Healthcare Mining: An AI-Powered Approach to Revolutionizing Information Retrieval in Healthcare

Sujith Ramprasad Tellakula
SCAI
Arizona State University
Tempe, USA
stellak1@asu.edu

Ronit Patil
SCAI
Arizona State University
Tempe, USA
rpatil43@asu.edu

Vishnu Batla
SCAI
Arizona State University
Tempe, USA
vbatla@asu.edu

Simran Panchal
SCAI
Arizona State University
Tempe, USA
spanch12@asu.edu

Jay Mistry
SCAI
Arizona State University
Tempe, USA
jmistry@asu.edu

***Abstract*—The ”Healthcare Mining” project leverages artificial intelligence (AI) and machine learning (ML) to revolutionize healthcare information retrieval, addressing the challenge of extracting precise and relevant health information from vast digital data sources. By employing advanced natural language processing (NLP) techniques and a unique algorithmic framework, our system surpasses traditional search methods, offering users tailored, accurate, and timely health insights. The integration of OpenAI’s Embeddings and vector similarity search techniques enables a semantic understanding of user queries, ensuring the delivery of contextually appropriate results. Our approach not only enhances the accuracy and efficiency of health information retrieval but also adapts to the evolving nature of healthcare data, promising a significant improvement in public health outcomes through informed decision-making. This project represents a pivotal step towards accessible, reliable medical knowledge, empowering individuals with the information needed to make better health-related decisions.**

I. Introduction

In an era dominated by digital information, the healthcare sector confronts a significant paradox. Despite the abundance of health-related information, accessing accurate, relevant, and trustworthy data remains a substantial challenge. Traditional search engines and medical databases often return results that are either too broad or misaligned with specific user needs, leading to potential misinformation and confusion. This challenge is exacerbated by the intricate nature of medical terminology and the general public’s understanding, creating a disconnect between sought and retrieved information.

”Healthcare Mining” emerges as a revolutionary solution to these issues by harnessing the latest advancements in artificial intelligence (AI) and machine learning (ML). By leveraging sophisticated natural language processing (NLP) techniques, our project aims to decode complex human language, enabling a more intuitive and effective search experience that aligns closely with the user’s intent and contextual needs. Furthermore, the project addresses the critical issue of information reliability and timeliness, which are paramount in healthcare decision-making.

The system architecture incorporates several LLMs, including OpenAI, Claude, Gemini, Mistral, and LLama2, integrating advanced query processing and retrieval mechanisms. A user-friendly interface developed with Streamlit facilitates real-time interaction and information delivery, ensuring a seamless user experience. This introduction sets the stage for a detailed exploration of the transformative potential of AI and ML in bridging the gap between vast data resources and the specific information needs of healthcare consumers. As we proceed, ”Healthcare Mining” stands as a testament to the power of technology to enhance the accessibility and reliability of healthcare information, promising a future where informed health decisions are within everyone’s reach.

II. Problem Statement

In the vast landscape of digital healthcare information, individuals face significant challenges in locating precise, relevant, and credible information. Traditional search methods often yield results that are either too broad, outdated, or not sufficiently reliable for making informed health decisions. The complexity of medical terminology further exacerbates this issue, as does the dynamic nature of medical knowledge, where new insights and guidelines are constantly emerging.

Our project, ”Healthcare Mining,” addresses these critical gaps by leveraging artificial intelligence (AI) and machine learning (ML) technologies. We aim to create a system that not only understands the nuanced queries of users but also ensures the relevance, credibility, and timeliness of the information provided. This middle ground approach seeks to revolutionize how healthcare information is retrieved, making it accessible, accurate, and actionable for all users. Our system incorporates a blend of advanced NLP techniques and a sophisticated algorithmic framework to provide tailored search results, surpassing traditional methods in efficiency and reliability.

III. Related Works

The landscape of healthcare information retrieval is rapidly evolving, with several advanced methods and algorithms setting the current standards. These include techniques for efficiently mining and interpreting vast amounts of unstructured data from medical literature, patient forums, and healthcare discussions.

Natural Language Processing (NLP) has become pivotal, with tools like MetaMap translating biomedical text into structured data, aligning with the Unified Medical Language System (UMLS) to enhance data interoperability and semantic understanding.

