International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

# Personalized QA System with Contextual Memory Using GooglePalm

**Mudavath Hanmanthu[1], Priyanshu Kumar[2], Raj Priyanshu[3], Vanga Hareesh Reddy[4], Mrs. S. J. Shruthi Rani Yadav[5]**

[1,2,3,4]Final Year Student, Department of Computer Science and Engineering Malla Reddy College of Engineering & Technology, Hyderabad, India.
[5]Assistant Professor Department of Computer Science and Engineering Malla Reddy College of Engineering & Technology Hyderabad, India.

**Abstract**
Personalized Q&A System with Contextual memory using GooglePalm, LangChain in Ed-Tech Industry aims to build a CQA(Conversational Question Answer) system which is a interactive search systems that effectively serve information by interacting with users. Despite its effectiveness, challenges exist as human annotation is time consuming, inconsistent, and not scalable. To address this issue and investigate the applicability of large language models in conversational question-answering (CQA) simulation, we propose a simulation framework that employs Langchain, Google Palm LLMs. Here we will add a csv file which consists of frequently asked questions. Furthermore, we conduct extensive analyses to thoroughly examine the LLM performance by benchmarking state-of-the-art reading
comprehension models on datasets. Our results reveal that the Service Provider LLM generates lengthier answers that tend to be more accurate and complete.
This is an end to end LLM project based on Google Palm and Langchain. We are building a Q&A system for an ed-learning company. In particular, noting that it takes time for the user to speak, threading related to database searches is performed while the user is speaking.

**Keywords:** Langchain, Large Language Model, Google Palm, Contextual Memory, Deep Learning, Generative AI.

## 1. Introduction

In current situation, we came across various problems in Ed-Tech Industry conversational question answering system like CQA system with no past memory and it required a human effort for maintenance and answering the questions for repeated questions. Therefore, it leads to time consuming and less efficient. We have reached a practical and realistic phase in human-support dialogue agents by developing a large language model (LLM).
However, when requiring expert knowledge or anticipating the utterance content using the massive size of the dialogue database, we still need help with the utterance content's effectiveness and the efficiency of its output speed, even if using LLM. Our proposed system will overcome all of these problems and it will be an efficient CQA system of LLM. It will be more efficient and flexible to use and enables to add queries and answers if the answer for certain question is not available to the user. Therefore, we propose a

framework that uses LLM asynchronously in the part of the system that returns an appropriate response and in the part that understands the user's intention and searches the database. In particular, noting that it takes time for the user to speak, threading related to database searches is performed while the user is speaking.

Our aim is to propose a framework that uses LLM asynchronously in the part of the system that returns an appropriate response and in the part that understands the user's intention and searches the database.

## 2. Literature Review

1. **Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fasst(July 10-12,2023:Konya,Turkey):**

This study focused on the utilization of Large Language Models (LLMs) for the rapid development of applications, with a spotlight on LangChain, an open-source software library. LLMs have been rapidly adopted due to their capabilities in a range of tasks, including essay composition, code users. The crux of the study centers around LangChain, designed to expedite the development of bespoke AI applications using LLMs. LangChain has been widely recognized in the AI community for its ability to seamlessly interact with various data sources and applications. But The application is developed without contextual memory.

2. **End–To-End LLM Project Using LangChain, GooglePalm in Pharma Industry (Nov 7, 2023, Charanrio ):**

They have built an end-to-end LLM project using LangChain, Google PaLM, and a SQL database to answer questions about a pharma company's sales data. Our application is able to convert a natural language question into a SQL query, execute the query on the database, and return the answer to the user.

This is just one example of how LangChain, Google PaLM, and SQL databases can be used to build powerful LLM- powered applications. With LangChain, you can easily combine LLMs with other components to build complex applications that can solve a wide range of problems.

3. **Let the LLMs talk: Simulating Human-to-Human Conversational QA via Zero-Shot LLM-to-LLM interactions (5 Dec 2023,Zahra Abbasiantaeb, Yifei yuan, Evangelos kanoulas):**

To replicate human-to-human conversations, existing work uses human annotators to play the roles of the questioner and answerer. Despite its effectiveness, challenges exist as human annotation is time-consuming, inconsistent, and not scalable.

