Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 6, April : 2024

AUGMENTED QUESTION-ANSWERING SYSTEM FOR BIOMEDICAL LITERATURE RETRIEVAL USING LANGUAGE MODELS AND PUBMED INTEGRATION

Dr. Arun M, Associate Professor, Department of Electronics and Communication Engineering, KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India.
Dhilip S, Nithish K, Dhanush R, Rohith Krishnan R, UG Scholars, Department of Electronics and Communication Engineering, KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India.

Abstract
In this research, a biological Question Answering (QA) system is presented, with an emphasis on extracting pertinent data from PubMed articles. The system makes use of a number of methods, such as Firebase integration, PubMed article fetching, and language models (LM). The Haystack framework, which makes it easier to orchestrate components within a pipeline, serves as the foundation for the design. Through a Streamlit interface, users can engage with the system by asking questions on biomedical subjects and getting responses produced by the language model. A PubMedFetcher class for retrieving articles from PubMed, prompt templates for directing the creation of LMs, and Firebase for storing queries and answers are the main parts of the system. The goal of the QA system is to facilitate users' efficient access to pertinent biomedical information.
Keywords: Biomedical, Question Answering, PubMed, Language Models, Haystack Framework.

I.Introduction
It probably addresses the difficulties and developments in the search for biomedical literature, with an emphasis on the contribution of artificial intelligence (AI) to enhanced search capabilities. Topics including natural language processing methods for reading biomedical articles, AI algorithms for retrieving literature, and the incorporation of AI into PubMed and other platforms for more effective literature searches might all be covered [1]. It examines how retrieval-augmented language models, or RALMs, are used in clinical care. RALMs combine the capabilities of language models with information retrieval techniques, presumably to help practitioners find pertinent medical information [2].
This presents the latest developments in automated question answering for the biomedical field. It might span a range of techniques and approaches, including machine learning algorithms, knowledge graphs, and natural language processing methods, that are used to create automated question answering systems that are especially designed for biomedical applications [3]. It is probably a thorough analysis that looks at many strategies and difficulties in the area of biological question answering. It could go over different kinds of biomedical queries, approaches to addressing them, and the constraints and potential paths of biomedical quality assurance systems [4].
The difficulties of manual curation in genomic databases are covered in this work. The limitations of manual curation approaches are probably brought to light, and the necessity for automated or semi-automated techniques to annotate genomic data more precisely and effectively is investigated [5]. Based on log data, this paper may examine how PubMed users search. To enhance the user experience, it might investigate trends in search queries, how users interact with search results, and how satisfied users are with the PubMed platform [6]. Evidently, it looks into how well pretrained language models work to enhance output on subsequent tasks. It could examine methods like prompt and head tuning to adjust pretrained models for certain assignments [7].
This seems to cover methods for keeping up on coronavirus research developments. It may include data analysis methodologies and tactics for conducting effective literature searches in order to quickly monitor and synthesize new research findings relevant to the COVID-19 pandemic [8]. The significance of developing well-structured clinical questions in evidence-based medicine is probably covered in this publication. It might offer templates or instructions for formulating clinical inquiries that make it easier to find pertinent data [9].

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 6, April : 2024
The Cochrane Handbook for Systematic Reviews of Interventions, an extensive manual for performing systematic reviews in the medical field, is probably being presented here. It might go over how to formulate research questions, find pertinent studies, evaluate the quality of studies, and synthesize data to guide clinical practice [10].

II.Literature Review

In 2024 Qiao Jin, Robert Leaman, Zhiyong Lu proposed a system PubMed and Beyond: Biomedical Literature Search in the Age of Artificial Intelligence. The difficulties presented by the abundance of biomedical literature accessible are highlighted. While highlighting PubMed as the main resource for biomedical literature, it also points out that it is not very good at meeting specialized information demands. One enhancement noted in PubMed is the switch from recency-based to relevance-based ranking; nonetheless, the system is mainly designed to handle brief keyword-based queries without requiring extensive research. The study highlights the necessity of specialized literature search engines, especially in situations where quick distribution of fresh information is essential, such as the COVID-19 epidemic. It analyses several web-based literature search tools categorized by their specific functionalities, including evidence-based medicine, precision medicine, semantic search, literature recommendation, and literature mining. The tools' varied approaches to query processing and result presentation are intended to meet particular information needs. The paper's overall goal is to help researchers and clinicians retrieve information more effectively by offering insights into the current state of biomedical literature search tools.