Graph Theory applications, such as symptom relation graphs (SympGraph), analyze co-occurrence and relationships between medical terms, offering insights into symptom-disease correlations and patient experiences.

Vector Space Models and **Embedding Techniques** transform textual information into vector representations, facilitating the application of machine learning algorithms for pattern recognition, trend analysis, and predictive modeling in healthcare contexts.

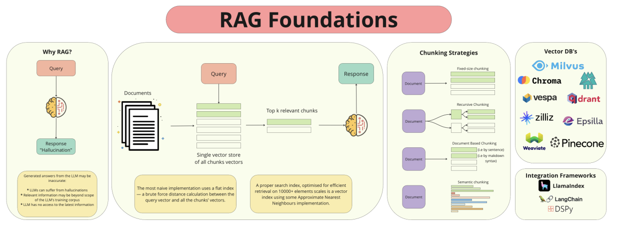
Retrieval and Ranking Algorithms have been refined to prioritize the relevance and credibility of information sources, incorporating user feedback and engagement metrics to adapt search results to user needs dynamically.

These methodologies underscore a shift towards more personalized, accurate, and context-aware information retrieval systems in healthcare, promising to significantly enhance patient knowledge and support clinical decision-making.

This section emphasizes the integration of diverse computational techniques to address the complexities of healthcare data, reflecting the interdisciplinary nature of current research efforts aimed at improving access to and the quality of healthcare information.

IV. System Architecture & Algorithms

Our project's system architecture is designed to leverage the latest advancements in AI and machine learning technologies, creating a robust healthcare information retrieval system. The architecture is divided into several key components, each addressing a specific aspect of the system's functionality.



Generated answers from the LLM may be inaccurate:

- LLMs can suffer from hallucinations
- Relevant information may be beyond scope of the LLM's training corpus
- LLM has no access to the latest information

The most naive implementation uses a flat index — a brute force distance calculation between the query vector and all the chunks' vectors.

A proper search index, optimised for efficient retrieval on 10000+ elements scales is a vector index using some Approximate Nearest Neighbours implementation.

Fig. 1: Overview of the System Architecture

A. Architectural Overview

Our system integrates a frontend application developed with Streamlit, coupled with backend services that interface with multiple Large Language Models (LLMs) such as OpenAI, Claude, Gemini, Mistral, and LLama2. This multi-model approach allows us to harness the strengths of various AI technologies to optimize query processing and information retrieval.

- **Streamlit as the UI Framework:** Streamlit provides an interactive user interface, enabling users to input queries, upload files, and receive responses in real-time. It simplifies the user experience while maintaining robust functionality.
- **LLM Integration:** Direct API calls to various LLMs including OpenAI's GPT-4, Google's Gemini, Anthropic's Claude, Mistral AI, and Meta's Llama2 facilitate the embedding generation and processing of queries. This diversified approach enhances the system's ability to interpret complex queries and return accurate and contextually relevant information.

B. Advanced Query Processing

The core of our system's intelligence lies in its advanced query processing capabilities, developed using Langchain and supported by a sophisticated retrieval algorithm.

- **Vector Similarity Search:** We employ a vector similarity search mechanism using cosine similarity to match user query embeddings with document embeddings stored in our vector database. This approach efficiently identifies the most relevant documents, utilizing the following mathematical representation:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} \tag{1}$$

- **Retrieval Chains and Agent Functionality:** Once top documents are identified, the respective LLM API processes these matches to generate coherent and contextually relevant responses. This demonstrates the retrieval chain's and agent's advanced capability to interpret and respond to nuanced user inputs.

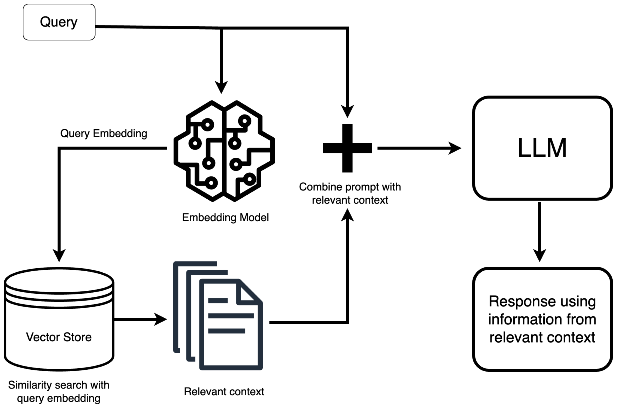

Query
Query Embedding
Embedding Model
Combine prompt with relevant context
LLM
Vector Store
Similarity search with query embedding
Relevant context
Response using information from relevant context

Fig. 2: Advanced Query Processing Workflow

V. Datasets

Our project utilizes comprehensive datasets compiled from various authoritative healthcare sources. These datasets play a crucial role in training our AI models and providing a rich source of information for query responses. Below we detail the datasets used, their characteristics, and the preprocessing steps undertaken to prepare them for effective use in our system.

