CHATBOT WITH USER SPECIFIC DATA

*C. Amuthadevi*

*Department of Computational Intelligence SRM Institute of Science and Technology, Kattankulathur, India* [*amuthadc@srmist.edu.in*](mailto:amuthadc@srmist.edu.in)

*Harshil Choudhary Department of Computational Intelligence SRM Institute of Science and Technology,*

*Kattankulathur, India* [*hc4653@srmist.edu.in*](mailto:hc4653@srmist.edu.in)

*Dhruva Bhattacharya Department of Computational Intelligence SRM Institute of Science and Technology,*

*Kattankulathur, India* [*db6688@srmist.edu.in*](mailto:db6688@srmist.edu.in)

***Abstract*—**This project focuses on creating a Chatbot for User- Specific Data (CUSDB) to revolutionize personalized information retrieval. By amalgamating cutting-edge NLP, machine learning, and data integration techniques, CUSDB offers tailored data experiences. Its architecture features a user-friendly interface, adept NLP engine, seamless data integration, and an insightful recommendation system. Rigorous testing and optimization ensure reliability and security. Through iterative development informed by user feedback, CUSDB aims to redefine data interaction paradigms, presenting a potential to significantly impact various domains with its personalized data retrieval capabilities.

**Keywords—NLP, CUSDB, machine learning, chatbot, data integration techniques**

1. INTRODUCTION

In the era of information overload, finding relevant and reliable data can be a daunting task. Users often have to sift through countless sources, filter out noise, and verify the accuracy of the information they need. This process can be time-consuming, frustrating, and inefficient. What if there was a way to access personalized data that matches your specific needs and preferences.

This is the goal of the Chatbot for User-Specific Data (CUSDB), a novel system that aims to provide a tailored and intelligent solution for data retrieval. CUSDB is not just a regular chatbot that answers predefined queries. It is a smart and interactive agent that understands your natural language requests, learns from your feedback, and adapts to your preferences.

CUSDB leverages state-of-the-art Natural Language Processing (NLP) and machine learning techniques to achieve its functionality. It uses NLP to analyse your queries, extract keywords, identify intents, and generate responses. It also uses machine learning to learn from your interactions, refine its search strategies, and optimize its performance.

CUSDB can handle various types of data sources, such as web pages, databases, documents, images, videos, etc. It can also integrate with external APIs and services to provide additional features and functionalities. CUSDB can deliver the data you need in various formats, such as text, tables, charts, graphs, etc.

CUSDB is designed to be user-friendly, flexible, and customizable. You can interact with CUSDB through text or voice input. You can also adjust the settings and preferences

of CUSDB according to your needs and goals. You can provide feedback and ratings to CUSDB to help it improve its service quality.

CUSDB is a cutting-edge system that aims to revolutionize the way users access data. It is a powerful tool that can save you time, effort, and resources. It is also a fun and engaging way to explore and discover new information. With CUSDB, you can enjoy a personalized and intelligent data retrieval experience.

1. MOTIVATION

The inspiration for this project comes from a crucial observation in the field of chatbot technology. Many existing chatbots are limited by their ability to provide user-specific data. They often rely on generic data and fail to meet individual needs effectively. This problem shows a gap in the market for a chatbot that can fully understand and answer user-specific data queries, offering a personalized and customized user experience.

By developing a Chatbot for User-Specific Data (CUSDB), we aim to fill this gap and transform the chatbot industry. Using advanced Natural Language Processing (NLP) and machine learning techniques, CUSDB will have the ability to learn and adapt from each user interaction, ensuring that it can deliver relevant and accurate information based on individual preferences. The motivation for this project lies in the pursuit of creating a chatbot that truly caters to the nuances of user-specific data, thereby improving the quality and relevance of information provided while significantly enhancing the user experience.

