Application of Large Language Models (LLMs) for Electronic Health Records summarisation and patient search

Leilu Du K24003142@kcl.ac.uk

Background

- **LLMs** have **strong generalization** and emergent capabilities[1], have provided new impetus for the intelligent development of medical care.
- Improving the information retrieval efficiency of doctors could optimize the clinical decision-making process[2], and **enhance the practicality** and **usability** of electronic health records (EHRs).
- **RAG (Retrieval-Augmented Generation)** could inject external knowledge[3] into the model, enhancing **accuracy and relevance**.
- LLMS have **"hallucinations"**[4], and the accuracy problem is significant.

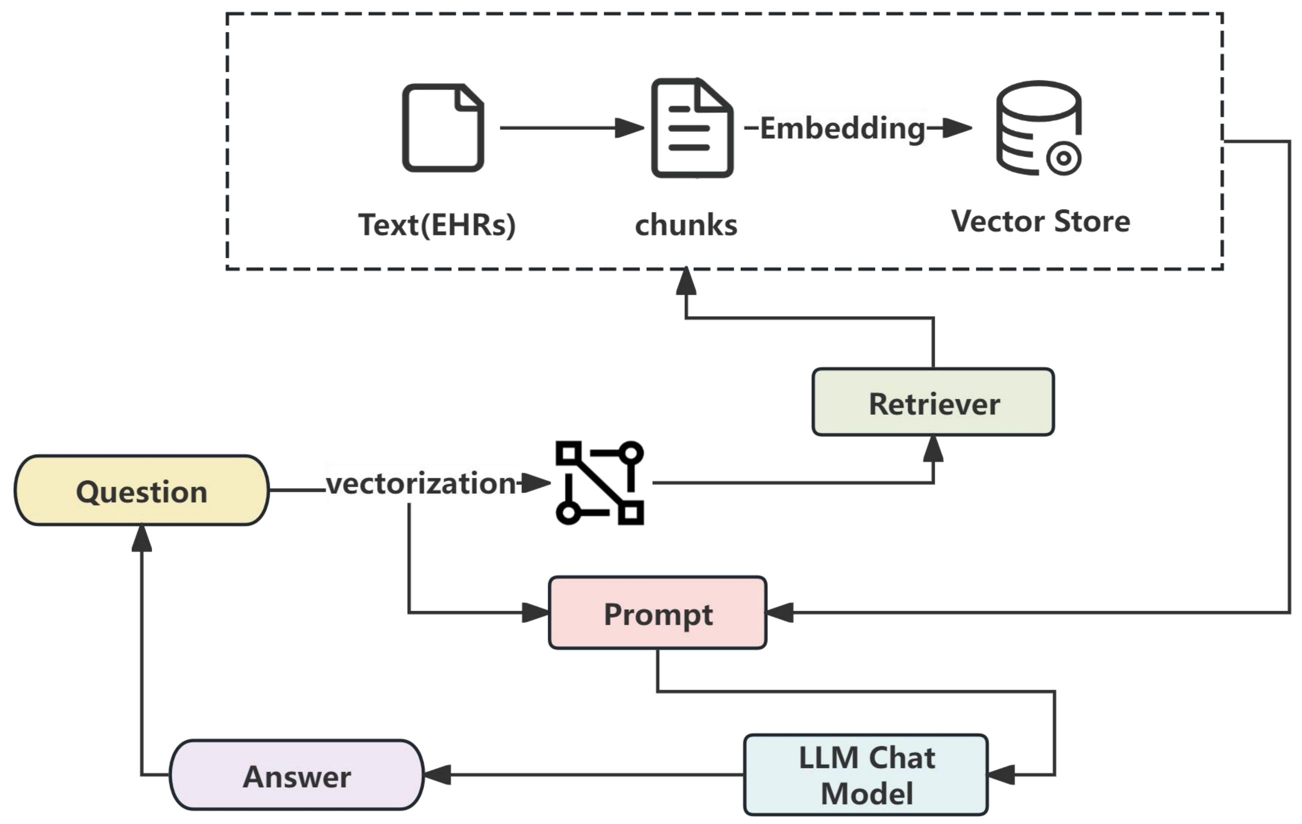
Aim:

Develop a **RAG question-answering system** that combines unstructured EHRs data of patients.

Data: MIMIC-IV

Ethics Statement:
The MIMIC-IV database used in this study is publicly available and de-identified medical data. All uses follow the official data usage agreement and relevant ethical training has been completed.

Study Design



- **Data preprocessing:**
  - Select 1000 sets of subset data
  - Remove unnecessary Spaces and line breaks
- **Embedding：**
  - Using the MiniLM embedding model
  - Storing in a vector database
- **RAG Pipeline：**
  - Retrieve relevant text chunks using a LlamaIndex-based retriever
  - Improve relevance via Cohere Rerank
  - Combine query and context into a prompt for Cohere-chat-model to generate an evidence-based answer
- **RAG Evaluation:**
  - 10 manually designed questions were tested
  - Answers were manually reviewed for accuracy and citation relevance

Result

Take the medical record data of the patient with note-id 16255266-DS-3 as an example. Examples are illustrated with text snippets from clinical notes and corresponding QA pairs.

Name: ___. Unit No: ___
Admission Date: ___ Discharge Date: ___
Date of Birth: ___ Sex: M
Service: CARDIOTHORACIC
Allergies:
___ recorded as having No Known Allergies to Drugs
Attending: ___
Chief Complaint:
exertional chest pain
...