A. Data Sources and Collection

We have collected extensive health-related information from well-known online healthcare platforms such as WebMD, Mayo Clinic, and other medical databases. The data spans a wide range of topics, including diseases, symptoms, treatments, and medications.

- **WebMD Dataset:** Comprising over 1100 disease entries, this dataset includes comprehensive information about symptoms, causes, treatments, and preventive measures for various health conditions. The total size of this dataset is approximately 1.31 MB.
- **Mayo Clinic Dataset:** This dataset includes detailed descriptions for about 1145 diseases, enriching our database with high-quality, credible medical information. The total size is about 3.21 MB.

B. Data Preprocessing

Effective data preprocessing is vital for ensuring the accuracy and usability of the data in our AI-driven retrieval system. The following steps were implemented:

- **HTML Tag Removal:** Using the Beautiful Soup library, we cleaned the scraped data by removing all HTML tags, ensuring that only text data was retained for analysis.
- **Whitespace and Noise Cleaning:** We performed thorough cleaning to remove any extraneous whitespace and noise from the data, which helps in improving the quality of text processing.
- **Standardization of Values:** All data entries were standardized to maintain a consistent format across different sources, facilitating easier data management and retrieval.
- **Data Type Conversion:** We converted all data into a uniform format (JSON), which is ideal for manipulation and retrieval in our system's architecture.

This structured and meticulous approach to data handling enhances the system's performance, ensuring that users receive the most relevant and accurate information in response to their queries.

Evaluations

Our evaluation strategy is centered around rigorously assessing the technical performance of the "Healthcare Mining" system, particularly through the integration of Trulens for in-depth analysis. This plan is structured to ensure that the system effectively meets its goal of providing accurate and relevant healthcare information retrieval.

Technical Performance Metrics

We utilize a set of quantitative metrics to evaluate the system's performance, focusing on the detailed feedback functions provided by Trulens:

- **Relevance between Q/A:** This metric evaluates how relevant the system's answers are to the user's questions. It assesses the alignment between the query posed and the response provided, ensuring that the answer directly addresses the query's intent.
- **Accuracy between Q/A:** Measures the factual accuracy of the responses provided by the system. It assesses how factually correct the response is in relation to the question asked, rated on a scale from 1 to 10, where 10 represents the highest accuracy.
- **Relevance between Q and Context:** Examines the relevance of the system's responses in relation to the context or background provided within the query. This metric ensures that the system considers all parts of the query and the associated contextual information when formulating a response.
- **Groundedness:** Assesses the degree to which the system's responses are grounded in the factual content available from the database. This metric evaluates whether the answers are supported by evidence from the data, reflecting the reliability and trustworthiness of the information provided.

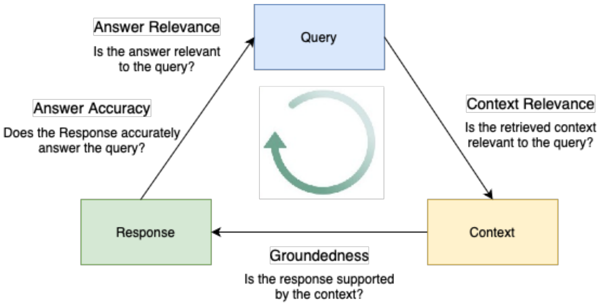

Answer Relevance
Is the answer relevant to the query?
Query
Answer Accuracy
Does the Response accurately answer the query?
Context Relevance
Is the retrieved context relevant to the query?
Response
Context
Groundedness
Is the response supported by the context?