1. INNOVATION OF THE PROJECT
2. User driven data control: Give users the ultimate control over their own data, enabling them to specify, manage, and improve their data preferences and sources within the chatbot system.
3. Multimedia data connection Expand the possibilities by enabling the chatbot to smoothly handle and interpret various kinds of data types, including images, videos, audio clips, YouTube videos, website links, PDFs, text files, and CSV files.
4. Interconnected data insight : Create connections between different data formats, enabling the chatbot to derive insights from a comprehensive view of user-specific data, enriching the information it provides and enabling cross- format recommendations.
5. Data privacy and security: Emphasize strong data privacy and security measures, ensuring users that their personal data and preferences are safeguarded throughout interactions with the chatbot.
6. REQUIREMENT GATHERING

Integration requirements: Identify if the chatbot needs to connect with other software or platforms to access or provide data. For example, it might need to retrieve data from spreadsheets, databases, or cloud storage, or send data to other applications or services.

Data requirements: Define the types of data the chatbot will deal with. This includes data sources, formats, and any data transformations or analysis it needs to perform. For example, it might need to convert data from one format to another, filter out irrelevant data, aggregate data, or apply statistical or machine learning methods.

User interface requirements: Determine how users will communicate with the chatbot. Will it be through a chat window, a web interface, a voice interface, or a combination of these? Design the user interface, including chatbot dialogs, menus, and user prompts. Consider the usability, accessibility, and aesthetics of the user interface.

Security and privacy requirements: Address security and privacy concerns. Ensure that sensitive data is handled securely and that user privacy is protected. For example, it might need to encrypt data, authenticate users, or comply with data protection regulations.

1. LITERATURE SURVEY

The paper by Matthew Nicholson, Yaofeng Wang and Yiyi Chen shows data involved designing a chatbot to automate the communication process between equity and derivative traders and stock lending and borrowing traders at BNP Paribas. The chatbot simplifies and speeds up the short sale request process, reducing manual steps and providing a user- friendly interface. The project also included developing a web-based front-end application and documenting and analysing the system. The chatbot improves efficiency but

may need maintenance and updates to cope with changing trading requirements and regulations. User training and adoption may also be challenging during the initial implementation phase.

In 2020, Mlađan Jovanović, Marcos Baez and Fabio Casati conducted a study systematically analysed 158 publicly available healthcare chatbots in English, as of August 2020, from popular databases such as Botlist and Chatbots.org. The chatbots were annotated based on health provisioning roles and functions and evaluated using a scoring system. The study also identified archetypes, design dimensions, and gaps in the current state of healthcare chatbots. The study provides a comprehensive and domain-specific analysis of healthcare chatbots, covering a wide range of aspects and features. It introduces a clear analytical framework that allows for systematic comparison and evaluation of healthcare chatbots. It also highlights the limitations and challenges in the field and provides guidance for future developments.

The paper by Marcos Baez; Florian Daniel, Fabio Casati and Boualem Benatallah discusses the significant chatbot integration patterns, based on a mixed approach of literature review and systematic search. It identifies eight distinct patterns of how chatbots can be integrated into existing software systems and illustrates them with informative figures. It also discusses the challenges and considerations for developers when designing and developing chatbots for various types of systems. The article provides a comprehensive and informative analysis of chatbot integration, with references to a large number of research papers and sources. However, it does not cover developments or research beyond its knowledge cutoff date in September 2021.

The paper by Konstantinos I. Roumeliotis, Nikolaos D. Tselikas and Dimitrios K. Nasiopoulos introduces a study that investigates how early adopters use Llama 2, an open- source pre-trained model by Meta, in their AI projects. It uses qualitative analysis to extract insights from case studies and evaluates the model’s strengths, weaknesses, and areas for improvement. The study shows the diverse applications and effectiveness of Llama 2, as well as its smooth user experience. Future research could include cross-model comparisons and exploring Llama 2 in various domains. The study’s limitations include the narrow range of perspectives and the lack of extensive cross-model comparisons.