In 2023 Cyril Zakka, Akash Chaurasia, Rohan Shad, Alex R. Dalal, Jennifer L. Kim, Michael Moor Almanac: Retrieval-Augmented Language Models for Clinical Medicine. It draws attention to the difficulties that large language models (LLMs) face, including the creation of factually false claims and the reinforcement of societal prejudices. The authors present Almanac, a framework intended to investigate medical LLMs' involvement in clinical practice while protecting patient confidentiality and safety. With in-text citations for validation, Almanac provides precise and thorough answers to clinical queries by utilizing established knowledge repositories and external resources. The research suggests factuality, completeness, and safety as the three main goals for assessing these models. In order to provide correct and trustworthy answers, grounded large-language models have been empirically evaluated by board-certified doctors and resident physicians on a new dataset of clinical scenarios. This evaluation opens the door for the regulated use of these models in the healthcare industry.

In 2021 Krishanu Das Baksi proposed a system Recent Advances in Automated Question Answering in Biomedical Domain This is especially important in the fields of healthcare, medicine, and biology, where practitioners and scientists frequently formulate information needs as questions in natural language. It highlights the difficulties faced by domain specialists in staying current with the state of knowledge and the problems caused by the quickly expanding PubMed database of biomedical literature. The article emphasizes the paradigm change toward evidence-based medicine and the necessity of providing clear and precise responses to doctors' concerns, many of which have a time constraint. In order to shorten the time spent looking for information and support clinical decision making, it makes the case for the creation of effective and precise biomedical question answering (QA) systems. The review examines datasets, approaches, and systems created during the last ten years with a focus on recent developments in end-to-end question answering systems in the biomedical field. It provides an overview of the paper's structure and contains sections on broad domain quality assurance (QA), the intricacies of biomedical QA, datasets, various biomedical QA system types, limitations, and future directions for the discipline. With an emphasis on factoid open domain QA systems, the paper's overall goal is to shed light on the developments and difficulties facing automated question answering in the biomedical field.

In 2020 Qiao Jin, Zheng Yuan, Guangzhi Xiong, Qianlan Yu, Huaiyuan Ying, Chuanqi Tan proposed a system Biomedical Question Answering: A Survey of Approaches and Challenges. It also discusses the significance of efficient biomedical knowledge acquisition. It draws attention to the difficulties and developments in QA research, with a special emphasis on the Biomedical Question Answering System challenge and the Text Retrieval Conference (TREC) QA Track as driving forces behind contemporary QA research. The development of large-scale Pre-trained Language Models (PLMs) and their influence on biological question answers are discussed, as well as the evolution of Question Answering models—particularly those based on deep learning. The study summarizes the primary obstacles that current Biomedical Question Answering systems must overcome, in spite of tremendous advancements. These obstacles include concerns about evaluation, questions about explainability, problems with dataset size and annotation, problems with fairness and bias, and underutilization of domain knowledge. The survey methodically goes over the different BQA techniques, strategies, and datasets that have been put forth recently. In summary, the report ends with a discussion of possible future avenues for BQA research and a summary of its problems.