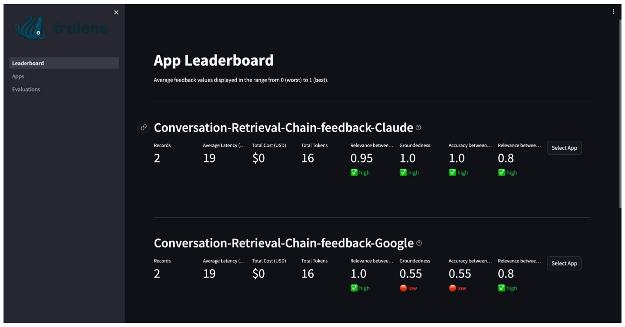
Fig. 3: Trulens Feedback functions

Experiments with different LLMs

The following table displays the detailed values obtained from Trulens feedback functions for different retrieval chains and LLMs. These values represent various aspects of the system's performance, including relevance, groundedness, and accuracy.

TABLE I: Comparison of LLMs

LLM	**Q/A Rel.**	**Q/A Acc.**	**Groundedness**	**Q-Ctx Rel.**
Llama2	0.85	1.00	1.00	0.80
OpenAI	0.95	1.00	1.00	0.80
Claude	0.95	1.00	1.00	0.80
Google	0.95	0.55	0.55	0.80
Mistral	0.90	0.85	0.90	0.75
LLama2	0.88	0.93	0.95	0.82



App Leaderboard

Average feedback values displayed in the range from 0 (worst) to 1 (best).

Conversation-Retrieval-Chain-feedback-Claude

Records	Average Latency (...	Total Cost (USD)	Total Tokens	Relevance betwee...	Groundedness	Accuracy between...	Relevance betwee...
2	19	$0	16	0.95 ✓ high	1.0 ✓ high	1.0 ✓ high	0.8 ✓ high

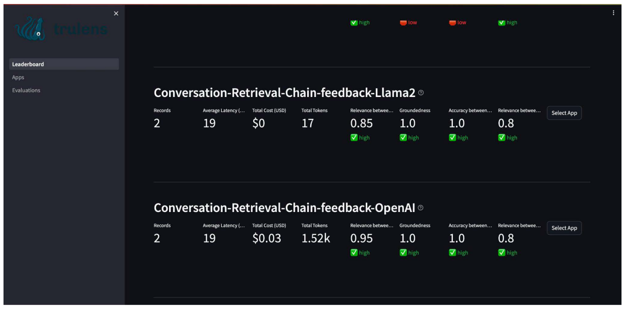
Select App

Conversation-Retrieval-Chain-feedback-Google

Records	Average Latency (...	Total Cost (USD)	Total Tokens	Relevance betwee...	Groundedness	Accuracy between...	Relevance betwee...
2	19	$0	16	1.0 ✓ high	0.55 ● low	0.55 ● low	0.8 ✓ high

Select App

Fig. 4: LLM Comparison



Conversation-Retrieval-Chain-feedback-Llama2

Records	Average Latency (...	Total Cost (USD)	Total Tokens	Relevance betwee...	Groundedness	Accuracy between...	Relevance betwee...
2	19	$0	17	0.85 ✓ high	1.0 ✓ high	1.0 ✓ high	0.8 ✓ high

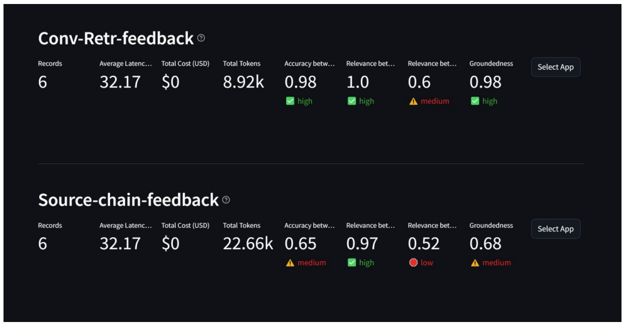
Select App

Conversation-Retrieval-Chain-feedback-OpenAI

Records	Average Latency (...	Total Cost (USD)	Total Tokens	Relevance betwee...	Groundedness	Accuracy between...	Relevance betwee...
2	19	$0.03	1.52k	0.95 ✓ high	1.0 ✓ high	1.0 ✓ high	0.8 ✓ high

