

LLM and RAG Powered Chatbot for the College of Computer Science and Mathematics at the University of Mosul

¹Ban Sharief Mustafa, ²Yusuf Ersayyem Madhi

¹Instructor Dr, Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

²Student, Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

Abstract - This paper presents a proposal for the design and implementation of a chatbot for the website of the College of Computer Science and Mathematics at the University of Mosul. The chatbot uses a Large Language Model (LLM) supported by Retrieval Augmented Generation (RAG) technology, the first of its kind at the university. This innovative approach to designing the chatbot combines the strengths of pre-trained language models and RAG technology, enabling the chatbot to generate informative and accurate responses to a wide range of user inquiries. The chatbot is designed to provide immediate support and act as a virtual assistant for students, faculty, staff and visitors to the college website. It covers various topics related to the College of Computer Science and Mathematics at the University of Mosul, including faculty profiles, research activities, events, academic programs, admission requirements and other college-related matters.

Keywords: Chatbot, Hallucination, Large Language Models (LLMs), Retrieval Augmented Generation (RAG), Natural Language Processing (NLP).

I. INTRODUCTION

Chatbots, also known as conversational agents, have become ubiquitous in our digital world and are permeating various domains such as websites, messaging apps, and phone systems. These computer programs are designed to engage in human-like conversations, effectively simulating the experience of interacting with a real person [1]. The primary goal of chatbots is to provide users with a seamless interface to access information, complete tasks, or receive assistance. Chatbots can be implemented in various ways, including text-based interactions, voice commands, or a hybrid approach that combines the two [2]. Text-based chatbots rely on natural language processing (NLP) techniques to interpret user input and generate appropriate responses, while voice-based chatbots leverage speech recognition techniques to facilitate spoken conversations [3].

II. MOTIVATION

The College of Computer Science and Mathematics at the University of Mosul requires an easy, efficient and informative way to answer inquiries from faculty members, students or visitors on its website. The aim of this paper is to design a chatbot system for the College of Computer Science and Mathematics at the University of Mosul. This chatbot will be the first of its kind and will integrate seamlessly with the website of the College of Computer Science and Mathematics at the University of Mosul and enable user interaction for comprehensive support. It leverages large language models (LLMs)[9] and RAG technology to provide an advanced user experience. It leverages pre-trained LLMs and fine-tunes these models on college-specific data, including event and activity descriptions, faculty profiles and other college-specific information. It also takes advantage of the Retrieval Augmented Generation (RAG) technique, which is a technique used in large language models (LLM) to enhance their cognitive capabilities by incorporating external information during response generation. The Retrieval Augmented Generation technique improves the quality and accuracy of responses by retrieving relevant information and addressing the problem of hallucination in large language models and the inability to access the correct data [4] and [5] Then the information for the automated simulator can be updated dynamically based on the information on the college website and respond to user inquiries

III. PROPOSED SYSTEM

Designing a chatbot that interacts with users requires a systematic, comprehensive and precise approach. Figure 1 illustrates the design of a chatbot for the website of the College of Computer Science and Mathematics at the University of Mosul. This diagram illustrates the basic stages of implementation and user interface design. The goal is to create an automated chatbot that is able to understand user inquiries, provide relevant information, and provide an exceptional user experience. The following is the general

diagram of the proposed automated chatbot for the college website, shown in Figure 1.

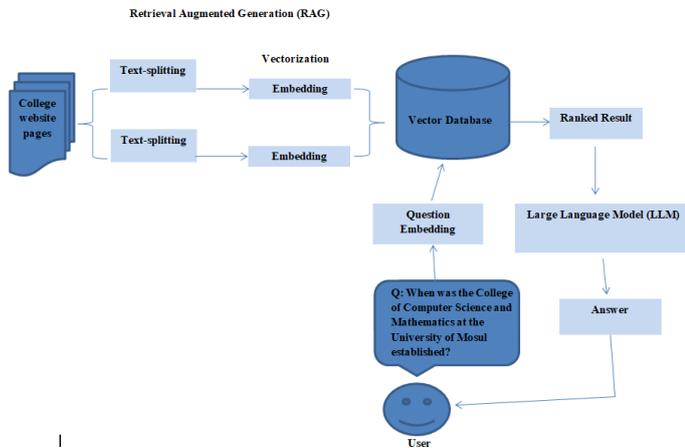


Figure 1: Chatbot Design Diagram

IV. HOW RAG WORKS IN THE DIAGRAM

1. Information retrieval: The system retrieves relevant information from the website in response to a user query using BeautifulSoup, a Python library for parsing HTML and XML documents [6]. The text is divided into chunks and then converted into embeddings, a powerful technique in natural language processing (NLP) that represents words as vectors of real numbers [7].
2. Semantic search: The question embedding is compared to the knowledge base embeddings. This is done using a semantic search algorithm that aims to find the most relevant documents based on their meaning and not just keywords [8].
3. Large Language Model (LLM): The retrieved data is then fed into a large language model, the LLM uses this information to generate a personalized response to the user query.
4. RAG-based chatbots are more informative and comprehensive than traditional chatbots because they have access to up-to-date and accurate information. This allows for more useful and relevant responses to user queries.

