



UNI – A Retrieval Augmented Generation Powered Virtual Assistant for College Related Queries

Sanjana J.^{*1}, Mahadev Prasad Y. N.¹, Srinivas B.¹, Sharanya S.¹, Madhusudhan M.¹

¹Department of Information Science and Engineering, MIT Thandavapura, India

Emails: sanjana.4mn20is027@gmail.com; mahadevapasady@gmail.com; srinivasb20016@gmail.com; sharanyashivanna121@gmail.com; madhu1435sudhan@gmail.com

Abstract

This paper unveils an advanced chatbot engineered to cater specifically to college-related inquiries. Harnessing the power of BARD and incorporating a wake word activation system with automatic speech recognition, the chatbot offers an enhanced user experience marked by both linguistic sophistication and spoken command initiation. The methodology encompasses the nuanced process of pre-training on diverse corpora, fine-tuning to optimize responsiveness to college-specific queries, and the seamless integration of intent classification and entity recognition. These facets collectively empower the chatbot to understand and respond effectively to the intricacies of user inputs. A comprehensive knowledge base is curated to ensure not only accurate information retrieval but also to foster a depth of contextual understanding. This project signifies a pioneering leap in providing an innovative, user-friendly, and ethically driven solution for addressing college-related queries through natural language interactions. By showcasing practical advancements in chatbot technology tailored to the educational landscape, this research contributes to the evolving landscape of intelligent virtual assistants.

Keywords: Virtual assistant; Large language model; Natural language queries; Entity recognition; Wake word activation; Intent classification; Retrieval augmented generation

1. Introduction

In recent years, the integration of artificial intelligence and natural language processing technologies has witnessed a transformative impact on human-computer interactions. Within the educational domain, the development of intelligent virtual assistants has emerged as a significant frontier, offering innovative solutions to address user inquiries and enhance overall user experiences. This paper introduces a pioneering project focused on the implementation of a sophisticated chatbot specifically tailored to cater to college-related queries, positioning itself at the intersection of advanced natural language processing techniques and the unique demands of the educational landscape.

The advent of Large Language Models (LLMs), has opened new avenues for crafting virtual assistants capable of nuanced language understanding and generation. This project harnesses the power of LLMs to deliver a virtual assistant that not only comprehends user inputs with sophistication but also responds in a manner that replicates human-like interactions. A distinctive feature of this virtual assistant lies in its utilization of a wake word activation system, coupled with automatic speech recognition, allowing users to initiate interactions effortlessly through spoken commands.

By addressing the unique challenges posed by college-related inquiries, this project not only contributes to the growing body of research in natural language processing and virtual assistant technologies but also presents a practical application at the forefront of educational technology. The ensuing sections of this paper will delve into the literature reviewed, VA characteristics, and methodology of the implemented chatbot, shedding light on its potential implications and advancements in the landscape,

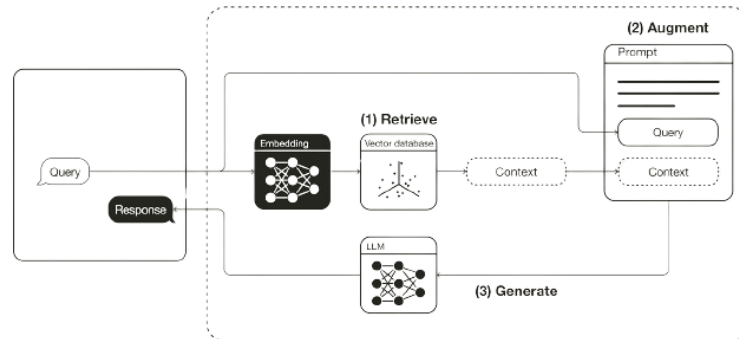


Figure 1: Retrieval Augmented Generation

2. Related Work

The integration of Natural Language Processing (NLP) into virtual assistants has witnessed substantial advancements, particularly within educational contexts. This literature review synthesizes key insights from relevant studies that explore the implementation of virtual assistants for addressing college-related inquiries.

A. Understanding Chatbots

In their comprehensive overview of chatbot technology, Adamopoulou and Moussiades (2020) provide foundational insights into the broader landscape of chatbots. While not specifically focused on college-related FAQs, this source lays the groundwork for understanding the evolution, features, and applications of chatbots, setting the stage for a deeper exploration in educational settings.

B. Virtual Assistance in any context

Annette et al. (2020) present insights into virtual assistance across diverse contexts, emphasizing the adaptability of these technologies. While not specific to colleges, this study contributes valuable perspectives on the broader applications of virtual assistants. Understanding virtual assistance across various domains can inform the design and deployment strategies for a college-related virtual assistant.