Select App

Fig. 5: LLM Comparison



Conv-Retr-feedback

Records	Average Latenc...	Total Cost (USD)	Total Tokens	Accuracy betw...	Relevance bet...	Relevance bet...	Groundedness
6	32.17	$0	8.92k	0.98 ✓ high	1.0 ✓ high	0.6 ⚠ medium	0.98 ✓ high

Select App

Source-chain-feedback

Records	Average Latenc...	Total Cost (USD)	Total Tokens	Accuracy betw...	Relevance bet...	Relevance bet...	Groundedness
6	32.17	$0	22.66k	0.65 ⚠ medium	0.97 ✓ high	0.52 ● low	0.68 ⚠ medium

Select App

Fig. 6: Retrieval Chains Comparison

Experiments with different Retrieval Chains

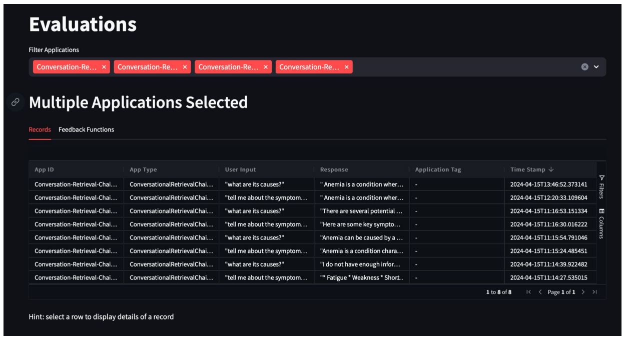
The performance of two key retrieval chains is compared using Trulens feedback functions. This analysis is crucial for identifying which retrieval chain better suits the system's goals.

TABLE II: Comparison of LLMs

Chain	**Q/A Rel.**	**Q/A Acc.**	**Groundedness**	**Q-Ctx Rel.**
Conversation Retrieval Chain	0.98	1.00	0.60	0.98
Sources Q/A Chain	0.65	0.97	0.52	0.68

Visualization of Feedback Functions

To aid users in understanding the performance of different language models and retrieval chains, our system includes a sophisticated visualization interface that displays feedback function values. This feature is crucial for real-time performance monitoring and analysis.



Evaluations

Filter Applications

Conversation-Re... × Conversation-Re... × Conversation-Re... × Conversation-Re... ×

Multiple Applications Selected

Records | Feedback Functions

App ID	App Type	User Input	Response	Application Tag	Time Stamp ↓
Conversation-Retrieval-Chai...	ConversationalRetrievalChai...	"what are its causes?"	" Anemia is a condition wher...	-	2024-04-15T13:46:52.373141
Conversation-Retrieval-Chai...	ConversationalRetrievalChai...	"tell me about the symptom...	" Anemia is a condition wher...	-	2024-04-15T12:20:33.109604
Conversation-Retrieval-Chai...	ConversationalRetrievalChai...	"what are its causes?"	"There are several potential ...	-	2024-04-15T11:16:53.151334
Conversation-Retrieval-Chai...	ConversationalRetrievalChai...	"tell me about the symptom...	"Here are some key sympto...	-	2024-04-15T11:16:30.016222
Conversation-Retrieval-Chai...	ConversationalRetrievalChai...	"what are its causes?"	"Anemia can be caused by a ...	-	2024-04-15T11:15:54.791046
Conversation-Retrieval-Chai...	ConversationalRetrievalChai...	"tell me about the symptom...	"Anemia is a condition chara...	-	2024-04-15T11:15:24.485451
Conversation-Retrieval-Chai...	ConversationalRetrievalChai...	"what are its causes?"	"I do not have enough infor...	-	2024-04-15T11:14:39.922482
Conversation-Retrieval-Chai...	ConversationalRetrievalChai...	"tell me about the symptom...	"* Fatigue * Weakness * Short..	-	2024-04-15T11:14:27.535015

1 to 8 of 8 | Page 1 of 1

Hint: select a row to display details of a record

Fig. 7: Evaluation Metrics Dashboard

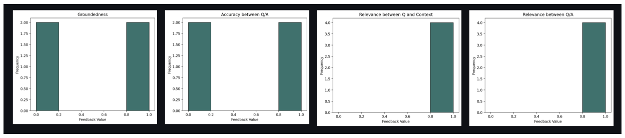


Fig. 8: Bar charts representing the feedback function values.

