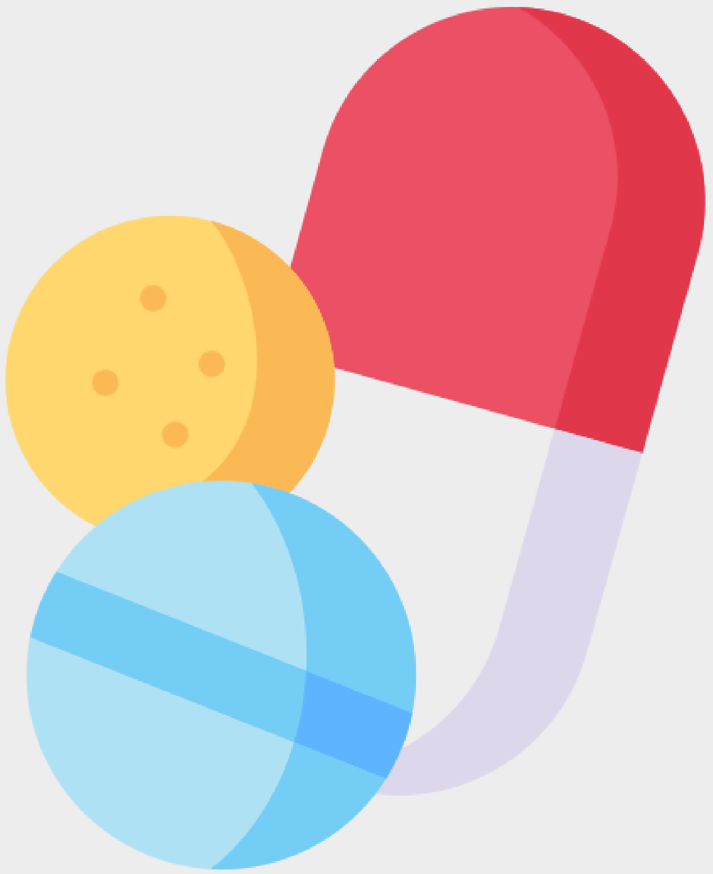
MediQ

TEAM MEMBERS

Artificial Intelligence & Machine Learning

Vishwa M	- 221501179	Viswa V	- 221501180
Yamuna S	- 221501181	Yudeeswaran V	- 221501182
Yuvansandar J	- 221501184	Yuvashree A	- 221501185
Jaya Karthick R	- 221501507	Yogeswaran S	- 221501511

Harish A - 221501518



MENTORS

Mr.SelvaKumar K

Mrs. Sangeetha K

PROBLEM STATEMENT

Accessing prescribing information from drug labels is often difficult because:

- Drug labels are lengthy and complex.
- Critical details are hard to locate quickly (dosage, interactions, contraindications).
- Unstructured PDF formats make automated extraction challenging.
- Patients and healthcare providers require clear, reliable summaries instead of technical jargon.
- Lack of contextual understanding in existing search tools prevents meaningful follow-up queries.
- Absence of citations reduces trust and traceability of extracted information.

OBJECTIVES

- To Build a PDF ingestion pipeline to extract drug prescribing details and precautions.
- To Enable intent recognition and vector search for accurate query resolution.
- To Provide reliable answers with citations from source documents.
- To Support patient-specific advice on interactions and administration.
- To Deliver a user-friendly, scalable chatbot with session management.