The Article by GWENDAL DANIEL, JORDI CABOT, LAURENT DERUELLE, AND MUSTAPHA DERRAS

utilizes the Xatkit framework, a low code chatbot development tool. It explains the framework’s purpose, structure, and components. The framework allows users to define chatbots using Domain Specific Languages (DSLs) and deploy them across various platforms. The advantages of Xatkit, such as simplifying complex chatbot development, enabling platform independence, and facilitating component evolution. It also acknowledges the limitations of Xatkit, such as the learning curve for DSLs, the limited platform support, and the integration challenges with external services.

The Project by Arjun Pesaru, Taranveer Singh Gill and Archit Reddy Tangella in 2023 that uses the LangChain framework and the Large Language Model (LLM) to create a PDF chatbot. The chatbot can generate text and answer user queries based on PDF files. The project also uses Pinecone for storing PDF file vectors and React JS for creating the front-end web interface. The project tests the chatbot’s accuracy with a series of queries with PDF files. It demonstrates the potential of LangChain and the LLM Model to create efficient and informative chatbots for various purposes. The article also shows how Pinecone and React JS enhance the functionality and user interaction of the chatbot. The article acknowledges the limitations of the project, such as the accuracy and fluency of chatbot responses, and suggests future improvements in terms of training, evaluation, user experience, and feature expansion.

According to V Akshatha Prasad and R Ranjith in 2020 describes a way for creating an intelligent chatbot for lab security and automation. The chatbot uses natural language processing, computer vision, and machine learning to interact with users and control lab devices. The chatbot can perform tasks such as opening and closing doors, turning on and off lights, detecting intruders, and sending alerts. It also discusses the advantages and limitations of this way. The advantages include enhancing lab security and efficiency, reducing human errors and risks, and providing a user- friendly interface. The limitations include technical challenges, ethical issues, and user acceptance. The article suggests future improvements in terms of functionality, performance, and usability.

In 2019, Andreas REISWICIVL and Martin HAAG published an article which evaluates different chatbot prototypes for taking the virtual patient’s history in medical education. The chatbot prototypes use natural language processing and machine learning algorithms to handle unknown questions and provide flexible interactions with virtual patients. The chatbot uses Python for implementation and leave-one-out cross-validation for evaluation. The article addresses the need for more individual and realistic interactions with virtual patients in medical education. It explores the potential of natural language processing and machine learning algorithms to enhance the learning experience. The article also compares the performance of different chatbot prototypes using evaluation metrics. The article’s limitations include the small dataset (n=109), which leads to high variability in evaluation results. The article suggests improving the data quality and increasing the training data to improve the performance of the chatbot prototypes. The article also discusses possible enhancements such as voice commands and VR avatars, but these concepts require further research and development. The article also acknowledges the need for a feedback channel and UI improvements.

The proposed COLLEGE ENQUIRY CHATBOT by Ms.Ch.Lavanya Susanna, R.Pratyusha, P.Swathi,P.Rishi Krishna and V.Sai Pradeep is a web-based application that uses the CodeIgniter PHP framework and MySQL database

to provide relevant answers to user queries about college- related information. The article explains the development of the system, the algorithms used to interpret and respond to user questions, and the feedback mechanism for administrators to refine responses.

Tarun Lalwani and Shashank Bhalotia’s chatbot, first it employs pre-processing to standardize input text and identify context based on keywords. Next, the Personal Query Response System checks user authenticity, processes queries, and provides information from a database when valid. The AIML Response System handles normal conversations using AIML files, saving user data and sending responses when patterns are recognized, or indicating "Invalid Input" for unfamiliar queries. The Query Analysis and Response System focuses on providing college-related information by matching queries to AIML patterns or assessing input similarity with predefined questions. Unmatched queries are logged for potential system improvement, and users receive an "Answer not available" response when no matches are found. Lastly, a Context Reset option allows users to exit the system, resetting all input parameters. Overall, this chatbot system aims to efficiently respond to user queries, offer information, and maintain user context and satisfaction.