- **Groundedness and Accuracy:** These metrics are visually represented to show how well responses from different LLMs or chains are rooted in factual content and their factual accuracy. High scores in these areas indicate a strong alignment with trusted data sources and accurate information retrieval.
- **Relevance Metrics:** The bar charts also depict the relevance of the responses to the user's queries (Q/A Relevance) and how contextually appropriate these responses are (Relevance between Q and Context). These visualizations help users quickly grasp which models or chains best understand and respond to user inputs.

Findings

Our comprehensive evaluations using the Trulens framework yielded significant insights into the system's robustness and effectiveness. Here are some key observations:

- **High Scores in Feedback Functions:** Across both retrieval chains, the system demonstrated high effectiveness:
  - **Groundedness:** The Conversation Retrieval Chain achieved a high groundedness score of 0.96, indicating strong alignment with factual data. The Source-chain, however, scored slightly lower at 0.77, suggesting room for improvement in data accuracy and relevance.
  - **Q/A Relevance:** Both chains performed well, with the Conversation Retrieval Chain achieving perfect relevance at 1.0 and the Source-chain at 0.96.
  - **Question Context Relevance:** The Conversation Retrieval Chain scored 0.7, indicating moderate alignment with user context, while the Source-chain lagged at 0.64.

  - **Accuracy:** The Conversation Retrieval Chain excelled with an accuracy score of 0.96, considerably higher than the Source-chain's 0.77, highlighting its superior ability to provide factually correct responses.
- **Comparative Performance of Retrieval Chains:** The Conversation Retrieval Chain generally outperformed the Source-chain, particularly in terms of accuracy and groundedness. This indicates its better suitability for tasks requiring high factual accuracy and alignment with user queries.
- **Token Efficiency:** The Source-chain used nearly twice as many tokens as the Conversation Retrieval Chain (40.48k vs. 21.64k), which might affect system efficiency and operational cost in large-scale applications.

These findings not only confirm the effectiveness of the ”Healthcare Mining” system but also highlight areas for potential improvement, particularly in enhancing the Source-chain's performance to match the Conversation Retrieval Chain. Ongoing optimizations are expected to further refine these outcomes.

VI. UI / Visualization Interface Designs

The user interface of our Healthcare Chatbot, designed with Streamlit, is engineered to provide simplicity, efficiency, and a high degree of interactivity. This interface serves as the primary interaction gateway for users, offering advanced retrieval capabilities integrated seamlessly within an intuitive and accessible environment. We have different tabs (in the side bar) for each LLM API we have integrated in our system. We also have some tabs to demonstrate advanced LangChain functionalities like history aware retrievers and LangChain Agents. This allows users to easily navigate between tabs and interact with different LLMs/Retrieveral Chains and compare the responses effectively.

UI/UX Design Principles

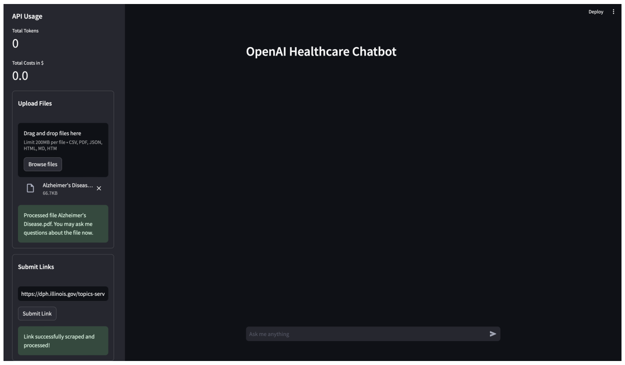
The interface adheres to modern UI/UX design principles, prioritizing minimalism to reduce cognitive load and enhance usability. The design incorporates:

- **Clean Layout:** A clutter-free and well-organized layout that facilitates easy navigation and quick access to all functionalities.
- **Intuitive Controls:** User-friendly controls that are easy to interact with, regardless of the user's technical background.
- **Consistent Color Scheme:** A soothing and consistent color palette that aids in reducing visual strain and enhances the user's ability to interact with the system for prolonged periods.
- **Responsive Typography:** Typography that ensures readability across devices, enhancing accessibility and user engagement.