C. Machine Learning and Chatbots

Suta et al. (2020) provides an insightful overview of machine learning in chatbots, directly relevant to the incorporation of NLP in virtual assistants. The study explores the intersection of machine learning and chatbot development, offering insights into methodologies that can boost the performance of a virtual assistant for college-related queries.

D. Other College Enquiry Chatbots

Phalle et al. (2021) contribute a study exploring an AI powered web-based interactive college enquiry chatbot. Although not explicitly centered on FAQs, the research likely offers practical insights into deploying a chatbot in the context of college inquiries. This source is instrumental in understanding the synergy between AI and web-based platforms for handling user queries related to college information. In their work on an AI chatbot tailored for college inquiries, Tiwari et al. (2023) directly address the theme of

interest. The study's findings are expected to provide valuable insights into the challenges, successes, and considerations when implementing an AI chatbot specifically designed for handling college related FAQs.

E. Retrieval Augmented Generation

Li et al. (2023) presents a survey on retrieval-augmented text generation, offering insights into advanced techniques that involve the integration of retrieval mechanisms with text generation. This survey informs the enhancement of virtual assistant capabilities, especially in the context of information retrieval for FAQs. Shuster et al. (2023) explore the application of retrieval augmentation in conversation to reduce hallucination. This research is valuable for understanding how retrieval mechanisms can improve the accuracy and reliability of responses generated by virtual assistants, a crucial consideration for handling college-related queries.

F. NLP and STT Applications with Chatbots

Inupakutika et al. (2021) focus on the integration of NLP and speech-to-text applications with chatbots. This study is particularly pertinent to the implementation of a virtual assistant for college-related FAQs, addressing the fusion of NLP and speech recognition technologies. Understanding the synergy between these technologies significantly contributes to the effectiveness of the implemented system.

In conclusion, the amalgamated insights from these studies form a comprehensive understanding of chatbot and virtual assistant technologies, with a specific focus on their applications in handling college-related inquiries. These findings lay the groundwork for future research endeavors, providing a robust foundation for refining and optimizing virtual assistants tailored for addressing FAQs in a college setting.

3. Chatbot Characteristics

Recognizing the essential attributes of a chatbot holds significance in the design process. These characteristics have been identified by studying people's expectations for chatbots, employing a methodology that involves comparing past human-human and human-chatbot interactions. As indicated in a chatbot survey, the evolution of chatbot development has progressed from simple pattern matching and "Q&A" styles to a more human-like approach to engaging in and sustaining conversations. This evolution highlights the expectation for advanced chatbots not only to answer questions but also to learn and enhance their capabilities with each interaction, ultimately demonstrating adaptability across various contexts. To craft an intelligent chatbot capable of such functionalities, the essential characteristics and capabilities are outlined as follows:

A. Communication Using Natural Language

Chatbots are expected to comprehend and reply in a manner consistent with human natural language. Presently, these chatbots predominantly employ Machine Learning (ML) in conjunction with Natural Language Processing (NLP) within the realm of Artificial Intelligence (AI). It has been observed that the structure of information-seeking chatbots typically comprises three key components: Natural Language Understanding (NLU) for categorizing the user's intent, Dialogue Management (DM) for determining the user's intent, and Natural Language Generation (NLG) for crafting responses in natural language. Following the identification of pertinent answer categories, the chatbot is required to: 1) formulate a suitable, personalized response in natural language, and 2) deliver the response promptly without any delay.

B. Reliability

Chatbots are expected to offer information that is not only reliable but also verified by authoritative sources. Ensuring the accuracy and credibility of the information is crucial to instill confidence in users and provide them with trustworthy responses. By relying on verified data from recognized authorities, chatbots contribute to a more secure and reliable exchange of information, promoting user satisfaction and confidence in the accuracy of the provided content.

C. Security

Safeguarding the integrity of the chatbot system, security features must be in place to exclusively grant system or database access to authorized users. Validation and sanitization of user inputs constitute essential measures to counter prevalent security threats like SQL injection. These proactive security protocols fortify the chatbot against potential vulnerabilities, reinforcing its capability to ensure a secure and trusted user experience.

4. Methodology

The methodology for implementing the virtual assistant for college-related FAQs involves a systematic process that encompasses the following key components:

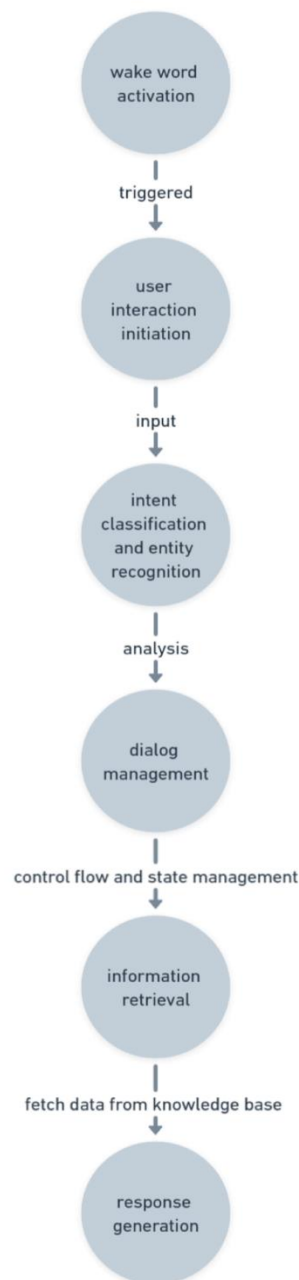


Figure 2: Flowchart

A. Wake Word Activation

The chatbot incorporates a wake word activation system to listen for a specific trigger phrase. This system employs automatic speech recognition (ASR) to convert spoken words into text, allowing the chatbot to identify when a user wants to initiate a conversation.

B. User Interaction Initiation

Upon detecting the wake word, the chatbot system recognizes that a user wants to interact. The ASR system processes the spoken command, converting it into a textual representation for further analysis.

C. Intent Classification and Entity Recognition

The pre-trained language model performs intent classification to understand the user's goal or purpose in the interaction. It categorizes user inputs into specific intent categories, such as asking for information, making a request, or expressing a preference. Simultaneously, entity recognition identifies and extracts specific details or entities from the user's input. For example, it recognizes faculty names, departments, or other relevant information.

D. Dialog Management

The chatbot uses a dialog management system, potentially based on state machines or rule-based approaches, to organize the flow of conversation. It maintains context, manages conversation states, and generates coherent responses based on the identified intent and extracted entities.

E. Information Retrieval

When the user's query requires specific information, the chatbot retrieves relevant data from its knowledge base. This could include details about colleges, admission processes, course offerings, and other related information.

F. Response Generation

The language model generates a natural language response based on the identified intent, extracted entities, and retrieved information. The response is crafted to provide accurate and contextually relevant answers to the user's query.

5. Conclusion

In this paper, we have presented an innovative project focused on the implementation of a virtual assistant tailored specifically for college-related FAQs, leveraging advanced Natural Language Processing (NLP) techniques. The development of the chatbot, rooted in the capabilities of BARD and a wake word activation system, aims to address the unique challenges posed by inquiries within the educational domain. Our methodology encompasses comprehensive pre-training, finetuning, and the integration of intent classification and entity recognition to enhance the chatbot's language understanding capabilities. The literature review provided a contextual foundation, drawing from studies that explore the broader landscape of chatbot technology, the integration of NLP in virtual assistants, and recent advancements in retrieval-augmented text generation. By incorporating insights from diverse sources, we have positioned our project within the evolving field of intelligent virtual assistants, specifically tailored for college related queries. The continuous improvement strategy, involving real-world deployment and user feedback analysis, ensures the adaptability and responsiveness of the chatbot to the evolving needs of users. Ethical considerations embedded in the design prioritize user privacy, reflecting a commitment to responsible AI deployment.

6. Future Enhancements

As we look ahead, there are several avenues for future research and development:

1. **Enhanced Knowledge Base:** Further enrichment of the knowledge base with dynamic, real-time updates to accommodate evolving information and changes within the college environment.
2. **Multi-Modal Integration:** Exploration of multi-modal capabilities, including image and document processing, to enable the chatbot to handle a broader range of user inquiries.
3. **Trustworthiness and Explainability:** Incorporation of mechanisms to enhance the trustworthiness and explainability of the chatbot's responses, addressing user concerns and ensuring transparency in the decision-making process.
4. **User Experience Optimization:** Iterative enhancements to the user interface and experience to make interactions more intuitive, personalized, and aligned with user preferences.
5. **Collaboration with Educational Institutions:** Collaborative efforts with educational institutions to integrate the chatbot seamlessly into existing systems, providing tailored support for students, faculty, and staff.
6. **Integration with Academic Resources:** Integration with academic resources such as course catalogs, syllabi, and academic calendars to provide more comprehensive and accurate responses to academic inquiries.
7. **Advanced Conversational Capabilities:** Exploration of advanced conversational capabilities, including sentiment analysis and emotional intelligence, to enable more nuanced and empathetic interactions.

