

AI-Driven Protocol Review: Accelerating and Simplifying Review

Aditya Moogi

Clinical Data Insights Analyst, Novo Nordisk Service Centre India Private Ltd

Introduction

- Clinical trials rely on documents like protocols, plans, and manuals that contain valuable knowledge but remain static.
- The increasing volume and complexity of these documents make traditional access and interpretation methods unsustainable.
- A new framework is introduced using Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG).
- This framework converts passive documents into interactive knowledge interfaces.
- The focus shifts from building tools to rethinking workflows, capabilities, and cognitive bottlenecks in clinical trials.

Problem Statement

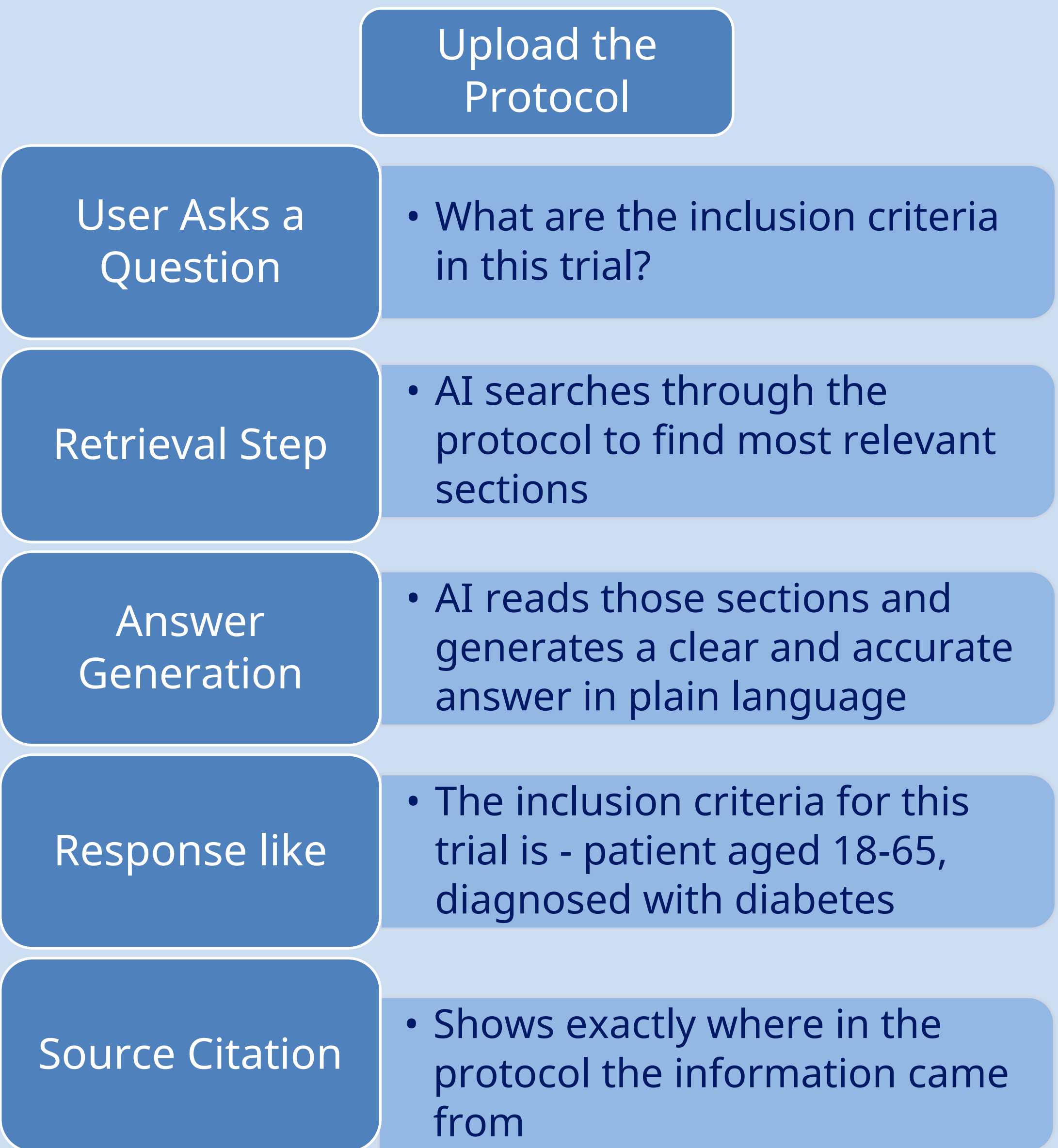
- Protocols are long, complex documents, making quick information retrieval challenging for data managers, monitors, site staff etc.
- Teams spend significant time searching, interpreting and cross-referencing documents often relying on manual review.
- Misunderstanding or overlooking a protocol requirement can lead to delays, deviations and patient safety concerns.
- Growing trial complexity and document volume make manual review become increasingly challenging and prone to error.

