University of Texas at Austin City Vermont rajesh.kohli@utexas.com

Abstract

We present a risk-taking project that implements a retrieval-augmented QA system over clinical sepsis guidelines. We ingest a PDF, chunk it, embed with OpenAI embeddings, index with FAISS, and build a chat-based interface. We demonstrate grounded, accurate answers to clinical queries. Our code is available at `<your-repo-url>`.

Retrieval-Augmented Question Answering over Sepsis Management Guidelines

Rajesh Kohli

April 29, 2025

1 Introduction

Sepsis—a life-threatening organ dysfunction caused by a dysregulated host response to infection—remains one of the leading causes of mortality in hospitalized patients, with reported mortality rates between 30% and 80% for severe sepsis and septic shock [1]. Clinicians face an overwhelming amount of ever-evolving guideline recommendations (e.g., fluid resuscitation, vasopressor initiation, antimicrobial timing), making rapid, accurate retrieval of the latest evidence challenging. Delays or deviations from protocolized care bundles have been shown to significantly worsen patient outcomes [2].

In this work, we present a risk-taking pilot project that implements a retrieval-augmented question-answering (QA) system over institutional sepsis management guidelines. By ingesting the full PDF, chunking by token count, embedding text with OpenAI's embedding model, and indexing with FAISS, our system can ground clinician queries in exact guideline snippets. We demonstrate high-fidelity answers to real clinical questions—such as mortality ranges and vasopressor thresholds—directly sourced from the document. This approach may shorten the time to guideline access and reduce cognitive load during critical decision-making.

2 Related Work

Discuss 2–3 retrieval-augmented methods in healthcare or similar domains (e.g., Retrieval-Augmented Generation, ClinicalBERT retrieval pipelines).

3 Methodology

- PDF ingestion with PyPDF2
- Token-based chunking (500 tokens, 50 overlap)
- OpenAI embeddings (text-embedding-ada-002)
- FAISS indexing (L2 Flat)
- Chat completion prompt design

Include a workflow figure (e.g., a simple block diagram).

4 Results

Show sample Q&A sessions:

- Mortality range → "30–80%"
- Vasopressor timing → "MAP¡65mmHg after fluids"
- ...

5 Conclusion & Future Work

Summarize and propose:

- Improving chunking via smarter overlap
- Adding UI (web app)
- Extending to other guidelines

References

[1] Andrew Rhodes, Lewis E Evans, Waleed Alhazzani, and et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive care medicine*, 43:304–377, 2017.

[2] Emanuel Rivers, Ba Nguyen, Steve Havstad, and et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *New England Journal of Medicine*, 345(19):1368–1377, 2001.