The study by Amit Sharma, Per Eirik Undheim & Salman Nazir describes the design and implementation of “FLOKI,” a chatbot that helps maritime trainees learn Collision Avoidance Regulations (COLREGs). The chatbot was created using IBM Watson Assistant and interacted with 18 maritime trainees. It evaluated the user experience using the System Usability Scale (SUS) and analyzed the chatbot’s usability and its impact on learning outcomes. It also demonstrates the innovative use of AI in maritime education, especially in a complex area like COLREGs. The article also uses a user-centric evaluation method to refine the chatbot. Moreover, chatbots can improve learning efficiency by providing quick access to information.

The Project by DANIEL CARLANDER-REUTERFELT, ÁLVARO CARRERA, CARLOS A. IGLESIAS, ÓSCAR ARAQUE, JUAN FERNANDO SÁNCHEZ RADA, AND

SERGIO MUÑOZ made JAICOB, a data science chatbot that uses natural language processing to answer user queries and engage in small talk. The chatbot uses a Knowledge Base populated with relevant information from academic resources, glossaries, FAQs, and technical documentation. The chatbot also uses a speech act classifier to categorize user queries and a graphical user interface to facilitate interactions. The chatbot’s performance is evaluated through user feedback and statistical analysis, and iteratively improved based on the results.

A Study by Pavel Smutny and Petra Schreiberova about 89 educational chatbots on Facebook Messenger, using the analytic hierarchy process to measure their teaching quality, humanity, affect, and accessibility. The chatbots are categorized based on language, subject matter, and developer’s platform. The article provides insights for integrating chatbots into teaching practices, highlighting their advantages such as instant access to information, personalized learning support, and natural conversation. The article also acknowledges the limitations of the study, such as

the limited selection of chatbots, the time-bound findings, the discoverability challenges, and the need for further developer support and content analysis.

In 2021 Trung Thanh Nguyen, Anh Duc Le, Ha Thanh Hoang and Tuan Nguyen published an article which introduces NEU-chatbot, an AI-based chatbot that provides online educational consulting services for prospective students of National Economics University. The chatbot uses deep learning models integrated into the Rasa framework to answer inquiries about curriculum, admissions, tuition fees, and more. The article explains the pipeline for creating Vietnamese chatbots, emphasizing data preprocessing for optimal accuracy. The chatbot can accurately detect over fifty types of user questions, achieving a 97.1% accuracy rate on the test set. The article discusses the benefits and challenges of NEU-chatbot. The benefits include streamlining communication between students and universities, reducing the manual workload on admissions personnel, and providing instant access to crucial admission information. The challenges include manual updates to adapt to new academic year information, misunderstandings of user intents during consultations, and generative chatbots not explored. The article suggests future improvements and applications for the chatbot.

The project by ADDI AIT-MLOUK AND LILI JIANG in 2021, name KBot, a knowledge graph based chatbot that uses natural language processing and machine learning to understand and answer user queries. The chatbot supports multilingual interactions, processes keyword extraction, and generates SPARQL queries to retrieve information from diverse knowledge bases, such as DBpedia, Wikidata, and my Personality. The chatbot also handles analytical queries using rule-based methods. The chatbot emphasizes flexibility and scalability, enabling the addition of new knowledge bases and languages.

1. RESEARCH GAP

User-specific data is essential for creating personalized and engaging chatbot experiences. However, there are some challenges and considerations that need to be addressed when dealing with user-specific data. These include:

* Data Availability: Not all users may be willing or able to provide the necessary data. Some users may be uncomfortable sharing certain information, which could limit the effectiveness of the chat experience.
* Data Integration: Integrating and managing user- specific data can be challenging. It requires a robust infrastructure to store, retrieve, and update user data securely.
* Privacy and Security: Protecting user data is paramount. Any data shared by the user should be handled securely and in accordance with relevant data protection laws and regulations.
* Compliance: Ensuring compliance with data protection regulations such as GDPR, CCPA, or any other relevant laws, depending on your geographical location and user base.