To address this issue and investigate the applicability of large language models (LLMs) in CQA simulation, they proposed a simulation framework that employs zero-shot learner LLMs for simulating questioner-answerer interactions. But they have developed without contextual memory.

3. **A Review on Large Language Models : Architectures, Applications, Taxanomies, Open Issuses and Challenges(Mohaimenul Azam Khan Raiaan,Md.saddam Hossain Mukta,Kaniz Fatema,Nur Mohammed Fahad, Sadman Sakib ,Sep 2023):**

Researchers from various fields have conducted exhaustive studies on the rise of LLMs, shedding lighton their remarkable advancements, diverse applications, and potential to revolutionize tasks from text generation and comprehension to demonstrating reasoning skills. Collectively, these studies contribute to our comprehension of LLMs significant role in shaping the landscape of AIdriven language processing and problem-solving. They provides final conclusion as The field of LLMs has witnessed a remarkable

evolution and expansion, resulting in extraordinary capabilities in natural language processing (NLP) and various applications in various areas. Based on neural networks and the transformative transformer architecture, these LLMs have revolutionized our approach to machine language comprehension and generation. The thorough review of this research has provided an insightful overview of LLMs, encompassing their historical development, architectural foundations, training methods, and vast advancement resources. It has also examined the various applications of LLMs in disciplines such as healthcare, education, social sciences, business, and agriculture, demonstrating their potential to address real-world issues.

## 3. TECHNOLOGIES USED

### 3.1 DEEP LEARNING

Large datasets of labeled data and neural network architectures that automatically extract features from the data while learning them directly from the data are used to train deep learning models.

A computer model learns to carry out categorization tasks directly from images, text, or sound using deep learning. Modern precision can be attained by deep learning models, sometimes even outperforming human ability. A sizable collection of labeled data and multi-layered neural network architectures are used to train models. Deep learning models are sometimes referred to as deep neural networks because the majority of deep learning techniques use neural network topologies.

### 3.2 GOOGLE PALM

**PaLM** (**Pathways Language Model**) is a 540 billion parameter transformer -based Large language model developed by Google AI. Researchers also trained smaller versions of PaLM, 8 and 62 billion parameter models, to test the effects of model scale.

PaLM is capable of a wide range of tasks, including commense reasoning, arithematic Reasoning, joke explanation, code generation, and translation. When combined with chain-of-thought-prompting, PaLM achieved significantly better performance on datasets requiring reasoning of multiple steps, such as word problems and logic based questions.

### 3.3 LANGCHAIN

LangChain is an open-source framework that allows you to connect large language models to your applications. LangChain has a diverse and vibrant ecosystem that brings various providers under one roof, including google's PaLM 2 large language model. LangChain provides clean and simple code and the ability to swap models with minimal changes.

## 4. METHODOLOGY

### 4.1 Problem Definition

Clearly define the objectives and scope of the personalized QA system.

### 4.2 Data collection

Gather a diverse dataset of questions and corresponding answers. If available, collect user preferences or profiles to personalize responses.

### 4.3 Data Preprocessing

Clean the dataset to remove noise, duplicates, and irrelevant information. Perform any necessary text preprocessing tasks such as tokenization.

**4.4 Feature Selection:** Extract relevant features from the questions and user profiles. If incorporating user preferences, transform them into feature vectors.

**4.5 Model Selection:** Choose GooglePalm as the language model backbone for generating responses.GooglePalm is a large language model that can be used to generate text. Decide on the architecture for personalized response generation, considering factors like user preferences and contextual information. Selecting the GooglePalm model involves assessing its architecture, size, pre-training, fine-tuning capabilities, compatibility with the task, availability, performance, customization options, and community support. By carefully considering these factors, developers can choose the most suitable GooglePalm model for building a personalized question-answer system that meets the project requirements and user needs. Selecting the GooglePalm model involves considering several factors to ensure it aligns with the requirements and constraints of the project. Here's an overview of the key considerations in the model selection process.