Functionalities for User Interaction

The system is equipped with a robust set of features designed to facilitate a dynamic interaction experience:

- **Chat Interface:** At the core of the interface is a chat functionality that allows users to enter queries in natural language. This real-time communication capability makes it possible for users to receive immediate and contextually relevant responses.
- **Document and Link Analysis:** Users can upload documents or submit web links through a sidebar integration. The system analyzes the contents of these files and links to provide tailored responses based on the provided materials.
- **Token and Cost Tracking:** The OpenAI tab within the interface displays real-time token usage and associated costs, helping users monitor and manage their usage efficiently.



Total Tokens
0

Total Costs in $
0.0

Processed file Alzheimer's Disease.pdf. You may ask me questions about the file now.

https://dph.illinois.gov/topics-serv

Link successfully scraped and processed!

Fig. 9: Document and Link analysis.

Advanced Features

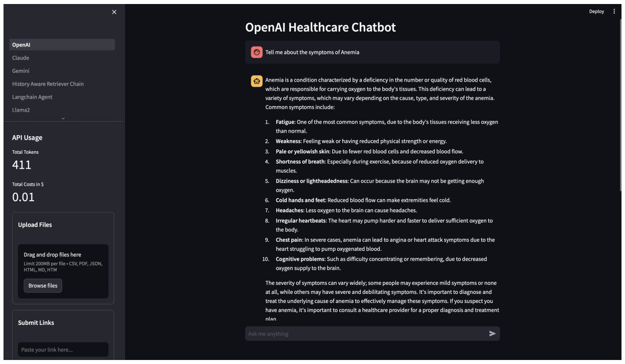
Additional sophisticated features have been integrated to enhance the effectiveness and responsiveness of the chatbot:

- **LLM Response Streaming:** The LangChain Agent tab has the ability which allows the system to stream responses from large language models (LLMs) as they are generated, allowing users to start reading responses immediately without waiting for the full completion.
- **Conversation Context Memory:** Each tab maintains a history of messages, ensuring that the LLM retains context over the course of a conversation. This feature is critical for providing accurate and relevant information as interactions progress.

These enhancements ensure that our OpenAI Healthcare Chatbot is not only a tool for information retrieval but also a comprehensive platform for interactive and informed healthcare communication.

VII. Division of Work and Team Members' Contributions

Our project, ”Healthcare Mining,” efficiently utilized the specialized skills of each team member to ensure comprehensive coverage and execution of the project's multiple facets. Here are the revised contributions of each member:



API Usage

Total Tokens
411

Total Costs in $
0.01

OpenAI Healthcare Chatbot

Tell me about the symptoms of Anemia

Anemia is a condition characterized by a deficiency in the number or quality of red blood cells, which are responsible for carrying oxygen to the body's tissues. This deficiency can lead to a variety of symptoms, which may vary depending on the cause, type, and severity of the anemia. Common symptoms include:

1. **Fatigue**: One of the most common symptoms, due to the body's tissues receiving less oxygen than normal.
2. **Weakness**: Feeling weak or having reduced physical strength or energy.
3. **Pale or yellowish skin**: Due to fewer red blood cells and decreased blood flow.
4. **Shortness of breath**: Especially during exercise, because of reduced oxygen delivery to muscles.
5. **Dizziness or lightheadedness**: Can occur because the brain may not be getting enough oxygen.
6. **Cold hands and feet**: Reduced blood flow can make extremities feel cold.
7. **Headaches**: Less oxygen to the brain can cause headaches.
8. **Irregular heartbeats**: The heart may pump harder and faster to deliver sufficient oxygen to the body.
9. **Chest pain**: In severe cases, anemia can lead to angina or heart attack symptoms due to the heart struggling to pump oxygenated blood.
10. **Cognitive problems**: Such as difficulty concentrating or remembering, due to decreased oxygen supply to the brain.

The severity of symptoms can vary widely; some people may experience mild symptoms or none at all, while others may have severe and debilitating symptoms. It's important to diagnose and treat the underlying cause of anemia to effectively manage these symptoms. If you suspect you have anemia, it's important to consult a healthcare provider for a proper diagnosis and treatment plan.

Fig. 10: Interactive conversation between a user and the system.