2.1 Emerging Trend

- **Artificial Intelligence (AI) Integration in Healthcare:** In recent times, there has been a notable surge in the adoption of AI in healthcare. Artificial intelligence (AI) technologies, such as machine learning and natural language processing, are being used in healthcare for a variety of purposes, such as patient management, therapy recommendation, and illness diagnosis. AI-powered diagnostic technologies, for example, can analyse medical images and accurately identify problems, which may help medical personnel diagnose patients more quickly and accurately. Furthermore, AI-powered chatbots and virtual assistants are being used to offer patients individualized healthcare guidance and assistance, expanding patient involvement and facilitating better access to healthcare services.
- **Growth of Remote Healthcare and Telemedicine:** Particularly in reaction to the COVID-19 pandemic, the use of telemedicine and remote healthcare services has increased dramatically. Telemedicine makes use of modern communication tools like video conferencing and smartphone apps to enable patients to consult with medical professionals from a distance. This trend has reduced the risk of viral transmission by reducing in-person visits to healthcare facilities and has made it easier for people, particularly those living in rural or underserved areas, to get medical care. Furthermore, wearable technology and remote monitoring systems have made it possible to continuously monitor patients' health parameters, giving people the ability to take charge of their health and enabling medical professionals to act quickly in an emergency.
- **Personalized Medicine and Genomic Healthcare:** Personalized medicine, sometimes referred to as precision medicine, is customizing medical care to each patient's unique attributes, including genetics, lifestyle, and surroundings. Thanks to developments in genomics and molecular biology, precision medicine techniques are now possible, allowing medical professionals to tailor treatments and interventions to individual patients' genetic profiles and illness states. Thanks to advancements in genome sequencing technologies, it is now possible to conduct thorough genomic analyses and identify genetic variations linked to specific disorders. Optimizing resource allocation in healthcare delivery, minimizing side effects, and improving treatment outcomes are all possible with this individualized approach to healthcare.
- **Blockchain Technology in Healthcare Data Management:** More and more research is being done on the possible uses of blockchain technology, which was first created for safe and transparent cryptocurrency transactions. Blockchain technology ensures the integrity, security, and privacy of sensitive medical data by providing decentralized and immutable data storage. Healthcare firms can improve patient consent management, reduce data breaches and illegal access, and expedite data sharing and interoperability among diverse systems by deploying blockchain-based technologies. Furthermore, smart contracts enabled by blockchain technology can streamline administrative overhead and increase operational efficiency in the healthcare industry by automating tasks like processing insurance claims.
- **Virtual Reality (VR) and Augmented Reality (AR) in Medical Education and Training:** An Overview Through immersive and interactive learning environments, augmented reality (AR) and

virtual reality (VR) technologies are transforming medical education and training. Medical professionals and students can practice practical skills in a safe and controlled environment by simulating surgical procedures, anatomical dissections, and patient interactions in virtual environments. Medical simulations powered by augmented reality and virtual reality provide learners with realistic and captivating learning experiences that help them retain information and improve their competencies. Furthermore, these technologies facilitate tele-mentoring and remote cooperation, linking students with professionals across the globe and encouraging ongoing professional growth.

III.Research Objectives

- To evaluate the effectiveness of biomedical question answering systems.
- To investigate user satisfaction and usability in healthcare settings.
- To explore the impact on evidence-based medicine practices.
- To address ethical and privacy concerns in biomedical AI.
- To propose strategies for system enhancement and future development.

IV.Design Implementation

Basically, there are eight steps in biomedical QA system which constitute of the entire methodology. They are User Input, Keyword Builder, Keyword LLM, Prompt Fetcher, Prompt Builder, LLM, Firebase, Streamlit (UI).