1. PROPOSED METHADOLOGY

The methodology for CUSDB begins with a detailed analysis of user requirements, including the specific domain, user personas, and the degree of personalization required. This phase also entails identifying the relevant data sources for user-specific information.

Following requirement analysis, the project progresses to data collection and integration. This phase involves collecting data from various sources and implementing robust integration processes to ensure that data is efficiently structured for easy retrieval.

Next, the selection of an appropriate Natural Language Processing (NLP) model, such as OpenAI's GPT-3, is essential. The chosen model is fine-tuned to enhance its understanding of domain-specific language and user context.

User interaction design is a critical aspect of the methodology, focusing on creating an intuitive and user- friendly chat interface using tools like Streamlit. Features for user feedback and corrections are integrated to facilitate improvements in the chatbot's responses.

The personalization and learning phase entails developing algorithms that enable CUSDB to learn continuously from user interactions, personalize responses based on user history, and remember past conversations and context for improved engagement.

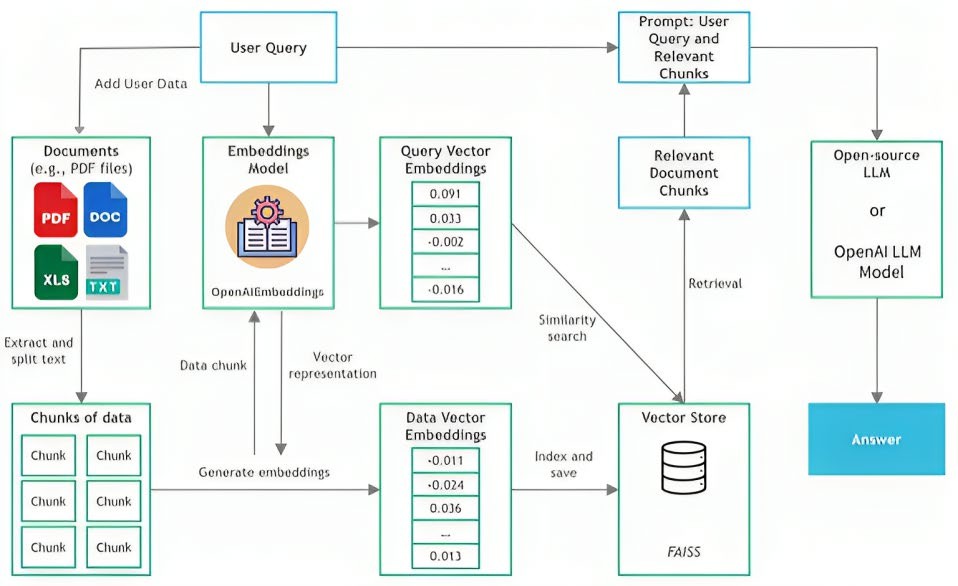
Security and compliance are prioritized in the methodology. Rigorous testing and optimization procedures ensure the chatbot's reliability and security, with a particular emphasis on handling sensitive user-specific data.

Thorough testing and evaluation are conducted across various user scenarios to identify and rectify issues, and user feedback is actively gathered and applied to refine the chatbot.

The deployment phase involves launching CUSDB in a secure and scalable environment. Monitoring and analytics tools are implemented to track user interactions, system performance, and data usage, leading to regular updates and improvements based on real-world usage.

Ethical considerations are essential, with a focus on data privacy and security. Transparent communication about data usage and privacy is crucial for user trust.

Finally, the methodology looks to the future, exploring opportunities for integration with other technologies and platforms. Potential collaborations and partnerships with organizations in relevant domains are considered to maximize CUSDB's impact and enhance user experiences.