Solution

- A Retrieval-Augmented Generation (RAG) based protocols instantly searchable, authoritative. Large Language Model (LLM) assistant that makes

References: Literature Review on RAG(Retrieval Augmented Generation) for Custom Domains, Chatbot with LLM and RAG in Action (Image), Rag: The Revolution of Retrieval-Augmented Text Generation

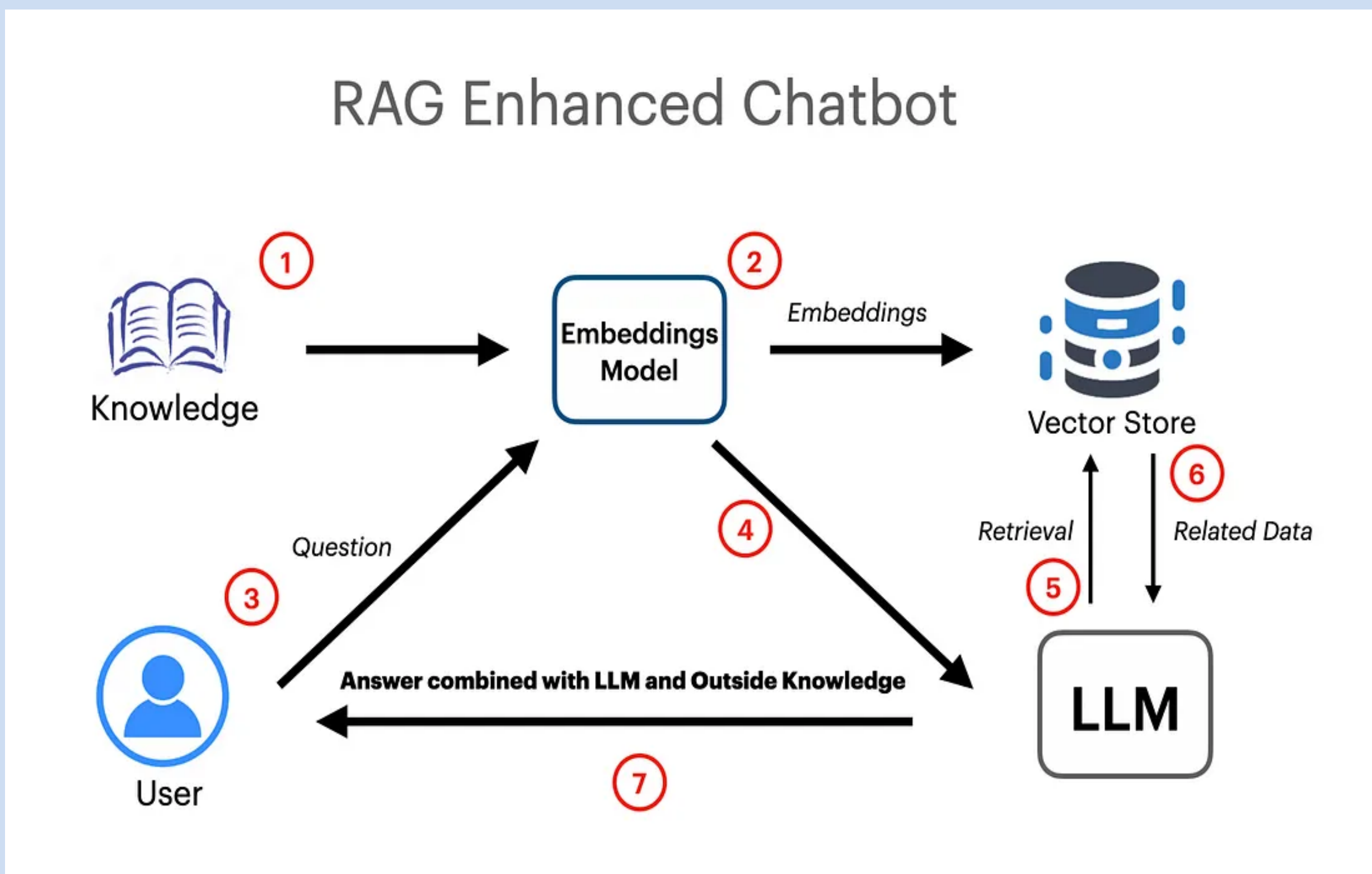
Disclaimer: The poster does not contain any proprietary or confidential information from Novo Nordisk. The views and opinions expressed are solely those of the presenter and not of Novo Nordisk.



Outcome

- Instead of manually searching through lengthy protocol documents, trial team can instantly retrieve accurate, context-specific answers.
- The system can assist in reviewing data entries, identifying inconsistencies, and suggesting corrections, thereby cutting down the time spent on manual review.
- By providing authoritative, source-cited responses, the system helps ensure that data handling aligns with protocol requirements and regulatory standards, reducing the risk of errors and deviations.
- Automated, intelligent responses to data-related queries streamline communication between the entire trial team, accelerating the resolution process and minimizing delays.

Technical Architecture



- **Document Upload & Segmentation:** Protocol uploaded & automatically broken down into logical and structured sections.
- **Text Embedding with LLMs:** Each is converted into vector embeddings using a pre-trained language model.
- **Vector Database Integration:** Embeddings are stored in a vector database (e.g. FAISS) for semantic search.
- **RAG Based Query Handling:** User submit a question > Relevant protocol chunks are retrieved > LLM generates a precise source linked answer.
- **Front-End & Feedback System:** A simple chat interface allows users to interact with the assistant. Feedback is collected to fine tune accuracy.

Conclusion

By leveraging RAG and LLMs, the AI Protocol Assistant transforms static documents into dynamic, searchable knowledge. It enables clinical teams to access accurate. Context aware answers instantly, enhancing protocol compliance, reducing ambiguity and supporting faster, smarter decision making across the trial life cycle.