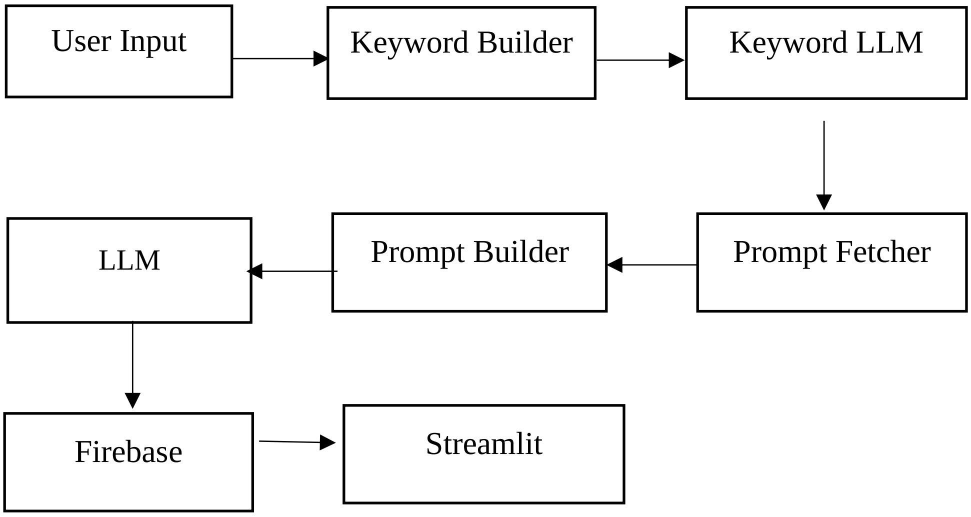

Fig 1: Biomedical QA System

4.1 User Input

The interface where users can enter their inquiries about biomedical subjects is referred to as the user input block. This block is intended to collect users' natural language questions, enabling them to express their information demands in a clear and comfortable manner. This block specifically makes use of the textbox component, a graphical user interface element that lets users enter text. In this instance, the textbox is set up to take inquiries from users about biomedical subjects. The textbox allows users to submit their inquiries directly, giving them a free-form way to convey their information demands.

The textbox component lets users engage with the system first. Following the user's entry and submission of their question, the system analyses the data and starts to retrieve and generate pertinent information in response to the user's request. Through direct user interaction with the biomedical Q&A system, this input mechanism makes it possible for users to obtain information and expertise in the biomedical domain quickly and effectively. Ultimately, the biomedical question answering system's user input block is essential for facilitating user engagement and information retrieval because it is the initial point of entry for system queries and the retrieval of pertinent biomedical knowledge and insights.

4.2 Keyword Builder

Within the biomedical question answering system, the Keyword Prompt Builder component in the given code plays a crucial role in creating prompts based on the user's input query. This part is intended to convert the user's query into an organized prompt format with the express purpose of producing applicable keywords that can be utilized to find relevant PubMed research articles. Using the user's query as input, the builder creates a prompt template that directs the creation of keywords. Usually, this template describes the intended output (which includes keywords generated from the question) and the format in which the user's inquiry should be presented. Using natural language processing techniques, the Keyword Prompt Builder analyses the user's question, pinpoints important ideas, and extracts pertinent terms or phrases that capture the substance of the inquiry. These keywords constitute the basis for focused searches in PubMed, allowing the engine to find research articles that closely match the user's information requirements. The Keyword Prompt Builder improves the efficacy and efficiency of the biomedical question answering system by organizing the user's query into a prompt and producing pertinent keywords. This allows for more precise and focused retrieval of medical literature to answer the user's questions.

4.3 Keyword LLM

The given code's Keyword LLM (Large Language Model) component generates keywords based on the prompt created by the Keyword Prompt Builder. This part makes use of a sizable language model that has been specially trained on biomedical text data to provide pertinent keywords that both meet the prompt's requirements and encapsulate the user's query.

The Keyword LLM uses natural language processing techniques to examine the input text after receiving the prompt from the Keyword Prompt Builder and extract important concepts, terms, and phrases that are suggestive of the user's information demands. Using the context and vocabulary it has learned about biological terms during training, the LLM creates a list of keywords that summarize the primary ideas or subjects included in the user's question.

The Keyword LLM makes use of sophisticated language modeling features like semantic coherence and contextual awareness to make sure that the generated keywords appropriately capture the essence of the user's query. The LLM can generate keywords that are both contextually appropriate and customized for the particular information retrieval task at hand by taking into account the prompt's context.

Following their generation, the keywords are utilized to create search terms or obtain pertinent papers from biomedical sources like PubMed. The Keyword LLM improves the biomedical question answering system's efficacy and efficiency by automating the keyword generating process. This allows the system to promptly and precisely discover pertinent material and offer insightful responses to users based on their inquiries.

4.4 Prompt Fetcher

Utilizing the keywords produced by the Keyword LLM, the Prompt Fetcher component of the supplied code is in charge of obtaining articles from PubMed, a database of biological literature. This part is essential because it acts as a link between the process of creating keywords and the retrieval of pertinent material to answer the user's inquiry.

Once the Prompt Fetcher receives the generated keywords from the Keyword LLM, it creates search queries based on these keywords and delivers them to the PubMed search engine. Next, a list of publications that PubMed believes are relevant to the supplied keywords is provided.

These articles are gathered by the Prompt Fetcher, who also gets them ready for additional processing inside the biomedical Q&A system. It gathers pertinent data from every article, including the title, abstract, and keywords, and arranges it in a way that makes it easily readable by system components that come after it.