Q:What is the patient's age and sex?(√)
A: Male and Unknown.
D: Date of Birth: ___ Sex: M

Q: What is the chief complaint on admission? (×)
A: The chief complaint on admission was acute blood loss anaemia and GI bleeding.
D: exertional chest pain

Q:What is the patient's past medical history?(√)
A:The patient's past medical history includes:
- HTN
- COPD

D: Past Medical History:
HTN
COPD

Q: What were the findings of the cardiovascular and respiratory examinations? (×)
A: A 62-year-old male with chest pain , DOE, and palpitations. Catheterization showed LM and 2VD. He underwent an off-pump CABG x 2.
D:Physical Exam:
___ yo M in NAD
HR 62 RR 24 BP 110/64
Lungs CTAB
Heart RRR no Murmur
Abdomen soft/NT
No varicosities
Pulses 2+ t/o
No carotid bruits
Pertinent Results:
___ 01:15PM BLOOD WBC-11.5* RBC-3.12* Hgb-9.7* Hct-28.4*
MCV-91 MCH-31.0 MCHC-34.1 RDW-13.4 Plt ___
___ 01:15PM BLOOD Plt ___
___ 04:56AM BLOOD ___ PTT-24.4 ___
___ 01:15PM BLOOD Glucose-108* UreaN-10 Creat-0.8 Na-140

- The system produced accurate and relevant answers for **6 out of 10 questions**, yielding an overall **accuracy of 60%**.

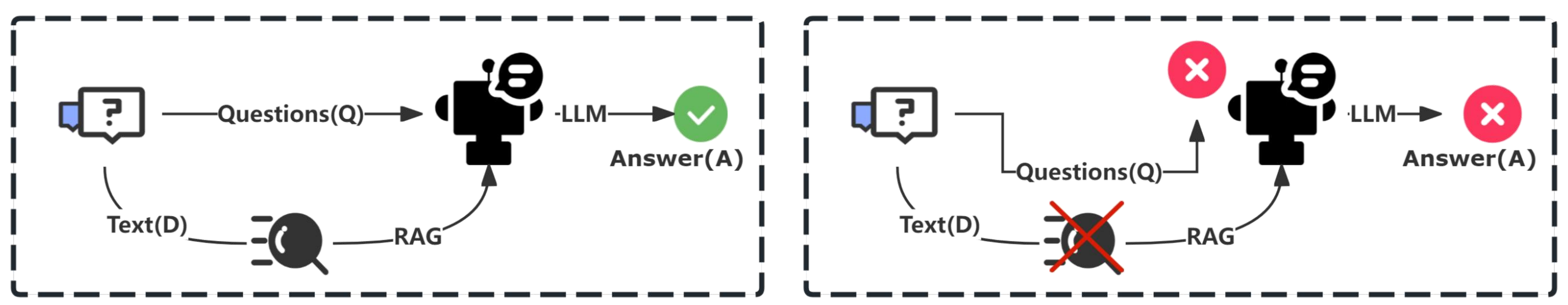


Diagram of potential failure points in the RAG process

Conclusion

- This project demonstrates the feasibility of using a RAG pipeline for clinical question answering based on EHRs.
- While the system performed **well on most queries**, several failure cases were observed where irrelevant or insufficient content was retrieved, even when the input questions were clearly defined.
- These issues are likely related to limitations **in the embedding model's ability** to capture clinical semantics, **suboptimal text chunking strategies**, or **retrieval mechanisms** that fail to align with the clinical context.

Discussion

Strengths:
- A **question-answering system** based on RAG was constructed.
- Preliminary experiments were conducted on real clinical discharge summary data.

Limitations:
- This study only used a subset of clinical data, and the data application was incomplete.
- The current question-answering system is still in the initial prototype stage and does not support multi-round conversations or **integration with structured data**. Its functions are not yet complete.
- the performance of the underlying language model remains **suboptimal,** with **occasional hallucinations** and **inaccurate responses** observed.

Future Work:
- Apply the current process to the **complete clinical discharge record dataset** to **enhance the system's coverage and generalization ability.**
- Continue to develop **a more complete question-answering system.**
- Combine **sparse retrieval** and **dense retrieval** to improve both keyword matching and semantic understanding during document retrieval.
- Explore how to **better integrate** this question-answering system into the clinical workflow.

References

[1] Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., Yang, L., Zhang, W., Jiang, J., & Cui, B. (2024). *Retrieval-Augmented Generation for AI-Generated Content: A Survey.* https://doi.org/10.48550/arxiv.2402.19473
[2] Ramadhan, A. J., Mohammed, S. Y., Aljanabi, M., Mijwil, M. M., Abotaleb, M., Alkattan, H., & Dutta, P. K. (2024). Enhancing EHR Analysis: Leveraging RAG-Enabled Generative AI for Clinical Data Summarization. *Library of Progress-Library Science, Information Technology & Computer, 44*. https://doi.org/10.48550/arXiv.2501.07391
[3] Rani, M., Mishra, B. K., Thakker, D., & Khan, M. N. (2024). To enhance graph-based retrieval-augmented generation (RAG) with robust retrieval techniques. *2024 18th International Conference on Open Source Systems and Technologies (ICOSST)*, 1-6. https://doi.org/10.1109/icosst64562.2024.10871140
[4] Lu, Y., Zhao, X., & Wang, J. (2024). ClinicalRAG: Enhancing Clinical Decision Support through Heterogeneous Knowledge Retrieval. *In Proceedings of the 1st Workshop on Towards Knowledgeable Language Models (KnowLLM 2024)*, 64–68. https://doi.org/10.18653/v1/2024.knowllm-1.6