EXISTING SYSTEMS

Author(s)/Title	Year	Technique	Description	Outcome
Med-Bot (PDF-Based Medical Assistant with LangChain & ChromaDB)	2025	PyTorch, LangChain, AutoGPT-Q, ChromaDB	PDF-based Q&A on medical literature. Focuses on efficient retrieval and context-aware answers.	Improved accuracy in extracting patient-specific knowledge from medical PDFs.
Community RAG + MedBot GitHub Projects	2025	RAG, FAISS, LangChain, Streamlit	Open-source medical chatbot prototypes for PDF Q&A.	Interactive retrieval-based answers; limited enterprise validation.
MedDoc-Bot (LLM for Pediatric Guideline PDFs)	2024	LLMs (Llama-2, Mistral, Meditron)	Pediatric guideline PDF chatbot; compared different LLMs for reliability.	Found Meditron most effective for healthcare-specific guidelines.
InfoGenie (PDF Information Extraction Chatbot)	2024	HuggingFace embeddings, Chroma, Seq2Seq	Extracts and answers queries from uploaded PDFs.	Provided structured answers; faced scalability challenges.
General-purpose PDF Chatbots (Community Projects)	2023	OpenAI embeddings, Supabase, Pinecone	General chatbot for any PDF Q&A. Popular on GitHub/Reddit.	Easy adoption for non-medical PDFs, lacked domain-specific accuracy.

LIMITATIONS OF EXISTING SYSTEM

S.no	Limitation
1	High resource consumption makes it difficult to scale for low-power environments.
2	Lacks medical domain fine-tuning, leading to unreliable answers for clinical queries.
3	Limited to pediatric hypertension—does not generalize to broader healthcare topics.
4	Struggles to extract accurate answers from tables or charts in medical PDFs.
5	Not optimized for medical data—often misinterprets technical drug-related terms.

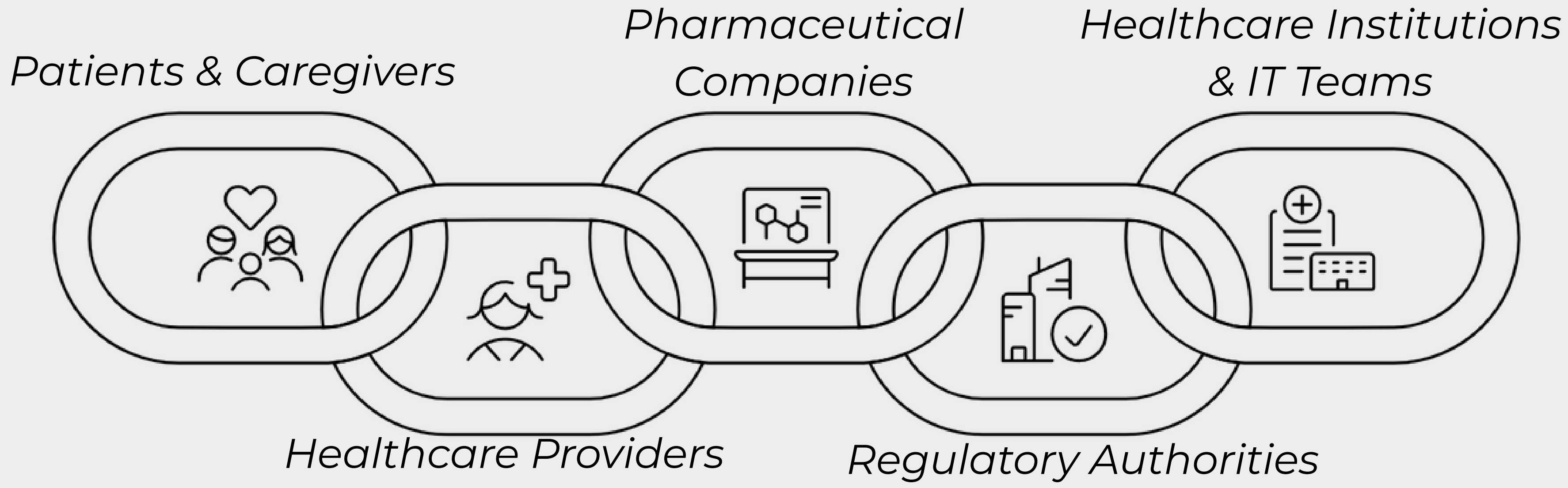
DOMAIN OVERVIEW

This project operates in the healthcare and pharmaceutical information domain, focusing on drug prescribing information. Patients, caregivers, and healthcare professionals often face difficulties in accessing accurate, structured, and up-to-date drug details from lengthy and complex medical PDFs. The challenge lies in bridging the gap between unstructured prescribing documents and the need for clear, reliable, patient-friendly medical knowledge.