The Prompt Fetcher may also use ranking or filtering algorithms to order the retrieved articles according to relevance or other factors. This guarantees that the user receives the most relevant material, enabling quick and easy access to important biomedical knowledge.

With everything taken into account, the Prompt Fetcher uses the generated keywords to find pertinent articles on PubMed, which is a crucial part of the biomedical question answering system's information retrieval process. It plays a crucial role in the system's pipeline, allowing users to efficiently meet their information demands by gaining access to and utilizing the abundance of biological literature included in PubMed.

4.5 Prompt Builder

The Prompt Builder component is in charge of creating a prompt that combines the user's query with the PubMed articles that were retrieved. By organizing the incoming data in a way that makes efficient information synthesis and creation possible, this component is essential to the biomedical question answering system's future processes.

The Prompt Builder takes the user's inquiry and the articles it has retrieved from PubMed and merges them to produce a coherent prompt that offers background information for answering. It creates a template that describes the format and content of the prompt, outlining the integration of the user's query with the articles that were retrieved.

The Prompt Builder evaluate the user's query and extract important terms or concepts by using natural language processing algorithms. It then makes sure that the generated prompt appropriately represents the user's information demands by incorporating these ideas into the prompt template.

Further, the question Builder arranges the data that is obtained from PubMed, including article titles, abstracts, and keywords, in a way that makes it simple to add the material to the question. This guarantees that the prompt offers pertinent background information and sources to help create answers based on the literature that was collected.

The Prompt Builder improves the effectiveness of the following phases in the biological question answering system by organizing the prompt in a clear and logical manner. In order to facilitate accurate and perceptive answers to biological questions, it offers a framework for generating responses that are informed by both the user's query and the pertinent literature collected from PubMed.

4.6 Large Language Models (LLM)

The component known as the LLM (Large Language Model) is in charge of producing a response based on the prompt that the Prompt Builder has created. To comprehend the prompt and produce a logical response, this component makes use of a sizable pre-trained language model.

The LLM uses its deep learning architecture to examine the input text and provide an answer after it receives the prompt. The model uses sophisticated natural language processing methods to understand the prompt's context and semantic meaning, including transformer architectures and attention processes.

Then based on its prediction of the most likely token sequence that follows the prompt, the LLM prepares a response. It explores the space of potential responses using probabilistic sample techniques, like beam search and top-k sampling, and chooses the best one based on context and learnt language patterns.

Furthermore, the LLM may apply modification or fine-tuning strategies to modify its responses according to the particular biological literature domain. In order to improve the model's performance on biomedical question answering tasks, it is necessary to either fine-tune its parameters on a related task or train the model on domain-specific data.

With all factors taken into account, the LLM functions as the central element of the biomedical Q&A system, producing precise and educational responses in response to user queries and pertinent material obtained from PubMed. The utilisation of natural language generation and understanding capabilities by the LLM enables the automated synthesis of biomedical knowledge, hence improving information accessible for doctors, researchers, and the wider public.

4.7 Firebase

The main use of Firebase is to keep questions and answers together in a Firebase database. Google offers a feature-rich platform called Firebase for creating online and mobile applications. It can be used for a wide range of application development needs because it provides a variety of services, such as hosting, cloud storage, authentication, and real-time databases.

Question-answer pairs are kept in the Firebase database using Firebase Storage, a feature of the Firebase platform. Cloud storage for user-generated material, including text documents, videos, and photos, is made safe and scalable with Firebase Storage. Within this framework, Firebase Storage functions as a repository for user-posted queries and the accompanying responses produced by the biological question answering system.

Through the use of Firebase Storage, the application may store question-answer pairs in a safe and dependable manner that works well with other Firebase services. Because Firebase takes care of the complicated tasks of data administration and storage, developers can concentrate on creating the essential features of the application.

4.8 Streamlit

Building dynamic web apps straight from Python scripts is made easier with the help of Streamlit, a robust Python tool. Developers may design user interfaces for data exploration, visualization, and interactivity that are responsive and easy to use with Streamlit. Without requiring in-depth knowledge of web programming, developers may quickly prototype and implement online apps with the help of the library's simple API.

A web application dedicated to biological question-answering is constructed using Streamlit. A title and instructions directing users on how to utilize the system are presented at the beginning of the application. Users can submit questions about biological subjects, and the application uses a number of different components to answer and store question-answer pairings, including language models, PubMed integration, and Firebase. To further showcase its features and promote user interaction, the application also offers sample questions users may pose.