## SYSTEM ARCHITECTURE

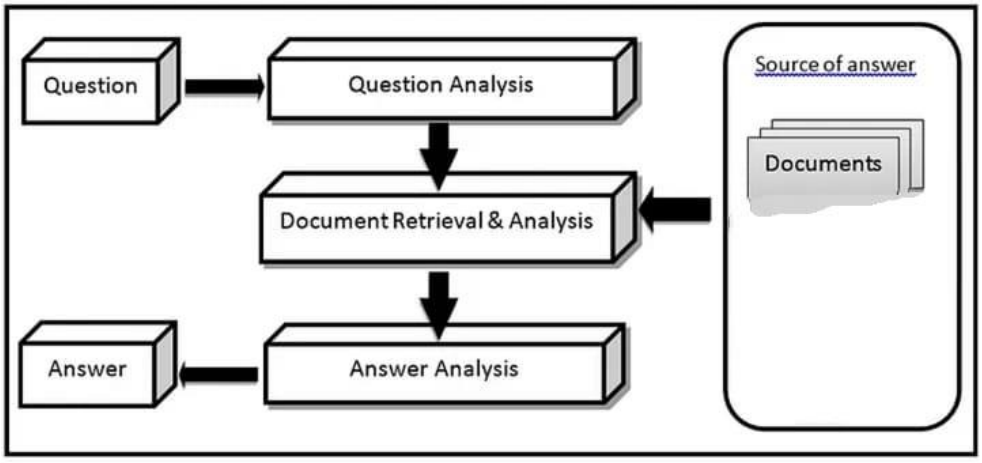



**Fig 1: System Architecture**