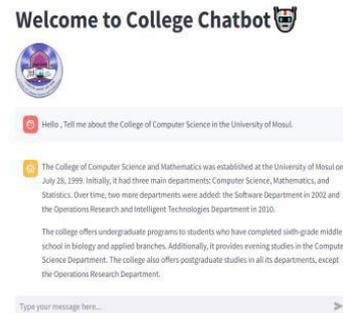
V. RESULTS AND DISCUSSIONS

The chatbot demonstrated acceptable performance in understanding user intent and providing relevant responses. It successfully interpreted a variety of queries, from general questions about the college and its programs to more specific requests for information. A key strength of the chatbot was its ability to access and retrieve information from the college website. It could provide quick links to important resources and even notify users of college events. The chatbot

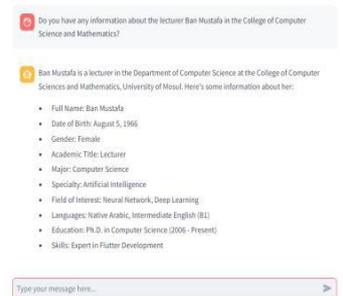
performed above expectations, demonstrating a deep understanding of the site's content and structure.

The following examples highlight the versatility and capabilities of the chatbot and user support program, which contribute significantly to the development of the college's educational landscape.

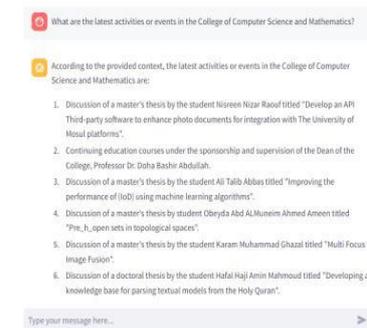
- Providing accurate responses



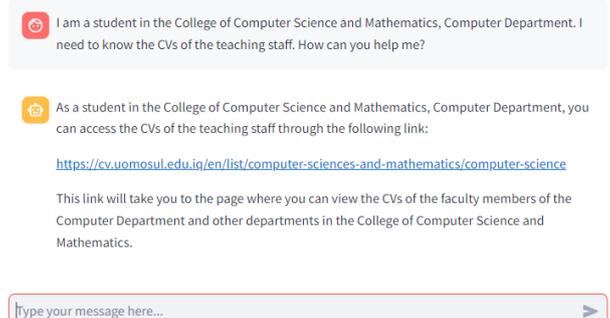
- Faculty and staff information



- Latest of event and activities in the College



- Research assistance



VI. CONCLUSION

The chatbot is designed to provide instant support and act as a virtual assistant for students, faculty, staff and visitors to the college website and covers various topics related to the College of Computer Science and Mathematics at the University of Mosul including faculty profiles, research activities, events, academic programs, admission requirements and other college matters.

REFERENCES

- [1] Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y.,... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.
- [2] Caldarini, G., Jaf, S. and McGarry, K., 2022. A literature survey of recent advances in chatbots. *Information*, 13(1), p.41.
- [3] Oktar, Y., Okur, E., & Turkan, M. (2020). Self-recognition in conversational agents. arXiv preprint arXiv:2002.02334.
- [4] Li, D., Yan, J., Zhang, T., Wang, C., He, X., Huang, L.,... & Huang, J. (2024). On the Role of Long-tail Knowledge in Retrieval Augmented Large Language Models. arXiv preprint arXiv:2406.16367.
- [5] Wang, S., Song, J. L. S., Cheng, J., Fu, Y., Guo, P., Fang, K.,... & Dou, Z. (2024). Domain RAG: A Chinese Benchmark for Evaluating Domain-specific Retrieval-Augmented Generation. arXiv preprint arXiv:2406.05654.
- [6] Patel, J. M., & Patel, J. M. (2020). Web scraping in python using beautiful soup library. Getting structured data from the internet: running web crawlers/scrapers on a big data production scale, 31-84.
- [7] Patil, R., Boit, S., Gudivada, V., & Nandigam, J. (2023). A survey of text representation and embedding techniques in nlp. *IEEE Access*, 11, 36120-36146.
- [8] Li, L., & Lv, J. (2024). Unlocking Insights: Semantic Search in Jupyter Notebooks. arXiv preprint arXiv:2402.13234.
- [9] Bharathi Mohan, G., Prasanna Kumar, R., Vishal Krishh, P., Keerthinathan, A., Lavanya, G., Meghana, M. K. U., ... & Doss, S. (2024). An analysis of large language models: their impact and potential applications. *Knowledge and Information Systems*, 1-24.
- [10] Ms.Ch.Lavanya Susanna, R.Pratyusha, P.Swathi, P.Rishi Krishna, V.Sai Pradeep "College Enquiry Chatbot" *International Research Journal of Engineering and Technology (IRJET)*.

Citation of this Article:

Ban Sharief Mustafa, & Yusuf Ersayyem Madhi. (2024). LLM and RAG Powered Chatbot for the College of Computer Science and Mathematics at the University of Mosul. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 8(10), 59-61. Article DOI <https://doi.org/10.47001/IRJIET/2024.810010>