Once deployed, the user-friendly interface of this Streamlit web app makes it simple for users to access and engage with the biomedical question-answering system. Streamlit is a great option for developing data-driven web applications and prototypes because of its ease of use and versatility, which let developers construct strong applications like this one with little to no work.

V.Result

The resultant product is a biological question answering system that lets users submit inquiries about biology, medicine, and healthcare and get responses based on pertinent scientific material. The system makes use of a number of natural language processing methods, such as answer creation through large language models (LLMs), keyword extraction, and article retrieval from PubMed.

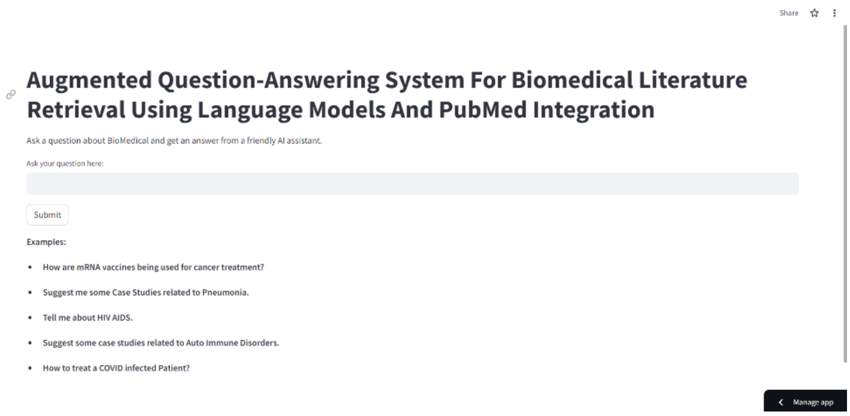

Fig 2: User Interface

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 6, April : 2024

A textbox is displayed to the user so they can enter their inquiry when interacting with the system via the Streamlit interface. After the query is submitted, the system reads the input, pulls out keywords to look up pertinent papers on PubMed, retrieves the articles, and uses an LLM to create an answer. The user can then promptly get the requested information by seeing the response shown back to them in the interface.

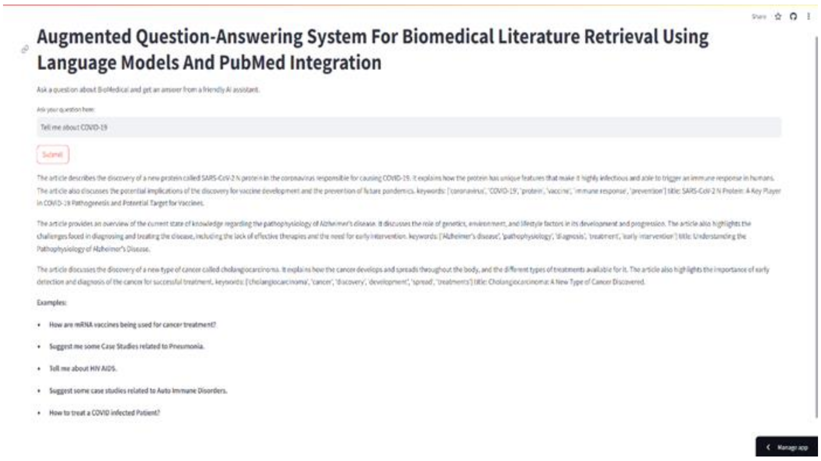

Fig 3: Question with Response

The output is a response to the user's query that is drawn from scientific publications that are listed on PubMed. The Streamlit interface presents the solution in an approachable manner that makes it easy to access and read. In addition, the system records and retrieves previous interactions for analysis and reference by storing the question-answer pairings in a Firebase database.