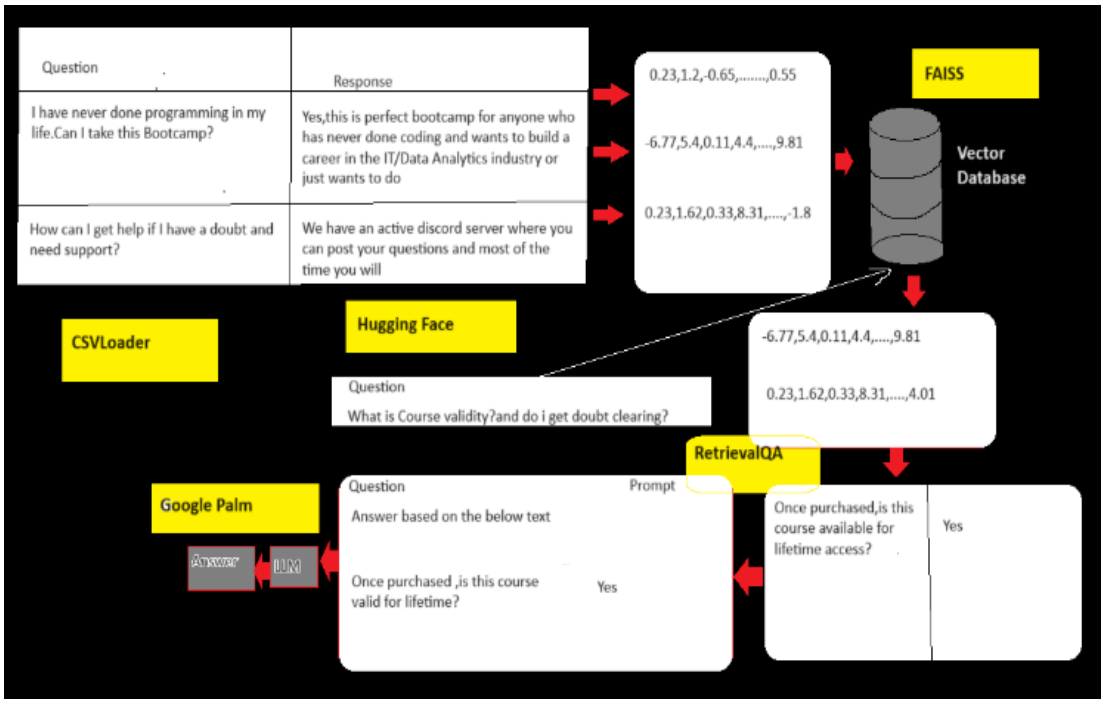
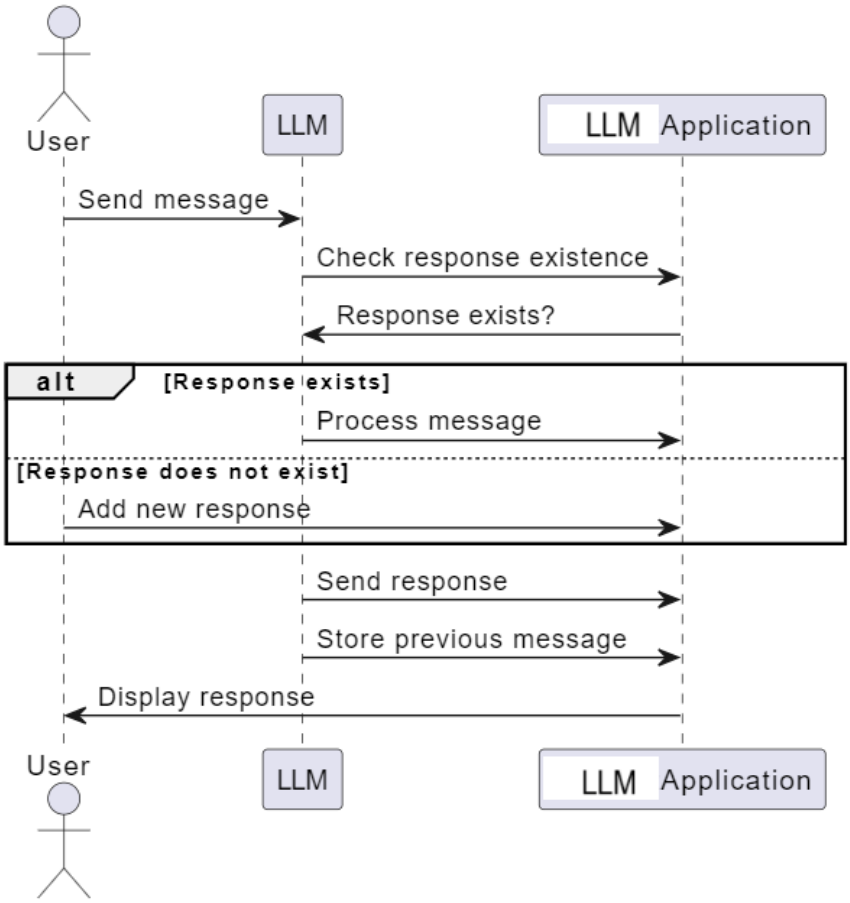
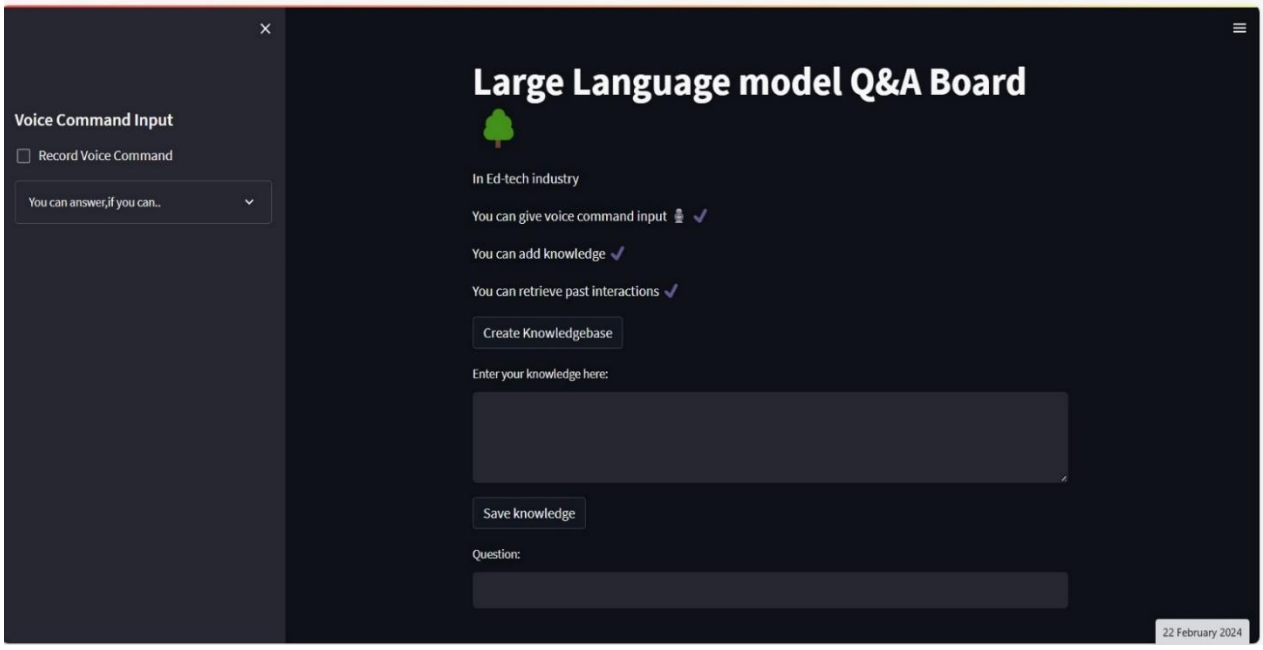