- **Sujith Ramprasad Tellakula** - Led the design and development of the user interface using Streamlit, which significantly enhanced real-time user interaction. Implemented the history aware retrieval chain and retrieval agents in Langchain to enhance query processing. Also worked on integrating various Large Language Models (LLMs) such as Claude, Gemini, Mistral, and Llama2 apart from OpenAI's GPT into the system, allowing for the comparison of these popular Large Language Models.
- **Ronit Patil** - Implemented various features in the chatbot along with different chains to figure out the best information retrieval method. Ingested the data by disease to prevent context being lost and minimize hallucinations. Integrated TruLens as an evaluation metric to analyze and score the responses given by the chatbot, as well as check for any potential issues that may have occurred during the RAG retrieval mechanism. Tested and refined the system to make sure the chatbot performs to the best of it's abilities. Implemented document upload and link input features to allow user to ask queries about the documents or web links.
- **Vishnu Batla** - Conducted thorough research on potentially integrating X-Ray models through multi-model LLM APIs and contributed key insights to develop the project's conceptual framework. Also recommended suitable models and helped with the integration of readily available APIs like Google's Gemini model and Anthropic's Claude.
- **Simran Panchal** - Employed Python libraries such as BeautifulSoup and Scrapy for data extraction from websites using web scraping methods. This involved sending an HTTP request to the website to fetch web data using Python. Efficiently parsed HTML and XML documents with BeautifulSoup, facilitating smooth navigation and extraction of specific information.
- **Jay Mistry** - Utilized Python tools like BeautifulSoup and Playwright to extract information from the popular forum WebMD. Successfully extracted over 1100 diseases and stored them in a JSON format. Employed precise data pre-processing techniques to filter out inconsistent and poor data.

These efforts collectively ensured that the ”Healthcare Mining” project not only met but exceeded its goals, providing a robust platform for accurate and efficient healthcare information retrieval.

VIII. Conclusions

The ”Healthcare Mining” project has successfully demonstrated how artificial intelligence can revolutionize information retrieval in the healthcare domain. Through the integration of advanced language models and custom-designed retrieval systems, our project has addressed significant challenges in accessing accurate, relevant, and trustworthy healthcare information.

- **Achievements:** We developed a highly interactive and user-friendly chatbot capable of understanding and processing complex medical queries with high accuracy. The system leverages state-of-the-art LLMs including OpenAI, Claude, Gemini, Mistral, and Llama2, ensuring comprehensive coverage and nuanced understanding of user inquiries. Additionally, the use of Trulens as an evaluation framework has significantly enhanced the system's capability to assess the relevance and accuracy of the information provided, ensuring our outputs meet the high standards required in healthcare.
- **Impact:** By simplifying the process of retrieving medical information, our project not only enhances user experience but also supports healthcare professionals and patients by providing quick access to essential information. This is expected to lead to better patient outcomes and more informed decision-making in clinical settings.
- **Future Directions:** Moving forward, the project can be expanded in several ways. First, by incorporating more diverse data sources and medical databases to broaden the scope of information available. Second, further refining the AI models to handle even more complex queries and support additional languages to increase accessibility. Lastly, integrating more personalized responses based on user history and preferences could make the system even more user-centric.

In conclusion, ”Healthcare Mining” stands as a testament to the potential of AI in healthcare. It underscores our commitment to pushing the boundaries of technology to serve critical needs in medical information retrieval. As we continue to refine and expand our system, we look forward to contributing further to the evolution of digital healthcare technologies.

The code for the system has been pushed to GitHub and can be found here.

References

[1] Langchain documentation. https://python.langchain.com/docs/get_started/introduction.
[2] Openai api documentation. https://platform.openai.com/docs/introduction.
[3] Streamlit documentation. https://docs.streamlit.io.
[4] Trulens evaluation with langchain quickstart. https://www.trulens.org/trulens_eval/langchain_quickstart/.
[5] Anand V. Saurkar, Kedar G. Pathare, and Shweta A. Gode. An overview on web scraping techniques and tools. 2018.

[6] Parikshit Sondhi, Jimeng Sun, Hanghang Tong, and Chengxiang Zhai. Sympgraph: A framework for mining clinical notes through symptom relation graphs. In *KDD'12 - 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1167–1175, September 2012. 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2012 ; Conference date: 12-08-2012 Through 16-08-2012.

[7] Illhoi Yoo, Jinbo Bi, and Xiaohua Hu, editors. *2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, San Diego, CA, USA, November 18-21, 2019*. IEEE, 2019.