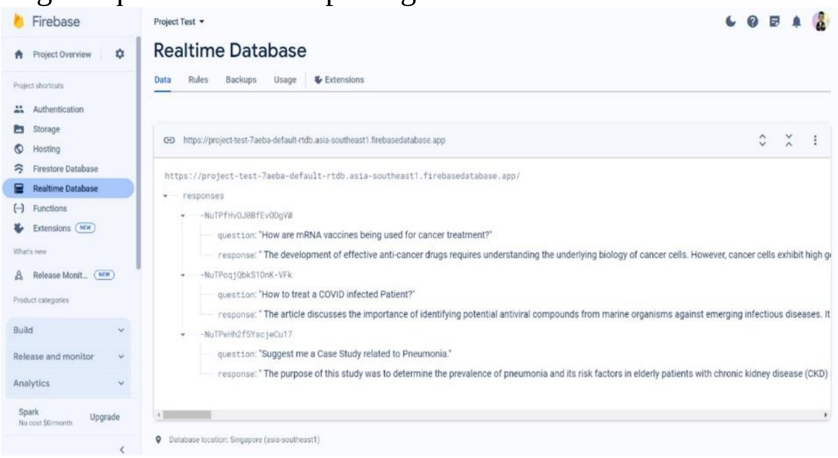

Fig 4: Firebase Database

VI. Conclusion

In conclusion, the provided biomedical question-answering system is an important development in the field of natural language processing methods to make scientific material more accessible and to give consumers quick, correct information. The system simplifies information retrieval for researchers, healthcare professionals, and the general public by combining components including answer creation using large language models (LLMs), keyword extraction, and article retrieval from PubMed. Users may simply ask questions using the Streamlit interface, and they will receive thorough responses that are based on reliable scientific sources.

Moreover, the system's ability is further enhanced by allowing the tracking and long-term analysis of user interactions through the storage of question-answer pairings in a Firebase database. This feature not only makes tracking user inquiries easier, but it also makes it possible to evaluate user engagement and system performance.

Overall, the biomedical question-answering system shows how natural language processing technology might help evidence-based decision-making in biomedical research and healthcare by expanding access to scientific knowledge. Improvements in data integration and machine learning are anticipated to extensively expand the functionalities and efficacy of these systems as the area develops, ultimately leading to better patient care and scientific research.

References

[1] Qiao Jin, Robert Leaman, Zhiyong Lu, (2024) PubMed and Beyond: Biomedical Literature Search in the Age of Artificial Intelligence, National Center for Biotechnology Information Vol 100 Issue:1 Feb 2024.

[2] Cyril Zakka, Akash Chaurasia, Rohan Shad, Alex R. Dalal, Jennifer L. Kim, Michael Moor, Kevin Alexander, Euan Ashley, Jack Boyd, Kathleen Boyd, Karen Hirsch, Curt Langlotz,Joanna Nelson, William Hiesinger, (2023) Almanac: Retrieval-Augmented Language Models for Clinical Medicine, National Center for Biotechnology Information Issue: 25 Jan 2024.

[3] Krishanu Das Baksi, (2021) Recent Advances in Automated Question Answering in Biomedical Domain Issue: 10 Nov 2021.

[4] Qiao Jin, Zheng Yuan, Guangzhi Xiong, Qianlan Yu, Huaiyuan Ying, Chuanqi Tan, Mosha Chen, Songfang Huang, Xiaozhong Liu, Sheng Yu (2020) Biomedical Question Answering: A Survey of Approaches and Challenges, Association for Computing Machinery Volume 55 Issue: 30 Nov 2020.

[5] Baumgartner W.A., Jr., Cohen K.B., Fox L.M., Acquaah-Mensah G., Hunter L. Manual curation is not sufficient for annotation of genomic databases. Bioinformatics, PubMed, 2007.

[6] Islamaj Dogan R., Murray G.C., Neveol A., Lu Z. Understanding PubMed user search behavior through log analysis. Database (Oxford), PubMed, 2009.

[7] Wei C., Xie S.M., Ma T.: Why Do Pretrained Language Models Help in Downstream Tasks? An Analysis of Head and Prompt Tuning, 2021.

[8] Chen Q., Allot A., Lu Z. Keep up with the latest coronavirus research, PubMed, 2020.

[9] Richardson W.S., Wilson M.C., Nishikawa J., Hayward R.S. The well-built clinical question: a key to evidence-based decisions. ACP J Club, PubMed, 1995.

[10] Higgins J.P., Thomas J., Chandler J., et al. Cochrane handbook for systematic reviews of Interventions, John Wiley & Sons, 2019.