**Fig2: Technical Architecture**

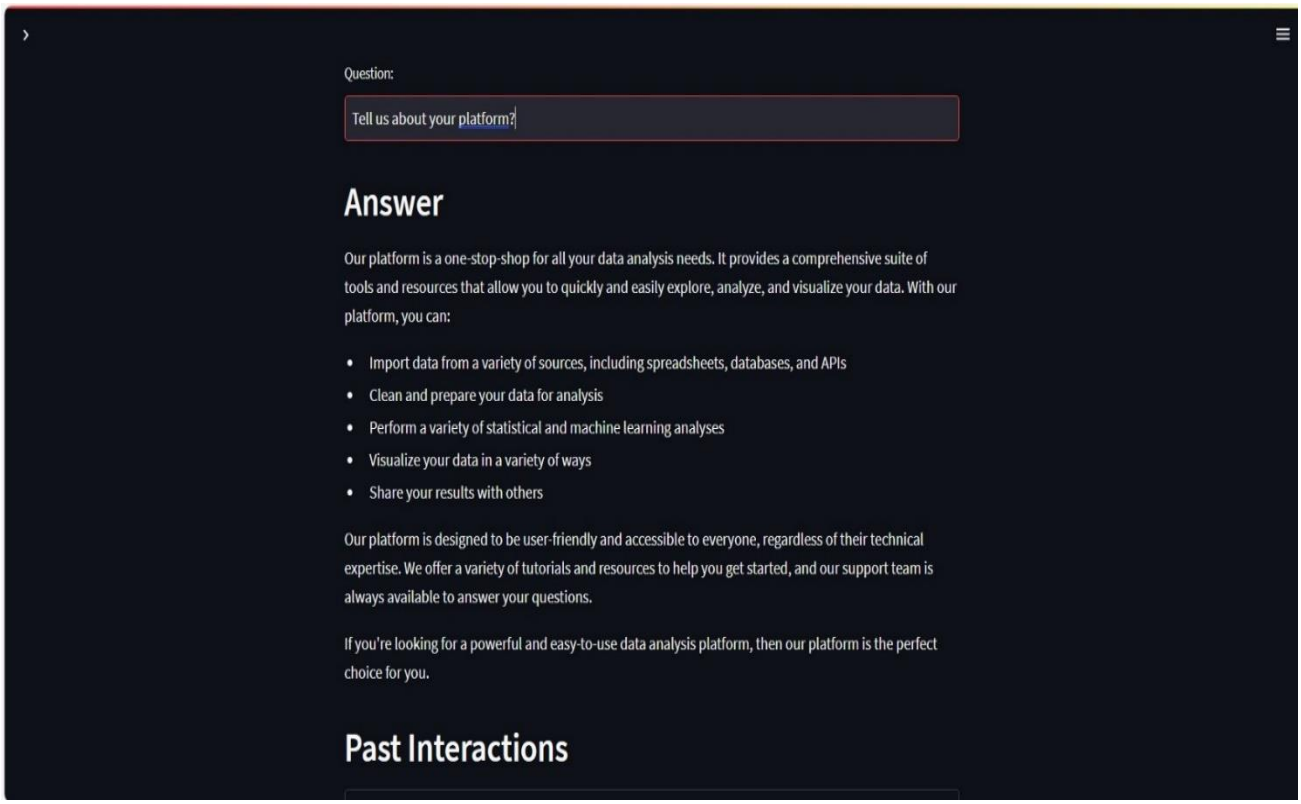
### 4.6 Implementation



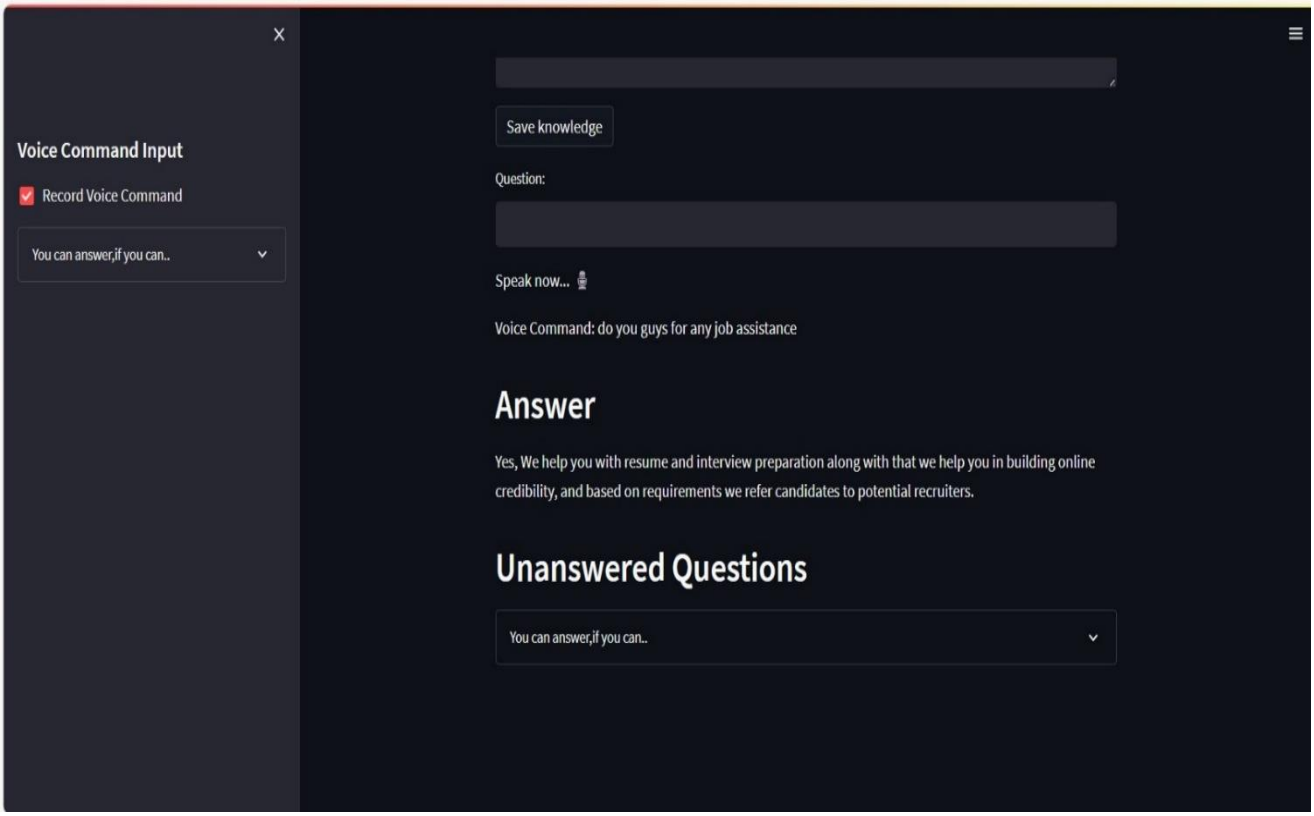

## 5. RESULT



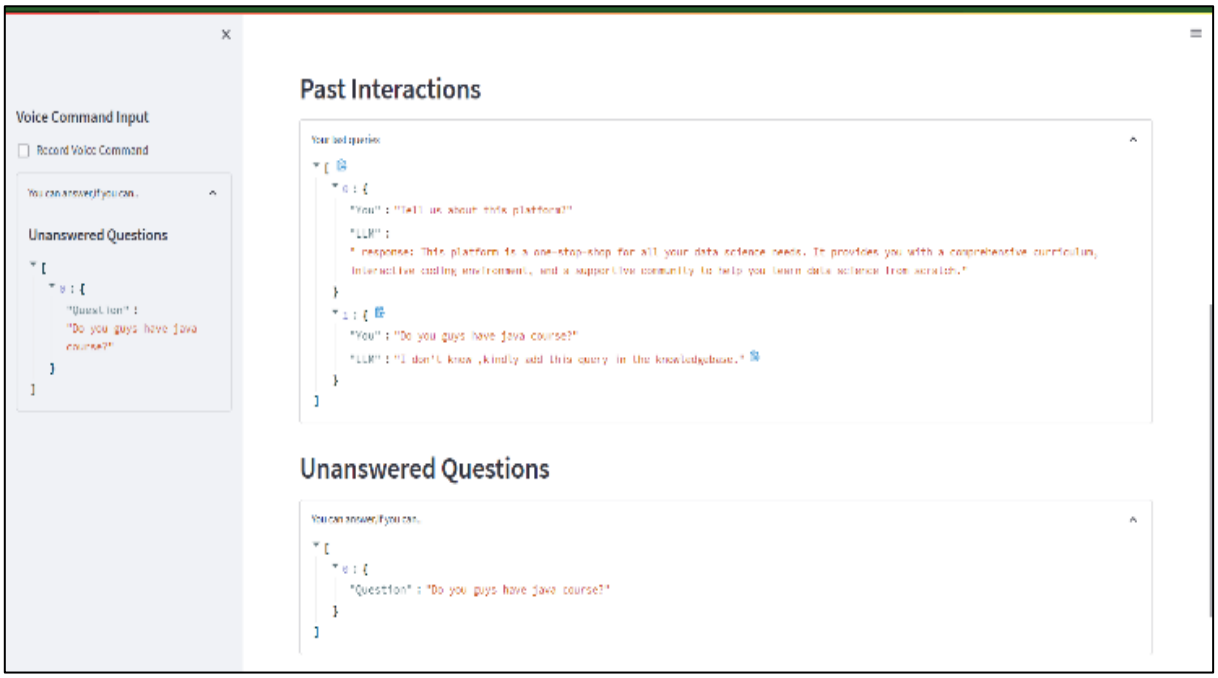

**Test Case 1: LLM without any questions. It is the initial phase of LLM (user interface)**




**Test Case 2: LLM with user questions and answers through text prompts.**




**Test Case 3: LLM with audio input questions and with it's answers.**




**Test Case 4: LLM with contextual memory**

## 6. CONCLUSION

This study focuses on Large Language Model (LLM) CQA Board tailored for the Ed-tech industry. It provides users with the capability to interact with the system via text input and voice commands, enabling them to ask questions, add knowledge, and retrieve past interactions. The system leverages the GooglePalm model, a state-of-the-art language model, to generate responses to user queries. Additionally, it incorporates features such as the ability to create a knowledge base and store past interactions for future reference.