CHALLENGES :

- Information Overload in lengthy prescribing PDFs
- Unstructured content (tables, figures, OCR text)
- Risk of misinformation or hallucinations from generic chatbots
- Limited personalization (age, gender, medical conditions)
- Heavy dependency on healthcare staff for repetitive queries

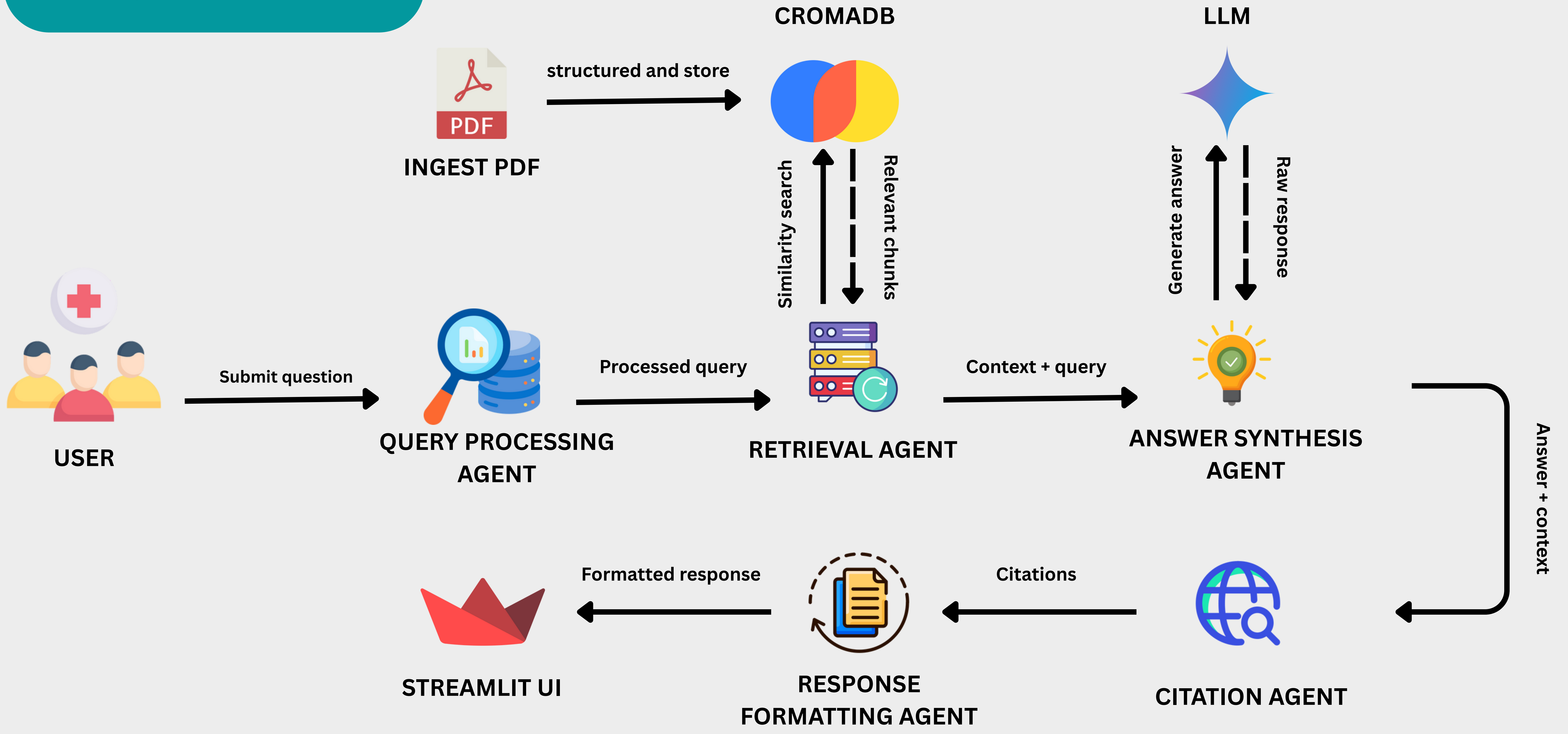
STAKEHOLDERS :



