

Evidence grounded RAG system for analysing Antidepressant safety literature with sentiment analysis

Siya Asthana
Supervisor: Dr Khalil Saadat, Dr Viraj Dawarka
GradEX

Final Year Project Poster | BSc Computer Science | Staffordshire University London Digital Institute

Problem:

Biomedical research on antidepressant safety is spread across papers, trial records and academic databases. This makes it difficult to quickly identify reliable evidence. General AI chatbots can summarise information, but their answers may be unsupported, unclear, or difficult to trace back to sources.



Design priority: Simple interface with citations and generated conclusions displayed parallelly to show their relationship and encourage the user to use their own judgment through further reading.

Technology Stack

Python

OpenAI API

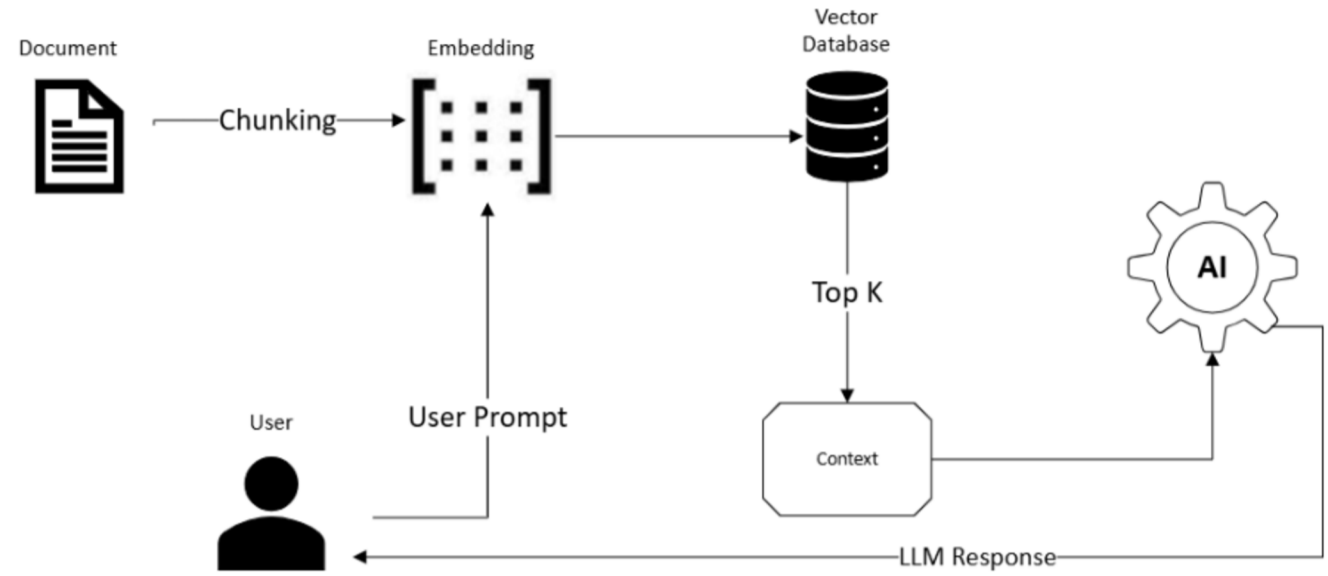
PubMed / ClinicalTrials

FAISS Vector DB

Streamlit

Excel

Figure 1: RAG architecture diagram



Aim:

To develop a safe and transparent AI based RAG system to search specific medical research sites to find relevant papers on antidepressant safety, do knowledge mining and present the output with sentiment analysis.

Objectives:

1. Improve accessibility to antidepressant safety
2. Reduce misinformation
3. Increase trust in AI systems
4. Demonstrate how RAG architecture gives safer biomedical research outputs than unconstrained LLMs
5. Enable safer interpretations of antidepressant safety by highlighting any uncertainties, conflicting evidence and limitations that are found.

Perform embedding on user query turning it to numerical format = Understanding the meaning behind the text to allow for semantic searching.

FAISS uses nearest number searching to identify the chunks closest aligned with the query embedding.

Top scoring chunks are passed to the LLM to be used for the output.