In conclusion, this project marks a significant step towards the development of intelligent virtual assistants tailored for educational settings. The outlined future works aim to propel the system's capabilities further, ensuring it remains at the forefront of addressing college-related queries through natural language interactions. The collaboration between technology and education holds vast potential, and the ongoing refinement of our virtual assistant seeks to unlock new possibilities for enhanced user support and engagement within the academic realm.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

- [1] In proceeding of 'AI and web-based interactive college enquiry chatbot', Parkar, Rohan and Payare, Yash and Mithari, Keyur and Nambiar, Jitesh and Gupta, Jaya, 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 1–5, 2021, IEEE
- [2] 'An overview of machine learning in chatbots', Suta, Prissadang and Lan, Xi and Wu, Biting and Mongkolnam, Pornchai and Chan, Jonathan H, International Journal of Mechanical Engineering and Robotics Research, 502–510, 2020
- [3] Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects, Hadi, Muhammad Usman and Qureshi, Rizwan and Shah, Abbas and Irfan, Muhammad and Zafar, Anas and Shaikh, Muhammad Bilal and Akhtar, Naveed and Wu, Jia and Mirjalili, Seyedali and others, Authorea Preprints, 2023,
- [4] A comprehensive solution to retrieval-based chatbot construction, Moore, Kristen and Zhong, Shenjun and He, Zhen and Rudolf, Torsten and Fisher, Nils and Victor, Brandon and Jindal, Neha, Computer Speech & Language, 10, 2023,
- [5] 'Virtual assistance in any context: A taxonomy of design elements for domain-specific chatbots', Janssen, Antje and Passlick, Jens and Rodríguez Cardona, Davinia and Breitner, Michael H, Business & Information Systems Engineering, 62, 211–225, 2020,
- [6] Pooja, Dr. Manish Kumar Mukhija, Satish Kumar Alaria. "Smart City's Security Model for Management of Image Data on Cloud." Journal of Cognitive Human-Computer Interaction, Vol. 2, No. 1, 2022 ,PP. 8-14.

- [7] Parvesh K,Tharun C,Prakash M. "Apparel Recommendation Engine Using Inverse Document Frequency and Weighted Average Word2vec." *Journal of Cognitive Human-Computer Interaction*, Vol. 1, No. 2, 2021 ,PP. 46-56.
- [8] 'Chatbot Development for Educational Institute', Shingte, Kshitija and Chaudhari, Anuja and Patil, Aditee and Chaudhari, Anushree and Desai, Sharmishta, Available at SSRN 3861241, 2021
- [9] 'AI Chatbot for College Enquiry', Tiwari, Rashmi and Khandelwal, Reema and Agrawal, Yash and Tiwari, Vinit and Bisen, Wani, journal=Available at SSRN 4422268, 2023
- [10] 'Web-based chatbot for frequently asked queries (FAQ) in hospitals', Mittal, Mamta and Battineni, Gopi and Singh, Dharmendra and Nagarwal, Thakur Singh and Yadav, Prabhakar, *Journal of Taibah University Medical Sciences*, 740–746, 2021,
- [11] Pooja,Dr. Manish Kumar Mukhija,Satish Kumar Alaria. "Smart City's Security Model for Management of Image Data on Cloud." *Journal of Cognitive Human-Computer Interaction*, Vol. 2, No. 1, 2022 ,PP. 8-14.
- [12] 'Integration of NLP and Speech-to-text Applications with Chatbots', Inupakutika, D and Nadim, M and Gunnam, GR and Kaghyan, S and Akopian, D and Chalela, P and Ramirez, AG, *Electronic Imaging*, 2021, 3, 35–1, 2021,
- [13] M. Sumithra,B. Buvaneswar,Jessica Judith S,,Dymphna Mary C,Punitha R,,Pavithraa S. "Innovation for Better Education System using Artificial Intelligence." *Journal of Cognitive Human-Computer Interaction*, Vol. 2, No. 1, 2022 ,PP. 19-28.
- [14]Sumithra.M,Vettri Chezian. P,Raj Kumar. K,Nithish Kumar. S. "Automated Attendance System using Real Time Face Recognition and MySQL Database." *Journal of Cognitive Human-Computer Interaction*, Vol. 2, No. 1, 2022 ,PP. 15-18.
- [15] 'A survey on retrieval-augmented text generation', Li, Huayang and Su, Yixuan and Cai, Deng and Wang, Yan and Liu, Lema, arXiv preprint arXiv:2202.01110, 2022
- [16] 'Retrieval augmentation reduces hallucination in conversation', Shuster, Kurt and Poff, Spencer and Chen, Moya and Kiela, Douwe and Weston, Jason, 2021,
- [17] RajmohanP,Rishisarvesh. U. S ,Ponkumar. G ,Naveen. R,Pragathish. R. S,Santhosh. P ,Arun. k. D ,Syed moineudeen. "Transformer Analyzer BOT." *Journal of Cognitive Human-Computer Interaction*, Vol. 3, No. 2, 2022 ,PP. 16-20.
- [18] 'TRAC: Trustworthy Retrieval Augmented Chatbot', Li, Shuo and Park, Sangdon and Lee, Insup and Bastani, Osbert, arXiv preprint arXiv:2307.04642, 2023
- [19] 'An overview of chatbot technology', Adamopoulou, Eleni and Moussiades, Lefteris, *IFIP international conference on artificial intelligence applications and innovations*, 373–383, 2020,