Impact on Stakeholders

1. Improved Accessibility: Simplified and structured access to complex prescribing information enhances decision-making and patient awareness.
2. Enhanced Trust & Transparency: Citations with sources build credibility and reduce misinformation risks.
3. Personalized Insights: Patient profile–based responses empower users with tailored medical advice.
4. Efficiency Gains: Reduces repetitive workload for healthcare staff, enabling them to focus on critical care.
5. Regulatory Alignment: Ensures drug information is consistent with approved prescribing labels.

FLOW DIAGRAM



TECH STACK

PROGRAMMING AND FRAMEWORKS

Streamlit

CrewAI

Python

DATA STORAGE AND RETRIEVAL

SQLite

Redis

ChromaDB

MACHINE LEARNIING AND NLP

PyTorch

Google Gemini API

SentenceTransformers

DOCUMENT PROCESS

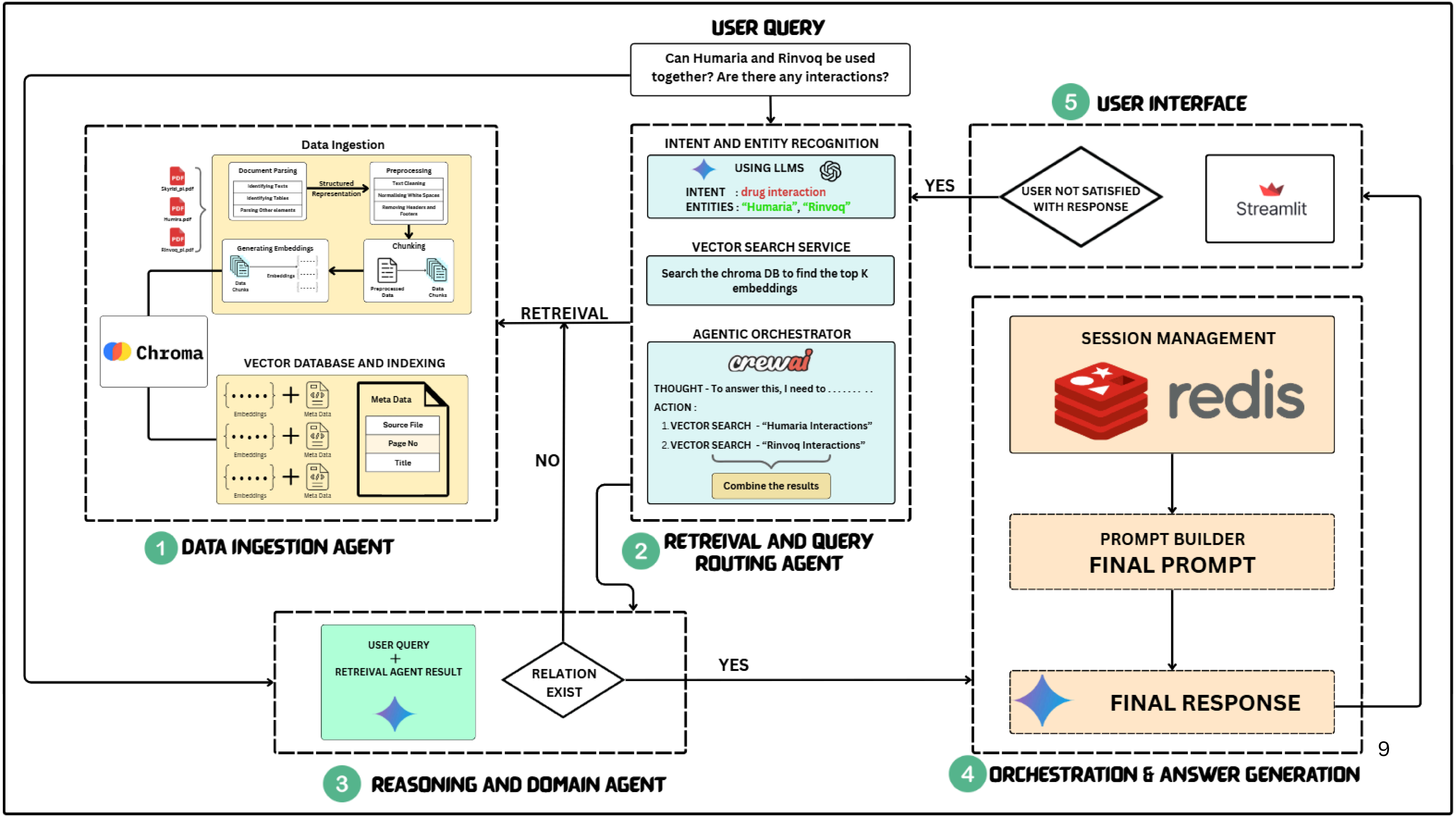
Pytesseract

PyMuPDF

Camelot

Pillow

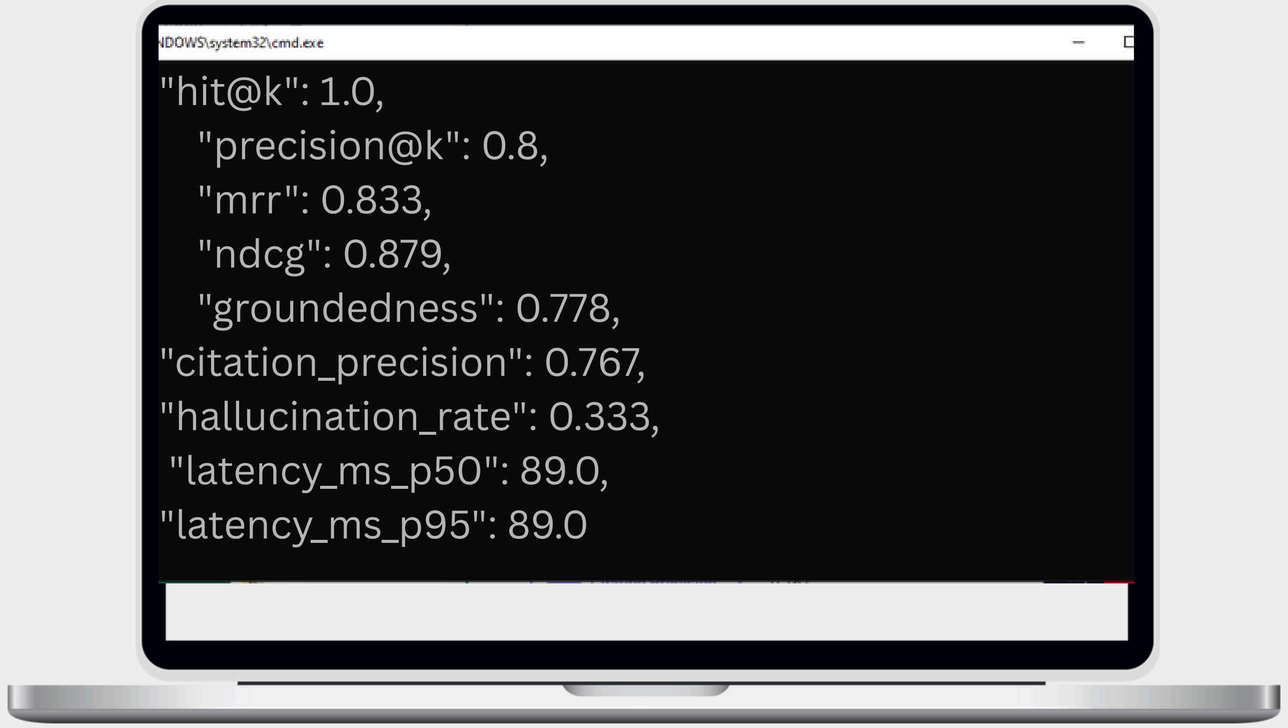
pdf2image

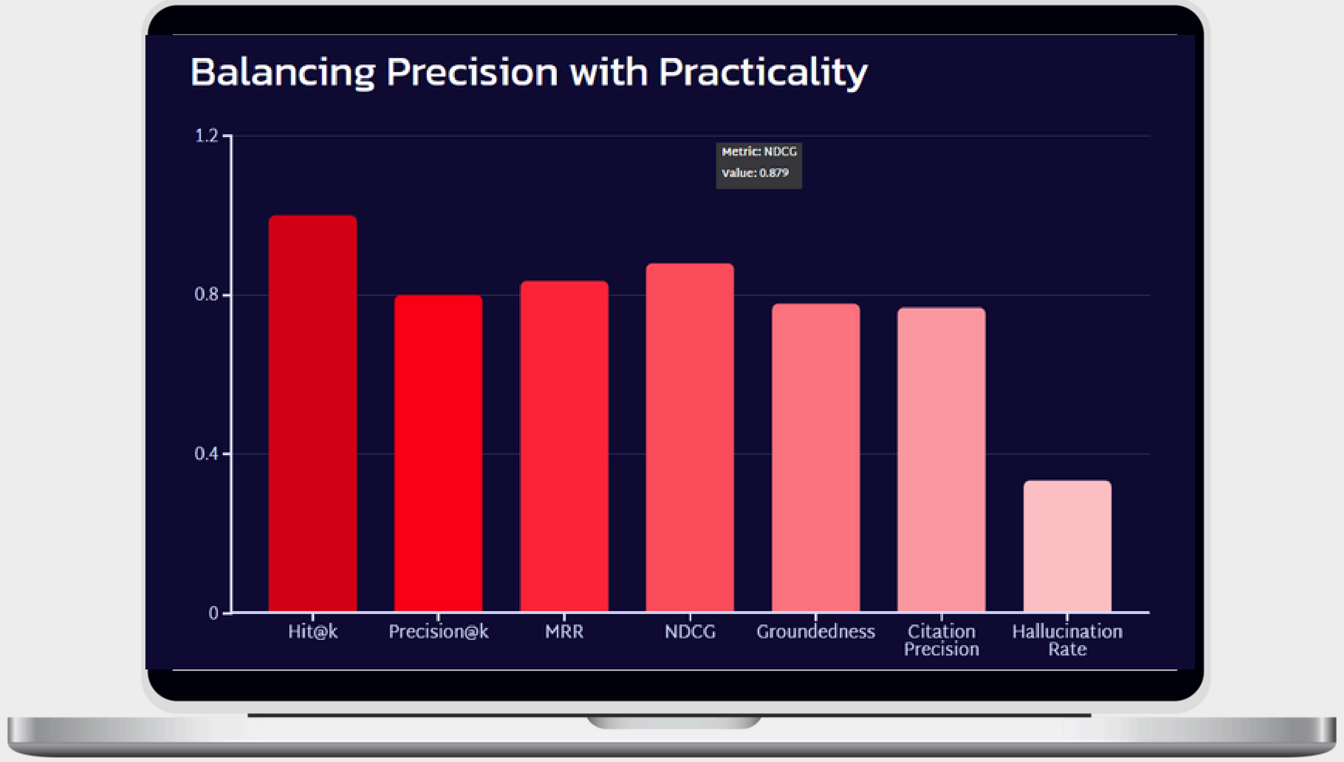


USER QUERY
Can Humaria and Rinvoq be used together? Are there any interactions?
Data Ingestion
Skyrizi_pi.pdf
Humira.pdf
Rinvoq_pi.pdf
Document Parsing
Identifying Texts
Identifying Tables
Parsing Other elements
Structured Representation
Preprocessing
Text Cleaning
Normalising White Spaces
Removing Headers and Footers
Chunking
Preprocessed Data
Data Chunks
Generating Embeddings
Data Chunks
Embeddings
Chroma
