Introducing ASHTRA

Unlocking the power of Research Knowledge



App Overview:

Our app redefines knowledge dissemination by converting complex research papers into captivating podcasts. Additionally, users can easily interact with the open-source code developed as part of the paper and interact seamlessly with any research paper through chat functionality.

Why Choose ASHTRA?

01

Stay Ahead of Trends

Easily gain knowledge from cutting-edge research and advancements to stay at the forefront of innovation

02

Simplify Digestion

Easily digest complex research papers with our user-friendly podcast format for convenient learning

Why Choose ASHTRA?

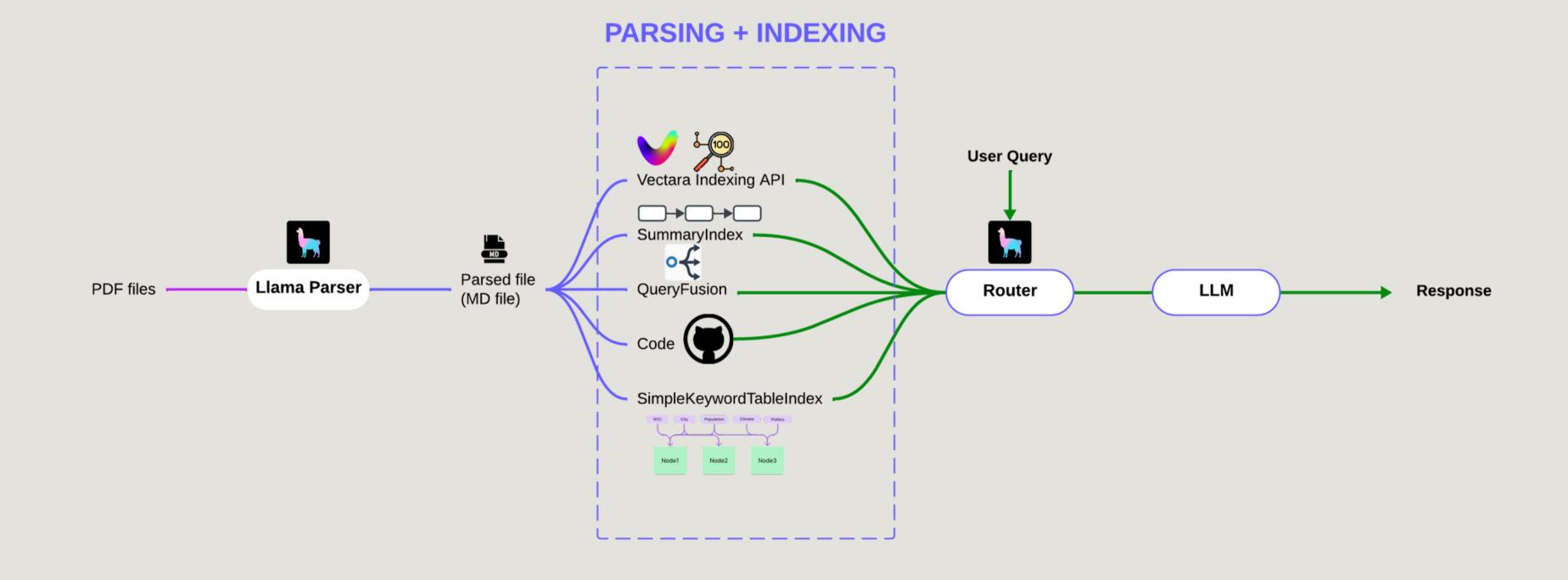
03

Bridging Theory and Practice

Users can integrate open-source code alongside the paper, accessible on platforms like GitHub for effortless understanding and implementation

APP ARCHITECTURE OVERVIEW

Building for Future Success



USER Query: What is the LLM cascade technique that is discussed in the paper Router Vectara llama index client

Response:

LLM cascade is a strategy discussed in the paper that involves sending a query to a list of LLM APIs sequentially. If one LLM API's response is reliable, then its response is returned, and no further LLMs in the list are needed.

The remaining LLM APIs are queried only if the previous APIs' generations are deemed insufficiently reliable. This approach significantly reduces query cost by utilizing relatively inexpensive and reliable LLMs first before resorting to more expensive options.



Tech Stack

- Web Page: Streamlit
- Voice to Text: Deepgram
- Language Models: OpenAl (LLM)
- Research Paper Parsing: Llama Parser



• RAG Building: Llama Index



- RAG as a Service: Vectara
- Open-source Serverless LLMs: Together Al together ai