In this project we use GooglePalm as the Large Language model and the LangChain is an open-source framework that allows you to connect large language models to your applications.

LangChain has a diverse and vibrant ecosystem that brings various providers under one roof, including Google's PaLM2 large language model. Langchain provides clean and single code and the ability to swap models with minimal changes.

## 7. ACKNOWLEDGEMENT

We would like to thank our guide, Mrs. S. J. Shruthi Rani Yadav, for her invaluable support and guidance throughout this project. Their expertise and encouragement have been instrumental in our success. We are truly grateful for their dedication and metorship.

## 8. REFERENCES

1. Ms. T Padmaja, Associate Professor, Department of Computer Science and Engineering, Malla Reddy College of Engineering and Technology, Hyderabad, India.
2. Mohaimenul Azam Khan Raiaan, Md. Saddam Hossain Mukta, Kaniz Fatema, Nur Mohammad Fahad,Sadman Sakib, Most.Marufatul Jannat Mim1,Jubaer Ahmad1Mohammed Eunus Ali3,and Sami Azam

3. (PDF) A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges (researchgate.net)
4. (PDF) Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast (researchgate.net)
5. END-TO-END LLM Project Using Langchain, Google PaLM in Pharma Industry | by Charanrio | Medium
6. How to create an End-to-End LLM project in the retail industry, Using LangChain, Google Palm – ArtificialIntelligencepedia
7. End-to-End Machine Learning Projects with Source Code | by Aman Kharwal | Coders Camp | Medium
8. https://aqibrehmanpirzada/End-to-End-LLM-QNA-with-Palm
9. Zahra Abbasiantaeb, Yifei yuan, Evangelos kanoula
10. Exciting Project Ideas Using Large Language Models (LLMs) - GeeksforGeeks
11. Google AI PaLM 2 – Google AI