Figure 2: Prototype Interface



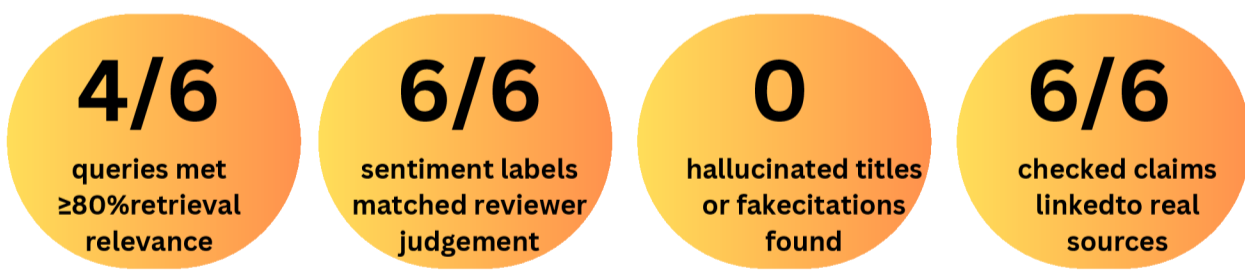
Testing:

Area	Example query
Drug safety	Fluoxetine adverse effects in young adults
Withdrawal/discontinuation	SSRI withdrawal symptoms
Comparative tolerability	Sertraline vs escitalopram safety comparison

Evaluation Methods:

1. Retrieval relevance
2. Sentiment Accuracy
3. Hallucination Rate
4. Traceability
5. Summary quality

Results:



Interpretation: The strongest finding was not perfect retrieval in every case, but consistent traceability. Even when retrieval relevance was lower, the system exposed sources and evidence for checking.

Limitations:

- Evaluation was small-scale
- Manual verification was carried out by 1 reviewer
- Retrieval depended on source quality
- Sentiment also simplified nuanced biomedical evidence.

Future work:

1. Larger evaluation
 - a. More queries, reviews + baseline LLMs
2. Usability testing
 - a. To test whether trust increased
3. Further sentiment categories
 - a. eg. mixed evidence, uncertain + safety concern
4. Sentence aware chunking boundaries
5. More complex configurations for the user
 - a. eg. study type, publication year, population group, antidepressant class

