attempt to implement any LangChain-based application using LlamaIndex
 - You can follow this tutorial "A Guide to Building a Full-Stack Web App with LlamaIndex" and see how this compares to LangServe. Link:
https://docs.llamaindex.ai/en/stable/understanding/putting_it_all_together/apps/fullstack_app_guide/
- Do this course link:
<https://learn.deeplearning.ai/courses/event-driven-agentic-document-workflows/lesson/ad4t8/what-are-agentic-document-workflows>

3. Advanced RAG

- What is the difference between Basic RAG and Advanced RAG:

"Advanced Retrieval-Augmented Generation: From Theory to LlamaIndex Implementation | by Leonie Monigatti | Feb, 2024 | Towards Data Science" Link:

<https://towardsdatascience.com/advanced-retrieval-augmented-generation-from-theory-to-llamaindex-implementation-4de1464a9930/>

- Implement this advanced RAG technique for optimizing retrieval using metadata:

"Advanced RAG: Optimizing Retrieval with Additional Context & MetaData using LlamaIndex🔗🔗 | by Akash Mathur " Link:

<https://akash-mathur.medium.com/advanced-rag-optimizing-retrieval-with-additional-context-metadata-using-llamaindex-aeaa32d7aa2f>

4. WEEKLY PROJECT

Build on your RAG implementation from day 1 which involved answering questions based on a file, and extend it to include document-lifecycle functionality, as follows:

- Your application will take an input path with multiple files, both .pdf and .docx.
- The pipeline should create embeddings on the vector store.
- In case new documents are added (or some documents are removed), the pipeline should update the embeddings accordingly in the vector store.
- Similarly, if an individual document changes, the embeddings should change as well.