VECTOR DATABASE AND INDEXING
Embeddings
Meta Data
Meta Data
Source File
Page No
Title
1 DATA INGESTION AGENT
RETRIEVAL
NO
INTENT AND ENTITY RECOGNITION
USING LLMS
INTENT : drug interaction
ENTITIES : "Humaria", "Rinvoq"
VECTOR SEARCH SERVICE
Search the chroma DB to find the top K embeddings
AGENTIC ORCHESTRATOR
crewai
THOUGHT - To answer this, I need to . . . .
ACTION :
1. VECTOR SEARCH - "Humaria Interactions"
2. VECTOR SEARCH - "Rinvoq Interactions"
Combine the results
2 RETREIVAL AND QUERY ROUTING AGENT
YES
5 USER INTERFACE
USER NOT SATISFIED WITH RESPONSE
Streamlit
SESSION MANAGEMENT
redis
PROMPT BUILDER
FINAL PROMPT
FINAL RESPONSE
USER QUERY + RETREIVAL AGENT RESULT
RELATION EXIST
YES
3 REASONING AND DOMAIN AGENT
4 ORCHESTRATION & ANSWER GENERATION

Evaluation **Metrics:**

Metric	Value	Description
Hit@k	1.0	Correct drug always retrieved
Precision@k	0.80	80% relevant results retrieved
MRR	0.833	Correct answer ranked near top
NDCG	0.879	High ranking quality of answers
Groundedness	0.778	Most answers well-supported by sources
Citation Precision	0.767	Citations were ~77% accurate
Hallucination Rate	0.333	~1 in 3 responses unverified
Latency (p50/p95)	89 ms	Fast responses

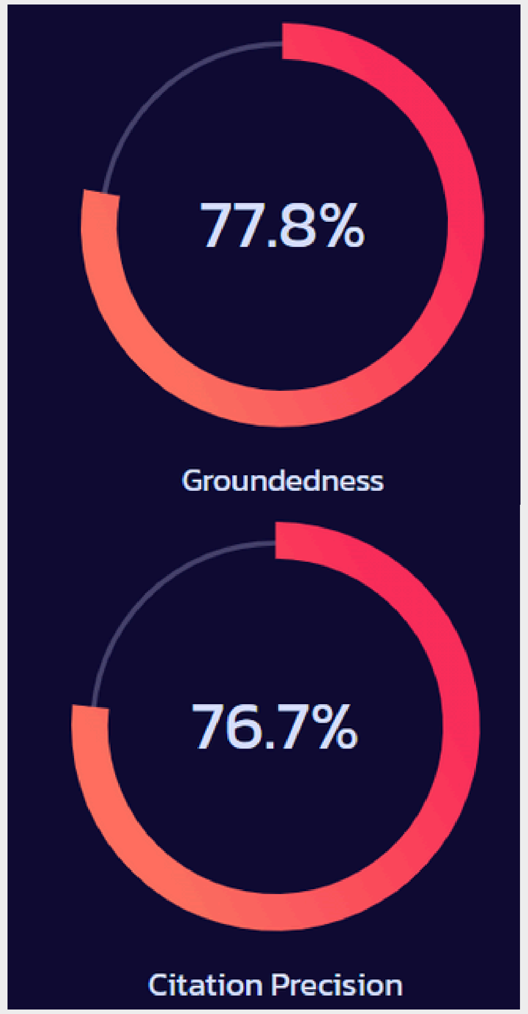




Accuracy
Ensuring correct drug information is delivered.

Efficiency
Providing timely responses to critical queries.

Trust
Building confidence in AI-driven medical tools.



Thank You