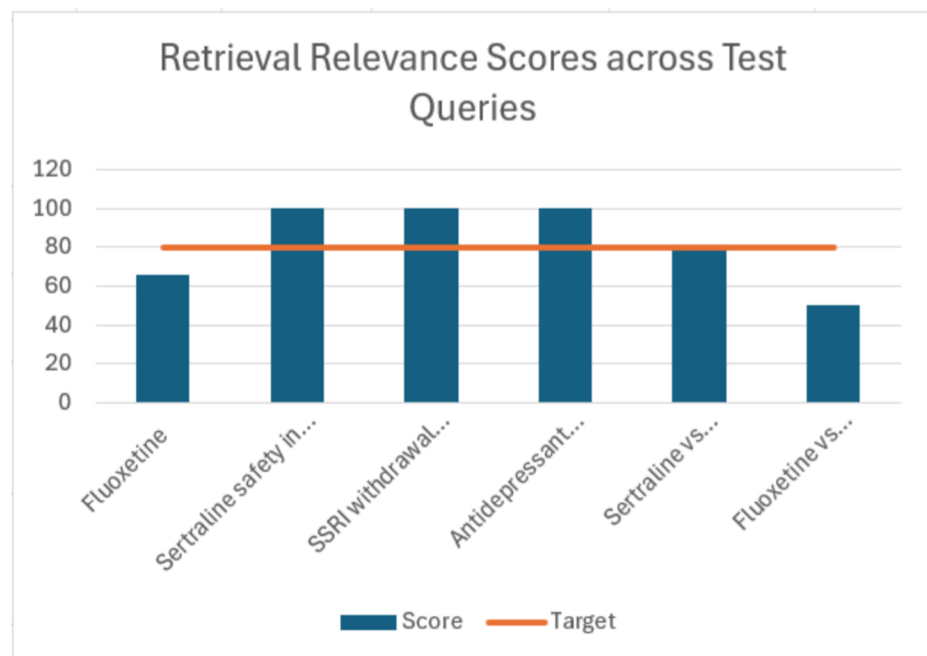
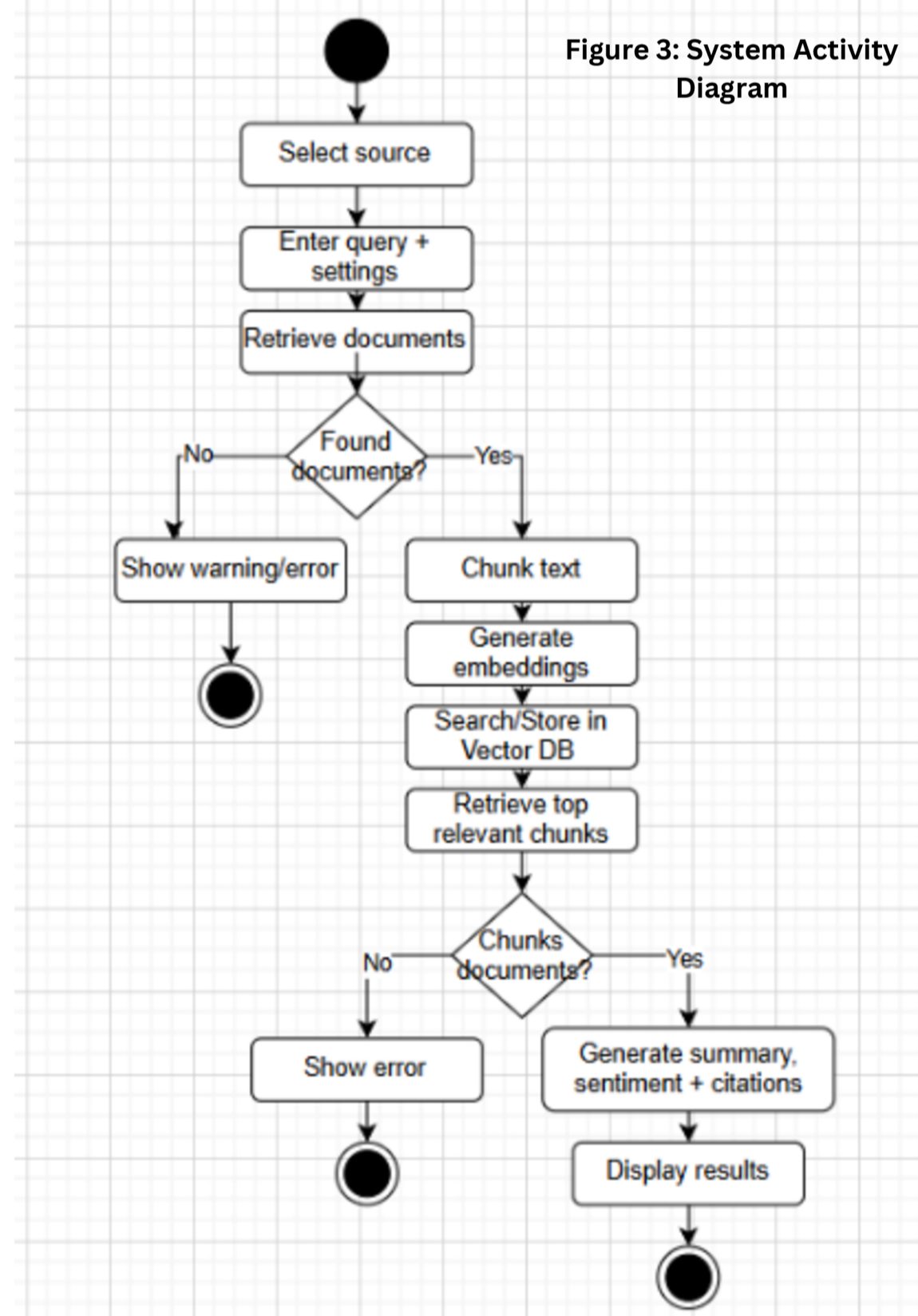


Figure 3: System Activity Diagram



Conclusion: RAG is not a perfect solution, but it provides a stronger foundation for trustworthy AI-assisted biomedical research support than an unconstrained LLM alone.

References:
Lewis, P. et al. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.
NCBI (n.d.) PubMed. National Center for Biotechnology Information.
ClinicalTrials.gov (n.d.) ClinicalTrials.gov. U.S. National Library of Medicine.
OpenAI (n.d.) OpenAI API Documentation.
Meta AI (n.d.) FAISS: A library for efficient similarity search.

Note: This project is for literature summarisation and research support only; it is not medical advice.