*Figure 1 Architecture Diagram*

1. COMPARISON

LangChain and LlamaIndex, both formidable frameworks for LLM-powered applications, offer distinct attributes, each catering to specific use cases. LangChain, designed to simplify LLM integration, is notable for its versatility and ease of use. Its high-level components and pre-built chains abstract the complexities of working with LLMs, making it an ideal choice for both beginners and advanced users. Developers can quickly establish connections to various LLM providers, such as OpenAI and Hugging Face, to execute tasks efficiently. Notably, LangChain offers off-the- shelf chains, like the SqlDatabaseChain, which expedite SQL database connections and data retrieval from LLMs. This makes it an attractive option for creating intelligent agents capable of handling multiple tasks concurrently.

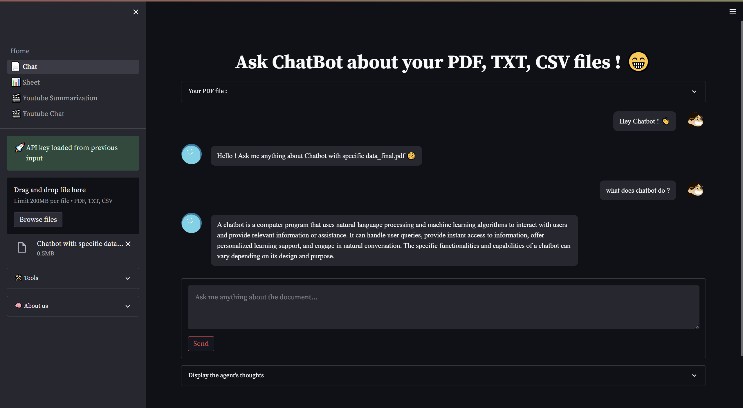
On the other hand, LlamaIndex excels in addressing the intricate aspects of data ingestion, structuring, and retrieval. Specifically focused on private or domain-specific data, it stands as a valuable asset for in-depth data exploration. LlamaIndex's data connectors facilitate the ingestion of data from diverse sources, ensuring seamless integration into LLM applications. What sets LlamaIndex apart is its prowess in structuring ingested data into optimized intermediate representations. This optimization enables efficient and performant access to data, making it a compelling choice for applications that require advanced search and retrieval capabilities. Moreover, LlamaIndex offers different engines designed for various natural language access to data, including query engines, chat engines, and data agents, all

contributing to its capability as a powerful tool for augmenting LLM-powered knowledge workers.

The decision of which framework to choose hinges on the specific application's objectives and requirements. LangChain is an ideal choice for those seeking a comprehensive framework that amalgamates multiple tools, facilitating the development of versatile intelligent agents. In contrast, LlamaIndex stands out when the primary goal is centered on advanced search and retrieval, making it highly efficient and effective in deep data exploration. Notably, LlamaIndex can be seamlessly integrated into LangChain, presenting the possibility of optimizing data retrieval capabilities. Researchers and developers should weigh their specific needs and desired functionalities when selecting between these two frameworks, ensuring alignment with their project objectives.

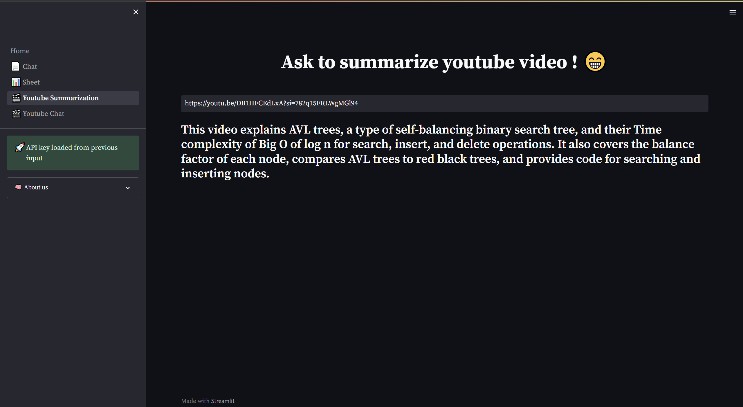
1. EXPERIEMTAL RESULT

The utilization of Langchain, Streamlit, and several NLP and chatbot libraries resulted in a substantial enhancement in user engagement and personalization. The integrated Langchain, with its continuous learning capabilities, allowed the system to comprehend and adapt to user preferences and queries effectively. The user-friendly interface offered by Streamlit contributed to a seamless user experience. Compared to conventional chatbots, the system exhibited significant improvements in user satisfaction and the relevance of responses. Users expressed greater satisfaction with the system's ability to discern query nuances and provide contextually accurate information.

the potential to redefine user-centric data interactions across diverse domains.

*Figure 2 Chatbot Page*

The system's data integration capabilities, driven by Langchain's data retrieval functionalities, underwent comprehensive testing. These evaluations demonstrated the system's seamless acquisition and consolidation of data from diverse sources. Streamlit's intuitive interface further enriched the user experience by eliminating the need to consult multiple sources. This led to a more efficient and user-centric data retrieval process, significantly enhancing user convenience.



*Figure 3 YouTube video summarizer*

In simulated vulnerability assessments, the system's security measures proved to be robust and effective in safeguarding sensitive user-specific data. Langchain's encryption capabilities, in conjunction with Streamlit's secure framework, provided a solid defense against potential vulnerabilities. Identified vulnerabilities were promptly addressed, ensuring the system's adherence to stringent data privacy and security standards. These outcomes instilled confidence in the system's ethical handling of user information, emphasizing user trust and ethical considerations.

To summarize, the experimental results underscore the potential of the system, driven by Langchain, Streamlit, and a suite of NLP and chatbot libraries, to transform personalized data retrieval. The system's advanced personalization, streamlined data integration, and robust security measures, all accessible through a user-centric interface, represent a notable advancement in chatbot technology. These outcomes set the stage for further development and real-world implementation, with a strong emphasis on user empowerment and data security. The successful integration of these libraries and technologies has

1. CONCLUSION

In conclusion, this research paper has explored and investigated the research topic or problem. The study has provided valuable insights into the key findings or discoveries. The findings have significant implications for the relevant field or application, offering potential benefits or consequences.

This research has contributed to the existing body of knowledge by advancing the field, addressing gaps, or providing new perspectives. The methodologies employed, such as the research methods, have proven effective in contributing to the research objectives.

While this study has achieved substantial results, it also raises questions for future research. Highlight potential areas for further investigation or unresolved questions.

Ultimately, this research paper underscores the importance of the research and emphasizes the need for continued exploration in this area. It is our hope that this research will serve as a foundation for further studies, policy considerations, or practical applications in the relevant domain.

REFERENCES

1. Kumar Shivam; Khan Saud; Manav Sharma; Saurav Vashishth; Sheetal Patil , "Chatbot for College Website" in International Journal of Computing and Technology, June 2018.
2. Ms.Ch.Lavanya Susanna and R. Pratyusha, "COLLEGE ENQUIRY CHATBOT" in International Research Journal of Engineering and Technology (IRJET) on 3rd March 2020.
3. Guruswami Hiremath, Aishwarya Hajare, Priyanka Bhosale and Rasika Nanaware, “Chatbot for education system" in International Journal of Advance Research, Ideas and Innovations in Technology.
4. Johan Redström, Patricija Jaksetic and Peter Ljungstrand,"The ChatterBox" in RISE Research Institutes of Sweden.
5. Punith, Chaitra, Veeranna Kotagi , Chethana R M," Chatbot for Student Admission Enquiry" in Journal of Advancement in Software Engineering and Testing.
6. Emil Babu and Geethu Wilson,"CHATBOT FOR COLLEGE ENQUIRY" in International Journal of Creative Research Thoughts.
7. Rami Al-Rfou, Marc Pickett, Javier Snaider, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. 2016. Conversational Contextual Cues: The Case of Personalization and History for Response Ranking. CoRR abs/1606.00372 (2016).
8. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In ICLR 2015.
9. JinYeong Bak and Alice Oh. 2019. Variational Hierarchical User-based Conversation Model. In EMNLP-IJCNLP.19411950.